ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

В.Н. ТУРЧИН

В.Н. ТУРЧИН

ТЕОРИЯ ВЕРОЯТНОСТЕЙ **И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Учебник для студентов высших учебных заведений

Издание второе, переработанное и дополненное

Утверждено Министерством образования и науки Украины

Дніпро . ЛІРА . 2018

УДК 519.2 (075.8)

Т89

Рецензенты:
А. В. Скороход, д-р физ.-мат. наук, проф., акад. НАН Украины (Институт математики НАН Украины),
Ю. С. Мишура, д-р физ.-мат. наук, проф. (Киевский национальный университет имени Тараса Шевченко),
В. В. Булдыгин д-р физ.-мат. наук, проф. (Национальный технический университет Украины "КПИ").

Утверждено Министерством образования и науки Украины как учебник для студентов высших учебных заведений (*письмо* №1/11 – 12196 *от* 30.07.2013)

ТУРЧИН В. Н.
Т89 **Теория вероятностей и математическая статистика.** Учебник для студентов высших учебных заведений. — Днепр, Издательство "Лира". — 2018. — 752 с.

ISBN 978-966-981-029-8

Учебник охватывает программный материал курсов "Теория вероятностей" и "Математическая статистика".

Изложены основные понятия и факты теории вероятностей и математической статистики. Теоретические положения проиллюстрированы многочисленными примерами. К каждой главе приведены задачи для самостоятельной работы.

Для студентов высших учебных заведений.

УДК 519.2 (075.8)

В оформлении обложки использована работа Камиля Писсарро "Бульвар Монмартр. После полудня, солнечно".

ISBN 978-966-981-029-8

100 лет
Днепропетровскому
национальному университету
1918–2018

Предисловие

Эта книга для тех, кто приступает к изучению теории вероятностей и математической статистики. Она написана на основе лекций по курсам “Теория вероятностей”, “Математическая статистика”, “Теория вероятностей и математическая статистика”, читаемых автором в Днепропетровском национальном университете для студентов механико-математического факультета, факультета прикладной математики и других факультетов.

Книга охватывает традиционную тематику курсов “Теория вероятностей” и “Математическая статистика”: стохастический эксперимент и его математическая модель, дискретная модель, схема независимых испытаний, случайная величина, ее распределение и числовые характеристики, свертка, сходимость распределений, характеристическая функция, предельные теоремы, оценивание параметров распределений, методы получения оценок, проверка статистических гипотез, параметрические и непараметрические критерии, простая линейная регрессия.

Изложение материала иллюстрируется многочисленными примерами и задачами. Задачи математической статистики в своем большинстве формулируются не в формально математических терминах (понятиях), а в естественно-научных — так, как они возникают во многих сферах деятельности человека: в физике, химии, биологии, генетике, медицине; в психологии, экологии, сельском хозяйстве; в астрономии, машиностроении, строительстве, геологии, металлургии; в экономике, лингвистике, педагогике, спорте и т. д.

Во втором издании добавилась глава “Линейная регрессия”, были исправлены замеченные опечатки и неточности и внесены изменения в изложение материала, коснувшиеся в основном глав 10, 12, 13, 14.

Учебником могут пользоваться студенты как механико-математических факультетов, факультетов прикладной математики и кибернетики университетов, так и технических, педагогических, экономических высших учебных заведений.

Автор благодарен сыну Евгению Турчину и дочери Наталье Олерих за тщательное прочтение рукописи книги и ее содержательное обсуждение.

Автор будет признателен всем, кто в той или иной форме выскажет свои пожелания, замечания и предложения относительно содержания книги. Предложения и пожелания просим присылать по адресу: Украина, 49010, Днепр-10, пр. Гагарина, 72, Днепровский национальный университет имени Олеся Гончара, механико-математический факультет, кафедра статистики и теории вероятностей, В. Н. Турчину или по адресу vnturchyn@gmail.com

Введение

Случайная природа явлений мира. Окружающий нас мир явлений и процессов носит случайный характер.

Случайными являются переходы атомов и молекул из одних энергетических состояний в другие, эти переходы сопровождаются поглощением или излучением света. Так что свет, окружающий нас всюду, по своей природе случаен.

Случайный характер носят ядерные реакции. Но именно они являются источником энергии звезд и собственного тепла планет; ядерные реакции на Солнце являются основой жизненных процессов на нашей планете.

Другим видом случайности в микромире, влияние которого трудно переоценить, является броуновское движение — хаотическое движение частиц под воздействием теплового движения молекул газов и жидкостей, в котором последние постоянно пребывают. Броуновское движение по своей природе случайно, но именно благодаря броуновскому движению происходит растворение веществ, идут химические реакции, протекают жизненно важные процессы в живых организмах.

Случайными являются передача наследственных признаков, солнечные вспышки, рождение новых и сверхновых звезд ...

Какова же природа случайности? Если исходить из современных физических представлений о строении Вселенной, то основным источником случайности является квантовая природа явлений микромира. Поэтому неудивительно, что многие фундаментальные процессы являются случайными по своей сути.

Изучение случайности. Можно ли точными методами изучать случайность? На первый взгляд кажется, что нет, что математика, известная непреложностью и достоверностью своих выводов, никак не может служить инструментом в изучении случайного. Но именно математикой была предложена количественная мера случайности — вероятность, которая дает возможность изучать случайные явления и процессы.

Раздел математики, занимающийся изучением случайных явлений, процессов и связанных с ними событий, носит название теории вероятностей.

Развитие теории вероятностей. Возникновение теории вероятностей обычно связывают с комбинаторными задачами азартных игр и относят к XVII веку. Азартные игры едва ли можно считать серьезным занятием, но именно они привели к задачам, которые невозможно было решить в рамках известных на то время математических моделей и которые тем самым стимулировали появление в математике новых идей и подходов. Первые общие методы решения таких задач, в частности, задачи о разделе ставки, встречаются у Паскаля (1623 – 1662), Ферма (1601 – 1665), Гюйгенса (1629 – 1695). Но обычно историю теории вероятностей начинают с работы Якоба Бернулли (1654 – 1705) “Искусство предположения” (опубликована в 1713 г.), в которой была приведена первая предельная теорема теории вероятностей — закон больших чисел.

Фундаментальный труд Лапласа (1749 – 1827) “Аналитическая теория вероятностей” придает теории вероятностей, в известном смысле, законченный вид. Но теория, построенная Лапласом на определении вероятности события как отношения числа исходов, благоприятствующих событию, к общему числу исходов, не могла охватить широкий круг наблюдаемых случайных явлений, и поэтому было понятным падение интереса к теории вероятностей, вплоть до непризнания ее математической дисциплиной.

Выдающиеся ученые предвидели важную роль науки, занимающейся изучением случайных явлений и процессов, но еще долго азартные игры, страхование и демография были теми областями, в которых возникали задачи, формировались понятия и создавались методы теории вероятностей. В конце XIX века и начале XX века появился ряд серьезных задач, вызванных нуждами естествознания и связанных с изучением случайных явлений, которые стали толчком к развитию теории вероятностей как большого, в значительной мере самостоятельного, раздела математики.

Основоположниками теории вероятностей как строгой математической науки были выдающиеся российские математики: П. Л. Чебышёв (1821 – 1894) и его ученики А. А. Марков (1856 – 1922) и А. М. Ляпунов (1857 – 1918). Число работ П. Л. Чебышёва по теории вероятностей невелико — всего четыре, но их роль трудно переоценить.

Основной вклад в становление и развитие современной теории вероятностей внесла плеяда блестящих математиков:

С. Н. Бернштейн, Э. Борель, Б. В. Гнеденко, А. Н. Колмогоров, П. Леви, Р. Мизес, А. Я. Хинчин. В частности, А. Н. Колмогоровым в 1933 году в книге “Основные понятия теории вероятностей” была предложена аксиоматика теории вероятностей, ныне общепринятая во всем мире. Эта аксиоматика не только охватывает классические области теории вероятностей, но и является основой для развития ее многочисленных новых разделов. Построенная на аксиоматике А. Н. Колмогорова как чисто математическая наука, теория вероятностей “работает” в приложениях так же хорошо, как, например, евклидова геометрия или ньютонова механика.

Большой вклад в развитие современной теории вероятностей и математической статистики внесли украинские математики: Б. В. Гнеденко, И. И. Гихман, И. Н. Коваленко, В. С. Королюк, А. В. Скороход, М. И. Ядренко.

Математическая статистика. Математическая статистика как самостоятельная математическая дисциплина зарождается в 20 – 30-е гг. XX века. В определенном смысле математическая статистика занимается задачами, обратными задачам теории вероятностей. Если основная задача теории вероятностей — изучение математических моделей стохастических экспериментов с целью прогнозирования случайных явлений и процессов, то задача математической статистики — выявление структуры вероятностных моделей по результатам стохастических экспериментов.

Случайные процессы. Теория случайных процессов выделилась из теории вероятностей сравнительно недавно. Традиционная теория вероятностей изучает мгновенно складывающиеся случайные ситуации, теория случайных процессов исследует процессы, находящиеся под воздействием случайных факторов и разворачивающиеся во времени, пространстве, ... Основными областями применения теории случайных процессов являются радио- и электротехника, кибернетика, экономика, биология, но влияния теории случайных процессов в той или иной степени, пожалуй, не избежала ни одна из естественных наук.

Приложения. Теорию вероятностей (вместе с примыкающими к ней статистикой и случайными процессами) справедливо относят к числу математических дисциплин, имеющих широкие и тесные связи с практикой. К достоинствам теории вероятностей в приложениях следует отнести то, что предлагаемые ею методы работают и доставляют качественную и количественную информацию в крайне неблагоприятных условиях, когда об изучаемом явлении, порождающих его факторах, известно мало, а то, что известно, имеет стохастическую природу. Этим объясняется широкое применение теории вероятностей и математической статистики в самых разнообразных областях — от исследований в области физики элементарных частиц и квантовой механики до социологии, экономики, информационных технологий. И каким бы странным это ни показалось, но чем глубже мы проникаем в суть явлений, тем больше роль, значение и широта применения вероятностно-статистических методов.

Глава 1

Стохастический эксперимент

Теория вероятностей изучает *стохастические* (*случайные*) *эксперименты*, точнее, количественные закономерности стохастических экспериментов, исследуя их математические модели.

Под стохастическим экспериментом мы будем понимать эксперимент, исход которого невозможно предсказать заранее (до проведения эксперимента), но который можно повторить (воспроизвести) в принципе неограниченное число раз, причем так, чтобы результаты предыдущих экспериментов не влияли на последующие (независимым образом).

Наша ближайшая задача — построить математическую модель стохастического эксперимента.

1.1 Примеры стохастических экспериментов

1. Эксперимент состоит в подбрасывании монеты. В результате эксперимента может выпасть либо герб, либо решетка. Заранее (до проведения эксперимента) мы не можем сказать, что выпадет. Эксперимент можно повторить независимым образом неограниченное число раз.

2. Подбрасывают пару монет. В результате проведения эксперимента на каждой из них может выпасть либо герб, либо решетка. Что выпадет — до проведения эксперимента предсказать невозможно. Эксперимент можно повторить независимым образом неограниченное число раз.

3. Эксперимент состоит в подбрасывании монеты до первого выпадения “герба”. Сколько подбрасываний будет проведено — заранее предсказать нельзя. Эксперимент можно повторить неограниченное число раз.

4. Эксперимент состоит в подбрасывании игральной кости (игральная кость — кубик, грани которого пронумерованы от 1 до 6). Предсказать наперед, какое число очков выпадет на грани игральной кости, невозможно. Эксперимент можно повторить независимым образом неограниченное число раз.

5. Эксперимент состоит в бросании наудачу точки на отрезок $[0;1]$. Заранее невозможно предсказать, попадет ли точка в заданный промежуток $[a;b]$ из $[0;1]$. Эксперимент можно повторить независимым образом неограниченное число раз.

6. Эксперимент состоит в определении срока службы изделия до первого отказа. Предсказать длительность безотказной работы заранее невозможно. Эксперимент можно повторить неограниченное число раз.

7. Частица принимает участие в броуновском движении — движется вследствие толчков молекул жидкости либо газа. Сами же молекулы пребывают в хаотическом тепловом движении. Наблюдая частицу в течение определенного промежутка времени, можно начертить траекторию ее движения. Вид этой траектории невозможно предвидеть заранее. Здесь мы имеем дело со стохастическим экспериментом, результатом которого является траектория движения броуновской частицы. Эксперимент можно повторить неограниченное число раз.

В качестве примера не стохастического эксперимента с непредсказуемым исходом можно рассмотреть выборы депутатов в парламент — результат выборов (будущий состав парламента) до их проведения предсказать нельзя, но повторить эксперимент невозможно.

1.2 Алгебра наблюдаемых событий стохастического эксперимента

Попытаемся выделить характерные свойства стохастического эксперимента, которыми его можно было бы описать, и в дальнейшем, при построении математической модели, постулировать.

Из опыта известно, что с каждым стохастическим экспериментом можно связать семейство событий, которые можно наблюдать (которые могут произойти) в данном эксперименте, будем их называть *наблюдаемыми событиями.* Наблюдаемые события стохастического эксперимента, как правило, обозначают большими латинскими буквами $A, B, C, \ldots$, а семейство всех наблюдаемых событий — готической буквой $\mathfrak{M}$.

В теории вероятностей мы будем интересоваться событиями только с одной стороны — произошло событие в эксперименте или нет и как часто событие происходит.

Примеры наблюдаемых событий. Стохастический эксперимент состоит в подбрасывании пары монет. Наблюдаемые события: A — "выпал хотя бы один герб", B — "монеты легли разными сторонами", C — "обе монеты легли одной стороной".

Стохастический эксперимент состоит в подбрасывании игральной кости. Наблюдаемые события: A — "выпало четное число очков", B — "выпало нечетное число очков", C — "выпало число очков, кратное трем", D — "выпало 6 очков".

Стохастический эксперимент состоит в определении срока службы данного прибора. Наблюдаемые события: "срок службы более 100 часов", "срок службы не превосходит 50 часов".

Стохастический эксперимент состоит в бросании наудачу точки на отрезок $[0, 1]$. Наблюдаемые события: "точка попала на отрезок $[0, 1/2]$", "точка попала на отрезок $[1/4, 3/4]$".

Алгебра наблюдаемых событий. Если наблюдаемые события A и B таковы, что всякий раз, когда происходит событие A, происходит и событие B, будем говорить, что событие A *влечет* событие B, и будем обозначать это так: $A \subset B$.

Наблюдаемые события A и B такие, что $A \subset B$ и $B \subset A$, (всякий раз, когда происходит A, происходит и B, и всякий раз, когда происходит B, происходит и A) в теории вероятностей неразличимы. Мы их отождествляем и записываем это так: $A = B$.

Заметим, что понятие "влечет", употребляемое по отношению к случайным событиям, отличается от общепринятого в обыденной жизни, где событие A, влекущее событие B, во-первых, происходит раньше во времени, и, во-вторых, выступает для B в качестве причины. В теории вероятностей причинная зависимость между A и B необязательна.

В семействе $\mathfrak{M}$ наблюдаемых событий стохастического эксперимента выделяются два — *достоверное* событие (будем обозначать его U), оно происходит при каждой реализации стохастического эксперимента, и *невозможное* — событие, которое не происходит ни при одной реализации стохастического эксперимента (его обозначают V). Например, в эксперименте, состоящем в подбрасывании игральной кости, достоверными являются события "число выпавших очков не превосходит шести", "число выпавших очков меньше десяти", невозможными являются события "число выпавших очков на грани игральной кости равно десяти", "число выпавших очков кратно семи". В теории вероятностей все достоверные события отождествляют и рассматривают только одно достоверное событие, все невозможные события также отождествляют и рассматривают только одно невозможное событие.

Далее, если событие A — наблюдаемое, то, естественно, событие, состоящее в том, что A не происходит, также наблюдаемое, будем его называть *противоположным событию* A, и обозначать $\overline{A}$ (при подбрасывании игральной кости событием, противоположным событию A — "выпало четное число очков", является событие B — "выпало нечетное число очков").

Если A и B — наблюдаемые события стохастического эксперимента, то событие, состоящее в том, что происходит хотя бы одно из событий A или B — наблюдаемое, будем его называть *суммой событий* A и B и обозначать $A + B$ или $A \cup B$.

Если A и B — наблюдаемые события стохастического эксперимента, то событие, состоящее в том, что происходит как событие A, так и событие B, также наблюдаемое, будем называть его *произведением событий* A и B и обозначать AB или $A \cap B$ (при подбрасывании пары монет произведением событий A — "выпал хотя бы один герб" и B — "обе монеты легли одной стороной" является событие "обе монеты легли гербами").

Для наблюдаемых событий A и B событие, состоящее в том, что A происходит, а B не происходит, также наблюдаемое, будем его называть *разностью событий* A и B и обозначать $A \setminus B$ (при подбрасывании пары монет разностью событий A — "выпал хотя бы один герб" и B — "монеты легли разными сторонами" является событие "на обеих монетах выпал герб").

События A и B будем называть *несовместными*, если $AB = V$ (при подбрасывании игральной кости события A — "выпало четное число очков" и событие B — "выпало нечетное число очков" несовместны).

Итак, класс $\mathfrak{M}$ наблюдаемых событий стохастического эксперимента вместе с каждым наблюдаемым событием A содержит событие $\overline{A}$, противоположное A, вместе с каждой парой наблюдаемых событий A и B содержит их сумму $A \cup B$. Такое семейство событий будем называть *алгеброй наблюдаемых событий.*

Проиллюстрируем еще раз введенные понятия суммы, произведения, противоположного события, несовместных событий на примере так называемой диаграммы Венна.

Пример 1.2.1 (диаграмма Венна). *Эксперимент состоит в бросании наудачу точки в квадрат и наблюдении событий, состоящих в попадании точки в те или иные его подмножества. Обозначим через A событие “точка попала в вертикальный прямоугольник”, а через B — “точка попала в горизонтальный прямоугольник”. Тогда наблюдаемые события A, B, $\overline{A}$, $A \cup B$, $A \cap B$, $B \setminus A$ состоят в попадании точки в соответствующие заштрихованные области* (*см. рис.* 1.2.1).

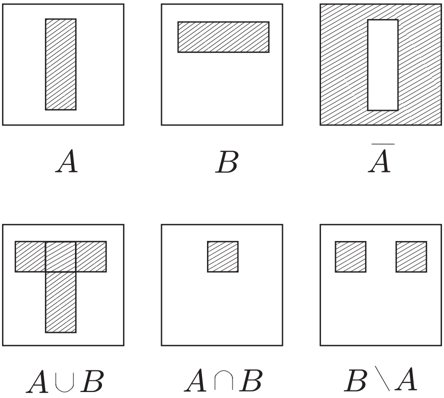

Рис. 1.2.1: Диаграмма Венна

1.3 Исходы стохастического эксперимента

Среди наблюдаемых событий стохастического эксперимента можно выделить “элементарные” (“неразложимые”, “неделимые”) события, которые мы будем называть исходами. Точнее, наблюдаемое событие E мы будем называть *исходом стохастического эксперимента*, если для любого другого наблюдаемого события A либо $E \subset A$, либо $E \cap A = V$ (два последних соотношения — формализованные условия “элементарности”, “неразложимости”). Совокупность всех исходов стохастического эксперимента будем обозначать $\mathcal{E}$.

Примеры

1. Стохастический эксперимент состоит в подбрасывании двух различимых монет. Исходы стохастического эксперимента: “на первой монете выпал герб, на второй — герб”, “на первой монете выпал герб, на второй — решетка" и т. д. Каждое из перечисленных событий для любого другого события либо влечет его, либо несовместно с ним — является исходом стохастического эксперимента.

2. Стохастический эксперимент состоит в подбрасывании игральной кости. Исходы стохастического эксперимента: "выпало 1 очко", "выпало 2 очка", ..., "выпало 6 очков". Каждое из этих событий для любого другого наблюдаемого события либо влечет его, либо несовместно с ним — является исходом стохастического эксперимента.

3. Стохастический эксперимент состоит в бросании точки на отрезок $[0,1]$. Возможные исходы стохастического эксперимента: "брошенная точка попала в точку a", $a \in [0,1]$.

В дальнейшем для нас принципиально важным будет следующее утверждение, устанавливающее связь между элементарными и наблюдаемыми событиями.

Каждое наблюдаемое событие A стохастического эксперимента можно описать некоторым подмножеством множества его исходов $\mathcal{E}$, а именно, подмножеством тех исходов E, которые влекут A:

$$A = \{E : E \in \mathcal{E}, \textit{ которые влекут } A\}. \tag{1.3.1}$$

Другими словами, каждое наблюдаемое событие A можно интерпретировать (рассматривать) как подмножество множества $\mathcal{E}$ исходов стохастического эксперимента.

Пример. *Стохастический эксперимент состоит в подбрасывании игральной кости. Описать множество исходов $\mathcal{E}$ стохастического эксперимента; описать событие A — "выпало четное число очков" как подмножество $\mathcal{E}$.*

Решение. Множество исходов $\mathcal{E} = \{$"выпало 1 очко", "выпало 2 очка", ..., "выпало 6 очков"$\}$.

Событие A — "выпало четное число очков" — описывается подмножеством $\{$"выпало 2 очка", "выпало 4 очка", "выпало 6 очков"$\}$ множества исходов $\mathcal{E}$.

Из определений операций сложения, умножения, взятия противоположного события следует, что

$$A + B = \{E : E \in \mathcal{E}, \text{которые влекут } A \text{ или } B\},$$

$$AB = \{E : E \in \mathcal{E}, \text{которые влекут } A \text{ и } B\},$$

$$\overline{A} = \{E : E \in \mathcal{E}, \text{которые влекут } \overline{A}\}.$$

1.4 Частота события

Каждому стохастическому эксперименту соответствует пара $\{\mathcal{E}, \mathfrak{M}\}$, в которой $\mathcal{E}$ описывает исходы эксперимента, $\mathfrak{M}$ — наблюдаемые события. Но пары $\{\mathcal{E}, \mathfrak{M}\}$ явно недостаточно для описания стохастического эксперимента. Как показывает опыт, в стохастическом эксперименте одни события происходят "чаще", другие происходят "реже" — каждому наблюдаемому событию свойственна своя частота появления — в длинной серии независимых экспериментов доля тех из них, в которых наблюдается данное событие, остается практически неизменной. Другими словами, если $k_n(A)$ — число тех экспериментов из n проведенных, в которых наступило событие A, то их доля

$$\frac{k_n(A)}{n} = \nu_n(A) \tag{1.4.1}$$

среди n, еще называемая *частотой события A в последовательности n экспериментов*, при больших n практически неизменна. На частоту $\nu_n(A)$ события A в последовательности n экспериментов естественно смотреть как на количественную меру частоты его появления в n экспериментах.

События A, у которых частота $k_n(A)/n$ больше, происходят чаще, события B, у которых частота $k_n(B)/n$ меньше, происходят реже. Например, в эксперименте, состоящем в подбрасывании симметричной игральной кости, событие A — "одно очко не выпало" происходит чаще события B — "одно очко выпало"; в эксперименте, состоящем в подбрасывании симметричной монеты до первого выпадения герба, событие D — "эксперимент закончился на 1-м шаге" (герб выпал при первом же подбрасывании) происходит гораздо чаще события C — "эксперимент закончился на 10-м шаге" (9 раз подряд выпала решетка, и только потом герб).

Непосредственно из определения частоты события в последовательности экспериментов, см. (1.4.1), получаем следующие её свойства:

1° для каждого события A

$$\nu_n(A) \geq 0$$

(частота неотрицательна);

2° для несовместных событий A и B

$$\nu_n(A+B) = \nu_n(A) + \nu_n(B)$$

(частота аддитивна);

3° для достоверного события U

$$\nu_n(U) = 1$$

(частота нормирована).

Заметим, что $\nu_n(A)$ может быть вычислена только после проведения n экспериментов и, вообще говоря, меняется, если провести другую серию экспериментов или изменить n.

Резюме. *Итак, с каждым стохастическим экспериментом мы связываем три объекта:*

1° *множество исходов* $\mathcal{E}$ *стохастического эксперимента;*

2° *алгебру наблюдаемых событий* $\mathfrak{M}$ *стохастического эксперимента;*

3° *частоту* $\nu_n(A)$ *наблюдаемого события* A *в последовательности экспериментов.*

Глава 2

Математическая модель стохастического эксперимента

Мы будем строить математическую модель стохастического эксперимента, вводя последовательно 1) модель совокупности исходов $\mathcal{E}$; 2) модель алгебры наблюдаемых событий $\mathfrak{M}$; 3) модель частоты события в последовательности экспериментов.

2.1 Пространство элементарных событий

В качестве математической модели совокупности $\mathcal{E}$ исходов стохастического эксперимента рассмотрим множество Ω, вообще говоря, произвольное, но такое, что между исходами E совокупности всех исходов $\mathcal{E}$ стохастического эксперимента и элементами ω множества Ω можно установить взаимно однозначное соответствие (другими словами, множество Ω такое, что каждый исход E данного стохастического эксперимента можно обозначить одним элементом ω из Ω и после этого в Ω не останется элементов). Взаимно однозначное соответствие между исходами из $\mathcal{E}$ и элементами из Ω обозначим через K:

$$K : E \to \omega = K(E). \tag{2.1.1}$$

Множество Ω называют *пространством элементарных событий* (*пространством исходов*) стохастического эксперимента, а его элементы ω — *элементарными событиями* (*исходами*).

Пространство элементарных событий Ω *является математической моделью совокупности исходов стохастического эксперимента*, в том смысле, что каждый исход E стохастического эксперимента описывается одним и только одним элементом ω из Ω.

Проведение стохастического эксперимента интерпретируется как случайный выбор точки ω из множества Ω.

Замечание. Для данного стохастического эксперимента можно построить не одно пространство элементарных событий Ω. Между этими пространствами естественным образом устанавливается взаимно однозначное соответствие. Относительно выбора Ω теория вероятностей не дает никаких правил и рекомендаций.

Пример 2.1.1 . *Стохастический эксперимент состоит в подбрасывании игральной кости. Предложить пространство элементарных событий этого стохастического эксперимента.*

Совокупностью исходов $\mathcal{E}$ стохастического эксперимента является {“выпало 1 очко”, “выпало 2 очка”, . . . , “выпало 6 очков”}.

В качестве пространства элементарных событий можно и естественно рассмотреть множество $\Omega = \{1, 2, 3, 4, 5, 6\}$, установив взаимно однозначное соответствие между $\mathcal{E}$ и Ω следующим образом: “выпало 1 очко” $\leftrightarrow$ 1, ”выпало 2 очка” $\leftrightarrow$ 2, . . . , “выпало 6 очков” $\leftrightarrow$ 6 (см. рис. 2.1.1).

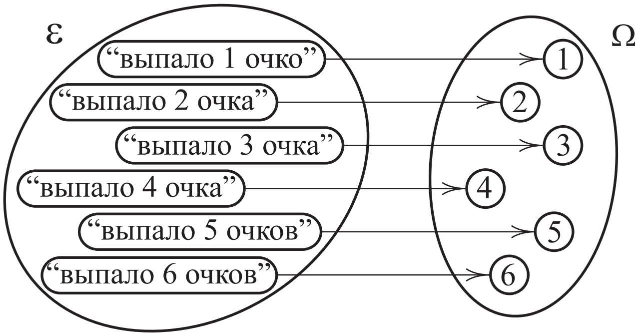

Рис. 2.1.1: Иллюстрация к построению Ω в примере 2.1.1

2.2 Алгебра множеств

Теперь построим математическую модель алгебры наблюдаемых событий стохастического эксперимента.

Каждое событие A из алгебры наблюдаемых событий $\mathfrak{M}$ описывается подмножеством множества $\mathcal{E}$ исходов стохастического эксперимента, а именно, подмножеством тех исходов, которые влекут событие A:

$$A = \{E : E \in \mathcal{E},\ \text{которые влекут } A\}.$$

Между множеством исходов $\mathcal{E}$ и множеством Ω существует взаимно однозначное соответствие

$$K : E \to \omega = K(E),$$

см. определение (2.1.1) отображения K и рисунок 2.2.1. Поэтому каждому наблюдаемому событию A из $\mathfrak{M}$ естественным образом ставится в соответствие подмножество M_A из Ω:

$$M_A = \{\omega : \omega \in \Omega,\ \text{для которых соответствующие } E \text{ влекут } A\}$$

(см. также рис. 2.2.1). В терминах отображения K

$$A \to K(A) = M_A,\ A \in \mathfrak{M}.$$

Класс всех подмножеств M_A $(A \in \mathfrak{M})$ множества Ω обозначим через $\mathfrak{A}$.

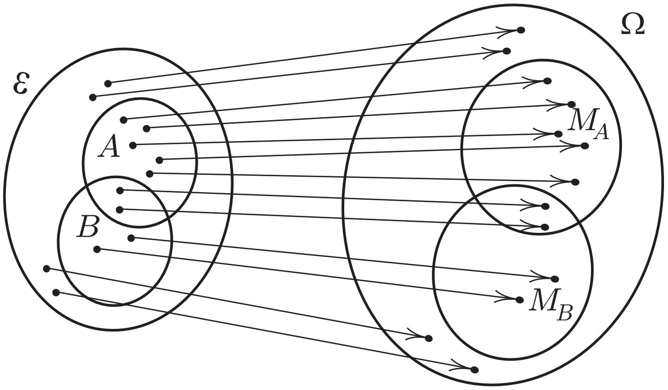

Рис. 2.2.1: Иллюстрация к определению M_A

Класс $\mathfrak{A}$ обладает следующими свойствами:

1° Для любого M_A из $\mathfrak{A}$ его дополнение $\overline{M}_A$ также принадлежит $\mathfrak{A}$, поскольку

$$\overline{A} \to K(\overline{A}) = M_{\overline{A}} = \overline{M}_A.$$

2° Для любых M_A и M_B из $\mathfrak{A}$ их объединение $M_A \cup M_B$ также принадлежит $\mathfrak{A}$, поскольку

$$A \cup B \to K(A \cup B) = M_{A \cup B} = M_A \cup M_B.$$

3° Для любых M_A и M_B из $\mathfrak{A}$ их пересечение $M_A \cap M_B$ также принадлежит $\mathfrak{A}$, поскольку

$$A \cap B \to K(A \cap B) = M_{A \cap B} = M_A \cap M_B.$$

Класс подмножеств, обладающий свойствами 1° и 2°, называют алгеброй.

Определение. Непустой класс $\mathfrak{A}$ подмножеств пространства Ω будем называть *алгеброй подмножеств* Ω, если:

1° вместе с любым C из $\mathfrak{A}$ его дополнение $\overline{C}$ также принадлежит $\mathfrak{A}$ (класс $\mathfrak{A}$ замкнут относительно операции дополнения);

2° вместе с любыми C и D из $\mathfrak{A}$ их объединение $C \cup D$ также принадлежит $\mathfrak{A}$ (класс $\mathfrak{A}$ замкнут относительно операции объединения).

Заметим, что из определения алгебры множеств следует, что она замкнута и относительно операции пересечения.

Таким образом, взаимно однозначное соответствие K (см. (2.1.1)) между элементами множеств $\mathcal{E}$ и Ω задает взаимно однозначное соответствие между событиями алгебры наблюдаемых событий $\mathfrak{M}$ и множествами некоторой алгебры $\mathfrak{A}$ подмножеств пространства Ω, причем оно обладает свойствами 1° – 3°. Поэтому *алгебру подмножеств* $\mathfrak{A}$ *пространства* Ω *естественно рассматривать в качестве математической модели семейства* (*алгебры*) *наблюдаемых событий* $\mathfrak{M}$, в том смысле, что каждое наблюдаемое событие из алгебры $\mathfrak{M}$ однозначно описывается некоторым (вполне определенным) подмножеством M_A из алгебры подмножеств $\mathfrak{A}$ пространства Ω, при этом сумме событий из $\mathfrak{M}$ соответствует объединение множеств из $\mathfrak{A}$, произведению событий из $\mathfrak{M}$ соответствует пересечение множеств из $\mathfrak{A}$, противоположному событию из $\mathfrak{M}$ соответствует дополнение множества из $\mathfrak{A}$, достоверному событию соответствует Ω, а невозможному — $\emptyset$.

Проведение эксперимента интерпретируется как случайный выбор точки ω из Ω. Если при этом $\omega \in M_A$, то мы говорим, что произошло событие A, в противном случае — событие A не произошло.

В дальнейшем множество M_A из алгебры $\mathfrak{A}$, подмножеств Ω, соответствующее событию $A \in \mathfrak{M}$, будем обозначать той же буквой A, а подмножества Ω из $\mathfrak{A}$ будем называть наблюдаемыми событиями, или просто событиями.

Итак, на данном этапе построения математической модели стохастического эксперимента под его математической моделью будем понимать пару $\{\Omega, \mathfrak{A}\}$, где Ω — пространство элементарных событий — множество, вообще говоря, любой природы (модель семейства исходов стохастического эксперимента), $\mathfrak{A}$ — алгебра подмножеств пространства Ω — модель семейства $\mathfrak{M}$ наблюдаемых событий.

Пример. *Стохастический эксперимент состоит в подбрасывании пары различимых монет.*

Предложить математическую модель семейства исходов стохастического эксперимента, описать событие A — "монеты легли разными сторонами" как подмножество пространства элементарных событий.

В качестве пространства элементарных событий Ω естественно рассмотреть $\Omega = \{\text{ГГ}, \text{ГР}, \text{РГ}, \text{РР}\}$. Событие A — "монеты легли разными сторонами" опишется так: $A = \{\text{ГР}, \text{РГ}\}$.

2.3 Вероятность

Для завершения построения математической модели стохастического эксперимента остается построить математическую модель частоты появления события.

В параграфе 1.4 в качестве количественной меры частоты наблюдаемого события A было предложено рассматривать частоту $\nu_n(A)$ наблюдаемого события A в последовательности n экспериментов. К сожалению эта количественная мера имеет недостаток: если провести другую серию экспериментов или изменить n, то $\nu_n(A)$, вообще говоря, изменится — одному и тому же наблюдаемому событию A, в зависимости от n или от серии экспериментов, ставятся в соответствие различные числа $\nu_n(A)$, что ставит под сомнение корректность использования $\nu_n(A)$ в качестве количественной мерой частоты события A. Но, как показывает опыт не одного поколения исследователей, *частота $\nu_n(A)$ наблюдаемого события A в последовательности экспериментов обладает замечательными свойствами: при больших n она практически неизменна — несущественно уклоняется от* *некоторого числа* (*частота устойчива*) *и с ростом n заметные уклонения $\nu_n(A)$ от него встречаются все реже* (*частота стабилизируется*). Например, в экспериментах, проведенных Ж. Бюффоном и К. Пирсоном и состоящих в подбрасывании монеты, частота выпадения герба в последовательности экспериментов вела себя, как указано в табл. 2.3.1.

Таблица 2.3.1

Экспериментатор	Число подбрасываний	Число выпадений герба	Частота
Ж. Бюффон	4040	2048	0,5080
К. Пирсон	12 000	6019	0,5016
К. Пирсон	24 000	12 012	0,5005

Свойства устойчивости и стабилизации частоты $\nu_n(A)$ в последовательности экспериментов дает основание предположить существование предела $\lim\limits_n \nu_n(A)$ (обозначим его $\nu(A)$). Предел

$$\lim_n \nu_n(A) = \nu(A),$$

если бы его существование было гарантировано, и был бы той количественной мерой частоты появления события A, которая для каждого события A определялась бы однозначно — не зависела от серии и от числа экспериментов в серии. При этом предел $\nu(A)$ обладал бы следующими свойствами:

1° для каждого события A

$$\nu(A) \geq 0$$

(неотрицательность ν);

2° для несовместных событий A и B

$$\nu(A + B) = \nu(A) + \nu(B)$$

(аддитивность ν);

3° если U — достоверное событие, то

$$\nu(U) = 1$$

(нормированность ν).

Но в рамках математической теории невозможно установить существование предела частоты $\nu_n(A)$ события A в последовательности экспериментов — это невозможно сделать в принципе.

И коль скоро, к сожалению, это так, поступим следующим образом. Постулируем для каждого наблюдаемого события A существование числа $\mathsf{P}(A)$, что представляется вполне естественным, учитывая экспериментально установленный факт устойчивости частоты $\nu_n(A)$ события A, причем так, чтобы набор чисел $\mathsf{P}(A)$, $A \in \mathfrak{M}$, обладал свойствами, которые имел бы

$$\lim_n \nu_n(A) = \nu(A),$$

если бы его существование было гарантировано (см. свойства 1° – 3° предела $\nu(A)$), а именно:

1° для каждого события A

$$\mathsf{P}(A) \geq 0$$

(неотрицательность P);

2° для несовместных событий A и B

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B)$$

(аддитивность P);

3° если U — достоверное событие, то

$$\mathsf{P}(U) = 1$$

(нормированность P).

Число $\mathsf{P}(A)$ будем называть вероятностью события A, а набор чисел $\mathsf{P}(A)$ ($A \in \mathfrak{M}$), точнее функцию $\mathsf{P}(A)$ на $\mathfrak{M}$, будем называть вероятностью на $\mathfrak{M}$. Существование вероятности постулируется для каждого наблюдаемого события вне зависимости от того, проводилась серия независимых экспериментов или нет.

Для формального определения вероятности нам понадобятся следующие определения.

Функция множества. Если каждому подмножеству A из некоторого класса $\mathfrak{A}$ подмножеств Ω поставлено в соответствие число $W(A)$, возможно равное $+\infty$ или $-\infty$, то будем говорить, что на классе $\mathfrak{A}$ задана *функция множества* со значениями в $[-\infty, +\infty]$.

Определение. Функцию множества W со значениями в $[-\infty, +\infty]$, заданную на классе подмножеств $\mathfrak{A}$, будем называть *конечно-аддитивной*, если она принимает бесконечные значения только одного знака и для любого конечного набора попарно непересекающихся множеств A_i ($A_i \in \mathfrak{A}, A_k \cap A_j = \emptyset, k \neq j$) таких, что $\bigcup\limits_{i=1}^{n} A_i \in \mathfrak{A}$,

$$W\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} W(A_i).$$

Определение. Функцию множества W со значениями в $[-\infty, +\infty]$, заданную на классе подмножеств $\mathfrak{A}$, будем называть *счетно-аддитивной*, если она принимает бесконечные значения только одного знака и для любого счетного набора попарно непересекающихся множеств A_i ($A_i \in \mathfrak{A}, A_k \cap A_j = \emptyset, k \neq j$) таких, что $\bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$

$$W\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} W(A_i).$$

Аддитивные функции естественным образом возникают на практике. Приведем несколько примеров:

1. Длина. Каждому промежутку $[a_i, b_i)$ прямой $\mathbb{R}^1$ ставится в соответствие его длина:

$$L([a_i, b_i)) = b_i - a_i.$$

Если при этом $[a_i, b_i) \cap [a_j, b_j) = \emptyset, i \neq j$, то

$$L\left(\bigcup_{i}[a_i, b_i)\right) = \sum_{i} L([a_i, b_i)).$$

2. Площадь. Каждой части Q_i области Q ставится в соответствие ее площадь $S(Q_i)$. Если при этом $Q_i \cap Q_j = \emptyset, i \neq j$, то

$$S\left(\bigcup_{i} Q_i\right) = \sum_{i} S(Q_i).$$

3. Число жителей. Каждой части S_i данного географического района S ($S_i \subset S$) ставится в соответствие количество $k(S_i)$ жителей этого района. Если при этом $S_i \cap S_j = \emptyset, i \neq j$, то

$$k\left(\bigcup_{i} S_i\right) = \sum_{i} k(S_i).$$

4. Расстояние. Каждой части t_i промежутка времени t ставится в соответствие расстояние $L(t_i)$, которое пролетел самолет. Если при этом $t_i \cap t_j = \emptyset, i \neq j$, то

$$L\left(\bigcup_i t_i\right) = \sum_i L(t_i).$$

5. Число телефонных вызовов. Каждой части t_i временного промежутка t ставится в соответствие число $n(t_i)$ телефонных вызовов на АТС за промежуток t_i. Если при этом $t_i \cap t_j = \emptyset, i \neq j$, то

$$n\left(\bigcup_i t_i\right) = \sum_i n(t_i).$$

6. Примеры неаддитивных функций множества. Рассмотрим металлический стержень, например, единичной длины. Один конец его поместим в точку 0, другой — в точку 1. Будем считать, что температура стержня установилась, например, 25 по Цельсию. Рассмотрим функцию множества T — каждой части стержня поставим в соответствие ее температуру. Тогда $T([0,1/2]) = 25$, $T([1/2,1]) = 25$, $T([0,1]) = 25$. При этом

$$25 = T([0,1/2] \cup [1/2,1]) \neq T([0,1/2]) + T([1/2,1]) = 25 + 25 = 50.$$

Так что температура как функция множества не аддитивна.

Цена ткани. Каждому отрезу (части) ткани ставится в соответствие число — его цена. Как правило, цена рулона (опта) меньше суммарной цены частей. Так что цена, вообще говоря, не является аддитивной функцией.

Вероятность. Изложенные выше соображения и факты приводят к следующему формальному определению вероятности как математической модели частоты события в последовательности экспериментов.

Определение. Неотрицательную счетно-аддитивную нормированную функцию множества, заданную на алгебре подмножеств $\mathfrak{A}$ пространства Ω, будем называть *вероятностью* на $\mathfrak{A}$.

Другими словами, вероятность P — это функция множества со значениями в $\mathbb{R}^1$, заданная на алгебре $\mathfrak{A}$ подмножеств пространства Ω, такая, что:

1° для каждого $A \in \mathfrak{A}$

$$\mathsf{P}(A) \geq 0$$

(вероятность неотрицательна);

2° для любых $A_i \in \mathfrak{A}, i = 1, 2, \ldots,$ таких, что $A_k \cap A_j = \emptyset$, $k \neq j$, и $\bigcup\limits_{j=1}^{\infty} A_j \in \mathfrak{A}$

$$\mathsf{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathsf{P}(A_j)$$

(вероятность счетно-аддитивна);

3° для достоверного события Ω

$$\mathsf{P}(\Omega) = 1$$

(вероятность — нормированная функция множества).

Свойство 1° вероятности (неотрицательность) — постулированное свойство неотрицательности частоты ($\nu_n(A) \geq 0$ для каждого события A).

Свойство 2° вероятности (счетная аддитивность) — постулированное свойство аддитивности частоты (для несовместных событий A и B значение $\nu_n(A + B) = \nu_n(A) + \nu_n(B)$), но усиленное — вероятность объединения непересекающихся событий равна сумме вероятностей не только для объединения конечного числа непересекающихся событий, но и для объединения их счетного числа.

Свойство 3° вероятности (нормированность) — постулированное свойство нормированности частоты: $\nu_n(U) = 1$.

Вероятности, как правило, обозначают большими латинскими буквами $\mathsf{P}, \mathsf{Q}, \mathsf{F}, \mathsf{G}$.

В частности, если $\Omega = \mathbb{R}^n$, то вероятность называют *вероятностным распределением* или *распределением вероятностей* на $\mathbb{R}^n$.

О задании вероятности на $\boldsymbol{\sigma}$-алгебре. Вероятность, заданную на алгебре $\mathfrak{A}$ подмножеств Ω, всегда, без дополнительных предположений, можно считать заданной на более широком классе $\mathfrak{F}$ подмножеств пространства Ω, содержащем алгебру $\mathfrak{A}$, на так называемой наименьшей σ-алгебре, содержащей алгебру $\mathfrak{A}$. Поэтому в дальнейшем мы всегда будем считать вероятность заданной на σ-алгебре.

Определение. Непустой класс $\mathfrak{F}$ подмножеств Ω будем называть *σ-алгеброй*, если

1° для каждого $A \in \mathfrak{F}$ и $\overline{A} \in \mathfrak{F}$,

2° для любых $A_i \in \mathfrak{F}, i = 1, 2, \ldots,$ и $\bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{F}$.

В дальнейшем *вероятностью будем называть неотрицательную счетно-аддитивную нормированную функцию множества*, заданную на σ-алгебре $\mathfrak{F}$ подмножеств Ω.

Определение. Тройку $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, где Ω — пространство элементарных событий (множество произвольной природы); $\mathfrak{F}$ — σ-алгебра подмножеств пространства Ω; P — вероятность на $\mathfrak{F}$, будем называть *вероятностным пространством.*

Вероятностное пространство $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ является математической моделью стохастического эксперимента. При этом Ω — пространство элементарных событий — является математической моделью семейства исходов стохастического эксперимента; $\mathfrak{F}$ — σ-алгебра подмножеств пространства Ω — является математической моделью алгебры наблюдаемых событий стохастического эксперимента (множества из $\mathfrak{F}$ еще называют событиями); P — вероятность на $\mathfrak{F}$ — является математической моделью частоты $\nu_n(A)$ события A $(A \in \mathfrak{M})$ в последовательности экспериментов.

Замечание 1. События — это те подмножества множества исходов Ω, для которых можно говорить о их вероятностях.

Замечание 2. Значения вероятности на одноточечных множествах $\{\omega\}$ будем обозначать через $\mathsf{P}(\omega)$, т. е. $\mathsf{P}(\{\omega\}) = \mathsf{P}(\omega)$.

Пример. *Бросают две различимые симметричные монеты. Предложить математическую модель этого стохастического эксперимента.*

В качестве пространства элементарных событий естественно рассмотреть $\Omega = \{\text{ГГ}, \text{ГР}, \text{РГ}, \text{РР}\}$, а в качестве σ-алгебры $\mathfrak{F}$ наблюдаемых событий — класс всех подмножеств Ω. Вероятность на $\mathfrak{F}$ зададим так. Сначала из соображений симметрии зададим вероятности элементарных событий: $P(\text{ГГ}) = 1/4$, $P(\text{ГР}) = 1/4$, $P(\text{РГ}) = 1/4$, $P(\text{РР}) = 1/4$. Для любого другого $A \in \mathfrak{F}$ (любого подмножества Ω) определим $P(A)$ как сумму вероятностей элементарных событий, входящих в A. Легко проверить, что определенная таким образом функция множества удовлетворяет всем аксиомам вероятности.

Тем самым математическая модель стохастического эксперимента — вероятностное пространство — построена.

О двойственной терминологии в теории вероятностей. В связи с приведенным выше построением теоретико-множественной модели стохастического эксперимента в теории вероятностей сложилась двойственная терминология — теоретико-множественная и теоретико-вероятностная. Основой словаря, связывающего эти две терминологии, является табл. 2.3.1.

Таблица 2.3.1. Словарь

Язык теории множеств	Язык теории вероятностей
Пространство Ω	Достоверное событие
ω — точка из Ω	Исход стохастического эксперимента
$\emptyset$ — пустое множество	Невозможное событие
A — подмножество Ω, $A \in \mathfrak{F}$	A — событие
Множество A является подмножеством B $(A, B \in \mathfrak{F})$	Событие A влечет событие B
C — объединение множеств A и B из $\mathfrak{F}$	C — сумма событий A и B
C — пересечение множеств A и B из $\mathfrak{F}$	C — произведение событий A и B
$\overline{A}$ — дополнение A $(A \in \mathfrak{F})$	$\overline{A}$ — событие, противоположное к A
C — разность множеств A и B $(A, B \in \mathfrak{F})$	C — разность событий A и B
A и B не пересекаются $(A, B \in \mathfrak{F})$	События A и B несовместны

О приложениях. Подчеркнем следующий важный факт. Допустим, что мы хотим применить теорию вероятностей для вычисления (определения) вероятности некоторого физического события. Тогда расчет, основанный на модели $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, будет верным, если адекватно (правильно) определены пространство элементарных событий Ω, σ-алгебра событий $\mathfrak{F}$ и вероятность на $\mathfrak{F}$. Задача выбора модели $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ не может быть решена в рамках теории вероятностей как математической науки (например, невозможно доказать, что вероятность выпадения "герба" при подбрасывании симметричной монеты равна $1/2$ — это эмпирический факт). Но когда выбор сделан, можно, по меньшей мере в принципе, найти вероятность любого события из σ-алгебры $\mathfrak{F}$.

Сделать заключение об адекватности выбора математической модели $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ стохастическому эксперименту, т. е. о правильности выбора множества элементарных событий Ω, σ-алгебры наблюдаемых событий $\mathfrak{F}$ и вероятности P на σ-алгебре $\mathfrak{F}$ можно, сравнивая выводы, полученные из модели $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, с результатами стохастического эксперимента. Такое сравнение можно сделать, в частности, используя методы математической статистики.

2.4 Примеры и задачи

Примеры

Пример 2.4.1. *В лифте 7 пассажиров, лифт останавливается на 10 этажах. Для каждого пассажира фиксируется номер этажа, на котором он вышел. Построить пространство элементарных событий Ω. Описать как подмножество Ω событие A — "все пассажиры вышли на разных этажах".*

Решение. Каждому пассажиру поставим в соответствие номер этажа, на котором он вышел. Пространство Ω состоит из последовательностей длиной 7, составленных из 10 чисел (номеров этажей, на которых останавливается лифт); событие A описывается последовательностями, состоящими из различных чисел.

Пример 2.4.2. *В урне находится один шар, о котором известно, что он или белый, или черный. В урну положили белый шар, а потом после тщательного перемешивания взяли один за другим оба шара. Предложить пространство элементарных событий Ω этого стохастического эксперимента. Описать как подмножества Ω события: A_1 — "первым выбран белый шар", A_2 — "вторым выбран белый шар".*

Решение. Можно предложить по меньшей мере два пространства элементарных событий. Если обозначить белый шар через W, черный — через B, то пространством элементарных событий стохастического эксперимента будет

$$\Omega = \{WW, WB, BW\}.$$

Если белый шар, который кладется в урну, пометить, например, звездочкой и обозначить через W^*, то пространством элементарных событий будет

$$\Omega^* = \{WW^*, W^*W, BW^*, W^*B\}.$$

Событие A_1 как подмножество Ω опишется так: $\{WW, WB\}$, а как подмножество Ω^* — так: $\{WW^*, W^*W, W^*B\}$. Событие A_2 как подмножество Ω опишется следующим образом: $\{WW, BW\}$, а как подмножество Ω^* — так: $\{WW^*, W^*W, BW^*\}$.

Задачи

2.1. Игральную кость подбрасывают дважды. Построить пространство элементарных событий Ω. Описать как подмножества Ω события: A — "сумма выпавших очков равна 8"; B — "хотя бы один раз выпадет 6"; C — "при первом подбрасывании выпадет четное число очков”; D — “при обоих подбрасываниях выпадет нечетное число очков”.

2.2. Подбрасывают монету, а после этого игральную кость. Построить пространство элементарных событий Ω. Описать как подмножества Ω события: A — “выпал герб”; B — “выпала цифра 5”.

2.3. Из цифр $1, 2, 3, 4, 5$ сначала выбирают одну, а потом из четырех оставшихся — другую. Построить пространство элементарных событий Ω. Описать как подмножества Ω события: A — “на первом шаге выбрана четная цифра”, B — “на втором шаге выбрана четная цифра”.

2.4. Из партии, содержащей N изделий, среди которых M бракованных, наудачу выбирают n изделий ($n \leq N-M, n \leq M$). Построить пространство элементарных событий Ω. Описать как подмножество Ω событие A — “среди выбранных изделий имеется ровно m бракованных”.

2.5. Игральную кость последовательно подбрасывают n раз. Предложить пространство элементарных событий Ω. Описать как подмножество Ω событие A — “выпадет n_1 единиц, n_2 двоек, $\ldots$, n_6 шестерок” $(n_1 + n_2 + \ldots + n_6 = n)$.

2.6. В наудачу выбранной группе, состоящей из r ($r \leq 12$) студентов, интересуемся их месяцами рождения. Предложить пространство элементарных событий Ω. Описать как подмножества Ω события: A — “хотя бы два студента родились в одном месяце”, B — “только один студент родился в сентябре”, C — “хотя бы один студент родился в сентябре”, D — “ни один студент не родился в сентябре”.

2.7. Подбрасывают три игральные кости. Предложить пространство элементарных событий Ω. Описать как подмножества Ω события: A — “единица выпадет только на одной игральной кости”, B — “шестерка выпадет только на двух костях”, C — “на всех костях выпадут разные грани”, D — “на костях выпадет одинаковое число очков”.

2.9. Каждая из n различимых частиц попадает в одну из m ячеек. Предложить пространство элементарных событий Ω. Описать как подмножество Ω событие A — “в первую ячейку попадет k_1 частиц, во вторую — k_2, и т. д., в m-ю — k_m”.

Решение. Каждой частице поставим в соответствие номер ячейки, в которой она оказалась. Ω — множество всех последовательностей длиной n, образованных числами $1, 2, \ldots, m$. Событие A описывается последовательностями "длиной-n, составленными из k_1 единиц, k_2 двоек, $\ldots$, k_m чисел m.

Глава 3

Свойства вероятности

3.1 Следствия из определения вероятности

Приводимые далее свойства вероятности непосредственно следуют из определения вероятности как неотрицательной счетно-аддитивной нормированной функции множества.

Теорема. *Вероятность невозможного события равна нулю:*

$$\mathsf{P}(\emptyset) = 0.$$

Д о к а з а т е л ь с т в о. Из равенства $\Omega = \Omega \cup \emptyset \cup \emptyset \cup \ldots$, в силу счетной аддитивности вероятности имеем

$$\mathsf{P}(\Omega) = \mathsf{P}(\Omega \cup \emptyset \cup \emptyset \cup \ldots) = \mathsf{P}(\Omega) + \mathsf{P}(\emptyset) + \mathsf{P}(\emptyset) + \ldots$$

Поскольку $\mathsf{P}(\Omega) = 1$ (свойство нормированности вероятности), то ряд в правой части сходится абсолютно, что возможно только когда $\mathsf{P}(\emptyset) = 0$.

Следствие 1. *Вероятность является конечно-аддитивной функцией множества:* если $A_i \cap A_j = \emptyset, i \neq j$, $i, j = 1, 2, \ldots, n$, то

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i).$$

Утверждение достаточно установить для двух событий. Если $A \cap B = \emptyset$, то из очевидного равенства

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \ldots,$$

счетной аддитивности вероятности и равенства $\mathsf{P}(\emptyset) = 0$ получаем

$$\mathsf{P}(A \cup B) = \mathsf{P}(A \cup B \cup \emptyset \cup \emptyset \cup \ldots) =$$

$$= \mathsf{P}(A) + \mathsf{P}(B) + \mathsf{P}(\emptyset) + \mathsf{P}(\emptyset) + \ldots = \mathsf{P}(A) + \mathsf{P}(B).$$

Следствие 2 (ограниченность вероятности). *Вероятность любого события A не превосходит* 1:

$$\mathsf{P}(A) \leq 1.$$

Действительно, из равенства $\Omega = A \cup \overline{A}$ и аддитивности вероятности имеем $1 = \mathsf{P}(A) + \mathsf{P}(\overline{A})$. А так как вероятность неотрицательна, то

$$\mathsf{P}(A) \leq 1.$$

Теорема (о вероятности разности событий). *Если $A \subset B$ $(A, B \in \mathfrak{F})$, то*

$$\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A)$$

(*вероятность разности событий B и A $(A \subset B)$ равна разности их вероятностей*).

Доказательство. Так как $A \subset B$, то

$$B = A \cup (B \setminus A),$$

причем $A \cap (B \setminus A) = \emptyset$. Поэтому в силу аддитивности вероятности имеем

$$\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A).$$

Отсюда

$$\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A).$$

Следствие 1 (вероятность противоположного события). *Вероятность события $\overline{A}$, противоположного событию A, равна* $1 - \mathsf{P}(A)$:

$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A).$$

Достаточно заметить, что $\overline{A} = \Omega \setminus A$.

Следствие 2 (монотонность вероятности). *Если $A \subset B$, то*

$$\mathsf{P}(A) \leq \mathsf{P}(B).$$

Так как вероятность неотрицательна, то

$$\mathsf{P}(B) - \mathsf{P}(A) = \mathsf{P}(B \setminus A) \geq 0.$$

Поэтому
$$\mathsf{P}(A) \le \mathsf{P}(B).$$

Теорема (сложения). *Для любых событий A и B*
$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

Доказательство. Представим множество $A \cup B$ в виде
$$A \cup B = A \cup (B \setminus (A \cap B)),$$
при этом $A \cap (B \setminus (A \cap B)) = \emptyset$. Отсюда, в силу аддитивности вероятности и теоремы о вероятности разности событий, получаем
$$\mathsf{P}(A \cup B) = \mathsf{P}(A \cup (B \setminus (A \cap B))) =$$
$$= \mathsf{P}(A) + \mathsf{P}(B \setminus (A \cap B)) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

Из теоремы сложения (последовательно применяя ее) получаем.

Следствие 1.
$$\mathsf{P}\left(A_1 \cup A_2 \cup \ldots \cup A_{n-1} \cup A_n\right) =$$
$$= \sum_{1 \le i_1 \le n} \mathsf{P}\left(A_{i_1}\right) - \sum_{1 \le i_1 < i_2 \le n} \mathsf{P}\left(A_{i_1} \cap A_{i_2}\right) + \ldots +$$
$$+ (-1)^{k-1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mathsf{P}\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right) + \ldots +$$
$$+ (-1)^{n-1} \sum_{1 \le i_1 < i_2 < \ldots < i_n \le n} \mathsf{P}\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_n}\right).$$

Следствие 2 (свойство полуаддитивности вероятности). *Для любых событий A_i, $i = 1, 2, \ldots, n$*
$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathsf{P}(A_i).$$

Доказательство. Для двух событий A_1 и A_2 из равенства
$$\mathsf{P}(A_1 \cup A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1 \cap A_2)$$
имеем
$$\mathsf{P}(A_1 \cup A_2) \le \mathsf{P}(A_1) + \mathsf{P}(A_2)$$
Из последнего неравенства для любого n имеем
$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathsf{P}(A_i).$$

3.2 Непрерывность вероятности

Конечно-аддитивная функция множества, вообще говоря, не является счетно-аддитивной (это иллюстрирует приводимый далее пример). В связи с этим возникает вопрос: какие условия необходимо наложить на конечно-аддитивную функцию, чтобы она стала счетно-аддитивной?

Пример конечно-аддитивной, но не счетно-аддитивной функции. Пусть Ω — множество рациональных чисел промежутка $[0; 1]$. Для каждой пары чисел $a, b \in \Omega$ $(a \leq b)$ через $A_{a,b}$ обозначим множество рациональных чисел промежутка с концами a и b, в частности, может быть и $a = b$ (концы могут как принадлежать, так и не принадлежать промежутку). Конечные объединения непересекающихся множеств $A_{a,b}$ образуют алгебру (обозначим ее через $\mathfrak{A}$).

На алгебре $\mathfrak{A}$ определим функцию множества P. Для множества $A_{a,b}$ значение

$$\mathsf{P}\left(A_{a,b}\right) = b - a.$$

Значение P на множестве

$$B = \bigcup_{i=1}^{n} A_{a_i,b_i}, \; A_{a_i,b_i} \cap A_{a_j,b_j} = \emptyset, i \neq j,$$

определим равенством

$$\mathsf{P}(B) = \mathsf{P}\left(\bigcup_{i=1}^{n} A_{a_i,b_i}\right) = \sum_{i=1}^{n} \mathsf{P}\left(A_{a_i,b_i}\right) = \sum_{i=1}^{n} (b_i - a_i).$$

Так определенная на $\mathfrak{A}$ функция множества P конечно-аддитивна (в этом убеждаемся непосредственной проверкой), но не счетно-аддитивна.

Счетное множество Ω рациональных чисел промежутка $[0; 1]$ можно представить в виде объединения счетного числа одноточечных множеств:

$$\Omega = \bigcup_{i=1}^{\infty} \{r_i\}.$$

Если предположить, что P счетно-аддитивна, то

$$\mathsf{P}(\Omega) = \mathsf{P}\left(\bigcup_{i=1}^{\infty} \{r_i\}\right) = \sum_{i=1}^{\infty} \mathsf{P}(\{r_i\}).$$

Но, тогда, учитывая, что $\mathsf{P}(A_{a,b}) = b - a$, с одной стороны

$$\mathsf{P}(\Omega) = \sum_{i=1}^{\infty} \mathsf{P}(\{r_i\}) = \sum_{i=1}^{\infty} (r_i - r_i) = 0.$$

а с другой,

$$\mathsf{P}(\Omega) = \mathsf{P}(A_{0,1}) = 1 - 0 = 1$$

Так что предположение о счетной аддитивности P приводит к противоречию: $1 = 0$.

Приводимое далее свойство непрерывности конечно-аддитивной функции множества эквивалентно свойству счетной аддитивности.

Определение. Последовательность множеств $\{A_i\}$ будем называть *возрастающей*, если

$$A_i \subset A_{i+1}, \; i = 1, 2, \ldots$$

Последовательность множеств $\{A_i\}$ будем называть *убывающей*, если

$$A_{i+1} \subset A_i, \;\; i = 1, 2, \ldots$$

Теорема 3.2.1 (непрерывность вероятности). *Пусть* P *— неотрицательная конечно-аддитивная нормированная функция множества, заданная на σ-алгебре* $\mathfrak{F}$ *подмножеств пространства* Ω*. Следующие четыре утверждения являются эквивалентными:*

1° P *— счетно-аддитивна;*

2° *для любой возрастающей последовательности* $\{A_i\}$ $(A_i \in \mathfrak{F})$

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_n \mathsf{P}(A_n);$$

3° *для любой убывающей последовательности* $\{A_i\}$ $(A_i \in \mathfrak{F})$

$$\mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_n \mathsf{P}(A_n);$$

4° *для любой убывающей последовательности* $\{A_i\}$ $(A_i \in \mathfrak{F})$ *такой, что* $\bigcap\limits_{i=1}^{\infty} A_i = \emptyset$

$$\lim_n \mathsf{P}(A_n) = 0.$$

Доказательство. Достаточно доказать, что $1^\circ \Rightarrow 2^\circ \Rightarrow \Rightarrow 3^\circ \Rightarrow 4^\circ \Rightarrow 1^\circ$.

Покажем, что из $1^\circ \Rightarrow 2^\circ$.

Пусть P — счетноаддитивная функция и $\{A_i\}$ — возрастающая последовательность множеств. Очевидно,

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty}(A_i \setminus A_{i-1}), \text{ где } \quad A_0 = \emptyset,$$

причем

$$(A_k \setminus A_{k-1}) \cap (A_j \setminus A_{j-1}) = \emptyset, \ k \neq j.$$

В силу счетной аддитивности P

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathsf{P}\left(\bigcup_{i=1}^{\infty}(A_i \setminus A_{i-1})\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i \setminus A_{i-1}) =$$

$$= \lim_n \sum_{i=1}^{n} \mathsf{P}(A_i \setminus A_{i-1}) = \lim_n \sum_{i=1}^{n}(\mathsf{P}(A_i) - \mathsf{P}(A_{i-1})) = \lim_n \mathsf{P}(A_n).$$

Установим, что из $2^\circ \Rightarrow 3^\circ$.

Пусть имеет место утверждение 2°. Если $\{A_i\}$ — убывающая последовательность, то $\{\overline{A_i}\}$ — возрастающая последовательность множеств и поэтому для $\{\overline{A_i}\}$ (в силу 2°) имеем

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} \overline{A_n}\right) = \lim_n \mathsf{P}(\overline{A_n}).$$

А так как

$$\lim_n \mathsf{P}(\overline{A_n}) = \lim_n(1 - \mathsf{P}(A_n)) = 1 - \lim_n \mathsf{P}(A_n),$$

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} \overline{A_n}\right) = \mathsf{P}\left(\overline{\bigcap_{n=1}^{\infty} A_n}\right) = 1 - \mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right),$$

то

$$\mathsf{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_n \mathsf{P}(A_n).$$

Из $3^\circ \Rightarrow 4^\circ$, поскольку утверждение 4° — частный случай утверждения 3°.

Покажем, что из $4^\circ \Rightarrow 1^\circ$.

Пусть имеет место утверждение 4° и $\{A_i\}$ — последовательность непересекающихся множеств из $\mathfrak{F}$ ($A_i \cap A_j = \emptyset$, $i \neq j$). Поскольку P — конечно-аддитивная функция, то

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathsf{P}\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup \left(\bigcup_{i=n}^{\infty} A_i\right)\right) =$$

$$= \mathsf{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathsf{P}\left(\bigcup_{i=n}^{\infty} A_i\right) = \sum_{i=1}^{n-1} \mathsf{P}(A_i) + \mathsf{P}\left(\bigcup_{i=n}^{\infty} A_i\right).$$

Так что

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{n-1} \mathsf{P}(A_i) + \mathsf{P}\left(\bigcup_{i=n}^{\infty} A_i\right). \tag{3.2.1}$$

Множества

$$B_n = \bigcup_{i=n}^{\infty} A_i, \ n = 1, 2, \ldots,$$

образуют убывающую последовательность, причем $\bigcap\limits_{n=1}^{\infty} B_n = \emptyset$. Поэтому в силу предположения 4°

$$\lim_n \mathsf{P}\left(\bigcup_{i=n}^{\infty} A_i\right) = 0.$$

Для завершения доказательства достаточно в равенстве (3.2.1) перейти к пределу при $n \to \infty$.

Следствие. *Вероятность является счетно-полуаддитивной функцией множества: для любой последовательности событий A_i, $i = 1, 2, \ldots$*

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathsf{P}(A_i). \tag{3.2.2}$$

Доказательство. Для любого конечного n

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathsf{P}(A_i) \tag{3.2.3}$$

(см. следствие из теоремы сложения). Далее, события

$$B_n = \bigcup_{i=1}^{n} A_i, n = 1, 2, \ldots,$$

образуют монотонно возрастающую последовательность и

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{n=1}^{\infty} B_n.$$

Поэтому переходя в равенстве (3.2.3) к пределу при $n \to \infty$ и пользуясь свойством непрерывности вероятности получим (3.2.2).

Замечание. Теорема имеет место не только для нормированной функции множества ($\mathsf{P}(\Omega) = 1$), но и для ограниченной ($\mathsf{P}(\Omega) < \infty$).

3.3 Условная вероятность

Условная частота. Пусть A и B — наблюдаемые события стохастического эксперимента. Проведем стохастический эксперимент n раз, при этом $k_n(B)$ раз произошло событие B, и $k_n(AB)$ раз — событие A вместе с событием B. Частоту события A в последовательности тех экспериментов, в которых произошло событие B:

$$k_n(AB)/k_n(B)$$

будем называть *условной частотой* события A относительно события B в последовательности n экспериментов и будем обозначать $\nu_n(A/B)$. Так что по определению

$$\nu_n(A/B) = \frac{k_n(AB)}{k_n(B)}.$$

При фиксированном B условная частота $\nu_n(A/B)$, как функция A, обладает всеми свойствами частоты:

1° для каждого события A

$$\nu_n(A/B) \geq 0$$

(условная частота неотрицательна);

2° для несовместных событий A_1 и A_2

$$\nu_n((A_1 + A_2)/B) = \nu_n(A_1/B) + \nu_n(A_2/B)$$

(условная частота аддитивна);

3° если U — достоверное событие, то

$$\nu_n(U/B) = 1$$

(условная частота нормирована).

Так же, как частоте $\nu_n(A)$ события A в последовательности экспериментов соответствует вероятность $\mathsf{P}(A)$, являющаяся математической моделью частоты $\nu_n(A)$, частоте $\nu_n(A/B)$ соответствует вероятность $\mathsf{P}(A/B)$, являющаяся математической моделью условной частоты $\nu_n(A/B)$. Вероятность $\mathsf{P}(A/B)$ будем называть условной вероятностью события A относительно события B.

Очевидно,

$$\nu_n(A/B) = \frac{k_n(AB)}{k_n(B)} = \frac{k_n(AB)/n}{k_n(B)/n} = \frac{\nu_n(AB)}{\nu_n(B)}.$$

Поэтому, определяя условную вероятность $\mathsf{P}(A/B)$ как математическую модель условной частоты $\nu_n(A/B)$ события A относительно B, естественно потребовать, чтобы условная и безусловная вероятности были связаны соотношением

$$\mathsf{P}(A/B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

А если вероятность P уже задана, то на последнее равенство можно смотреть как на определение условной вероятности события A относительно события B.

Определение. Пусть $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ — вероятностное пространство, $B \in \mathfrak{F}$, $\mathsf{P}(B) > 0$. *Условной вероятностью* $\mathsf{P}(A/B)$ события A относительно события B будем называть отношение $\mathsf{P}(A \cap B)/\mathsf{P}(B)$, т. е.

$$\mathsf{P}(A/B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

Для условной вероятности $\mathsf{P}(A/B)$ события A относительно B также используют обозначение $\mathsf{P}_B(A)$.

Вероятность как количественная мера прогноза. В математической модели $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ стохастического эксперимента вероятность $p = \mathsf{P}(A)$ события A является математической моделью частоты $\nu_n(A)$ появления события A в последовательности независимых экспериментов. На вероятность $p = \mathsf{P}(A)$ события A часто смотрят как на количественную меру прогноза появления события A, в том смысле, что в достаточно "длинной" серии экспериментов частота $\nu_n(A)$ появления события A будет "близка" к p. Условную вероятность $p' = \mathsf{P}_B(A)$ события A относительно события B можно интерпретировать как количественную меру прогноза появления события A, если известно, что произошло событие B. В длинной серии экспериментов условная частота $\nu_n(A/B)$ события A, вычисленная по совокупности тех экспериментов, в которых произошло событие B, будет близка к p'.

Дальтонизм и пол (вероятностная интерпретация). Часть людей не различает цвета (как правило, красный и зеленый) — страдает дальтонизмом. Процент людей, страдающих дальтонизмом, составляет 2,625%. Известно, что дальтонизм связан с полом, а именно — процент дальтоников среди мужчин равен 5%, среди женщин — 0,25%. Этим данным можно дать следующую вероятностную интерпретацию. Рассматривается стохастический эксперимент: случайно выбирают человека и классифицируют его по полу (M — "мужчина", F — "женщина") и по отношению к дальтонизму (D — "дальтоник", $\overline{D}$ — "не дальтоник"). Тогда в длинной серии экспериментов частота события D "близка" к $0,02625$. Частота события D, если произошло событие M — частота дальтоников, подсчитанная среди мужчин — "близка" к 0,05, а частота события D, если произошло событие событие F — частота дальтоников, подсчитанная среди женщин — "близка" к 0,0025. Эти числа естественно рассматривать в качестве вероятностей (и количественной меры прогноза) соответствующих событий в эксперименте, состоящем в случайном выборе человека и классификации его по полу и по отношению к дальтонизму:

$$\mathsf{P}(D) = 0{,}02625;\ \mathsf{P}(D/M) = 0{,}05;\ \mathsf{P}(D/F) = 0{,}0025.$$

Как видно, прогноз события D — "дальтоник" существенно меняется в зависимости от дополнительной информации вида "произошло событие M", "произошло событие F".

Пространство $\{B, \mathfrak{F}_B, \mathsf{P}_B\}$. Пусть $\mathfrak{F}$ — σ-алгебра подмножеств Ω, $B \in \mathfrak{F}$. Обозначим через $\mathfrak{F}_B$ класс подмножеств пространства Ω вида $B \cap A$, $A \in \mathfrak{F}$ ($\mathfrak{F}_B$ состоит из подмножеств σ-алгебры $\mathfrak{F}$ содержащихся в B). Легко проверить, что класс $\mathfrak{F}_B$ является σ-алгеброй подмножеств множества B.

Теорема 3.3.1 (о пространстве $\{B, \mathfrak{F}_B, \mathsf{P}_B\}$). *Пусть $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ — вероятностное пространство, $B \in \mathfrak{F}$ — произвольное, но фиксированное множество и $\mathsf{P}(B) > 0$. Функция множества $\mathsf{P}_B(\cdot)$, заданная на $\mathfrak{F}$ равенством*

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)},$$

является вероятностью на $\mathfrak{F}$.

Тройка $\{B, \mathfrak{F}_B, \mathsf{P}_B\}$ является вероятностным пространством.

Доказательство. Проверим, что P_B удовлетворяет трем аксиомам вероятности.

Очевидно, P_B — неотрицательная и нормированная.

Убедимся, что P_B — счетно-аддитивная. Пусть последовательность множеств $\{A_i\}$ из $\mathfrak{F}$ такая, что $A_i \cap A_j = \emptyset$ при $i \neq j$, тогда

$$\mathsf{P}_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{\mathsf{P}\left(\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \cap B\right)}{\mathsf{P}\,(B)} = \frac{\mathsf{P}\left(\bigcup\limits_{i=1}^{\infty} (A_i \cap B)\right)}{\mathsf{P}\,(B)} =$$

$$= \sum_{i=1}^{\infty} \frac{\mathsf{P}(A_i \cap B)}{\mathsf{P}(B)} = \sum_{i=1}^{\infty} \mathsf{P}_B(A_i).$$

Чтобы убедиться, что $\{B, \mathfrak{F}_B, \mathsf{P}_B\}$ является вероятностным пространством, достаточно заметить, что P_B является неотрицательной счетно-аддитивной нормированной функцией на $\mathfrak{F}_B$.

Теорема (формула умножения). *Для любых событий A и $B \in \mathfrak{F}$ таких, что* $\mathsf{P}(A) > 0, \mathsf{P}(B) > 0$,

$$\mathsf{P}(A \cap B) = \mathsf{P}(A/B)\mathsf{P}(B) = \mathsf{P}(B/A)\mathsf{P}(A).$$

Это очевидным образом следует из равенств

$$\mathsf{P}(A/B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}, \quad \mathsf{P}(B/A) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(A)}.$$

Теорема допускает обобщение. *Пусть* $A_1, A_2, \ldots, A_n$ *— такие события, что* $\mathsf{P}(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$, *тогда*

$$\mathsf{P}(A_n \cap A_{n-1} \cap \ldots \cap A_2 \cap A_1) =$$

$$= \mathsf{P}(A_n/(A_{n-1} \cap \ldots \cap A_2 \cap A_1))\times$$

$$\times\mathsf{P}(A_{n-1}/(A_{n-2} \cap A_{n-3} \cap \ldots \cap A_2 \cap A_1)) \times \ldots \times \mathsf{P}(A_2/A_1)\mathsf{P}(A_1).$$

Формула полной вероятности. Будем говорить, что события $B_1, B_2, \ldots, B_n$ образуют *полную группу* событий, если, во-первых,

$$B_i \cap B_j = \emptyset,\ i \neq j.$$

и, во-вторых,

$$\bigcup_{i=1}^{n} B_i = \Omega.$$

Теорема (формула полной вероятности). *Пусть* B_i, $i = 1, 2, \ldots, n$, *— полная группа событий, причем* $\mathsf{P}(B_i) > 0$, $i = 1, 2, \ldots, n$. *Тогда для любого события* A

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(A/B_i)\mathsf{P}(B_i). \tag{3.3.1}$$

Доказательство. Воспользовавшись тем, что $\Omega = \bigcup\limits_{i=1}^{n} B_i$, представим событие A в виде

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} B_i\right) = \bigcup_{i=1}^{n} (A \cap B_i).$$

причем

$$(A \cap B_i) \cap (A \cap B_j) = \emptyset,\ \ i \neq j,$$

поскольку $B_i \cap B_j = \emptyset,\ i \neq j$. Поэтому

$$\mathsf{P}(A) = \mathsf{P}\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) = \sum_{i=1}^{n} \mathsf{P}(A \cap B_i) = \sum_{i=1}^{n} \mathsf{P}(A/B_i)\mathsf{P}(B_i).$$

З а м е ч а н и е. При вычислении $\mathsf{P}(A)$ зачастую проще вычислить $\mathsf{P}(A/B_i)$ и $\mathsf{P}(B_i)$, $i = 1, 2, \ldots, n$, для некоторой полной группы событий $\{B_i\}$, а затем воспользоваться формулой полной вероятности (3.3.1).

Формула полной вероятности справедлива и для счетной полной группы событий.

Теорема (формулы Байеса). *Пусть $B_i, i = 1, 2, \ldots, n$, — полная группа событий, причем $\mathsf{P}(B_i) > 0$, $i = 1, 2, \ldots, n$. Тогда для любого события A $(\mathsf{P}(A) > 0)$*

$$\mathsf{P}(B_i/A) = \frac{\mathsf{P}(A/B_i)\mathsf{P}(B_i)}{\sum\limits_{k=1}^{n} \mathsf{P}(A/B_k)\mathsf{P}(B_k)}, \; i = 1, 2, \ldots, n.$$

Д о к а з а т е л ь с т в о.

$$\mathsf{P}(B_i/A) = \frac{\mathsf{P}(B_i \cap A)}{\mathsf{P}(A)} = \frac{\mathsf{P}(A/B_i)\mathsf{P}(B_i)}{\sum\limits_{k=1}^{n} \mathsf{P}(A/B_k)\mathsf{P}(B_k)}, \; i = 1, 2, \ldots, n.$$

Пример 3.3.1. *На фоне шума на вход радиолокационного устройства с вероятностью p $(0 < p < 1)$ поступает сигнал. Если поступает сигнал (всегда на фоне шума), то устройство регистрирует наличие сигнала с вероятностью p_1, если только шум, то устройство регистрирует наличие сигнала с вероятностью p_2. Известно, что устройство зарегистрировало сигнал. Какова вероятность того, что на вход радиолокационного устройства поступил сигнал?*

Р е ш е н и е. Введем обозначения: A_c — на вход попал сигнал (с шумом), A — на вход попал только шум, B_c — устройство зарегистрировало наличие сигнала, B — устройство зарегистрировало шум. Необходимо вычислить $\mathsf{P}(A_c/B_c)$. Воспользуемся формулой Байеса. В качестве полной группы событий рассмотрим события A_c и A. Тогда

$$\mathsf{P}(A_c/B_c) = \frac{\mathsf{P}(B_c/A_c)\mathsf{P}(A_c)}{\mathsf{P}(B_c/A_c)\mathsf{P}(A_c) + \mathsf{P}(B_c/A)\mathsf{P}(A)}.$$

По условию задачи $\mathsf{P}(A_c) = p$, $\mathsf{P}(A) = 1 - p$, $\mathsf{P}(B_c/A_c) = p_1$, $\mathsf{P}(B_c/A) = p_2$. Следовательно, искомая вероятность

$$\mathsf{P}(A_c/B_c) = \frac{p_1 p}{p_1 p + p_2(1 - p)}.$$

3.4 Независимые события

Понятие независимости событий, случайных величин, случайных векторов играет в теории вероятностей исключительную роль. По сути, благодаря ему, теория вероятностей получила право на существование как самостоятельная наука. Сначала введем понятие независимости событий.

Независимые события мы связываем с независимыми стохастическими экспериментами. Под независимыми стохастическими экспериментами будем понимать эксперименты, для которых информация об исходе одного стохастического эксперимента не меняет прогноз (вероятность) появления событий в другом — если известно, что в одном эксперименте произошло событие A, то это никак не сказывается на прогнозе появления события B в другом стохастическом эксперименте. В противном случае эксперименты считаем зависимыми.

Примеры независимых и зависимых стохастических экспериментов. Стохастический эксперимент 1° состоит в подбрасывании пары монет, стохастический эксперимент 2° — игральной кости. Информация об исходе одного стохастического эксперимента не меняет прогноз (вероятность) появления событий в другом стохастическом эксперименте. Если известен исход эксперимента 1°, например, на обеих монетах выпал герб, то это никак не меняет вероятности наступления событий в стохастическом эксперименте 2° — вероятность выпадения шестерки остается равной $1/6$, четного числа очков — $1/2$, ...

В качестве примера зависимых стохастических экспериментов можно рассмотреть эксперименты с классификацией по полу и дальтонизму. Наудачу выбирается человек. Стохастический эксперимент 1° — классификация человека по полу (M — "мужчина", F — "женщина"); стохастический эксперимент 2° — классификация по отношению к дальтонизму (D — "дальтоник", $\overline{D}$ — "не дальтоник"). Известно, что вероятность $\mathsf{P}(D)$ человеку быть дальтоником равна 0,02625, вероятность $\mathsf{P}(D/M)$ мужчине быть дальтоником равна 0,05, а женщине — $\mathsf{P}(D/F) = 0{,}0025$. Информация об исходе стохастического эксперимента 1° меняет прогноз появления событий в стохастическом эксперименте 2°, а именно, если известно, что в эксперименте 1° произошло событие M, вероятность события D в эксперименте 2° увеличивается (с 0,02625 до 0,05), а если произошло событие F, вероятность события D уменьшается (с 0,02625 до 0,0025).

Заметим, что пару стохастических экспериментов можно рассматривать как один (составной) стохастический эксперимент, состоящий в проведении двух. Понятие независимости вводится для событий в одном и том же вероятностном пространстве — составного стохастического эксперимента (хотя об этом специально не упоминается).

Формализуем описанное понятие независимости событий. Количественно прогноз события A мы описываем вероятностью $\mathsf{P}(A)$, а если известно, что произошло событие B, — условной вероятностью $\mathsf{P}(A/B)$. Поэтому независимость прогноза появления события A от события B естественно описать равенством

$$\mathsf{P}(A/B) = \mathsf{P}(A),$$

а учитывая, что $\mathsf{P}(A/B) = \mathsf{P}(A \cap B)/\mathsf{P}(B)$, это можно записать в симметричной форме:

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B).$$

Изложенные соображения приводят к следующему определению.

Определение. Пусть $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ — вероятностное пространство. События $A, B \in \mathfrak{F}$ будем называть *независимыми*, если

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B).$$

З а м е ч а н и е. Независимость событий связана с независимостью стохастических экспериментов, но явно в определении независимых событий стохастические эксперименты не участвуют.

Теорема. *События A и B ($\mathsf{P}(B) > 0$) независимы тогда и только тогда, когда*

$$\mathsf{P}(A/B) = \mathsf{P}(A).$$

Д о к а з а т е л ь с т в о. Если события A и B независимы, то

$$\mathsf{P}(A/B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A)\mathsf{P}(B)}{\mathsf{P}(B)} = \mathsf{P}(A).$$

Если $\mathsf{P}(A/B) = \mathsf{P}(A)$, то

$$\mathsf{P}(A \cap B) = \mathsf{P}(A/B)\mathsf{P}(B) = \mathsf{P}(A)\mathsf{P}(B).$$

Последнее как раз и означает, что события A и B независимы.

Теорема. *Если события A и B независимы, то события A и $\overline{B}$ независимы.*

Доказательство.

$$\mathsf{P}(A \cap \overline{B}) = \mathsf{P}\left(A \cap (\Omega \backslash B)\right) =$$
$$= \mathsf{P}(A \backslash (A \cap B)) = \mathsf{P}(A) - \mathsf{P}(A \cap B) =$$
$$= \mathsf{P}(A) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(\overline{B}),$$

т. е. события A и $\overline{B}$ независимы.

Следствие. *Если события A и B независимы, то события $\overline{A}$ и $\overline{B}$ независимы.*

Определение. События $A_1, A_2, \ldots, A_n$ будем называть *независимыми в совокупности* (*независимыми*), если для любого набора $\{i_1, i_2, \ldots, i_k\}$, $k = 2, 3, \ldots, n$, различных индексов из $\{1, 2, \ldots, n\}$

$$\mathsf{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathsf{P}(A_{i_1})\mathsf{P}(A_{i_2}) \ldots \mathsf{P}(A_{i_k}).$$

Определение. События $A_1, A_2, \ldots, A_n$ будем называть *попарно независимыми,* если для каждой пары s, k $(s \neq k)$ события A_s и A_k независимы.

Замечание. События, независимые в совокупности, являются попарно независимыми, но наоборот, вообще говоря, нет (см. пример С. Н. Бернштейна (пример 4.4.2)).

Пример 3.4.1. *Работает m радиолокационных станций, каждая из которых за один цикл осмотра обнаруживает объект с вероятностью p (независимо от других циклов и станций). За время T каждая станция успевает сделать n циклов осмотра. Найти вероятность следующих событий: A — “за время T объект будет обнаружен” (хотя бы одной станцией), B — “за время T объект будет обнаружен каждой станцией”.*

Решение. Обозначим через A_{ji} событие, состоящее в том, что j-я станция обнаруживает объект в i-м цикле, $j = 1, 2, \ldots, m$; $i = 1, 2, \ldots, n$; A_j — объект обнаружен j-й станцией, $j = 1, 2, \ldots$ $\ldots, m$. Очевидно,

$$A_j = \bigcup_{i=1}^{n} A_{ji}; \;\; A = \bigcup_{j=1}^{m} A_j = \bigcup_{j=1}^{m} \bigcup_{i=1}^{n} A_{ji};$$

$$B = \bigcap_{j=1}^{m} A_j = \bigcap_{j=1}^{m} \left(\bigcup_{i=1}^{n} A_{ji} \right).$$

Отсюда, учитывая независимость событий $A_{ji}, j = 1, 2, \ldots, m$; $i = 1, 2, \ldots, n$, имеем:

$$\mathsf{P}(A) = 1 - \mathsf{P}(\overline{A}) = 1 - \mathsf{P}\left(\overline{\bigcup_{j=1}^{m}\bigcup_{i=1}^{n} A_{ji}}\right) =$$

$$= 1 - \mathsf{P}\left(\bigcap_{j=1}^{m}\bigcap_{i=1}^{n} \overline{A}_{ji}\right) = 1 - (1-p)^{nm};$$

$$\mathsf{P}(B) = \mathsf{P}\left(\bigcap_{j=1}^{m}\left(\bigcup_{i=1}^{n} A_{ji}\right)\right) = \prod_{j=1}^{m}\mathsf{P}\left(\bigcup_{i=1}^{n} A_{ji}\right) =$$

$$= \prod_{j=1}^{m}\left(1 - \mathsf{P}\left(\overline{\bigcup_{i=1}^{n} A_{ji}}\right)\right) = \prod_{j=1}^{m}\left(1 - \mathsf{P}\left(\bigcap_{i=1}^{n} \overline{A}_{ji}\right)\right) =$$

$$= \prod_{j=1}^{m}(1 - (1-p)^n) = (1 - (1-p)^n)^m.$$

3.5 Примеры и задачи

Примеры

Пример 3.5.1. *Пусть* $\{A_n\}$ — *последовательность непересекающихся множеств* $(A_i \bigcap A_j = \emptyset, i \neq j)$ *и*

$$B_n = \bigcup_{i=n}^{\infty} A_i, \ n = 1, 2, \ldots$$

Доказать, что $\bigcap\limits_{n=1}^{\infty} B_n = \emptyset$.

Р е ш е н и е. Предположим, что $\bigcap\limits_{n=1}^{\infty} B_n \neq \emptyset$, и пусть $\omega \in \bigcap\limits_{n=1}^{\infty} B_n$. Тогда $\omega \in B_1 = \bigcup\limits_{i=1}^{\infty} A_i$ и, следовательно, ω принадлежит хотя бы одному из множеств A_i, $i = 1, 2, \ldots$, а поскольку $A_i \cap A_j = \emptyset$, $i \neq j$, то только одному множеству, обозначим его через A_{n_0}. Поэтому

$$\omega \notin \bigcup_{i=n_0+1}^{\infty} A_i = B_{n_0+1}.$$

А последнее противоречит предположению $\omega \in \bigcap\limits_{n=1}^{\infty} B_n$.

Задачи

3.1. Пусть $\{A_n\}$ — последовательность множеств и

$$B_1 = A_1,\ B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i, n = 1, 2, \ldots$$

Доказать, что множества B_i попарно не пересекаются и

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

3.2. Радиолокационная станция ведет наблюдение за космическим объектом, который может создавать помехи. Если объект не создает помехи, то за один цикл осмотра станция обнаруживает его с вероятностью p_0, если создает — с вероятностью p_1 ($p_1 < p_0$). Вероятность создания помех на протяжении цикла осмотра равна p и не зависит от того, как и когда создавались помехи в других циклах. Найти вероятность обнаружения объекта хотя бы один раз на протяжении n циклов осмотра.

3.3. При рентгеновском обследовании вероятность обнаружить заболевание у больного туберкулезом равна $1-\beta$. Вероятность принять здорового человека за больного равна α. Пусть доля больных туберкулезом по отношению ко всему населению составляет γ.

Найти вероятность того, что человек здоров, если он был признан больным при обследовании.

Указание. Рассмотрим события: S — “человек болен”, $\overline{S}$ — “человек здоров”, R_S — “рентгеновский анализ положителен”, $R_{\overline{S}}$ — “рентгеновский анализ отрицателен”. Согласно формуле Байеса искомая вероятность

$$\mathsf{P}(\overline{S}/R_S) = \alpha(1-\gamma)/(\alpha(1-\gamma) + (1-\beta)\gamma).$$

Глава 4

Дискретное вероятностное пространство

4.1 Построение дискретного вероятностного пространства

Математической моделью стохастического эксперимента является вероятностное пространство $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, где Ω — пространство элементарных событий (математическая модель исходов стохастического эксперимента); $\mathfrak{F}$ — σ-алгебра подмножеств пространства Ω (модель алгебры наблюдаемых событий); P — вероятность (модель частоты $\nu_n(A)$ события A в последовательности экспериментов). Чаще всего математическая модель стохастического эксперимента строится по эмпирическим данным — обычно для некоторого класса наблюдаемых событий (в каждом эксперименте такой класс свой) известны их частоты. У нас, как правило, не будет трудностей с выбором Ω и σ-алгебры $\mathfrak{F}$ на Ω, чего нельзя сказать о задании вероятности — неотрицательной счетно-аддитивной нормированной функции. Но если Ω конечно или счетно (кратко будем говорить — *дискретно*), вероятность задать просто — достаточно задать вероятности элементарных событий.

Теорема 4.1.1 (вероятность на дискретном $\boldsymbol{\Omega}$). *Пусть Ω — дискретно и P — неотрицательная функция точки на Ω такая, что*

$$\sum_{\omega\in\Omega} P(\omega) = 1. \tag{4.1.1}$$

Тогда функция множества

$$\mathsf{P}(A) = \sum_{\omega \in A} P(\omega), \tag{4.1.2}$$

заданная на σ-алгебре всех подмножеств Ω, является вероятностью.

Доказательство. Прежде всего заметим, что определение (4.1.2) функции множества корректно: в силу абсолютной сходимости ряда (4.1.1) для каждого $A \subset \Omega$ абсолютно сходится и ряд $\sum\limits_{\omega \in A} P(\omega)$.

Убедимся, что $\mathsf{P}(\cdot)$ является вероятностью на σ-алгебре всех подмножеств Ω, т. е. является неотрицательной, счетно-аддитивной, нормированной функцией множества.

Неотрицательность $\mathsf{P}(A)$, $A \subset \Omega$, следует из равенства (4.1.2) и неравенств $P(\omega) \geq 0$, $\omega \in \Omega$.

Для любой счетной последовательности непересекающихся множеств A_i, учитывая свойство ассоциативности абсолютно сходящихся рядов, имеем

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{\omega \in \bigcup\limits_i A_i} P(\omega) = \sum_{i=1}^{\infty}\left(\sum_{\omega \in A_i} P(\omega)\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i),$$

т. е. $\mathsf{P}(\cdot)$ счетно-аддитивна.

Далее, согласно (4.1.2) для каждого $A \subset \Omega$

$$\mathsf{P}(A) = \sum_{\omega \in A} P(\omega),$$

в частности, и для $A = \Omega$

$$\mathsf{P}(\Omega) = \sum_{\omega \in \Omega} P(\omega),$$

поэтому учитывая (4.1.1) имеем

$$\mathsf{P}(\Omega) = 1,$$

т. е. $\mathsf{P}(\cdot)$ — нормированная функция множества.

Вероятность, задаваемая равенством

$$\mathsf{P}(A) = \sum_{\omega \in A} P(\omega)$$

на σ-алгебре всех подмножеств дискретного Ω (на самой "богатой" σ-алгебре), оказывается определенной и на любой другой σ-алгебре $\mathfrak{S}$ пространства Ω (поскольку каждая σ-алгебра подмножеств пространства Ω содержится в σ-алгебре всех подмножеств множества Ω). Поэтому, если Ω дискретно, нет необходимости выделять в пространстве Ω σ-алгебру, на которой задается вероятность. В связи с этим вероятностное пространство, когда Ω дискретно, обозначают парой $\{\Omega, \mathsf{P}\}$.

Определение. Если Ω дискретно, то вероятностное пространство $\{\Omega, \mathsf{P}\}$ будем называть *дискретным.*

Вероятность P, заданная на дискретном Ω (на множестве всех подмножеств дискретного Ω) называют *дискретной* (*атомической*) вероятностью.

Вероятность $\mathsf{P}(\cdot)$, заданная на σ-алгебре всех подмножеств дискретного пространства Ω посредством функции точки $P(\omega)$:

$$\mathsf{P}(A) = \sum_{\omega \in A} P(\omega),$$

определена, в частности, и на одноточечных подмножествах $\{\omega\}$ (элементарных событиях):

$$\mathsf{P}(\{\omega\}) = \sum_{\omega \in \{\omega\}} P(\omega) = P(\omega),$$

и ее значение на одноточечном множестве $\{\omega\}$ совпадает со значением $P(\omega)$ функции P в точке ω. Поэтому на значение $P(\omega)$ можно смотреть как на вероятность $\mathsf{P}(\{\omega\})$ элементарного события ω, при этом $\mathsf{P}(\{\omega\})$ будем обозначать через $\mathsf{P}(\omega)$.

Теорема 4.1.1, по существу, утверждает, что *в случае дискретного Ω для задания вероятности на σ-алгебре всех подмножеств Ω достаточно задать вероятности элементарных событий* (эти события образуют довольно узкий класс). Вероятность $\mathsf{P}(A)$ любого другого события A по вероятностям $\mathsf{P}(\omega)$ элементарных событий $\{\omega\}$ вычисляют так:

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega)$$

— *вероятность* $\mathsf{P}(A)$ *события* A *равна сумме вероятностей элементарных событий, входящих в событие* A. В общем случае так просто задать вероятность не удается.

На дискретном Ω равенство

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega),\ A \subset \Omega,$$

устанавливает взаимно однозначное соответствие между неотрицательными функциями точки

$$\mathsf{P} : \omega \to \mathsf{P}(\omega),\ \omega \in \Omega,$$

такими, что

$$\sum_{\omega \in \Omega} P(\omega) = 1,$$

и неотрицательными счетно-аддитивными нормированными функциями множества

$$\mathsf{P} : A \to \mathsf{P}(A) = \sum_{\omega \in A} P(\omega),\ A \subset \Omega,$$

— вероятностями на классе всех подмножеств Ω.

Поэтому неотрицательную функцию точки $\mathsf{P}(\omega),\ \omega \in \Omega$, на дискретном Ω, для которой $\sum\limits_{\omega \in \Omega} P(\omega) = 1$, также называют дискретной (атомической) вероятностью на Ω.

Дискретное распределение. Пусть X — дискретное (конечное либо счетное) подмножество $\mathbb{R}^n$. Вероятность, заданную на классе всех подмножеств дискретного множества $X \subset \mathbb{R}^n$, еще называют *дискретным вероятностным распределением на* $X \subset \mathbb{R}^n$ (кратко — на $\mathbb{R}^n$).

Согласно теореме 4.1.1, *вероятностное распределение* P *на дискретном множестве* $X \subset \mathbb{R}^n$ *задается своими значениями в точках* $x \in X \subset \mathbb{R}^n$:

$$\mathsf{P}(A) = \sum_{x \in A} \mathsf{P}(x),\ A \subset X. \tag{4.1.3}$$

Поэтому неотрицательную функцию точки

$$\mathsf{P} : x \to \mathsf{P}(x),\ x \in X \subset \mathbb{R}^n, \tag{4.1.4}$$

со значениями в $\mathbb{R}^1$ на дискретном подмножестве $X \subset \mathbb{R}^n$ такую, что

$$\sum_{x \in X} \mathsf{P}(x) = 1,$$

также называют *дискретным вероятностным распределением на* $X \subset \mathbb{R}^n$ (на $\mathbb{R}^n$).

Точку $x \in X$, такую, что

$$\mathsf{P}(x) > 0,$$

называют *атомом распределения* P, а число $\mathsf{P}(x)$ называют *массой атома* x.

Дискретное вероятностное распределение удобно интерпретировать как единичную массу, сосредоточенную в точках дискретного подмножества X: в точке x сосредоточена масса $\mathsf{P}(x)$, $x \in X$.

Представим, что бусины (i-я бусина с массой p_i) единичной суммарной массы ($\sum\limits_i p_i = 1$) рассыпают на прямой $\mathbb{R}^1$. Точки $x_i \in \mathbb{R}^1$, в которых остановились бусины, образуют дискретное множество $X \subset \mathbb{R}^1$ атомов, масса атома x_i равна массе p_i бусины, которая оказалась в точке x_i.

Бусины, находящиеся в точках $x_i \in X \subset \mathbb{R}^1$, задают на X функцию

$$\mathsf{P} : x_i \to \mathsf{P}(x_i) = p_i, \ x_i \in X,$$

которая представляет собой дискретное (атомическое) вероятностное распределение на $X \subset \mathbb{R}^1$ (на $\mathbb{R}^1$).

Если бусины суммарной единичной массы рассыпают на плоскости, получаем дискретное вероятностное распределение на $\mathbb{R}^2$.

О задании вероятностей элементарных событий. А как определяются (откуда берутся) вероятности элементарных событий? Их мы задаем сами, чаще всего исходя из эмпирических данных о стохастическом эксперименте. Теория вероятностей никаких рекомендаций по определению вероятностей элементарных событий не дает. А на естественно возникающий здесь вопрос: “адекватно ли (правильно ли) выбраны вероятности элементарных событий, а вместе с ними и математическая модель?” отвечает опыт, как правило, на языке математической статистики, занимающейся проверкой адекватности модели опыту (стохастическому эксперименту).

Пример. *Подбрасывают игральную кость. Построить математическую модель (вероятностное пространство) этого* *стохастического эксперимента. Вычислить вероятность события A — “выпало четное число очков”.*

Р е ш е н и е. В качестве пространства элементарных событий Ω естественно рассмотреть множество $\Omega = \{1, 2, \ldots, 6\}$. Пространство Ω описывает возможные исходы стохастического эксперимента. Элемент i $(i = 1, 2, \ldots, 6)$ пространства Ω описывает исход “выпало i очков”. Поскольку Ω дискретно, то для построения вероятностного пространства $\{\Omega, \mathsf{P}\}$ достаточно задать вероятности элементарных событий.

События “выпало i очков”, $i = 1, 2, \ldots 6$, являются наблюдаемыми, поэтому можно говорить о частоте их появления в последовательности экспериментов. Исходя из этих частот, т. е. из опытных данных, мы и задаем (приписываем) элементарным событиям вероятности. Если кость изготовлена так, что при ее подбрасывании грани появляются “одинаково часто”, то каждому элементарному событию естественно приписать вероятность $1/6$ — это не что иное, как отражение опытного факта, — ни из каких теорем и утверждений не следует, что

$$\mathsf{P}(i) = 1/6, \ i = 1, 2, \ldots, 6.$$

Еще говорят, что мы задаем вероятности $\mathsf{P}(i) = 1/6$, $i = 1, 2, \ldots$ $\ldots, 6$, исходя из соображений симметрии, но последнее опять-таки отражает опытные данные о поведении симметричной игральной кости как такой, грани которой выпадают одинаково “часто”. Разумеется, подбрасываемая игральная кость не обязательно симметричная. Представим, что кость изготовлена так, что грани $1, 2, 3, 4, 5$ выпадают одинаково часто, а грань 6 — так часто, как остальные вместе взятые. Тогда элементарным событиям $1, 2, \ldots, 6$ естественно приписать вероятности следующим образом:

$$\mathsf{P}(1) = 1/10, \mathsf{P}(2) = 1/10, \ldots, \mathsf{P}(5) = 1/10, \mathsf{P}(6) = 1/2.$$

И опять-таки это только отражение опытного факта и ни из каких теорем не следует.

Далее, если известны вероятности элементарных событий, то вероятность любого другого события A определяется равенством

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega).$$

В частности, вероятность события A — “выпало четное число очков”, описываемого подмножеством $\{2, 4, 6\}$ пространства Ω, вычисляется так:

$$\mathsf{P}(A) = \mathsf{P}(\{2, 4, 6\}) = \mathsf{P}(2) + \mathsf{P}(4) + \mathsf{P}(6).$$

Для симметричной игральной кости

$$\mathsf{P}(A) = \mathsf{P}(2) + \mathsf{P}(4) + \mathsf{P}(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

а для несимметричной

$$\mathsf{P}(A) = \mathsf{P}(2) + \mathsf{P}(4) + \mathsf{P}(6) = \frac{1}{10} + \frac{1}{10} + \frac{1}{2} = \frac{7}{10}.$$

Если Ω дискретно, исходным "материалом" для построения вероятностного пространства являются вероятности, заданные на классе $\mathfrak{K}$ элементарных событий.

При подбрасывании игральной кости класс $\mathfrak{K}$ образуют события: "выпало i очков", $i = 1, 2, \ldots, 6$. Вероятности на классе $\mathfrak{K}$ мы задаем, исходя из эмпирических данных о поведении игральной кости при ее подбрасывании. Далее "достраиваем" вероятность, заданную на классе $\mathfrak{K}$, до вероятности на алгебре всех подмножеств Ω так, чтобы выполнялось свойство аддитивности вероятности, поэтому вероятность каждого другого наблюдаемого события A определяется следующим образом:

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega).$$

4.2 Классическая модель

Определение. Дискретное вероятностное пространство $\{\Omega, \mathsf{P}\}$, элементарные события которого равновероятны:

$$\mathsf{P}(\omega_i) = \mathsf{P}(\omega_j)$$

для всех i, j, будем называть *классической моделью.*

Далее через $n(C)$ будем обозначать число элементов конечного множества C.

Теорема 4.2.1. *Если $\{\Omega, \mathsf{P}\}$ — классическая модель, то Ω конечно и вероятности элементарных событий*

$$\mathsf{P}(\omega_i) = \frac{1}{n(\Omega)}, \ i = 1, 2, \ldots, n(\Omega).$$

Действительно, классическая модель — частный случай дискретной модели, поэтому

$$\sum_{\omega_i \in \Omega} \mathsf{P}(\omega_i) = 1,$$

а так как все $\mathsf{P}(\omega_i)$, $\omega_i \in \Omega$, равны между собой, то $\mathsf{P}(\omega_i) = \mathsf{P}(\omega_1)$, $\omega_i \in \Omega$, и, следовательно,

$$\sum_{\omega_i \in \Omega} \mathsf{P}(\omega_1) = 1. \tag{4.2.1}$$

Ряд с общим членом равным константе $\mathsf{P}(\omega_1)$ сходится в двух случаях: если $\mathsf{P}(\omega_1) = 0$, но это противоречит равенству $\sum\limits_{\omega_i \in \Omega} \mathsf{P}(\omega_1) = 1$, или число $n(\Omega)$ членов ряда конечно. Поэтому равенство (4.2.1) можно записать так:

$$1 = \sum_{\omega_i \in \Omega} \mathsf{P}(\omega_1) = n(\Omega)\mathsf{P}(\omega_1).$$

Отсюда

$$\mathsf{P}(\omega_i) = \mathsf{P}(\omega_1) = \frac{1}{n(\Omega)}, \ i = 1, 2, \ldots, n(\Omega).$$

Формула классической вероятности. В классической модели, как и в любом дискретном вероятностном пространстве, для каждого события A

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega),$$

и к тому же $\mathsf{P}(\omega) = 1/n(\Omega)$ для всех $\omega \in \Omega$, поэтому

$$\mathsf{P}(A) = \sum_{\omega \in A} \frac{1}{n(\Omega)} = n(A)\frac{1}{n(\Omega)} = \frac{n(A)}{n(\Omega)}.$$

Так что *в классической модели вероятность события A равна отношению числа $n(A)$ элементарных событий, входящих в A, к общему числу $n(\Omega)$ элементарных событий,* т. е.

$$\mathsf{P}(A) = \frac{n(A)}{n(\Omega)}. \qquad (4.2.2)$$

Вероятность P, задаваемую равенством (4.2.2), называют *классической вероятностью*, а формулу (4.2.2) — *формулой классической вероятности.* Исходы (элементарные события), входящие в событие A, еще называют исходами, *благоприятствующими событию A*. Формулу классической вероятности часто читают так: *вероятность события A равна отношению числа исходов, благоприятствующих событию A, к общему числу исходов.*

Пример (гипотеза Д'Аламбера). *Бросают две симметричные монеты. Вычислить вероятности следующих событий: A — "обе монеты выпали гербом", B — "обе монеты выпали решеткой", C — "монеты выпали разными сторонами".*

Замечательный французский ученый и математик Д'Аламбер считал, что указанные события равновероятны и, следовательно, вероятность каждого из них равна 1/3. Между тем и до Д'Аламбера и после существовал правильный подход к решению этой задачи: монеты следует различать. Тогда в качестве пространства элементарных событий естественно рассмотреть $\Omega = \{\text{ГГ}, \text{ГР}, \text{РГ}, \text{РР}\}$. А поскольку монеты симметричны, то каждому исходу следует приписать вероятность, равную 1/4. Поэтому

$$\mathsf{P}(A) = \mathsf{P}(\text{ГГ}) = 1/4,\ \mathsf{P}(B) = \mathsf{P}(\text{РР}) = 1/4,$$

$$\mathsf{P}(C) = \mathsf{P}\{\text{РГ}, \text{ГР}\} = \mathsf{P}(\text{РГ}) + \mathsf{P}(\text{ГР}) = 1/4 + 1/4 = 1/2.$$

По-видимому, логически обосновать, почему не прав Д'Аламбер, полагая, что события A, B, C равновероятны, невозможно. В физике элементарных частиц встречаются ситуации, в которых скорее прав Д'Аламбер.

В рассматриваемом же примере в пользу второй модели говорит опыт. Регистрируя результаты сравнительно длинной серии экспериментов, не представляет труда с помощью статистических методов убедиться, что в качестве модели описанного стохастического эксперимента следует принять модель:

$$\Omega = \{\text{ГГ}, \text{ГР}, \text{РГ}, \text{РР}\},$$

$$\mathsf{P}(\text{ГГ}) = 1/4,\ \mathsf{P}(\text{РГ}) = 1/4,\ \mathsf{P}(\text{ГР}) = 1/4,\ \mathsf{P}(\text{РР}) = 1/4,$$

а не модель Д'Аламбера.

Теорема (вычисление условной вероятности в классической модели). *Если* $\{\Omega, \mathsf{P}\}$ *— классическая модель и* $\mathsf{P}(B) > 0$, *то* $\{B, \mathsf{P}_B\}$ *также классическая модель и условная вероятность* $\mathsf{P}_B(A)$ *события* A *относительно события* B *равна отношению числа* $n(A \cap B)$ *элементарных событий, входящих в* $A \cap B$, *к числу* $n(B)$ *элементарных событий, входящих в* B:

$$\mathsf{P}_B(A) = \frac{n(A \cap B)}{n(B)}.$$

Доказательство. Для каждого $\omega_i \in B$

$$\mathsf{P}_B(\{\omega_i\}) = \frac{\mathsf{P}(\{\omega_i\} \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(\{\omega_i\})}{\mathsf{P}(B)} = \frac{1/n(\Omega)}{n(B)/n(\Omega)} = \frac{1}{n(B)},$$

поэтому $\{B, \mathsf{P}_B\}$ — классическая модель. Осталось заметить, что

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{n(A \cap B)/n(\Omega)}{n(B)/n(\Omega)} = \frac{n(A \cap B)}{n(B)}.$$

Пример 4.2.1 (задача Льюиса Кэррола). *В урне находится один шар, относительно которого известно, что он либо белый, либо черный (с вероятностью* $1/2$*). В урну положили белый шар, а затем, тщательно перемешав шары, извлекли наудачу один шар, который оказался белым. Какова вероятность вынуть после этого из урны белый шар?*

Решение 1. Эксперимент состоит в последовательном извлечении из урны двух шаров. Обозначим белый шар через W, черный — B. Множеством всех исходов стохастического эксперимента является

$$\Omega = \{WW, WB, BW\}.$$

Например, пара WB описывает исход: “первым извлечен белый шар, вторым — черный”. Пусть событие A_1 — “белый шар вынут первым”, событие A_2 — “белый шар извлечен вторым”. Необходимо вычислить $\mathsf{P}(A_2/A_1)$. События A_1, A_2, $A_1 \cap A_2$ как подмножества Ω опишутся так:

$$A_1 = \{WW, WB\},\ \ A_2 = \{WW, BW\},\ A_1 \cap A_2 = \{WW\}.$$

Поскольку в Ω входят три элементарных события, в A_1 — два, в $A_1 \cap A_2$ — входит одно элементарное событие, то

$$\mathsf{P}(A_2/A_1) = \frac{\mathsf{P}(A_2 \cap A_1)}{\mathsf{P}(A_1)} = \frac{1/3}{2/3} = \frac{1}{2}.$$

Р е ш е н и е 2. Белый шар, который положили в урну, пометим звездочкой и обозначим через W^*. Тогда множеством всех исходов стохастического эксперимента является

$$\Omega^* = \{WW^*, W^*W, BW^*, W^*B\}.$$

События A_1, A_2, $A_1 \cap A_2$ как подмножества Ω^* опишутся так:

$$A_1 = \{WW^*, W^*W, W^*B\},\ A_2 = \{WW^*, W^*W, BW^*\},$$

$$A_1 \cap A_2 = \{WW^*, W^*W\}.$$

Поэтому

$$\mathsf{P}(A_2/A_1) = \frac{\mathsf{P}(A_2 \cap A_1)}{\mathsf{P}(A_1)} = \frac{2/4}{3/4} = \frac{2}{3}.$$

Два решения — два разных ответа. Нехорошо.

В о п р о с 1. Какое из решений неверное?

В о п р о с 2. Где допущена ошибка?

Попытайтесь сначала ответить на вопросы 1 и 2, не читая далее.

О т в е т к п р и м е р у 4.2.1. Решение 1 неверное. Модель не классическая, а вероятности событий вычисляются как в классической модели.

Поскольку сначала в урне находился белый или черный шар (с вероятностью $1/2$), то после того, как в урну положили белый шар, в урне находятся с вероятностью $1/2$ шары одного цвета и с вероятностью $1/2$ — разного. Поэтому в этой модели элементарным событиям необходимо приписать вероятности так:

$$\mathsf{P}(WW) = 1/2, \mathsf{P}(WB) = 1/4, \mathsf{P}(BW) = 1/4.$$

Тогда

$$\mathsf{P}(A_1 \cap A_2) = \mathsf{P}(WW) = 1/2;$$

$$\mathsf{P}(A_1) = \mathsf{P}(\{WW, WB\}) = \mathsf{P}(WW) + \mathsf{P}(WB) = 1/2 + 1/4 = 3/4,$$

$$\mathsf{P}(A_2/A_1) = \frac{\mathsf{P}(A_2 \cap A_1)}{\mathsf{P}(A_1)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

В решении 2 каждому исходу приписываем вероятность $1/4$. Заметим, что состав урны после вложения одного шара и до выбора шаров такой: два белых шара, один из которых помеченный (с вероятностью $1/2$), или один белый (помеченный), другой черный (с вероятностью $1/2$).

Комментарий к примеру. Пример, в частности, иллюстрирует тот факт, что для одного и того же стохастического эксперимента можно предложить разные пространства элементарных событий, а вместе с тем и разные вероятностные пространства (модели) стохастического эксперимента, адекватно его описывающие. В примере было предложено два пространства элементарных событий:

$$\Omega = \{WW, WB, BW\};\ \Omega^* = \{WW^*, W^*W, BW^*, W^*B\}.$$

Распределение вероятностей на пространстве элементарных событий Ω:

$$\mathsf{P}(WW) = \frac{1}{2},\ \mathsf{P}(WB) = \frac{1}{4},\ \mathsf{P}(BW) = \frac{1}{4},$$

а на пространстве Ω^*:

$$\mathsf{P}(WW^*) = \frac{1}{4},\ \mathsf{P}(W^*W) = \frac{1}{4},\ \mathsf{P}(W^*B) = \frac{1}{4},\ \mathsf{P}(BW^*) = \frac{1}{4}.$$

Обе модели адекватно описывают стохастический эксперимент. Заметим, что одна из них классическая, другая — нет.

4.3 Основные понятия комбинаторики

В классической модели задача вычисления вероятности события A сводится к подсчету отношения числа $n(A)$ исходов, входящих в A к числу $n(\Omega)$ всех исходов Ω. При подсчете этих чисел важную роль играют методы комбинаторики (раздел математики, изучающий конечные множества).

Правило умножения (основной принцип комбинаторики). Пусть A и B — произвольные множества. Каждые два элемента $a \in A$ и $b \in B$ определяют упорядоченную пару элементов (a, b). Множество всех упорядоченных пар (a, b), $a \in A$, $b \in B$ называют *декартовым произведением множеств* A и B и обозначают $A \times B$.

Правило умножения. *Число $n(A \times B)$ элементов декартова произведения $A \times B$ конечных множеств A и B равно произведению $n(A)n(B)$ числа $n(A)$ элементов множества A и числа $n(B)$ элементов множества B:*

$$n(A \times B) = n(A)n(B).$$

Доказательство. Декартово произведение $A \times B$ представим в виде объединения:

$$A \times B = (\{a_1\} \times B) \cup (\{a_2\} \times B) \cup \ldots \cup (\{a_{n(A)}\} \times B)$$

непересекающихся множеств $\{a_i\} \times B$, $i = 1, 2, \ldots, n(A)$. Число элементов каждого из множеств $\{a_i\} \times B$ равно $n(B)$. Поэтому

$$n(A \times B) = n(\{a_1\} \times B) + n(\{a_2\} \times B) + \ldots + n(\{a_{n(A)}\} \times B) =$$

$$= n(B) + n(B) + \ldots + n(B) = n(A)n(B).$$

Пример. *Пусть $A = \{1, 2, 3\}$ и $B = \{4, 5, 6, 7\}$. Найти число элементов в $A \times B$.*

Решение. Согласно правилу умножения

$$n(A \times B) = n(A)n(B) = 3 \cdot 4 = 12.$$

Декартовым произведением множеств $A^{(1)}, A^{(2)}, \ldots, A^{(k)}$ будем называть множество упорядоченных последовательностей

$$(a^{(1)}, a^{(2)}, \ldots, a^{(k)}),\ a^{(i)} \in A^{(i)},\ i = 1, 2, \ldots, k,$$

и будем обозначать его

$$A^{(1)} \times A^{(2)} \times \ldots \times A^{(k)}.$$

Используя метод математической индукции, устанавливаем следующее *общее правило умножения.*

Число $n(A^{(1)} \times A^{(2)} \times \ldots \times A^{(k)})$ элементов декартова произведения $A^{(1)} \times A^{(2)} \times \ldots \times A^{(k)}$ множеств $A^{(1)}, A^{(2)}, \ldots, A^{(k)}$ равно произведению $n(A^{(1)})n(A^{(2)}) \ldots n(A^{(k)})$ числа элементов $n(A^{(1)}), n(A^{(2)}), \ldots, n(A^{(k)})$ этих множеств:

$$n(A^{(1)} \times A^{(2)} \times \ldots \times A^{(k)}) = n(A^{(1)})n(A^{(2)}) \ldots n(A^{(k)}).$$

Часто правило умножения удобно формулировать терминах действий.

Пусть необходимо выполнить одно за другим k действий. Если первое действие можно выполнить n_1 способами, второе — n_2 способами и так до k-го действия, которое можно выполнить n_k способами, то все k действий вместе могут быть выполнены $n_1 n_2 \ldots n_k$ способами.

Последнее утверждение — не что иное как правило умножения, сформулированное в терминах действий. В самом деле, обозначим через через $A^{(1)}$ набор способов выполнить действие 1, через $A^{(2)}$ — действие 2, …, через $A^{(k)}$ — действие k (ясно, что $n(A^{(i)}) = n_i$, $i = 1, 2, \dots, k$). Тогда элемент $(a^{(1)}, a^{(2)}, \dots, a^{(k)})$ из $A^{(1)} \times A^{(2)} \times \dots \times A^{(k)}$ задает способ выполнить все k действий вместе (в указанном порядке), при этом

$$n(A^{(1)} \times A^{(2)} \times \dots \times A^{(k)}) = n(A^{(1)})n(A^{(2)}) \dots n(A^{(k)}) = n_1 n_2 \dots n_k.$$

Так что число способов выполнить k действий вместе равно $n_1 n_2 \dots n_k$.

Пример. *В каждую клетку прямоугольной таблицы, состоящей из m строк и n столбцов, записывают числа $+1$ или -1. Сколько таблиц можно получить таким образом?*

Решение. Заполнить таблицу $m \times n$ числами $+1$ и -1 значит выполнить mn действий (по числу клеток). Каждое действие можно выполнить двумя способами. Поэтому в силу правила умножения число способов заполнить таблицу (а, следовательно, и искомое число таблиц) равно 2^{mn}.

А сколько существует таблиц у которых произведение чисел по строкам и произведение по столбцам равно $+1$?

“Способ”. Во многих задачах комбинаторики необходимо ответить на вопрос: “Сколько существует способов выполнить то или иное действие, упорядочить то или иное множество и т. д.?” При этом, прежде чем подсчитывать число способов, необходимо выяснить, что именно представляет собой способ или как (чем) его можно описать. Прежде чем отвечать на вопрос “Сколько?” необходимо ответить на вопрос “Что? Что будем считать?” В рассматриваемых нами задачах способ образовать, составить, упорядочить, и т. д., как правило, можно описать элементами того или иного множества или последовательностью выполнения тех или иных действий.

Замечание. Вычисляя число $n(A)$ элементов множества A, удобно пользоваться таким фактом: если между множествами A и B установлено взаимно однозначное соответствие, то

$$n(A) = n(B),$$

часто число элементов множества B подсчитать проще.

Далее множество, состоящее из n элементов, коротко будем называть *n-элементным множеством.*

Упорядоченные множества. Мы будем различать множества и упорядоченные множества.

Определение. Множество будем называть *упорядоченным,* если каждому его элементу приписан номер, причем так, что различным элементам приписаны различные номера.

Упорядоченные множества различны, если они отличаются либо своими элементами, либо их порядком.

Конечное множество можно упорядочить следующим способом: *записать все элементы множества в список* $a, b, c, \ldots, f$, *и каждому элементу приписать его номер в списке, или, что то же, расположить элементы множества на занумерованных местах и каждому элементу приписать номер места, на котором он оказался.* Как правило, так и будем поступать.

Перестановки. n-Элементные упорядоченные подмножества данного n-элементного множества будем называть его *перестановками.*

Различные перестановки данного множества отличаются только порядком элементов, но не самими элементами.

Число всех перестановок n-элементного множества обозначают P_n.

Пример 4.3.1 . *Выписать все перестановки множества* $\Omega = \{a, b, c\}$.

Решение. $(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a)$ — все перестановки множества $\{a, b, c\}$, заметим, что $P_3 = 6$.

Теорема. *Число* P_n *всех перестановок* n*-элементного множества равно* $n!$, *т. е.*

$$P_n = n!$$

Доказательство. Каждой перестановке n-элементного множества соответствует способ его упорядочить (упорядочить — расположить элементы множества на n занумерованных местах и каждому элементу приписать номер места, на котором он оказался) и наоборот — способ упорядочить множество задает перестановку. Поэтому число перестановок n-элементного множества равно числу способов его упорядочить. Упорядочить n-элементное множество — расположить его элементы на n местах, можно, выполнив n действий: действие первое — заполнить первое место (одним из n элементов), действие второе — заполнить второе место (одним из $(n-1)$ оставшихся элементов) и т. д. По правилу умножения все n действий можно выполнить $n \cdot (n-1) \cdots 2 \cdot 1$ способами. Следовательно, n-элементное множество Ω можно упорядочить $P_n = n!$ способами.

Пример. *Сколькими способами можно упорядочить множество чисел* $1, 2, \ldots, 2n$ *так, чтобы четные числа получили четные номера?*

Решение. Расположим $2n$ чисел $1, 2, \ldots, 2n$ на $2n$ местах, причем так, чтобы четные числа заняли места с четными номерами (при этом нечетные числа займут места с нечетными номерами). Выполним это в два действия. Действие первое — расположить n четных чисел на n четных местах (упорядочить n-элементное множество), можно выполнить $n!$ способами. Действие второе — расположить n нечетных чисел на n нечетных местах — можно выполнить $n!$ способами. Два действия вместе согласно правилу умножения можно выполнить $n! \cdot n!$ способами.

Сочетания. k-Элементные подмножества n-элементного множества будем называть *сочетаниями из n элементов по k.*

Сочетания из n элементов по k различны, если они отличаются хотя бы одним элементом. Сочетания из n элементов по k, состоящие из одних и тех же элементов, неразличимы.

Число всех k-элементных подмножеств n-элементного множества (число сочетаний из n элементов по k) будем обозначать C_n^k.

Пример 4.3.2. *Пусть $\Omega = \{a, b, c\}$. Выписать все сочетания из 3 элементов по 1, из 3 по 2.*

Решение. $\{a\}, \{b\}, \{c\}$ — все 1-элементные подмножества 3-элементного множества $\{a, b, c\}$ (сочетания из 3 элементов по 1), заметим, что $C_3^1 = 3$. $\{a, b\}, \{a, c\}, \{b, c\}$ — все 2-элементные подмножества 3-элементного множества (сочетания из 3 элементов по 2), C_3^2 равно 3.

Из определения следует, что, например, $\{a, b\}$ и $\{b, a\}$ — одно и то же сочетание.

Теорема. *Число C_n^k всех k-элементных подмножеств n-элементного множества (число сочетаний из n элементов по k) равно* $n!/(k!(n-k)!)$, т. е.

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Доказательство. Подсчитаем число перестановок n-элементного множества двумя способами. С одной стороны, оно равно $n!$ С другой стороны, это число можно получить так: разобьем n-элементное множество на два подмножества, содержащих соответственно k и $(n-k)$ элементов, выбрав k-элементное подмножество (это можно сделать C_n^k способами). Затем каждое из этих подмножеств упорядочим, первое — $k!$ способами, второе — $(n-k)!$ способами и таким образом получим n-элементное упорядоченное множество. Всего выполнено три действия, поэтому согласно правилу умножения число n-элементных упорядоченных множеств равно $C_n^k k!(n-k)!$ Так что $n! = C_n^k k!(n-k)!$ Отсюда

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Пример 4.3.3 (“шахматный город”). *Рассмотрим прямоугольную сетку квадратов — “шахматный город”, состоящий из $m \times n$ квадратных кварталов, разделенных $n-1$ “горизонтальными” и $m-1$ “вертикальными” улицами* (*см. рис.* 4.3.1). *Сколько существует на этой сетке различных кратчайших путей, ведущих из левого нижнего угла* (*точки* $(0,0)$) *в правый верхний угол* (*точку* (m,n))*?*

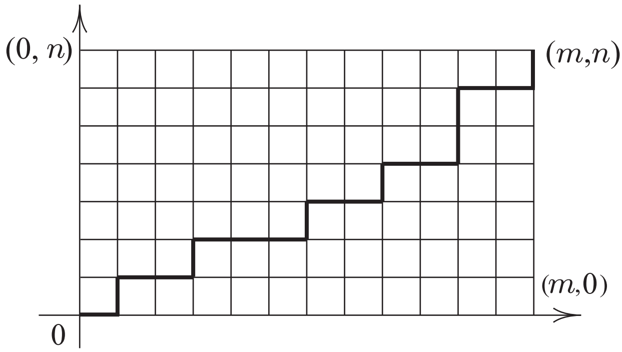

Рис. 4.3.1: “Шахматный город”

Р е ш е н и е. Обозначим буквой “Г” горизонтальный отрезок пути, буквой “В” — вертикальный. Каждый кратчайший путь из $(0,0)$ в (m,n) состоит из n вертикальных отрезков и m горизонтальных. Он полностью задается последовательностью длиной $n+m$, составленной из m букв “Г” и n букв “В”, и наоборот. Поэтому число кратчайших путей из $(0,0)$ в (m,n) равно числу последовательностей длиной $n+m$, составленных из m букв “Г” и n букв “В”. Каждая такая последовательность однозначно задается выбором m мест для буквы “Г” из $n+m$ мест (оставшиеся места заполняются буквами “В”). Поэтому число последовательностей из m букв “Г” и n букв “В” равно C_{n+m}^m.

Размещения. k-Элементные упорядоченные подмножества данного n-элементного множества будем называть *размещениями* из n элементов по k.

Размещения из n элементов по k различны, если они отличаются или своими элементами, или порядком элементов.

Число всех различных размещений из n элементов по k будем обозначать A_n^k.

Пример 4.3.4. *Пусть* $\Omega = \{a, b, c\}$. *Выписать все размещения из* 3 *элементов по* 2.

Решение. $(a, b), (b, a), (a, c), (c, a), (b, c), (c, b)$ — все 2-элементные упорядоченные подмножества 3-элементного множества $\{a, b, c\}$ (размещения из 3 элементов по 2), A_3^2 равно 6.

Теорема. *Число* A_n^k *всех упорядоченных* k*-элементных подмножеств* n*-элементного множества* (*число размещений из* n *элементов по* k) *равно* $n!/(n-k)!$, *т. е.*

$$A_n^k = n!/(n-k)!$$

Доказательство. k-Элементные упорядоченные подмножества n-элементного множества можно получить двумя действиями: сначала выбрать k-элементное подмножество n-элементного множества (C_n^k способами), а затем упорядочить его ($k!$ способами). Согласно правилу умножения эти два действия вместе можно выполнить $C_n^k \cdot k!$ способами. Так что число

$$A_n^k = C_n^k \cdot k! = n!/(n-k)!$$

Разбиения на подмножества, перестановки с повторениями. *Разбиением* n-элементного множества Ω на m попарно непересекающихся подмножеств, содержащих соответственно $k_1, k_2, \ldots, k_m$ элементов $(k_1 + k_2 + \ldots + k_m = n)$, будем называть *упорядоченный набор*

$$(B_1, B_2, \ldots, B_m)$$

из m *попарно непересекающихся подмножеств множества* Ω, *содержащих соответственно* $k_1, k_2, \ldots, k_m$ *элементов.*

Очевидно, имеет место представление

$$B_1 \cup B_2 \cup \ldots \cup B_m = \Omega.$$

Два разбиения множества на m попарно непересекающихся подмножеств, содержащих соответственно $k_1, k_2, \ldots, k_m$ элементов, различны, если хотя бы в одной паре соответствующих k_j-элементных подмножеств $(j = 1, 2, \ldots, m)$ имеются различные элементы.

Число всех разбиений n-элементного множества на m попарно непересекающихся подмножеств, содержащих соответственно $k_1, k_2, \ldots, k_m$ элементов ($k_1+k_2+\cdots+k_m = n$), будем обозначать $C_n(k_1, k_2, \ldots, k_m)$.

Пример 4.3.5. *Привести все возможные разбиения множества $\Omega = \{a, b, c, d\}$ на 3 непересекающихся подмножества (B_1, B_2, B_3), содержащих соответственно $k_1 = 1$ элементов (в множестве B_1), $k_2 = 2$ элементов (в множестве B_2), $k_3 = 1$ элементов (в множестве B_3).*

Решение.

$(\{a\}, \{b, c\}, \{d\}), (\{a\}, \{c, d\}, \{b\}), (\{a\}, \{b, d\}, \{c\}),$
$(\{b\}, \{a, c\}, \{d\}), (\{b\}, \{c, d\}, \{a\}), (\{b\}, \{a, d\}, \{c\}),$
$(\{c\}, \{a, b\}, \{d\}), (\{c\}, \{a, d\}, \{b\}), (\{c\}, \{b, d\}, \{a\}),$
$(\{d\}, \{a, b\}, \{c\}), (\{d\}, \{a, c\}, \{b\}), (\{d\}, \{b, c\}, \{a\}),$
число $C_4(1, 2, 1)$ равно 12.

Заметим, что например, $(\{a\}, \{b, c\}, \{d\})$ и $(\{d\}, \{b, c\}, \{a\})$ являются различными разбиениями множества $\Omega = \{a, b, c, d\}$.

Теорема. *Число $C_n(k_1, k_2, \ldots, k_m)$ всех разбиений n-элементного множества Ω на m попарно непересекающихся подмножеств $B_1, B_2, \ldots, B_m$, содержащих соответственно $k_1, k_2, \ldots$ $\ldots, k_m$ элементов ($k_1+k_2+\cdots+k_m = n$), равно $n!/(k_1!k_2!\ldots k_m!)$,* т. е.

$$C_n(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1!k_2!\ldots k_m!}.$$

Доказательство. Подсчитаем число перестановок n-элементного множества Ω. С одной стороны, это число, как известно, равно $n!$ С другой стороны, его можно получить следующим образом: разбить n-элементное множество на m непересекающихся подмножеств $B_1, B_2, \ldots, B_m$, содержащих соответственно $k_1, k_2, \ldots, k_m$ элементов (это можно сделать $C_n(k_1, k_2, \ldots, k_m)$ способами). А затем упорядочить каждое из них — соответственно $k_1!, k_2!, \ldots, k_m!$ способами. Тогда согласно правилу умножения число n-элементных упорядоченных подмножеств множества Ω равно произведению $C_n(k_1, k_2, \ldots, k_m)k_1!k_2!\ldots k_m!$ Отсюда имеем:

$$n! = C_n(k_1, k_2, \ldots, k_m)k_1!k_2!\ldots k_m!$$

Поэтому

$$C_n(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1!k_2!\ldots k_m!}.$$

Числа $C_n(k_1, k_2, \ldots, k_m)$ называются *полиномиальными коэффициентами.*

Определение. Последовательность длиной n, составленную из k_1 элементов (букв) a_1, k_2 элементов (букв) $a_2, \ldots, k_m$ элементов (букв) a_m $(k_1 + k_2 + \ldots + k_m = n)$, будем называть *перестановкой с повторениями* (*словом*) длиной n из элементов (букв) $a_1, a_2, \ldots, a_m$ в количестве $k_1, k_2, \ldots, k_m$ соответственно.

Например, слово “статистика” — перестановка с повторениями составленная из двух букв “а”, двух букв “и”, одной буквы “к”, двух букв “с”, трех букв “т”.

Две перестановки с повторениями различны, если у перестановок хотя бы на одном месте (из n упорядоченных мест) расположены различные буквы.

Пример 4.3.6. *Выписать все перестановки с повторениями (слова) длиной* 4*, составленные из букв* $\{a, b\}$*, в которых буквы* a*,* b *встречаются по два раза.*

Решение. (a, a, b, b), (a, b, a, b), (b, a, a, b), (b, a, b, a), (b, b, a, a), (a, b, b, a).

Теорема. *Число всех перестановок с повторениями (слов) длиной* n*, которые можно составить из* k_1 *элементов* a_1*,* k_2 *элементов* a_2 *и т. д.,* k_m *элементов* a_m $(k_1+k_2+\cdots+k_m = n)$ *равно*

$$C_n(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1!k_2!\ldots k_m!}.$$

Доказательство. Расположив k_1 элементов a_1, k_2 элементов a_2, и т. д., k_m элементов a_m на n занумерованных числами $1, 2, \ldots, n$ местах и приписав каждому элементу номер места, на котором он оказался, получим перестановку с повторениями. Очевидно, перестановка с повторениями определяется указанием k_1 мест для элемента a_1, k_2 мест для элемента a_2, и т. д., k_m мест для элемента a_m, т. е. разбиением n-элементного множества занумерованных мест на m непересекающихся подмножеств, содержащих соответственно $k_1, k_2, \ldots, k_m$ элементов. Число таких разбиений (а вместе с ними и перестановок с повторениями)

$$C_n(k_1, k_2, \ldots, k_m) = \frac{n!}{k_1!k_2!\ldots k_m!}.$$

Замечание. Составляя слова, необходимо: 1° указать длину n слова; 2° указать набор $\{a_1, a_2, \ldots, a_m\}$ различных букв, используемых в записи слова; 3° указать число k_j — повторений каждой буквы a_j, используемой в записи слова $(j = 1, 2, \ldots, m)$, $k_1 + k_2 + \ldots + k_m = n$.

Количество слов, которые при этом можно получить, равно $C_n(k_1, k_2, \dots, k_m)$.

Теорема (полиномиальная формула).

$$(a_1+a_2+\dots+a_m)^n = \sum_{k_1+k_2+\dots+k_m=n} C_n(k_1, k_2, \dots, k_m) a_1^{k_1} a_2^{k_2} \dots a_m^{k_m},$$

суммирование ведется по всем последовательностям $k_1, k_2, \dots, k_m$*, для которых* $k_1 + k_2 + \dots + k_m = n$.

Доказательство. По определению

$$(a_1 + a_2 + \dots + a_m)^n =$$

$$= (a_1 + a_2 + \dots + a_m)(a_1 + a_2 + \dots + a_m) \dots (a_1 + a_2 + \dots + a_m)$$

равно сумме всех слагаемых вида $d_1 d_2 \dots d_n$, где каждый сомножитель d_i $(i = 1, 2, \dots, n)$, равен или a_1, или a_2, и т. д. или a_m. Другими словами, каждое слагаемое $d_1 d_2 \dots d_n$ определяется словом длиной n, составленным из букв $a_1, a_2, \dots, a_m$. Каждому такому слову соответствует произведение $d_1 d_2 \dots d_n = a_1^{k_1} a_2^{k_2} \dots a_m^{k_m}$. Количество слагаемых вида $a_1^{k_1} a_2^{k_2} \dots a_m^{k_m}$ равно числу слов длиной n, которые можно составить из k_1 букв a_1, k_2 букв a_2, и т. д., k_m букв a_m, т. е. $C_n(k_1, k_2, \dots, k_m)$. Тем самым теорема доказана.

Модель Максвелла–Больцмана. Эта модель описывает размещение различимых частиц по ячейкам.

Теорема. *Пусть* n *различимых частиц распределяется по* m *ячейкам (областям пространства). Тогда число всех размещений частиц равно* m^n*. Число размещений частиц, в которых первая ячейка содержит* k_1 *частиц, вторая —* k_2 *частиц и т. д.,* m*-я —* k_m *частиц* $(k_1 + k_2 + \dots + k_m = n)$*, равно* $C_n(k_1, k_2, \dots, k_m)$.

Доказательство. Для определенности занумеруем частицы числами $1, 2, \dots, n$, а ячейки (области пространства) — номерами $1, 2, \dots, m$.

Оборот “распределить (разместить) n частиц по m ячейкам” обозначает — каждой частице приписать номер ячейки, в которой она окажется. Так что каждому размещению n частиц по m ячейкам соответствует последовательность длиной n, составленная из чисел $1, 2, \dots, m$, другими словами соответствует слово $(i_1, i_2, \dots, i_n)$ длиной n, составленное из номеров ячеек $1, 2, \dots, m$ (i_1 — номер ячейки, в которой оказалась частица с номером 1, i_2 — номер ячейки, в которой оказалась частица с номером 2 и т. д., i_n — номер ячейки, в которой оказалась частица с номером n). И наоборот, каждая такая последовательность (слово) задает распределение n различимых частиц по m ячейкам. Поэтому число всех размещений n частиц по m ячейкам равно m^n — числу слов длиной n, составленных из чисел от 1 до m. А число размещений n частиц по m ячейкам, у которых в первой ячейке находится k_1 частиц, во второй — k_2 частиц и т. д., в m-й — k_m частиц равно числу слов длиной n, составленным из k_1 элементов (букв) “1”, k_2 элементов (букв) “2” и т. д., k_m элементов (букв) “m”, т. е. равно $C_n(k_1, k_2, \ldots, k_m)$.

Сочетания с повторениями. Набор из n элементов, каждый из которых принадлежит одному из m типов, будем называть *сочетанием из m элементов по n с повторениями.*

Сочетание с повторениями задается указанием числа элементов каждого типа.

Два сочетания из m элементов по n различны, если они отличаются количеством элементов хотя бы одного типа.

Число всех сочетаний из m элементов по n с повторениями будем обозначать f_m^n.

Пример 4.3.7. *Выписать все сочетания с повторениями из 4 элементов a, b, c, d по 2.*

Р е ш е н и е. $aa, bb, cc, dd, ab, ac, ad, bc, bd, dc$, заметим, что число $f_4^2 = 10$.

Из определения следует, что, например, ab и ba — одно и то же сочетание с повторениями.

Теорема 4.3.1. *Число f_m^n сочетаний из m элементов по n с повторениями равно C_{n+m-1}^{m-1}:*

$$f_m^n = C_{n+m-1}^{m-1}.$$

Д о к а з а т е л ь с т в о. Каждому сочетанию из m элементов по n с повторениями поставим в соответствие последовательность из n нулей и $m-1$ единиц следующим образом: сначала выпишем нули в количестве, равном числу элементов первого типа, входящих в сочетание с повторениями, затем единицу; далее выписываем нули в количестве, равном числу элементов второго типа, затем единицу и т. д. (после нулей, соответствующих элементам m-го типа, единицу не записываем). Наоборот, каждой последовательности из n нулей и $m-1$ единиц соответствует сочетание из m элементов по n с повторениями (сначала набираем элементы первого типа в количестве равном числу нулей до первой единицы, затем — элементы второго типа в количестве, равном числу нулей между первой и второй единицей, и т. д.). Поэтому число сочетаний из m элементов по n с повторениями равно числу последовательностей, составленных из n нулей и

$m-1$ единиц, количество же таких последовательностей равно C_{n+m-1}^{m-1}.

Теорема 4.3.2 . *Если $n \geq m$, то число сочетаний из m элементов по n с повторениями, у которых элемент каждого типа встречается хотя бы один раз, равно C_{n-1}^{m-1}.*

Доказательство. Установим соответствие между сочетанием из m элементов по n с повторениями и последовательностями из n нулей и $m-1$ единиц, как это было описано в доказательстве теоремы 4.3.1. Сочетаниям из m элементов по n с повторениями, у которых элемент каждого типа встречается хотя бы один раз, соответствуют последовательности из n нулей и $m-1$ единиц, у которых никакие две единицы не расположены рядом. Такие последовательности можно получить, выбрав из $n-1$ промежутков между нулями $m-1$ промежуток и расположив в них единицы. Последнее можно сделать C_{n-1}^{m-1} способами.

Пример. *Кость домино — набор из двух чисел, каждое из которых выбрано из множества $\{0, 1, 2, \ldots, 6\}$. Кость можно рассматривать как сочетание из $m = 7$ элементов по $n = 2$ с повторениям. Число таких сочетаний (а вместе с ними и число костей домино) равно*

$$f_7^2 = C_{2+7-1}^{7-1} = C_8^6 = \frac{8!}{2!6!} = \frac{7 \cdot 8}{2} = 28.$$

Теорема 4.3.3. *Для целого $n > 0$ число решений уравнения*

$$x_1 + x_2 + \ldots + x_m = n \tag{4.3.1}$$

в целых неотрицательных числах равно C_{n+m-1}^{m-1}, а в целых положительных (при $n \geq m$) равно C_{n-1}^{m-1}.

Доказательство. Решением уравнения

$$x_1 + x_2 + \ldots + x_m = n$$

в целых неотрицательных числах является последовательность $(x_1, x_2, \ldots, x_m)$ целых неотрицательных чисел такая, что $x_1 + x_2 + \ldots + x_m = n$. Каждая такая последовательность задает сочетание из m элементов по n с повторениями и наоборот. В самом деле, сочетание из m элементов по n с повторениями однозначно задается числом x_1 элементов первого типа, числом x_2 элементов второго типа и т. д., числом x_m элементов m-го типа в него входящих, т. е. задается последовательностью $(x_1, x_2, \ldots, x_m)$ неотрицательных целых чисел, такой, что

$x_1 + x_2 + \ldots + x_m = n$, и наоборот, — сочетание из m элементов по n с повторениями однозначно определяет последовательность $(x_1, x_2, \ldots, x_m)$ неотрицательных целых чисел, таких, что $x_1 + x_2 + \ldots + x_m = n$ (x_1 — число элементов первого типа, x_2 — второго и т. д., x_m — m-го типа). Поэтому искомое число решений равно числу C_{n+m-1}^{m-1} сочетаний из m по n с повторениями.

Количество решений уравнения (4.3.1) в целых положительных числах (при условии $n \geq m$) равно числу сочетаний из m элементов по n с повторениями, у которых элемент каждого типа встречается хотя бы один раз, т. е. равно C_{n-1}^{m-1}.

Следствие. Между множеством сочетаний из m элементов по n с повторениями и множеством решений $(x_1, x_2, \ldots, x_m)$ уравнения $x_1+x_2+\ldots+x_m = n$ в целых неотрицательных числах существует взаимно однозначное соответствие.

Пример. *Сколько существует различных частных производных порядка n у бесконечно дифференцируемой функции m переменных?*

Решение. Частная производная порядка n бесконечно дифференцируемой функции $f(x_1, x_2, \ldots, x_m)$ от m переменных определяется числом дифференцирований по каждой переменной и не зависит от порядка дифференцирования. Если, скажем, по первой переменной мы продифференцируем k_1 раз, по второй — k_2 раз, и т. д., по m-й — k_m раз (причем $k_1 + k_2 + \ldots + k_m = n$), то получим частную производную $\dfrac{\partial^n f(x_1, x_2, \ldots, x_m)}{\partial x_1^{k_1} \partial x_2^{k_2} \ldots \partial x_m^{k_m}}$. Число таких производных равно числу решений уравнения $k_1 + k_2 + \ldots + k_m = n$ в целых неотрицательных числах, а последнее равно

$$f_m^n = C_{n+m-1}^{m-1}.$$

Модель Бóзе–Эйнштейна. Эта модель описывает размещение неразличимых частиц по ячейкам.

Теорема 4.3.4. *Пусть n неразличимых частиц распределяется по m ячейкам (областям пространства). Число всех возможных размещений частиц равно*

$$f_m^n = C_{n+m-1}^{m-1}.$$

Если $n \geq m$, то число тех размещений, у которых каждая ячейка содержит хотя бы одну частицу, равно C_{n-1}^{m-1}.

Доказательство. Размещение n неразличимых частиц по m ячейкам задается указанием числа x_1 частиц, попавших в 1-ю ячейку, x_2 — во 2-ю ячейку и т. д., числа x_m частиц, попавших в

m-ю ячейку, т. е. задается последовательностью $(x_1, x_2, \ldots, x_m)$ целых неотрицательных чисел такой, что

$$x_1 + x_2 + \ldots + x_m = n,$$

другими словами, задается решением $(x_1, x_2, \ldots, x_m)$ уравнения $x_1+x_2+\ldots+x_m = n$, в целых неотрицательных числах (и наоборот). А число таких решений согласно теореме 4.3.3 равно f_m^n.

Сочетания с повторениями и слова. Пусть имеется слово длиной n, составленное из m букв (элементов): x_1 букв a_1, x_2 букв a_2 и т. д., x_m букв a_m $(x_1 + x_2 + \ldots + x_m = n)$. "Ссыпав" буквы слова в урну, получим набор из n элементов, в котором число элементов первого типа равно x_1, второго — x_2, и т. д., m-го типа — x_m, т. е. получим сочетание из m элементов по n с повторениями (одно) с заданным числом элементов каждого типа.

Наоборот. Пусть мы имеем сочетание из m элементов по n с повторениями, в котором число элементов первого типа равно x_1, второго — x_2, и т. д., m-го типа — x_m (имеем набор из n элементов "ссыпанных" в урну: x_1 элементов a_1, x_2 элементов a_2 и т. д., x_m элементов a_m). Разместив элементы на n занумерованных местах (упорядочив их) получим слово. Число всех таких слов равно $C_n(x_1, x_2, \ldots, x_m)$.

4.4 Примеры и задачи

В предлагаемых далее задачах, прежде чем подсчитывать число элементов того или иного множества, необходимо определить, *что именно представляют собой эти элементы* (*т. е., что подсчитывать*). Прежде чем отвечать на вопрос "*Сколько?*", необходимо ответить на вопрос "*Что? Что будем считать?*"

Примеры

Пример 4.4.1. *Симметричную игральную кость подбрасывают шесть раз. Построить математическую модель этого стохастического эксперимента. Описать событие A — "выпадут все шесть граней" и вычислить его вероятность.*

Р е ш е н и е. В качестве пространства элементарных событий Ω стохастического эксперимента естественно рассмотреть множество последовательностей длиной шесть, образованных числами $1, 2, \ldots, 6$. Например, последовательность $(6, 1, 6, 3, 2, 4)$ описывает исход стохастического эксперимента, состоящий в том, что при первом подбрасывании игральной кости выпала 6, при втором — 1, ..., при шестом подбрасывании выпала 4.

Событие A — "выпали все грани" опишется подмножеством пространства Ω, состоящим из последовательностей, в записи которых встречается каждая из цифр $1, 2, \ldots, 6$.

Далее, поскольку каждый исход стохастического эксперимента ничем не хуже и не лучше, чем другой (кость симметрична), то естественно считать, что все элементарные события равновероятны (ни из каких теорем это утверждение не следует). Другими словами, в качестве математической модели данного стохастического эксперимента рассмотрим классическую модель. Тем самым вероятностное пространство (модель стохастического эксперимента) построено: предложено Ω и для каждого $\omega \in \Omega$ задана вероятность $\mathsf{P}(\omega)$: $\mathsf{P}(\omega) = 1/n(\Omega)$.

Теперь вычислим вероятность события A — "выпадут все шесть граней". Поскольку модель классическая, то в соответствии с формулой классической вероятности (см. (4.2.2)), вероятность события A равна отношению числа $n(A)$ элементарных событий, входящих в событие A, к числу $n(\Omega)$ всех элементарных событий.

Элементарных событий в Ω столько, сколько существует последовательностей длиной 6, которые можно построить из шести чисел (от 1 до 6). Согласно правилу умножения, их — 6^6. Элементарных событий, входящих в событие A, столько, сколько имеется перестановок из чисел $1, 2, \ldots, 6$ — их $6!$

Следовательно, искомая вероятность

$$\mathsf{P}(A) = \frac{6!}{6^6}.$$

Пример 4.4.2 (С. Н. Бернштейн). *Подбрасывают правильный тетраэдр, три грани которого окрашены соответственно в красный, синий и зеленый цвета, а в окраске четвертой имеются все три цвета. События: R — "красный", B — "синий", G — "зеленый" обозначают, что в окраске грани, соприкасающейся с поверхностью, имеются соответствующие цвета. Убедитесь, что события R, B, G попарно независимы, но не являются независимыми в совокупности.*

Решение. Поскольку тетраэдр правильный, то в качестве математической модели стохастического эксперимента естественно рассматривать классическую модель.

Каждый цвет присутствует в окраске двух граней, поэтому

$$\mathsf{P}(R) = 2/4 = 1/2, \ \mathsf{P}(B) = 2/4 = 1/2, \ \mathsf{P}(G) = 2/4 = 1/2.$$

В окраске только одной грани имеются три цвета, поэтому

$$\mathsf{P}(R\cap B\cap G)=1/4,\ \mathsf{P}(R\cap B)=1/4,$$

$$\mathsf{P}(R\cap G)=1/4,\ \mathsf{P}(B\cap G)=1/4.$$

Так что

$$\mathsf{P}(R\cap B)=\mathsf{P}(R)\mathsf{P}(B),\ \mathsf{P}(R\cap G)=\mathsf{P}(R)\mathsf{P}(G),$$

$$\mathsf{P}(B\cap G)=\mathsf{P}(B)\mathsf{P}(G),$$

т. е. события R, B, G попарно независимы. Но

$$\mathsf{P}(R\cap B\cap G)=\frac{1}{4}\neq\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\mathsf{P}(R)\mathsf{P}(B)\mathsf{P}(G).$$

Последнее означает, что события R, B, G не являются независимыми в совокупности.

Пример 4.4.3. *Подбрасывают три симметричные игральные кости. Найти вероятность выпадения хотя бы одной единицы, если известно, что на трех костях выпали разные грани.*

Решение. В качестве пространства элементарных событий Ω рассмотрим упорядоченные тройки, составленные из чисел $1, 2, \ldots, 6$. Поскольку кости симметричны, то естественно считать все элементарные события равновероятными. Тем самым вероятностное пространство стохастического эксперимента построено.

Пусть B — событие, состоящее в том, что на трех костях выпали разные грани, A — хотя бы на одной кости выпала единица. В задаче необходимо вычислить условную вероятность $\mathsf{P}(A/B)$.

По определению

$$\mathsf{P}(A/B)=\frac{\mathsf{P}(A\cap B)}{\mathsf{P}(B)}.$$

Далее,

$$\mathsf{P}(B)=\frac{6\cdot5\cdot4}{6^3};\ \mathsf{P}(A\cap B)=\frac{3\cdot5\cdot4}{6^3};$$

$$\mathsf{P}(A/B)=\frac{\mathsf{P}(A\cap B)}{\mathsf{P}(B)}=\frac{3\cdot5\cdot4}{6^3}\Big/\frac{6\cdot5\cdot4}{6^3}=\frac{1}{2}.$$

Вероятность $\mathsf{P}(A/B)$ можно также вычислить, воспользовавшись тем, что в классической модели $\{\Omega,\mathsf{P}\}$ условная вероятность $\mathsf{P}(A/B)$ равна отношению числа элементарных событий, входящих в $A \cap B$, к числу элементарных событий, принадлежащих B, т. е.

$$\mathsf{P}(A/B) = \frac{3 \cdot 5 \cdot 4}{6 \cdot 5 \cdot 4} = \frac{1}{2}.$$

Пример 4.4.4 (обобщение задачи Льюиса Кэррола). *В урне находится один шар, о котором известно, что он либо белый (с вероятностью 1/2), либо черный. В урну положили n белых шаров, а затем после тщательного перемешивания последовательно извлекли n шаров, которые оказались белыми. Какова вероятность следующему шару оказаться белым?*

Решение. Шар, который изначально находится в урне, обозначим через W, если он белый, и через B, если он черный (так же будем обозначать и события "в урне изначально белый шар", "в урне изначально черный шар"). Белые шары, которые кладут в урну, обозначим через W_i, $i = 1, 2, \ldots, n$ (будем считать их различимыми). Эксперимент состоит в последовательном извлечении из урны $(n+1)$ шаров. Исходы эксперимента — последовательности длиной $(n+1)$, составленные из букв $W_1, W_2, \ldots, W_n, W$ или из букв $W_1, W_2, \ldots, W_n, B$. Поскольку сначала в урне с вероятностью 1/2 находится белый или черный шар, то естественно считать модель классической. Пусть A_n — событие "первые n извлеченных шаров белые", A — "последний извлеченный шар белый". Ясно, что необходимо вычислить вероятность

$$\mathsf{P}(A/A_n) = \frac{\mathsf{P}(A \cap A_n)}{\mathsf{P}(A_n)}.$$

Вычислим вероятности $\mathsf{P}(A_n)$ и $\mathsf{P}(A \cap A_n)$ по формуле полной вероятности, в качестве полной группы событий рассмотрим события W, B. Имеем

$$\mathsf{P}(A_n/W) = 1, \ \mathsf{P}(A_n/B) = \frac{n!}{(n+1)!} = \frac{1}{n+1},$$

$$\mathsf{P}(A_n) = \mathsf{P}(A_n/W)\mathsf{P}(W) + \mathsf{P}(A_n/B)\mathsf{P}(B) =$$

$$= 1 \cdot \frac{1}{2} + \frac{1}{n+1} \cdot \frac{1}{2} = \frac{1}{2}\left(1 + \frac{1}{n+1}\right),$$

$$\mathsf{P}(A \cap A_n) = \mathsf{P}((A \cap A_n)/W)\mathsf{P}(W) + \mathsf{P}((A \cap A_n)/B)\mathsf{P}(B) =$$

$$= 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}.$$

Так что

$$\mathsf{P}(A/A_n) = \frac{1/2}{(1/2)(1+1/(n+1))} = \frac{n+1}{n+2}. \qquad (4.4.1)$$

Заметим, что если в урне находится черный шар, то вероятность того, что первые n вынутых шаров окажутся белыми

$$\mathsf{P}(A_n/B) = \frac{1}{n+1},$$

а вероятность того, что среди первых n вынутых шаров окажется черный

$$\mathsf{P}(\overline{A}_n/B) = 1 - \frac{1}{n+1} = \frac{n}{n+1},$$

т. е. если в урне имеется черный шар, то он скорее всего появится среди первых n вынутых.

И если первые n вынутых шаров оказались белыми, то скорее всего черного шара в урне и не было —

$$\mathsf{P}(B/A_n) = 1 - \mathsf{P}(A/A_n) = 1 - \frac{n+1}{n+2} = \frac{1}{n+2}$$

(см. (4.4.1)).

Поэтому естественно, что если среди первых n шаров окажутся только белые, то и $(n+1)$-й шар скорее всего будет белым:

$$\mathsf{P}(A/A_n) = \frac{n+1}{n+2}.$$

Полученные результаты вполне согласуются с нашей интуицией.

Заметим, что в частном случае, когда $n = 1$, мы имеем дело с задачей Льюиса Кэррола, при этом

$$\mathsf{P}(A/A_1) = \frac{n+1}{n+2} = \frac{2}{3},$$

что совпадает с ответом к этой задаче.

Пример 4.4.5. *Среди N экзаменационных билетов n “счастливых”. Студенты подходят за билетами один за другим. У кого больше вероятность взять “счастливый” билет — у того, кто подошел первым, или у того, кто подошел вторым?*

Решение. Обозначим через A_i^l событие "i-й студент взял счастливый билет", $i = 1, 2$; A_1^u — "1-й студент взял несчастливый билет". Вычислим $\mathsf{P}(A_1^l)$ и $\mathsf{P}(A_2^l)$.

По формуле классической вероятности

$$\mathsf{P}(A_1^l) = \frac{n}{N}.$$

Вероятность события $\mathsf{P}(A_2^l)$ вычислим по формуле полной вероятности, рассматривая в качестве полной группы событий A_1^l и A_1^u:

$$\mathsf{P}(A_2^l) = \mathsf{P}(A_2^l/A_1^l)\mathsf{P}(A_1^l) + \mathsf{P}(A_2^l/A_1^u)\mathsf{P}(A_1^u) = \\ = \frac{n-1}{N-1}\frac{n}{N} + \frac{n}{N-1}\frac{N-n}{N} = \frac{n}{N}.$$

Так что, вероятность взять счастливый билет для обоих студентов одинакова.

Замечание. Такой ответ мы получаем, если первый студент не объявляет, какой билет он вытянул (счастливый или несчастливый). В противном случае вероятность вытянуть счастливый билет вторым студентом равна

$$\mathsf{P}(A_2^l/A_1^l) = (n-1)/(N-1),$$

если первый вытянул счастливый билет, и

$$\mathsf{P}(A_2^l/A_1^u) = n/(N-1)$$

— если несчастливый.

Пример 4.4.6 (задача о разделе ставки). *Два игрока играют в справедливую игру — у обоих шансы выиграть каждую партию игры одинаковы, например, подбрасывают симметричную монету, если при этом монета легла гербом, партию выиграл первый игрок, если решеткой — второй. Игроки договорились, что игру выигрывает тот, кто первым выиграет* 6 *партий, при этом выигравший получает весь приз. На самом деле игра остановилась до того, как один из игроков выиграл* 6 *партий, скажем, первый выиграл* 5 *партий, а второй —* 3. *Как справедливо разделить приз?*

Задача о разделе ставки имеет давние корни. Впервые ее опубликовал в 1494 *г. в Венеции известный математик Фра Лука Пачоли, но имеются основания считать, что задача имеет арабское происхождение. На ее решение безрезультатно было потрачено немало усилий знаменитых математиков. Неправильное решение — приз следует разделить в отношении* 2 *к* 1, *дал Никколо Тарталья* (1499 – 1557)*, хотя он был достаточно гениален, чтобы в математической дуэли за одну ночь найти формулу корней кубического уравнения. Правильное решение независимо друг от друга получили Паскаль и Ферма в* 1654 *г. Это открытие выглядело настолько важным, что многие считают* 1654 *г. годом рождения теории вероятностей.*

Решение. Паскаль и Ферма рассматривали задачу о разделе ставки как вероятностную.

Пусть вместо игроков, которые прервали игру, ее продолжают два новых игрока (один за первого, другой — за второго, мы по-прежнему будем называть их первым и вторым игроками). Они играют до выигрыша одним из игроков шести партий. При этом 6 партий может выиграть как первый игрок, так и второй. Вот только с разными вероятностями. Поэтому естественно считать справедливым раздел приза пропорционально вероятностям выиграть 6 партий соответственно первым и вторым игроками при продолжении игры.

Игра продолжается не более трех партий. Второй игрок выигрывает игру, если в трех последовательных подбрасываниях монеты выпадет три решетки. Поэтому вероятность выигрыша игры вторым игроком равна $1/8$ — вероятности выпадения трех решеток. Вероятность выигрыша игры первым игроком равна $7/8$. Следовательно, приз необходимо разделить в отношении $7:1$.

Задачи

4.1. Для уменьшения общего количества игр $2n$ команд (разных по силе) необходимо разделить на две подгруппы по n команд каждая.

1. Какова вероятность того, что две самые сильные команды окажутся: а) в разных подгруппах; б) в одной подгруппе?

2. Какова вероятность того, что четыре самые сильные команды окажутся: а) в разных подгруппах (по две в каждой); б) в одной подгруппе; в) в разных подгруппах, причем в одной подгруппе три команды, в другой — одна?

Решение. Будем различать подгруппы: подгруппа 1° и подгруппа 2°. Исход эксперимента — упорядоченная пара n-элементных подмножеств $2n$-элементного множества (пространство элементарных событий — множество всех таких пар). При этом выбор одного n-элементного подмножества пары (одной из подгрупп, например, подгруппы 1°) определяет состав другого подмножества (подгруппы 2°). Поэтому всего исходов C_{2n}^n. Модель классическая.

Событие A — "две самые сильные команды оказались в разных подгруппах" — описывается упорядоченными парами

n-элементных подмножеств, в состав каждого из которых входит ровно одна из наиболее сильных команд. Поэтому

$$\mathsf{P}(A) = C_2^1 C_{2n-2}^{n-1}/C_{2n}^n.$$

Событие B — "две самые сильные команды оказались в одной подгруппе" описывается упорядоченными парами n-элементных подмножеств, причем в состав одного n-элементного подмножества (первого или второго) входят две самые сильные команды, в состав другого — ни одной. Поэтому сначала выберем подгруппу (1° или 2°), в состав которой войдут две самые сильные команды (C_2^1 способами), а затем сформируем эту подгруппу — C_{2n-2}^{n-2} способами. Отсюда

$$\mathsf{P}(B) = C_2^1 C_{2n-2}^{n-2}/C_{2n}^n.$$

4.2. Девять пассажиров наудачу рассаживаются в три вагона. Найти вероятность того, что:

а) в каждый вагон сядут по три пассажира;

б) в один вагон сядут четыре, другой — три, а в третий — два пассажира.

Решение. Оборот "девять пассажиров рассаживаются в три вагона" обозначает, что каждому пассажиру приписывается номер вагона, в котором он оказался. Исход — последовательность (слово) длиной 9, составленная из чисел $1, 2, 3$; всего исходов 3^9. Модель классическая.

а) Событие "в каждый вагон сядут по три пассажира" описывается словами длиной 9, составленными из трех 1, трех 2, трех 3; таких слов $9!/(3!3!3!)$. Поэтому искомая вероятность равна $\dfrac{9!/(3!3!3!)}{3^9}$.

б) Событие "в один вагон сядут четыре, в другой — три, третий — два пассажира" описывается последовательностями длиной 9, образованными цифрами (буквами) $1, 2, 3$, причем одна цифра встречается четыре, другая три, третья два раза. Всего таких последовательностей $3!9!/(4!3!2!)$. В самом деле, первым действием выберем цифру, которая встретится четыре раза (тремя способами), потом цифру, которая встретится три раза (двумя способами), цифра, встречающаяся два раза, может быть выбрана одним способом. Вторым действием из выбранных цифр образуем слово длиной 9 с заданным числом букв каждого типа ($9!/(4!3!2!)$ способами).

Так что искомая вероятность равна $\dfrac{3!9!/(4!3!2!)}{3^9}$.

4.3. Доказать, что вероятнее получить при четырех подбрасываниях игральной кости хотя бы одну единицу, чем при 24 подбрасываниях пары игральных костей хотя бы один раз пару единиц. (Ответ известен как “парадокс де Мере”. Придворный кавалер и азартный игрок шевалье де Мере, современник Блеза Паскаля, считал эти вероятности равными и обвинял математиков в своих проигрышах.)

Решение. Вероятность появления хотя бы один раз единицы при четырех подбрасываниях игральной кости равна $1-5^4/6^4$, а появление пары $(1,1)$ хотя бы один раз в 24 подбрасываниях пары костей равна $1-35^{24}/36^{24}$. Необходимо доказать, что

$$1-5^4/6^4 > 1-35^{24}/36^{24}.$$

Для этого достаточно установить, что $(35/36)^{24} > (5/6)^4$ или $(35/36)^6 > 5/6$. Последнее следует, например, из неравенства $(1+x)^n \geq 1+nx$, $x>-1$, $n \geq 1$ при $n=6$, $x=-1/36$.

4.4. В лифте семь пассажиров. Лифт останавливается на десяти этажах. Вычислить вероятность того, что никакие два пассажира не выйдут на одном и том же этаже.

Ответ: $A_{10}^7/10^7$, что составляет около $0,06$.

4.5. Участник лотереи “Спортлото” из 49 названий видов спорта (обозначенных числами 1 – 49) должен назвать шесть. Полный выигрыш получает правильно указавший все шесть названий. Выигрыш получают и угадавшие не менее трех названий. Вычислить вероятность полного выигрыша в “Спортлото”. Вычислить вероятность того, что участник “Спортлото” угадает пять, четыре, три названия. Какова вероятность получить выигрыш в “Спортлото”?

Ответы: $\mathsf{P}(A_r) = \{$ участник угадал r видов $\}$, $r=3,4,5,6$; $\mathsf{P}(A_3)=0,017650$, $\mathsf{P}(A_4)=0,000969$, $\mathsf{P}(A_5)=0,000018$, $\mathsf{P}(A_6)=$ $=0,00000007151$.

4.6. Вычислить вероятность того, что четырехзначный номер случайно выбранного автомобиля в большом городе: а) состоит из различных цифр; б) имеет ровно две одинаковые цифры; в) имеет две пары одинаковых цифр; г) имеет ровно три одинаковые цифры; д) состоит из одинаковых цифр.

Ответы: а) $\dfrac{10\cdot 9\cdot 8\cdot 7}{10^4}$; б) $\dfrac{10C_9^2 4!/(2!1!1!)}{10^4}$.

в) Четырехзначный автомобильный номер — упорядоченная последовательность длиной 4 (слово), образованная из цифр $0,1,2,\ldots,9$; всего таких последовательностей 10^4. Номера, имеющие две пары одинаковых цифр, — это последовательности (слова) длиной 4, образованные двумя разными цифрами, которые встречаются по два раза, их всего $C_{10}^2 4!/(2!2!)$. (Сначала из 10 цифр выбираем две (C_{10}^2 способами), а затем из них составляем слово длиной 4, в которое каждая буква входит дважды ($4!/(2!2!)$ способами)). Поэтому искомая вероятность равна $\dfrac{C_{10}^2 4!/(2!2!)}{10^4}$.

г) $\dfrac{10 \cdot 9 \cdot 4!/(3!1!)}{10^4}$ (аналогично решению задачи (пункт в)); д) $1/10^3$.

4.7 (призовая игра). Вы участвуете в призовой игре (приз — автомобиль). Перед вами три закрытых двери. За одной из них автомобиль. Ведущий предлагает вам выбрать дверь. Вы выбираете. Прежде чем открыть выбранную вами дверь ведущий открывает одну из оставшихся дверей — за ней автомобиля нет. После этого ведущий предлагает вам два варианта последующих действий: открыть выбранную ранее вами дверь или открыть оставшуюся дверь. Если за открытой дверью автомобиль — он ваш.

Какой выбор сделать — открыть первоначально выбранную дверь или открыть оставшуюся?

Естественно, что предпочтительнее тот выбор, который обеспечивает большую вероятность выигрыша автомобиля.

Вычислите вероятность выигрыша автомобиля, если: 1) вы открываете первоначально выбранную дверь; 2) вы открываете оставшуюся дверь.

Решение. Вероятность того, что автомобиль находится за данной (выбранной вами) дверью, равна $1/3$, а следовательно, вероятность того, что за данной дверью автомобиля нет — автомобиль находится за оставшимися дверями — равна $2/3$ (вне зависимости от того открыл ведущий дверь или нет). Поэтому следует отказаться от первоначального выбора двери и выбрать оставшуюся дверь (одну дверь ведущий уже открыл и за ней автомобиля не оказалось).

Предложенное решение особенно “прозрачно”, если автомобиль находится за одной из n дверей (например, $n = 1000$).

Вероятность автомобилю находиться за выбранной вами дверью равна $1/n$, и следовательно, вероятность противоположного события — автомобиль находится за оставшимися дверями — равна $1 - 1/n$. При $n = 1000$ эта вероятность составляет $0,999$. Так что открыв оставшуюся дверь, вы почти наверняка уедете домой на автомобиле.

Глава 5

Дискретная случайная величина

Вероятностное распределение как модель стохастического эксперимента. Рассмотрим несколько примеров стохастических экспериментов.

Подбрасывают игральную кость и регистрируют число выпавших очков.

Подбрасывают монету и игральную кость и регистрируют число выпавших гербов и число выпавших очков.

На отрезок $[0, 1]$ наудачу бросают точку и регистрируют ее координату.

Регистрируют время безотказной работы каждого из n приборов.

В стохастических экспериментах, описанных в примерах, исходы "числовые": либо число, либо последовательность чисел (вектор).

Математической моделью стохастического эксперимента с "числовым" множеством исходов, как и каждого стохастического эксперимента, является вероятностное пространство $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, но частного вида: множеством исходов Ω является $\mathbb{R}^n$ или его часть. В качестве σ-алгебры $\mathfrak{F}$ подмножеств $\mathbb{R}^n$, как правило, рассматривается "стандартная" σ-алгебра, так называемая σ-алгебра борелевских множеств на $\mathbb{R}^n$ — наименьшая σ-алгебра, содержащая параллелепипеды, она обозначается через $\mathfrak{B}^n$ (подробнее о σ-алгебре $\mathfrak{B}^n$ в разд. 7.2. гл. 7). Вероятность на σ-алгебре борелевских подмножеств пространства $\mathbb{R}^n$ обычно называют *вероятностным распределением* на $\mathbb{R}^n$.

Вероятностное пространство $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ с $\Omega = \mathbb{R}^n$ и $\mathfrak{F} = \mathfrak{B}^n$ полностью определяется заданием вероятностного распределения (вероятности) на $\mathbb{R}^n$, поэтому вероятностное распределение на $\mathbb{R}^n$ рассматривают в качестве математической модели стохастического эксперимента с "числовым" исходом. В частности, если множество исходов $X \subset \mathbb{R}^n$ стохастического эксперимента конечно или счетно (кратко, дискретно), то математической моделью стохастического эксперимента является дискретное вероятностное распределение на $\mathbb{R}^n$.

Примеры

1. Подбрасывают симметричную игральную кость и регистрируют число выпавших очков.

В качестве математической модели этого стохастического эксперимента естественно рассмотреть вероятностное распределение (вероятность) на подмножестве $\{1, 2, \ldots, 6\}$ прямой $\mathbb{R}^1$, задаваемое равенствами

$$\mathsf{P}(i) = \frac{1}{6}, \; i = 1, 2, \ldots 6.$$

2. На отрезок $[a, b]$ наудачу бросают точку и регистрируют ее координату.

В качестве математической модели этого стохастического эксперимента естественно рассмотреть вероятностное распределение, заданное на наименьшей σ-алгебре, содержащей промежутки, и такое, что значение распределения на каждом промежутке $[a', b']$ из $[a, b]$ пропорционально длине промежутка, т. е.

$$\mathsf{P}([a', b']) = (b' - a')/(b - a), \; [a', b'] \subset [a, b].$$

Далее мы рассмотрим этот пример подробнее.

5.1 Случайная величина как функция на множестве исходов

Результат стохастического эксперимента с исходами в $\mathbb{R}^n$ естественно рассматривать и как величину, зависящую от вмешательства случая. Поэтому помимо вероятностного распределения на $\mathbb{R}^n$, как математической модели стохастического эксперимента с исходами в $\mathbb{R}^n$, в качестве его математической модели можно предложить и другую естественную и содержательную модель.

На интуитивном уровне под величиной, зависящей от вмешательства случая, мы понимаем величину, которая принимает те или иные значения в зависимости от исхода стохастического эксперимента. Исходы стохастического эксперимента описываются точками ω пространства Ω элементарных событий вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$. Поэтому, формализуя понятие величины, зависящей от исхода стохастического эксперимента, в качестве ее математической модели естественно рассмотреть функцию $\xi = \xi(\omega)$, заданную на множестве исходов (элементарных событий) Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$. Такие функции мы будем называть случайными величинами.

Примеры

1. Монету, вероятность выпадения герба которой p, подбрасывают независимым образом до первого выпадения герба и регистрируют число ξ выпавших решеток — число подбрасываний до первого выпадения герба.

Стохастический эксперимент, состоящий в подбрасывании монеты до первого выпадения герба, описывается вероятностным пространством $\{\Omega, \mathsf{P}\}$, где

$$\Omega = \{\text{Г}, \text{РГ}, \text{РРГ}, \ldots\},\ \mathsf{P}(\underbrace{\text{РР}\ldots\text{Р}}_{k}\text{Г}) = (1-p)^k p,\ k = 0, 1, \ldots$$

Число подбрасываний монеты до первого выпадения герба можно рассматривать как функцию $\xi = \xi(\omega)$, определенную на Ω и принимающую в точке $\omega = (\underbrace{\text{РР}\ldots\text{Р}}_{k}\text{Г})$ значение k, $k = 0, 1, \ldots$

2. Подбрасывают симметричную монету и симметричную игральную кость и регистрируют: а) ξ — число выпавших гербов, б) η — число выпавших очков, в) $\zeta = (\xi, \eta)$ — число выпавших гербов и число выпавших очков.

Стохастический эксперимент, состоящий в подбрасывании симметричной монеты и симметричной игральной кости, описывается вероятностным пространством $\{\Omega, \mathsf{P}\}$, точки ω из Ω — упорядоченные пары (si), где s принимает “значения” Г или Р, i принимает значения $1, 2, \ldots, 6$, а $\mathsf{P}(si) = 1/12$ для каждой пары (si).

Число выпавших гербов ξ, число выпавших очков η, число выпавших гербов и очков $\zeta = (\xi, \eta)$ можно рассматривать как функции $\xi = \xi(\omega)$, $\eta = \eta(\omega)$, $\zeta = \zeta(\omega) = (\xi(\omega), \eta(\omega))$ на Ω, заданные равенствами:

а) функция $\xi = \xi(\omega)$:

$$\xi(\text{Г}i) = 1,\ \xi(\text{Р}i) = 0,\ i = 1, 2, \ldots, 6;$$

б) функция $\eta = \eta(\omega)$:

$$\eta(\text{Г}i) = i,\ \eta(\text{Р}i) = i,\ i = 1, 2, \ldots, 6;$$

в) функция $\zeta = \zeta(\omega) = (\xi(\omega), \eta(\omega))$:

$$\zeta(\Gamma i) = (1, i),\ \zeta(\mathrm{P} i) = (0, i),\ i = 1, 2, \ldots, 6.$$

Резюме. *Итак, для стохастического эксперимента с “числовым” исходом можно предложить две математические модели:*

1) *вероятностное распределение на* $\mathbb{R}^n$*;*

2) *функцию* $\xi = \xi(\omega)$ *со значениями в* $\mathbb{R}^n$ *на множестве элементарных событий* (*исходов*) Ω *некоторого вероятностного пространства* $\{\Omega, \mathfrak{F}, \mathsf{P}\}$*.*

Случайная величина на вероятностном пространстве. Функция $\xi = \xi(\omega)$ на Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ со значениями в $\mathbb{R}^n$, предлагаемая (рассматриваемая) в качестве модели величины, находящейся под воздействием случайных факторов, не может быть произвольной. Она должна удовлетворять некоторым естественным требованиям, а именно: для множеств $B \subset \mathbb{R}^n$ из достаточно широкого класса (в частности, для параллелепипедов, а в случае $\mathbb{R}^1$ для отрезков) можно говорить о вероятности попадания значения ξ в B.

Если мы рассматриваем величину ξ, зависящую от случая, как функцию $\xi = \xi(\omega)$ точки на множестве исходов Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ (как функцию исхода ω стохастического эксперимента), то говорить о вероятности попадания ξ в множество $B \subset \mathbb{R}^n$ можно тогда, когда можно говорить о вероятности множества $\{\omega : \xi(\omega) \in B\}$. Но вероятность определена только для множеств из класса $\mathfrak{F}$ (из σ-алгебры событий $\mathfrak{F}$), поэтому чтобы функцию $\xi = \xi(\omega)$ можно было рассматривать в качестве модели величины, зависящей от вмешательства случая, множество $\{\omega : \xi(\omega) \in B\}$ должно принадлежать классу $\mathfrak{F}$ (должно быть событием).

Случайная величина на дискретном вероятностном пространстве. Для функции $\xi = \xi(\omega)$, заданной на Ω дискретного вероятностного пространства $\{\Omega, \mathsf{P}\}$, всегда можно говорить о вероятности попадания значения ξ в $B \subset \mathbb{R}^n$, поскольку для любого $B \subset \mathbb{R}^n$ множество $\{\omega : \xi(\omega) \in B\}$ является событием и, следовательно, его вероятность определена (в общем случае необходимо требовать, чтобы множество $\{\omega : \xi(\omega) \in B\}$ было событием). Поэтому любую функцию $\xi = \xi(\omega)$ на дискретном Ω можно рассматривать в качестве математической модели величины, зависящей от случая, такие функции будем называть случайными величинами.

Определение. *Случайной величиной* со значениями в $\mathbb{R}^n$ на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$ будем называть функцию $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$, заданную на Ω вероятностного пространства $\{\Omega, \mathsf{P}\}$.

Если $n > 1$, то случайную величину

$$\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$$

называют *векторной*, если $n = 1$ — *скалярной.*

Случайная величина (функция $\xi = \xi(\omega)$) на дискретном $\{\Omega, \mathsf{P}\}$ принимает не более чем счетное число значений, такая случайная величина называется *дискретной.*

Замечание. Если $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — векторная случайная величина со значениями в $\mathbb{R}^n$, то, очевидно, каждая из компонент ξ_j, $j = 1, 2, \ldots, n$, является случайной величиной со значениями в $\mathbb{R}^1$ (и наоборот).

Распределение дискретной случайной величины $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\omega})$. Напомним, что на дискретном $X \subset \mathbb{R}^n$ неотрицательная функция точки $\mathsf{Q}(x)$, $x \in X$ такая, что $\sum\limits_{x \in X} \mathsf{Q}(x) = 1$, задает на классе всех подмножеств X неотрицательную счетно-аддитивную нормированную функцию множества

$$\mathsf{Q}(B) = \sum_{x \in B} \mathsf{Q}(x), \ B \subset X.$$

(см. (4.1.3)). Как функция множества $\mathsf{Q}(B)$, $B \subset X$, так и функция точки $\mathsf{Q}(x)$, $x \in X$, такая, что $\sum\limits_{x \in X} \mathsf{Q}(x) = 1$, называется распределением на дискретном X.

Пусть $\xi = \xi(\omega)$ — случайная величина со значениями в $\mathbb{R}^n$ на дискретном $\{\Omega, \mathsf{P}\}$. Точку $x \in \mathbb{R}^n$ будем называть *возможным значением* случайной величины ξ, если

$$\mathsf{P}\{\xi = x\} = \mathsf{P}\{\omega : \xi(\omega) = x\} > 0.$$

Множество возможных значений случайной величины ξ будем обозначать через X ($X \subset \mathbb{R}^n$).

Случайная величина $\xi = \xi(\omega)$, определенная на вероятностном пространстве $\{\Omega, \mathsf{P}\}$, *всегда* задает на множестве X своих возможных значений вероятностное распределение, называемое распределением случайной величины ξ.

Определение. *Распределением дискретной случайной величины $\xi = \xi(\omega)$* (со значениями в $\mathbb{R}^n$) будем называть функцию

$$\mathsf{P}_\xi : x \to \mathsf{P}_\xi(x) = \mathsf{P}\{\xi = x\}, \ x \in X, \tag{5.1.1}$$

заданную на множестве X различных возможных значений случайной величины ξ и ставящую в соответствие каждому возможному значению $x \in X$ вероятность $\mathsf{P}_\xi(x) = \mathsf{P}\{\xi = x\}$, с которой это значение принимается.

Для функции точки

$$\mathsf{P}_\xi : x \to \mathsf{P}_\xi(x) = \mathsf{P}\{\xi = x\},\ x \in X,$$

всегда

$$\sum_{x \in X} \mathsf{P}_\xi(x) = 1.$$

Последнее равенство следует из свойства ассоциативности абсолютно сходящихся рядов и представления

$$\bigcup_{x \in X} \{\omega : \xi(\omega) = x\} = \Omega,$$

$$\{\omega : \xi(\omega) = x_i\} \cap \{\omega : \xi(\omega) = x_j\} = \emptyset,\ x_i \neq x_j,$$

так:

$$1 = \mathsf{P}\left(\bigcup_{x \in X} \{\omega : \xi(\omega) = x\}\right) = \sum_{x \in X} \mathsf{P}\{\omega : \xi(\omega) = x\} =$$

$$= \sum_{x \in X} \mathsf{P}\{\xi = x\} = \sum_{x \in X} \mathsf{P}_\xi(x).$$

Так что распределение $\mathsf{P}_\xi(x)$, $x \in X$, случайной величины является вероятностным распределением на множестве ее значений.

Распределение случайной величины $\xi = \xi(\omega)$

$$\mathsf{P}_\xi : x \to \mathsf{P}_\xi(x) = \mathsf{P}\{\xi = x\},\ x \in X,$$

— функция точки на множестве X ее возможных значений — задает на классе подмножеств пространства X неотрицательную счетно-аддитивную нормированную функцию множества

$$\mathsf{P}_\xi(B) = \sum_{x \in B} \mathsf{P}_\xi(x),\ B \subset X, \tag{5.1.2}$$

(см. (4.1.3)), которую мы также будем называть распределением случайной величины $\xi = \xi(\omega)$.

Далее, *для любого подмножества $B \subset X$ имеет место равенство*

$$\mathsf{P}_\xi(B) = \mathsf{P}\{\omega : \xi(\omega) \in B\}. \tag{5.1.3}$$

Оно следует из свойства ассоциативности абсолютно сходящихся рядов и представления

$$\{\omega : \xi(\omega) \in B\} = \bigcup_{x \in B} \{\omega : \xi(\omega) = x\},$$

$$\{\omega : \xi(\omega) = x_i\} \cap \{\omega : \xi(\omega) = x_j\} = \emptyset, \ x_i \neq x_j,$$

так

$$\mathsf{P}\{\omega : \xi(\omega) \in B\} = \mathsf{P}\left(\bigcup_{x \in B} \{\omega : \xi(\omega) = x\}\right) =$$

$$= \sum_{x \in B} \mathsf{P}\{\omega : \xi(\omega) = x\} = \sum_{x \in B} \mathsf{P}_\xi(x) = \mathsf{P}_\xi(B).$$

На равенство (5.1.3) можно смотреть как на определение распределения $\mathsf{P}_\xi(B)$, $B \subset X$, случайной величины ξ.

Случайная величина $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$ на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$ задает дискретное вероятностное распределение P_ξ на множестве своих значений X. Справедливо и обратное.

Теорема 5.1.1. *Пусть Q: $x \to \mathsf{Q}(x)$ — вероятностное распределение на дискретном $X \subset \mathbb{R}^n$. Существуют вероятностное пространство $\{\Omega, \mathsf{P}\}$ и случайная величина $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$ на $\{\Omega, \mathsf{P}\}$ такие, что распределение P_ξ случайной величины ξ совпадает с распределением Q.*

Достаточно рассмотреть в качестве Ω множество X, в качестве вероятности P — распределение Q, а случайную величину $\xi = \xi(\omega)$ определить как тождественное отображение $\xi = \xi(\omega) = \omega$. Тогда

$$\mathsf{P}_\xi(x) = \mathsf{P}\{\omega : \xi(\omega) = x\} = \mathsf{P}(x) = \mathsf{Q}(x),$$

т. е.

$$\mathsf{P}_\xi = \mathsf{Q}.$$

Теорема устанавливает связь между моделью случайной величины как вероятностным распределением на $\mathbb{R}^n$ и моделью случайной величины как функцией на вероятностном пространстве со значениями в $\mathbb{R}^n$ и, в частности, объясняет часто встречающийся оборот: "случайная величина определяется (задается) своим распределением".

Формы записи распределения случайной величины. Распределение

$$\mathsf{P}_\xi\colon x \to \mathsf{P}_\xi(x),\ x \in X \subset \mathbb{R}^n,$$

случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ со значениями в $\mathbb{R}^n$ в координатной форме записывается так:

$$\mathsf{P}_\xi\colon (x_1, x_2, \ldots, x_n) \to \mathsf{P}_\xi(x_1, x_2, \ldots, x_n),\ (x_1, x_2, \ldots, x_n) = x \in X,$$

где X — множество различных возможных значений случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, а

$$\mathsf{P}_\xi(x_1, x_2, \ldots, x_n) = \mathsf{P}\{(\xi_1, \xi_2, \ldots, \xi_n) = (x_1, x_2, \ldots, x_n)\}$$

— вероятности, с которыми $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ принимает значения $x = (x_1, x_2, \ldots, x_n)$.

Распределение

$$\mathsf{P}_\xi\colon (x_1, x_2, \ldots, x_n) \to \mathsf{P}_\xi(x_1, x_2, \ldots, x_n)$$

векторной случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ еще называют *совместным распределением* случайных величин $\xi_1, \xi_2, \ldots, \xi_n$.

Если $\xi = \xi(\omega)$ — случайная величина со значениями в $\mathbb{R}^1$, то ее распределение

$$\mathsf{P}_\xi : x_i \to \mathsf{P}_\xi(x_i),\ x_i \in X \subset \mathbb{R}^1$$

часто записывают в виде таблицы, в верхней строке которой указывают различные возможные значения случайной величины ξ, а в нижней — вероятности, с которыми они принимаются:

x_i	x_1	$\cdots$	x_n	$\cdots$
$\mathsf{P}_\xi(x_i)$	$\mathsf{P}_\xi(x_1)$	$\cdots$	$\mathsf{P}_\xi(x_n)$	$\cdots$

.

Если $\zeta = (\xi, \eta)$ — случайная величина со значениями в $\mathbb{R}^2$, то ее распределение

$$\mathsf{P}_\zeta : (x_i, y_j) \to \mathsf{P}_\zeta(x_i, y_j),\ (x_i, y_j) \in X \subset \mathbb{R}^2$$

удобно записывать в виде табл. 5.1.1.

Таблица 5.1.1. Совместное распределение случайных величин ξ и η

Значения ξ	Значения η				
	y_1	y_2	$\cdots$	y_k	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1k}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2k}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$
x_s	p_{s1}	p_{s2}	$\cdots$	p_{sk}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

.

В табл. 5.1.1 $p_{sk} = \mathsf{P}_\zeta(x_s, y_k)$, $s = 1, 2, \ldots,$ $k = 1, 2, \ldots$

Замечание. Когда говорят "заданы две случайные величины", "задано n случайных величин", имеется в виду, что задана векторная случайная величина со значениями в $\mathbb{R}^2$, со значениями в $\mathbb{R}^n$.

Пример 5.1.1. *Пусть ξ — число выпавших гербов при подбрасывании пары симметричных монет. Найти распределение случайной величины ξ.*

Решение. Стохастический эксперимент, состоящий в подбрасывании пары монет, описывается вероятностным пространством $\{\Omega, \mathsf{P}\}$, где $\Omega = \{\text{ГГ}, \text{ГР}, \text{РГ}, \text{РР}\}$, $\mathsf{P}(\text{ГГ}) = 1/4, \mathsf{P}(\text{ГР}) = 1/4, \mathsf{P}(\text{РГ}) = 1/4, \mathsf{P}(\text{РР}) = 1/4$ (равенство вероятностей исходов отражает свойство симметричности монеты).

Число ξ выпавших гербов — функция $\xi = \xi(\omega)$ на Ω:

$$\xi(\text{ГГ}) = 2,\ \xi(\text{ГР}) = 1,\ \xi(\text{РГ}) = 1,\ \xi(\text{РР}) = 0.$$

Вычислим $\mathsf{P}_\xi(i) = \mathsf{P}\{\xi = i\}$, $i = 0, 1, 2$:

$$\mathsf{P}\{\xi = 0\} = \mathsf{P}(\text{РР}) = 1/4, \mathsf{P}\{\xi = 2\} = \mathsf{P}(\text{ГГ}) = 1/4,$$

$$\mathsf{P}\{\xi = 1\} = \mathsf{P}(\{\text{ГР}, \text{РГ}\}) = \mathsf{P}(\text{РГ}) + \mathsf{P}(\text{ГР}) = 1/4 + 1/4 = 1/2.$$

Тем самым распределение случайной величины ξ получено. Его можно записать и в виде

x_i	0	1	2
$\mathsf{P}_\xi(x_i)$	1/4	1/2	1/4

.

Распределение функции от случайной величины. Функция $g(\zeta)$ от случайной величины ζ является случайной величиной. *По распределению случайной величины ζ всегда можно найти распределение любой функции $g(\zeta)$ от ζ.*

Теорема 5.1.2 (о распределении функции от $\boldsymbol{\zeta}$). *Пусть ζ — случайная величина со значениями в $\mathbb{R}^n$, заданная на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$, и*

$$\mathsf{P}_\zeta : x \to \mathsf{P}_\zeta(x),\ x \in X \subset \mathbb{R}^n,$$

ее распределение; g — функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^l$. Для любого подмножества $B \subset \mathbb{R}^l$

$$\mathsf{P}\{g(\zeta) \in B\} = \sum_{x: g(x) \in B} \mathsf{P}_\zeta(x). \tag{5.1.1}$$

В частности, если $\zeta = (\xi, \eta)$ — случайная величина со значениями в $\mathbb{R}^2$,

$$\mathsf{P}_\zeta : (x_i, y_j) \to \mathsf{P}_\zeta(x_i, y_j),\ (x_i, y_j) \in X \subset \mathbb{R}^2,$$

— ее распределение, $g = g(x, y)$ — функция на $\mathbb{R}^2$ со значениями в $\mathbb{R}^l$, то для любого $B \subset \mathbb{R}^l$

$$\mathsf{P}\{g(\zeta) \in B\} = \mathsf{P}\{g(\xi, \eta) \in B\} = \sum_{(x_i, y_j): g(x_i, y_j) \in B} \mathsf{P}_\zeta(x_i, y_j). \tag{5.1.2}$$

Доказательство. Воспользовавшись свойством ассоциативности абсолютно сходящихся рядов, и представлением

$$\{\omega : g(\zeta(\omega)) \in B\} = \{\omega : \zeta(\omega) \in g^{-1}(B)\} =$$

$$= \bigcup_{x: x \in g^{-1}(B)} \{\omega : \zeta(\omega) = x\} = \bigcup_{x: g(x) \in B} \{\omega : \zeta(\omega) = x\},$$

$$\{\omega : \zeta(\omega) = x_i\} \cap \{\omega : \zeta(\omega) = x_j\} = \emptyset,\ x_i \neq x_j,$$

получим

$$\mathsf{P}\{g(\zeta) \in B\} = \sum_{x: g(x) \in B} \mathsf{P}\{\omega : \zeta(\omega) = x\} =$$

$$= \sum_{x:g(x)\in B} \mathsf{P}\{\zeta = x\} = \sum_{x:g(x)\in B} \mathsf{P}_\zeta(x).$$

В терминах координат утверждение теоремы запишется так:

$$\mathsf{P}\{g(\xi_1, \xi_2, \ldots, \xi_n) \in B\} = \sum_{(x_1,x_2,\ldots,x_n):g(x_1,x_2,\ldots,x_n)\in B} \mathsf{P}_\zeta(x_1, x_2, \ldots, x_n).$$

Следствие 1. *Пусть ξ — случайная величина со значениями в $\mathbb{R}^n$, заданная на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$. Для каждого $B \subset \mathbb{R}^n$*

$$\mathsf{P}\{\xi \in B\} = \sum_{x:x\in B} \mathsf{P}_\xi(x) \tag{5.1.3}$$

или в терминах координат

$$\mathsf{P}\{(\xi_1, \xi_2, \ldots, \xi_n) \in B\} = \sum_{(x_1,x_2,\ldots,x_n):(x_1,x_2,\ldots,x_n)\in B} \mathsf{P}_\xi(x_1, x_2, \ldots, x_n)$$

Соотношение (5.1.3) получается из равенства (5.1.1), достаточно в качестве g рассмотреть $g(x) = x$ — тождественное отображение $\mathbb{R}^n$ на $\mathbb{R}^n$.

Следствие 2 (вычисление распределения компоненты вектора). *Пусть $\mathsf{P}_\xi(\underbrace{x_1, x_2, \ldots,}_{s-1} x_s, \ldots, x_n)$ — распределение вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_s, \ldots, \xi_n)$ (совместное распределение случайных величин $\xi_1, \xi_2, \ldots, \xi_s, \ldots, \xi_n$), тогда*

$$\mathsf{P}_{\xi_s}(x) = \mathsf{P}\{\xi_s = x\} = \sum_{(x_1,x_2,\ldots,x_s,\ldots,x_n):x_s=x} \mathsf{P}_\xi(x_1, x_2, \ldots, x_s, \ldots, x_n).$$

Доказательство. Обозначим

$$B_x = \underbrace{\mathbb{R}^1 \times \mathbb{R}^1 \times \ldots \times \mathbb{R}^1}_{s-1} \times \{x\} \times \mathbb{R}^1 \times \ldots \times \mathbb{R}^1.$$

$$\mathsf{P}_{\xi_s}(x) = \mathsf{P}\{\xi_s = x\} = \mathsf{P}\{(\xi_1, \xi_2, \ldots, \xi_s, \ldots, \xi_n) \in B_x\} =$$

$$= \sum_{(x_1,x_2,\ldots,x_s,\ldots,x_n)\in B_x} \mathsf{P}_\xi(x_1, x_2, \ldots, x_s, \ldots, x_n) =$$

$$= \sum_{(x_1,x_2,\ldots,x_s,\ldots,x_n):x_s=x} \mathsf{P}_\xi(x_1,x_2,\ldots,x_s,\ldots,x_n).$$

В частности, для случайных величин со значениями в $\mathbb{R}^2$ (часто встречающийся случай) имеем:

Следствие 3. *Пусть* $\mathsf{P}_\zeta(x_i,y_j)$ *— распределение вектора* $\zeta=(\xi,\eta)$ *(совместное распределение случайных величин* ξ,η*), тогда*

$$\mathsf{P}_\xi(x) = \sum_{y_j} \mathsf{P}_\zeta(x,y_j),$$

$$\mathsf{P}_\eta(y) = \sum_{x_i} \mathsf{P}_\zeta(x_i,y).$$

Действительно,

$$\mathsf{P}_\xi(x) = \mathsf{P}\{\xi=x\} = \mathsf{P}\{(\xi,\eta)\in(\{x\}\times\mathbb{R}^1)\} =$$

$$= \sum_{(x_i,y_j)\in\{x\}\times\mathbb{R}^1} \mathsf{P}_\zeta(x_i,y_j) = \sum_{(x_i,y_j):x_i=x} \mathsf{P}_\zeta(x_i,y_j) = \sum_{y_j} \mathsf{P}_\zeta(x,y_j).$$

Аналогично проверяется второе равенство.

Замечание. Следствия 2 и 3 утверждают, что по совместному распределению случайных величин всегда можно найти распределение каждой из них; обратное, вообще говоря, не имеет места.

Пример 5.1.2. *Два раза подбрасывают две симметричные монеты, пусть* ξ *— число выпавших гербов при первом подбрасывании,* η *— при втором. Найти распределение случайной величины* $\zeta=\min\{\xi,\eta\}$.

Решение. Стохастический эксперимент, состоящий в двукратном подбрасывании двух симметричных монет, описывается вероятностным пространством $\{\Omega,\mathsf{P}\}$, у которого исходы ω равновероятные последовательности “длиной 4”, составленные из букв Г и Р. Например, исход $\omega=$ (ГРГГ) означает, что при первом подбрасывании двух монет на первой монете выпал герб, на второй — решетка, а при втором подбрасывании — на обеих монетах выпал герб. Всего равновероятных исходов 16, поэтому вероятность каждого исхода равна $1/16$.

Сначала найдем распределение случайной величины (ξ,η). Оно приведено в таблице:

Значения ξ	Значения η		
	0	1	2
0	1/16	2/16	1/16
1	2/16	4/16	2/16
2	1/16	2/16	1/16

.

Например, $\mathsf{P}_{\xi,\eta}(0,1)$ получается так:

$$\mathsf{P}_{\xi,\eta}(0,1) = \mathsf{P}\{(\xi,\eta) = (0,1)\} = \mathsf{P}\{\omega : (\xi(\omega), \eta(\omega)) = (0,1)\} =$$

$$= \mathsf{P}\{(\text{РРГР}), (\text{РРРГ})\} = 1/16 + 1/16 = 1/8.$$

Аналогично вычисляются $\mathsf{P}_{\xi,\eta}(i,j)$ для других пар (i,j).

Далее, по распределению случайной величины (ξ,η) всегда можно найти распределение любой функции g от (ξ,η):

$$\mathsf{P}\{g(\xi,\eta) \in B\} = \sum_{(x_i,y_j):g(x_i,y_j)\in B} \mathsf{P}_{\xi,\eta}(x_i,y_j).$$

В частности, для функции $g(\xi,\eta) = \min\{\xi,\eta\}$ имеем:

$$\mathsf{P}\{\min\{\xi,\eta\} = k\} = \sum_{(i,j):\min\{i,j\}=k} \mathsf{P}_{\xi,\eta}(i,j), \ \ k = 0,1,2.$$

При $k = 0$

$$\mathsf{P}\{\min\{\xi,\eta\} = 0\} = \sum_{(i,j):\min\{i,j\}=0} \mathsf{P}_{\xi,\eta}(i,j) =$$

$$= \mathsf{P}_{\xi,\eta}(0,0) + \mathsf{P}_{\xi,\eta}(0,1) + \mathsf{P}_{\xi,\eta}(0,2) + \mathsf{P}_{\xi,\eta}(1,0) + \mathsf{P}_{\xi,\eta}(2,0) =$$
$$= \frac{1}{16} + \frac{2}{16} + \frac{1}{16} + \frac{2}{16} + \frac{1}{16} = \frac{7}{16}.$$

Аналогично,

$$\mathsf{P}\{\min\{\xi,\eta\} = 1\} = \sum_{(i,j):\min\{i,j\}=1} \mathsf{P}_{\xi,\eta}(i,j) =$$

$$= \mathsf{P}_{\xi,\eta}(1,1) + \mathsf{P}_{\xi,\eta}(2,1) + \mathsf{P}_{\xi,\eta}(1,2) = \frac{4}{16} + \frac{2}{16} + \frac{2}{16} = \frac{8}{16} = \frac{1}{2}.$$

$$\mathsf{P}\{\min\{\xi,\eta\}=2\}=\frac{1}{16}.$$

Так что случайная величина $\min\{\xi,\eta\}$ имеет распределение

x_i	0	1	2
$\mathsf{P}_\zeta(x_i)$	7/16	1/2	1/16

.

5.2 Независимые случайные величины

Интуитивно две случайные величины ξ и η независимы, если информация о значении, принятом одной случайной величиной (например, ξ приняла значение из множества A), не меняет прогноз относительно значений, принимаемых другой случайной величиной (например, η приняла значение из множества B).

Примеры

1. Подбрасывают две монеты и игральную кость, ξ — число выпавших гербов на монетах; η — число выпавших очков на грани игральной кости. Случайные величины ξ и η независимы (информация о значении, принятом одной случайной величиной, не меняет прогноз значений, принимаемых другой случайной величиной).

2. Подбрасывают две монеты: ξ — число гербов, выпавших на первой монете; η — число гербов, выпавших на обеих монетах. Случайные величины ξ и η не являются независимыми (информация о значении, принятом случайной величиной ξ, меняет прогноз значений, принимаемых случайной величиной η).

Учитывая интерпретацию вероятности и условной вероятности события как количественной меры прогноза его появления (см. замечание к определению условной вероятности в разд. 3.3 гл. 3), независимыми естественно назвать случайные величины, для которых

$$\mathsf{P}\{\eta\in B/\xi\in A\}=\mathsf{P}\{\eta\in B\},$$

или, что то же,

$$\mathsf{P}\{\eta\in B\}=\frac{\mathsf{P}\{\xi\in A,\eta\in B\}}{\mathsf{P}\{\xi\in A\}},$$

или, в симметричном виде,

$$\mathsf{P}\{\xi\in A,\eta\in B\}=\mathsf{P}\{\xi\in A\}\mathsf{P}\{\eta\in B\}.$$

Последнее равенство мотивирует следующее формальное определение независимых случайных величин.

Определение. Случайные величины ξ и η со значениями в $\mathbb{R}^1$ будем называть *независимыми*, если для любых множеств A и B из $\mathbb{R}^1$

$$\mathsf{P}\{\xi \in A, \eta \in B\} = \mathsf{P}\{\xi \in A\}\mathsf{P}\{\eta \in B\}.$$

Определение. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ будем называть *независимыми в совокупности* (*независимыми*), если для любых множеств $A_1, A_2, \ldots, A_n$ из $\mathbb{R}^1$

$$\mathsf{P}\{\xi_1 \in A_1, \xi_2 \in A_2, \ldots, \xi_n \in A_n\} =$$

$$= \mathsf{P}\{\xi_1 \in A_1\}\mathsf{P}\{\xi_2 \in A_2\} \ldots \mathsf{P}\{\xi_n \in A_n\}.$$

Определение. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ будем называть *попарно независимыми*, если независимы ξ_i и ξ_j для любых $i \neq j$; $i, j = 1, 2, \ldots, n$.

Теорема. *Пусть f и g — функции на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$. Если ξ и η — независимые случайные величины, то $f(\xi)$ и $g(\eta)$ — независимые случайные величины (функции от независимых случайных величин являются независимыми случайными величинами).*

Доказательство.

$$\mathsf{P}\{f(\xi) \in A, g(\eta) \in B\} = \mathsf{P}\{\xi \in f^{-1}(A), \eta \in g^{-1}(B)\} =$$

$$= \mathsf{P}\{\xi \in f^{-1}(A)\}\mathsf{P}\{\eta \in g^{-1}(B)\} = \mathsf{P}\{f(\xi) \in A\}\mathsf{P}\{g(\eta) \in B\}.$$

Совместное распределение независимых случайных величин. По распределению вектора (ξ, η) (по совместному распределению случайных величин ξ и η) всегда можно найти распределение каждой компоненты (распределение каждой из случайных величин ξ и η), см. следствие из теоремы 5.1.2. Обратное утверждение, вообще говоря, не имеет места, исключая важный часто встречающийся случай независимых случайных величин, когда совместное распределение ξ и η определяется распределениями ξ и η — а именно, равно их произведению.

Теорема 5.2.1 (о совместном распределении независимых случайных величин). *Совместное распределение $\mathsf{P}_{\xi,\eta}$ независимых случайных величин ξ и η равно произведению их распределений P_ξ и P_η, т. е.*

$$\mathsf{P}_{\xi,\eta}(x_i, y_j) = \mathsf{P}_\xi(x_i)\mathsf{P}_\eta(y_j) \tag{5.2.1}$$

для всех возможных значений x_i *и* y_j *случайных величин* ξ *и* η.

Справедливо и обратное утверждение: если совместное распределение случайных величин ξ *и* η *равно произведению их распределений, то случайные величины* ξ *и* η *независимы.*

Доказательство. Пусть ξ и η независимые случайные величины, т. е. для любых $A, B \subset \mathbb{R}^1$

$$\mathsf{P}\{\xi \in A, \eta \in B\} = \mathsf{P}\{\xi \in A\}\mathsf{P}\{\eta \in B\}.$$

Положив, в частности, $A = \{x_i\}, B = \{y_j\}$, получим:

$$\mathsf{P}\{\xi = x_i, \eta = y_j\} = \mathsf{P}\{\xi = x_i\}\mathsf{P}\{\eta = y_j\}$$

или

$$\mathsf{P}_{\xi,\eta}(x_i, y_j) = \mathsf{P}_\xi(x_i)\mathsf{P}_\eta(y_j)$$

— совместное распределение независимых случайных величин ξ и η равно произведению их распределений.

Пусть теперь

$$\mathsf{P}_{\xi,\eta}(x_i, y_j) = \mathsf{P}_\xi(x_i)\mathsf{P}_\eta(y_j)$$

для всех возможных значений x_i, y_j случайных величин ξ и η. Тогда для произвольных множеств A и B из $\mathbb{R}^1$

$$\mathsf{P}\{\xi \in A, \eta \in B\} = \mathsf{P}\{(\xi, \eta) \in A \times B\} =$$

$$= \sum_{(x_i,y_j)\in A\times B} \mathsf{P}_{\xi,\eta}(x_i, y_j) = \sum_{(x_i,y_j)\in A\times B} \mathsf{P}_\xi(x_i)\mathsf{P}_\eta(y_j) =$$

$$= \sum_{x_i\in A} \mathsf{P}_\xi(x_i) \sum_{y_j\in B} \mathsf{P}_\eta(y_j) = \mathsf{P}\{\xi \in A\}\mathsf{P}\{\eta \in B\}$$

(воспользовались равенством (5.1.3)).

Теорема доказана.

Аналогичная теорема имеет место и для n случайных величин, а именно:

Теорема 5.2.2. *Совместное распределение* P_ξ *независимых случайных величин* $\xi_1, \xi_2, \ldots, \xi_n$ *равно произведению их распределений* $\mathsf{P}_{\xi_1}, \mathsf{P}_{\xi_2}, \ldots, \mathsf{P}_{\xi_n}$, *т. е.*

$$\mathsf{P}_\xi(x_1, x_2, \ldots, x_n) = \mathsf{P}_{\xi_1}(x_1)\mathsf{P}_{\xi_2}(x_2)\ldots\mathsf{P}_{\xi_n}(x_n) \tag{5.2.2}$$

для всех возможных значений $x_1, x_2, \ldots, x_n$ *случайных величин* $\xi_1, \xi_2, \ldots, \xi_n$.

Справедливо и обратное утверждение: если совместное распределение случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, равно произведению их распределений, то случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы.

З а м е ч а н и е. Совместное распределение

$$\mathsf{P}_{\xi,\eta}(x_i, y_j)\colon (x_i, y_j) \to \mathsf{P}_\xi(x_i)\mathsf{P}_\eta(y_j) = p_i q_j$$

независимых случайных величин ξ и η соответственно с распределениями

$$\mathsf{P}_\xi\colon x_i \to \mathsf{P}_\xi(x_i) = p_i \text{ и } \mathsf{P}_\eta\colon y_j \to \mathsf{P}_\eta(y_j) = q_j$$

в табличной форме записывается в виде табл. 5.2.1.

Т а б л и ц а 5.2.1. Совместное распределение независимых случайных величин ξ и η

Значения ξ	Значения η				
	y_1	y_2	$\cdots$	y_k	$\cdots$
x_1	p_1q_1	p_1q_2	$\cdots$	p_1q_k	$\cdots$
x_2	p_2q_1	p_2q_2	$\cdots$	p_2q_k	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$
x_s	p_sq_1	p_sq_2	$\cdots$	p_sq_k	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

.

Пример 5.2.1. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, каждая с распределением*

$$\mathsf{P}\{\xi_l = k\} = \frac{\lambda^k}{k!}e^{-\lambda},\ \lambda > 0,\ k = 0, 1, \ldots,\ (l = 1, 2, \ldots, n).$$

Выписать совместное распределение случайных величин $\xi_1, \xi_2, \ldots, \xi_n$.

Р е ш е н и е. По распределениям $\mathsf{P}_{\xi_1}, \mathsf{P}_{\xi_2}, \ldots, \mathsf{P}_{\xi_n}$ независимых случайных величин $\xi_1, \xi_2 \ldots, \xi_n$ их совместное распределение

$$\mathsf{P}_\xi(k_1, k_2, \ldots, k_n) = \mathsf{P}\{\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_n = k_n\}$$

находим как произведение распределений этих случайных величин:

$$\mathsf{P}_\xi(k_1, k_2, \ldots, k_n) = \prod_{i=1}^{n} \mathsf{P}_{\xi_i}(k_i) = \prod_{i=1}^{n} \frac{\lambda^{k_i}}{k_i!} e^{-\lambda} = \frac{\lambda^{\sum\limits_{i=1}^{n} k_i}}{k_1! k_2! \ldots k_n!} e^{-n\lambda},$$

$k_1 = 0, 1, \ldots;\ k_2 = 0, 1, \ldots;\ \ldots;\ k_n = 0, 1, \ldots$

5.3 Математическое ожидание случайной величины

Случайная величина описывается своим распределением:

$$\mathsf{P}_\xi : x_i \to \mathsf{P}_\xi(x_i) = \mathsf{P}\{\xi = x_i\},\ x_i \in X,$$

задающим значения, принимаемые случайной величиной, и вероятности, с которыми эти значения принимаются. Информацию о распределении случайной величины мы получаем из эксперимента (независимых наблюдений случайной величины). Но зачастую для получения удовлетворительного представления о распределении случайной величины имеющихся наблюдений мало. В связи с этим возникает необходимость в описании распределения случайной величины характеристиками, которые, возможно, и не полностью его описывают, но дают достаточно хорошее представление о распределении, и которые можно оценить по экспериментальным данным. Такими характеристиками, в частности, являются математическое ожидание и дисперсия случайной величины. (Некоторые распределения математическим ожиданием и дисперсией задаются полностью.)

Определение. *Математическим ожиданием* (*средним*) случайной величины $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^1$, заданной на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$, будем называть сумму ряда

$$\sum_{\omega \in \Omega} \xi(\omega) \mathsf{P}(\omega),$$

если он абсолютно сходится. Обозначать математическое ожидание будем через $\mathsf{M}\xi$:

$$\mathsf{M}\xi = \sum_{\omega \in \Omega} \xi(\omega) \mathsf{P}(\omega).$$

Так что по определению математическое ожидание $\mathsf{M}\xi$ случайной величины $\xi = \xi(\omega)$ конечно.

Если ряд $\sum\limits_{\omega\in\Omega} \xi(\omega)\mathsf{P}(\omega)$ не сходится абсолютно, то будем считать, что математическое ожидание случайной величины ξ не существует. Но далее (в гл. 9) мы определим математическое ожидание и для случайной величины $\xi = \xi(\omega)$, у которой ряд $\sum\limits_{\omega\in\Omega} |\xi(\omega)|\mathsf{P}(\omega)$ расходится (правда, не для всех таких случайных величин).

В дальнейшем обороты "математическое ожидание определено" и "математическое ожидание существует" обозначают одно и то же.

Математическое ожидание случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ со значениями в $\mathbb{R}^n$ определяется равенством

$$\mathsf{M}\xi = (\mathsf{M}\xi_1, \mathsf{M}\xi_2, \ldots, \mathsf{M}\xi_n).$$

Свойства математического ожидания. Свойства математического ожидания получаются непосредственно из его определения.

Свойство 1. *Если $\xi = \xi(\omega) = c$ (c — константа), то*

$$\mathsf{M}\xi = \mathsf{M}c = c$$

(математическое ожидание константы равно этой же константе).

Доказательство.

$$\mathsf{M}c = \sum_{\omega\in\Omega} c\mathsf{P}(\omega) = c\sum_{\omega\in\Omega} \mathsf{P}(\omega) = c.$$

Свойство 2 (линейность математического ожидания). *Пусть ξ и η — случайные величины с математическими ожиданиями $\mathsf{M}\xi$ и $\mathsf{M}\eta$, a и b — константы. Тогда случайная величина $a\xi + b\eta$ также имеет математическое ожидание и*

$$\mathsf{M}(a\xi + b\eta) = a\mathsf{M}\xi + b\mathsf{M}\eta. \tag{5.3.1}$$

Доказательство. Сумма абсолютно сходящихся рядов есть абсолютно сходящийся ряд. При этом

$$a\mathsf{M}\xi + b\mathsf{M}\eta = a\sum_{\omega\in\Omega} \xi(\omega)\mathsf{P}(\omega) + b\sum_{\omega\in\Omega} \eta(\omega)\mathsf{P}(\omega) =$$

$$= \sum_{\omega\in\Omega}(a\xi(\omega) + b\eta(\omega))\mathsf{P}(\omega) = \mathsf{M}(a\xi + b\eta).$$

Следствие 1. *Математическое ожидание суммы случайных величин равно сумме их математических ожиданий:*

$$\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta.$$

Утверждение следует из равенства (5.3.1), если положить $a = 1,\ b = 1$.

Следствие 2. *Константа выносится из-под знака математического ожидания:*

$$\mathsf{M}a\xi = a\mathsf{M}\xi.$$

Утверждение получается из равенства (5.3.1), если положить $b = 0$.

Свойство 3. *Если*

$$|\xi| \le |\eta|,$$

то из существования $\mathsf{M}\eta$ *следует существование* $\mathsf{M}\xi$.

Достаточно заметить, что из неравенства $|\xi| \le |\eta|$ следует

$$\sum_{\omega\in\Omega}|\xi(\omega)|\mathsf{P}(\omega) \le \sum_{\omega\in\Omega}|\eta(\omega)|\mathsf{P}(\omega).$$

Свойство 4. *Если* $\mathsf{M}\xi$ *определено, то*

$$|\mathsf{M}\xi| \le \mathsf{M}|\xi|.$$

Доказательство очевидно.

Свойство 5. *Если* $\xi \ge 0$, *то*

$$\mathsf{M}\xi \ge 0.$$

Доказательство очевидно.

Следующая теорема показывает, как по распределению случайной величины можно вычислить математическое ожидание любой функции от нее (если только оно существует).

Напомним, что множество всех возможных значений дискретной случайной величины мы обозначили через X.

Если при суммировании не указаны пределы изменения индекса, то суммируют по всем возможным его значениям.

Теорема 5.3.1 (о вычислении $\mathsf{M}g(\xi)$ по распределению ξ). *Пусть $\xi = \xi(\omega)$ — случайная величина со значениями в $\mathbb{R}^1$ на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$ и*

$$\mathsf{P}_\xi : x_i \to \mathsf{P}_\xi(x_i),\ x_i \in X,$$

— ее распределение, тогда для любой функции $g = g(x)$ на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$

$$\mathsf{M}g(\xi) = \sum_{x_i} g(x_i)\mathsf{P}_\xi(x_i), \tag{5.3.2}$$

в предположении абсолютной сходимости одного из рядов:

$$\sum_{\omega\in\Omega} g(\xi(\omega))\mathsf{P}(\omega) \ \text{или}\ \sum_{x_i} g(x_i)\mathsf{P}_\xi(x_i).$$

Доказательство. Пусть ряд $\sum\limits_{\omega\in\Omega} g(\xi(\omega))\mathsf{P}(\omega)$ сходится абсолютно. Представив Ω в виде объединения

$$\Omega = \bigcup_{x_i}\{\omega : \xi(\omega) = x_i\}$$

попарно непересекающихся множеств $\{\omega : \xi(\omega) = x_i\}, i = 1, 2, \ldots$ (см. также рис. 5.3.1) и воспользовавшись свойством ассоциативности абсолютно сходящихся рядов, получим

$$\sum_{\omega\in\Omega} |g(\xi(\omega))|\mathsf{P}(\omega) = \sum_{\omega\in\bigcup\limits_{x_i}\{\omega:\xi(\omega)=x_i\}} |g(\xi(\omega))|\mathsf{P}(\omega) =$$

$$= \sum_{x_i}\left(\sum_{\omega:\xi(\omega)=x_i} |g(\xi(\omega))|\mathsf{P}(\omega)\right) = \sum_{x_i}\sum_{\omega:\xi(\omega)=x_i} |g(x_i)|\mathsf{P}(\omega) =$$

$$= \sum_{x_i} |g(x_i)| \sum_{\omega:\xi(\omega)=x_i} \mathsf{P}(\omega) = \sum_{x_i} |g(x_i)|\mathsf{P}\{\omega : \xi(\omega) = x_i\} =$$

$$= \sum_{x_i} |g(x_i)|\mathsf{P}_\xi(x_i),$$

т. е.

$$\sum_{\omega\in\Omega} |g(\xi(\omega))|\mathsf{P}(\omega) = \sum_{x_i} |g(x_i)|\mathsf{P}_\xi(x_i). \qquad (5.3.3)$$

Так что ряд $\sum\limits_{\omega\in\Omega} g(\xi(\omega))\mathsf{P}(\omega)$ сходится абсолютно тогда и только тогда, когда абсолютно сходится ряд $\sum\limits_{x_i} g(x_i)\mathsf{P}_\xi(x_i)$.

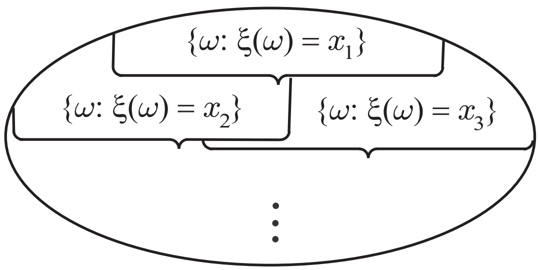

Рис. 5.3.1: К представлению Ω

В предположениях теоремы, повторяя выкладки, аналогичные приведенным выше, применительно к абсолютно сходящемуся ряду $\sum\limits_{\omega\in\Omega} g(\xi(\omega))\mathsf{P}(\omega)$, получим

$$\mathsf{M}g(\xi) = \sum_{\omega\in\Omega} g(\xi(\omega))\mathsf{P}(\omega) = \sum_{x_i} g(x_i)\mathsf{P}_\xi(x_i),$$

что и доказывает теорему.

Следствие. *Пусть $\xi = \xi(\omega)$ — случайная величина со значениями в $\mathbb{R}^1$ на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$ и*

$$\mathsf{P}_\xi : x_i \to \mathsf{P}_\xi(x_i),\ x_i \in X,$$

— ее распределение, тогда

$$\mathsf{M}\xi = \sum_{x_i} x_i\mathsf{P}_\xi(x_i), \qquad (5.3.4)$$

в предположении абсолютной сходимости одного из рядов

$$\sum_{\omega\in\Omega} \xi(\omega)\mathsf{P}(\omega) \ \text{ или } \ \sum_{x_i} x_i\mathsf{P}_\xi(x_i).$$

Достаточно в теореме в качестве функции $g(x)$ рассмотреть $g(x) = x$.

Аналогичная теорема о вычислении $\mathsf{M}g(\xi)$ по распределению ξ имеет место и для случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ со значениями в $\mathbb{R}^n$ (векторной случайной величины).

Теорема 5.3.2. *Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — случайная величина со значениями в $\mathbb{R}^n$ на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$ и*

$$\mathsf{P}_\xi : (x_1, x_2, \ldots, x_n) \to \mathsf{P}_\xi(x_1, x_2, \ldots, x_n),\ (x_1, x_2, \ldots, x_n) \in X,$$

ее распределение, тогда для любой функции $g = g(x_1, x_2, \ldots, x_n)$ на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$

$$\mathsf{M}g(\xi_1, \xi_2, \ldots, \xi_n) = \sum_{x_1, x_2, \ldots, x_n} g(x_1, x_2, \ldots, x_n)\mathsf{P}_\xi(x_1, x_2, \ldots, x_n), \tag{5.3.5}$$

в предположении, что ряд $\sum\limits_{x_1, x_2, \ldots, x_n} g(x_1, x_2, \ldots, x_n)\mathsf{P}_\xi(x_1, x_2, \ldots, x_n)$ или ряд $\sum\limits_{\omega\in\Omega} g(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))\mathsf{P}(\omega)$ сходится абсолютно.

Доказательство аналогично доказательству теоремы 5.3.1.

Пример 5.3.1. *Пусть ξ — число выпавших очков на игральной кости. Предположим, что кость изготовлена так, что распределение случайной величины ξ имеет вид:*

i	1	2	3	4	5	6
$\mathsf{P}_\xi(i)$	9/10	1/50	1/50	1/50	1/50	1/50

Вычислить математическое ожидание ξ.

Решение. Согласно теореме 5.3.1 о вычислении математического ожидания случайной величины по ее распределению имеем

$$\mathsf{M}\xi = 1 \cdot \frac{9}{10} + 2 \cdot \frac{1}{50} + 3 \cdot \frac{1}{50} + \ldots + 6 \cdot \frac{1}{50} = \frac{13}{10} = 1{,}3.$$

Если кость симметричная, то распределение случайной величины ξ имеет вид:

i	1	2	3	4	5	6
$\mathsf{P}_\xi(i)$	1/6	1/6	1/6	1/6	1/6	1/6

При этом

$$\mathsf{M}\xi = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3{,}5.$$

Если предположить, что кость изготовлена так, что распределение случайной величины ξ имеет вид:

i	1	2	3	4	5	6
$\mathsf{P}_\xi(i)$	1/50	1/50	1/50	1/50	1/50	9/10

то

$$\mathsf{M}\xi = 1 \cdot \frac{1}{50} + 2 \cdot \frac{1}{50} + \ldots + 5 \cdot \frac{1}{50} + 6 \cdot \frac{9}{10} = \frac{57}{10} = 5{,}7.$$

Пример 5.3.2. *Случайная величина ξ имеет распределение Пуассона с параметром λ:*

$$\mathsf{P}_\xi(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ k = 0, 1, \ldots$$

Вычислить

$$\mathsf{M}\frac{1}{\xi + 1}.$$

Решение. По распределению P_ξ случайной величины ξ всегда можно вычислить математическое ожидание $\mathsf{M}g(\xi)$ любой функции $g(\xi)$ от ξ (если только оно существует), см. теорему 5.3.1. В рассматриваемом примере $g(t) = 1/(t+1)$, а случайная величина ξ имеет распределение Пуассона с параметром λ. Поэтому

$$\mathsf{M}\frac{1}{\xi+1} = \sum_{k=0}^{\infty} \frac{1}{k+1} \mathsf{P}_\xi(k) = \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{\lambda^k}{k!} e^{-\lambda} =$$

$$= \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} = \frac{e^{-\lambda}}{\lambda}(e^{\lambda} - 1) = \frac{1 - e^{-\lambda}}{\lambda}.$$

Пример 5.3.3. *Дважды подбрасывают две симметричные монеты; ξ — число выпавших гербов при подбрасывании двух монет первый раз; η — второй.*

Вычислить $\mathsf{M}\sin\left(\frac{\pi}{6}\min\{\xi,\eta\}\right)$.

Решение. Распределение случайной величины $\zeta = (\xi,\eta)$ известно (см. пример 5.1.2):

Значения ξ	Значения η		
	0	1	2
0	1/16	2/16	1/16
1	2/16	4/16	2/16
2	1/16	2/16	1/16

.

По распределению $\zeta = (\xi,\eta)$ можно вычислить математическое ожидание (если только оно существует) любой функции от ζ, в частности, и $f(\zeta) = f(\xi,\eta) = \sin\left(\frac{\pi}{6}\min\{\xi,\eta\}\right)$, см. теорему 5.3.2. В нашем примере

$$\mathsf{M}\sin\left(\frac{\pi}{6}\min\{\xi,\eta\}\right) = \sin\left(\frac{\pi}{6}\min\{0,0\}\right)\cdot\frac{1}{16}+$$

$$+\sin\left(\frac{\pi}{6}\min\{0,1\}\right)\cdot\frac{2}{16}+\ldots+\sin\left(\frac{\pi}{6}\min\{2,2\}\right)\cdot\frac{1}{16} =$$

$$=\sin\left(\frac{\pi}{6}\cdot 0\right)\cdot\frac{1}{16}+\sin\left(\frac{\pi}{6}\cdot 0\right)\cdot\frac{2}{16}+\ldots+\sin\left(\frac{\pi}{6}\cdot 2\right)\cdot\frac{1}{16} = \frac{8+\sqrt{3}}{32}.$$

Теорема 5.3.3 (мультипликативное свойство математического ожидания). *Пусть ξ и η — независимые случайные величины с математическими ожиданиями $\mathsf{M}\xi$ и $\mathsf{M}\eta$. Тогда существует математическое ожидание произведения $\xi\eta$ и*

$$\mathsf{M}\xi\eta = \mathsf{M}\xi\cdot\mathsf{M}\eta$$

— математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий.

Доказательство. Из существования $\mathsf{M}\xi$ и $\mathsf{M}\eta$ и теоремы 5.3.1 следует абсолютная сходимость рядов

$$\sum_{x_i} x_i \mathsf{P}_\xi(x_i), \ \sum_{y_j} y_j \mathsf{P}_\eta(y_j) \tag{5.3.6}$$

и равенства

$$\mathsf{M}\xi = \sum_{x_i} x_i \mathsf{P}_\xi(x_i), \ \mathsf{M}\eta = \sum_{y_j} y_j \mathsf{P}_\eta(y_j).$$

Из абсолютной сходимости рядов (5.3.6) и теоремы о произведении абсолютно сходящихся рядов следует абсолютная сходимость ряда

$$\sum_{x_i, y_j} x_i y_j \mathsf{P}_\xi(x_i) \mathsf{P}_\eta(y_j)$$

и равенство

$$\left(\sum_{x_i} x_i \mathsf{P}_\xi(x_i)\right)\left(\sum_{y_j} y_j \mathsf{P}_\eta(y_j)\right) = \sum_{x_i, y_j} x_i y_j \mathsf{P}_\xi(x_i) \mathsf{P}_\eta(y_j),$$

которое, учитывая, что случайные величины ξ и η независимы:

$$\mathsf{P}_{\xi,\eta}(x_i, y_j) = \mathsf{P}_\xi(x_i) \mathsf{P}_\eta(y_j),$$

перепишем так

$$\left(\sum_{x_i} x_i \mathsf{P}_\xi(x_i)\right)\left(\sum_{y_j} y_j \mathsf{P}_\eta(y_j)\right) = \sum_{x_i, y_j} x_i y_j \mathsf{P}_{\xi,\eta}(x_i, y_j). \tag{5.3.7}$$

Из абсолютной сходимости ряда

$$\sum_{x_i, y_j} x_i y_j \mathsf{P}_{\xi,\eta}(x_i, y_j)$$

и теоремы 5.3.2 следует существование математического ожидания функции $g(\xi, \eta) = \xi\eta$ от случайного вектора (ξ, η) и равенство

$$\mathsf{M}\xi\eta = \sum_{x_i, y_j} x_i y_j \mathsf{P}_{\xi,\eta}(x_i, y_j).$$

Отсюда и из равенств (5.3.7) имеем

$$\mathsf{M}\xi\eta = \mathsf{M}\xi\mathsf{M}\eta.$$

Ковариация. Пусть ξ, η — случайные величины со значениями в $\mathbb{R}^1$ на $\{\Omega, \mathsf{P}\}$, у которых существуют $\mathsf{M}\xi^2$ и $\mathsf{M}\eta^2$. Поскольку

$$|\xi\eta| \leq \frac{1}{2}(\xi^2 + \eta^2),$$

и, в частности,

$$|\xi| \leq \frac{1}{2}(\xi^2 + 1),$$

то из существования $\mathsf{M}\xi^2$ и $\mathsf{M}\eta^2$ следует существование $\mathsf{M}\xi\eta$, $\mathsf{M}\xi$, $\mathsf{M}\eta$.

Заметим, что

$$(\mathsf{M}\xi\eta)^2 \leq \mathsf{M}\xi^2\mathsf{M}\eta^2 < \infty. \tag{5.3.8}$$

Неравенство следует из того, что квадратный относительно λ трехчлен $\mathsf{M}(\lambda\xi + \eta)^2$ при всех $\lambda \in \mathbb{R}^1$ неотрицателен, а следовательно, его дискриминант меньше или равен нуля.

Определение. Число

$$cov(\xi, \eta) = \mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) = \mathsf{M}\xi\eta - \mathsf{M}\xi\mathsf{M}\eta$$

называют *ковариацией* случайных величин ξ, η, а число

$$r(\xi, \eta) = \frac{\mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta)}{\sqrt{\mathsf{M}(\xi - \mathsf{M}\xi)^2\mathsf{M}(\eta - \mathsf{M}\eta)^2}}$$

называют *коэффициентом корреляции* случайных величин ξ, η.

Коэффициент корреляции по модулю не превосходит 1, что следует из неравенства (5.3.8).

Случайные величины ξ, η, для которых $cov(\xi, \eta) = 0$, называют *некоррелированными.*

Независимые случайные величины всегда некоррелированы. Если ξ и η независимы, то в силу мультипликативного свойства математического ожидания

$$cov(\xi, \eta) = \mathsf{M}\xi\eta - \mathsf{M}\xi\mathsf{M}\eta = \mathsf{M}\xi\mathsf{M}\eta - \mathsf{M}\xi\mathsf{M}\eta = 0.$$

Из некоррелированности случайных величин их независимость, вообще говоря, не следует.

Если случайные величины ξ, η линейно зависимы:

$$\eta = a\xi + b$$

(a, b — константы, $a \neq 0$), то коэффициент корреляции $r(\xi, \eta)$ равен $+1$ или -1.

Действительно,

$$r(\xi, \eta) = \frac{\mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta)}{\sqrt{\mathsf{M}(\xi - \mathsf{M}\xi)^2 \mathsf{M}(\eta - \mathsf{M}\eta)^2}} =$$

$$= \frac{\mathsf{M}(\xi - \mathsf{M}\xi)(a\xi + b - \mathsf{M}(a\xi + b))}{\sqrt{\mathsf{M}(\xi - \mathsf{M}\xi)^2 \mathsf{M}(a\xi + b - \mathsf{M}(a\xi + b))^2}} =$$

$$= \frac{a\mathsf{M}(\xi - \mathsf{M}\xi)^2}{|a|\sqrt{(\mathsf{M}(\xi - \mathsf{M}\xi)^2)^2}} = \frac{a}{|a|}.$$

Целочисленные случайные величины. Производящие функции. Важным классом случайных величин являются неотрицательные целочисленные случайные величины (принимают значения из множества $\{0, 1, 2, \ldots\}$). Метод производящих функций является наиболее общим при их изучении.

Определение. Пусть ξ — неотрицательная целочисленная случайная величина с распределением

$$\mathsf{P}\{\xi = k\} = p_k, \; k = 0, 1, \ldots$$

Функцию $P(t)$, определенную равенством

$$P(t) = \mathsf{M}t^{\xi} = \sum_{k=0}^{\infty} t^k p_k, \; |t| \leq 1,$$

будем называть *производящей функцией* целочисленной случайной величины ξ (распределения $\{p_k\}$).

Поскольку последовательность $\{p_j\}$ является вероятностным распределением, то производящая функция существует, по меньшей мере при $|t| \leq 1$.

Теорема 5.3.4. *Пусть ξ — неотрицательная целочисленная случайная величина и*

$$\mathsf{P}\{\xi = k\} = p_k, \; k = 0, 1, \ldots,$$

— ее распределение.

Если существует $\mathsf{M}\xi$*, то ее производящая функция*

$$P(t) = \sum_{k=0}^{\infty} t^k p_k, \ |t| \le 1, \tag{5.3.9}$$

дифференцируема на отрезке $[-1, 1]$ *и*

$$\mathsf{M}\xi = P'(1). \tag{5.3.10}$$

Если существует $\mathsf{M}\xi^2$*, то производящая функция* $P(t)$*,* $|t| \le 1$*, случайной величины* ξ *дважды дифференцируема на отрезке* $[-1, 1]$ *и*

$$\mathsf{M}\xi^2 = P'(1) + P''(1). \tag{5.3.11}$$

Доказательство. В самом деле, если

$$\sum_{k=0}^{\infty} k p_k < \infty,$$

то ряд

$$\sum_{k=1}^{\infty} k t^{k-1} p_k, \ |t| \le 1,$$

полученный почленным дифференцированием ряда $\sum\limits_{k=0}^{\infty} t^k p_k$, $|t| \le 1$, абсолютно сходится в точке $t = 1$, а, следовательно, сходится равномерно на отрезке $[-1, 1]$. Поэтому ряд (5.3.9) можно почленно дифференцировать на отрезке $[-1, 1]$, а именно

$$P'(t) = \sum_{k=1}^{\infty} k t^{k-1} p_k, \ |t| \le 1.$$

При этом в точке $t = 1$

$$P'(1) = \sum_{k=0}^{\infty} k p_k = \mathsf{M}\xi.$$

Если существует $\mathsf{M}\xi^2$, т. е.

$$\mathsf{M}\xi^2 = \sum_{k=1}^{\infty} k^2 p_k < \infty,$$

то ряд

$$\sum_{k=2}^{\infty} k(k-1)t^{k-2}p_k, \ |t| \le 1,$$

полученный почленным дифференцированием дважды ряда $\sum\limits_{k=0}^{\infty} t^k p_k, |t| \le 1$, абсолютно сходится в точке $t = 1$, а, следовательно, сходится равномерно на отрезке $[-1, 1]$. Поэтому ряд (5.3.9) можно дважды почленно дифференцировать на отрезке $[-1, 1]$, а именно

$$P''(t) = \sum_{k=2}^{\infty} k(k-1)t^{k-2}p_k, \ |t| \le 1.$$

При этом в точке $t = 1$

$$P''(1) = \sum_{k=2}^{\infty} k(k-1)p_k = \sum_{k=0}^{\infty} k^2 p_k - \sum_{k=0}^{\infty} kp_k = \mathsf{M}\xi^2 - \mathsf{M}\xi.$$

Отсюда, учитывая равенство (5.3.10), имеем, что

$$\mathsf{M}\xi^2 = P'(1) + P''(1).$$

Пример 5.3.4. *Пусть ξ — пуассоновская случайная величина с параметром θ:*

$$P\{\xi = k\} = \frac{\theta^k}{k!}e^{-\theta}, \ k = 0, 1, 2, \ldots$$

Найти производящую функцию случайной величины ξ. Вычислить $\mathsf{M}\xi$ и $\mathsf{M}\xi^2$.

Решение. По определению производящая функция случайной величины ξ

$$P(t) = \mathsf{M}t^{\xi} = \sum_{k=0}^{\infty} t^k \frac{\theta^k}{k!}e^{-\theta} = e^{t\theta}e^{-\theta} = \exp\{\theta(t-1)\}.$$

Далее, ряды $\sum\limits_{k=0}^{\infty} k\frac{\theta^k}{k!}e^{-\theta}$ и $\sum\limits_{k=0}^{\infty} k^2\frac{\theta^k}{k!}e^{-\theta}$ сходятся, поэтому $\mathsf{M}\xi$ и $\mathsf{M}\xi^2$ существуют и согласно теореме 5.3.4

$$\mathsf{M}\xi = P'(1), \ \mathsf{M}\xi^2 = P'(1) + P''(1).$$

Первая и вторая производные производящей функции:

$$P'(t) = \theta \exp\{\theta(t-1)\},\ P''(t) = \theta^2 \exp\{\theta(t-1)\}.$$

Поэтому

$$\mathsf{M}\xi = P'(1) = \theta,\ \mathsf{M}\xi^2 = P''(1) + P'(1) = \theta^2 + \theta.$$

Теорема 5.3.5 (единственности). *Различные вероятностные распределения на $\{0, 1, 2, \ldots\}$ имеют различные производящие функции.*

Доказательство. Пусть p_k, $k = 0, 1, \ldots$, и f_n, $n = 0, 1, \ldots$, — вероятностные распределения. Их производящие функции

$$P(t) = \sum_{k=0}^{\infty} t^k p_k,\ |t| \le 1,\ \ F(t) = \sum_{k=0}^{\infty} t^k f_k,\ |t| \le 1,$$

как степенные ряды, можно почленно дифференцировать неограниченное число раз внутри их промежутков сходимости (по меньшей мере при $|t| < 1$). Предположим, что

$$P(t) = F(t),\ |t| < 1. \tag{5.3.12}$$

Отсюда при $t = 0$ имеем равенство $P(0) = F(0)$, или $p_0 = f_0$. Почленно продифференцировав (5.3.12), получим

$$\sum_{k=1}^{\infty} k t^{k-1} p_k = P'(t) = F'(t) = \sum_{k=1}^{\infty} k t^{k-1} f_k.$$

Отсюда при $t = 0$ имеем, что $P'(0) = F'(0)$, а, следовательно, и $p_1 = f_1$ и т. д. Так что при всех $k = 0, 1, \ldots$

$$p_k = f_k.$$

Последнее обозначает, что распределения $\{p_k\}$ и $\{f_k\}$ совпадают.

Теорема 5.3.6 (свойство мультипликативности производящих функций). *Производящая функция суммы независимых случайных величин равна произведению производящих функций слагаемых.*

Доказательство. Для независимых случайных величин ξ и η в силу мультипликативного свойства математического ожидания

$$\mathsf{P}_{\xi+\eta}(t) = \mathsf{M}t^{\xi+\eta} = \mathsf{M}t^{\xi} t^{\eta} = \mathsf{M}t^{\xi}\mathsf{M}t^{\eta} = P_{\xi}(t) P_{\eta}(t).$$

Пример 5.3.5. *Доказать, что сумма независимых пуассоновских случайных величин является пуассоновской случайной величиной с параметром, равным сумме параметров слагаемых.*

Решение. Пусть ξ и η — пуассоновские случайные величины с параметрами θ_1 и θ_2, их производящие функции соответственно равны $\exp\{\theta_1(t-1)\}$ и $\exp\{\theta_2(t-1)\}$ (см. пример 5.3.4). Производящая функция суммы $\zeta = \xi + \eta$ независимых случайных величин

$$P_\zeta(t) = P_\xi(t)P_\eta(t) = \exp\{\theta_1(t-1)\}\exp\{\theta_2(t-1)\} =$$

$$= \exp\{(\theta_1+\theta_2)(t-1)\}$$

— в силу мультипликативного свойства производящих функций. С другой стороны, $\exp\{(\theta_1+\theta_2)(t-1)\}$ — производящая функция пуассоновской случайной величины с параметром $\theta_1 + \theta_2$. Поэтому согласно теореме единственности (теорема 5.3.5) распределение случайной величины $\xi+\eta$ совпадает с пуассоновским распределением с параметром $\theta_1+\theta_2$.

5.4 Неравенство Чебышёва, дисперсия случайной величины

Неравенство Чебышёва. *Если* $\mathsf{M}\xi^2 < \infty$, *то для каждого* $\varepsilon > 0$

$$\mathsf{P}\{|\xi - \mathsf{M}\xi| \geq \varepsilon\} \leq \frac{\mathsf{M}(\xi - \mathsf{M}\xi)^2}{\varepsilon^2}. \tag{5.4.1}$$

Доказательство.

$$\mathsf{M}(\xi-\mathsf{M}\xi)^2 = \sum_{\omega\in\Omega}(\xi(\omega)-\mathsf{M}\xi)^2\mathsf{P}(\omega) \geq$$

$$\geq \sum_{\omega:|\xi(\omega)-\mathsf{M}\xi|\geq\varepsilon}(\xi(\omega)-\mathsf{M}\xi)^2\mathsf{P}(\omega) \geq$$

$$\geq \varepsilon^2\sum_{\omega:|\xi(\omega)-\mathsf{M}\xi|\geq\varepsilon}\mathsf{P}(\omega) = \varepsilon^2\mathsf{P}\{|\xi-\mathsf{M}\xi|\geq\varepsilon\}.$$

Отсюда и следует неравенство Чебышёва.

Неравенство Чебышёва, в частности, дает ответ на вопрос “Насколько часто наблюдаются большие уклонения случайной величины ξ от своего среднего (математического ожидания) $\mathsf{M}\xi$?” и мотивирует введение величины $\mathsf{M}(\xi - \mathsf{M}\xi)^2$ в качестве количественной меры разброса случайной величины ξ относительно $\mathsf{M}\xi$.

Договоримся “большое” уклонение ξ от $\mathsf{M}\xi$ характеризовать некоторым числом ε: если $|\xi - \mathsf{M}\xi| \geq \varepsilon$, то уклонение “большое”, в противном случае — “малое” (в каждой конкретной ситуации ε, разумеется, свое). Из неравенства Чебышёва следует, что если величина $\mathsf{M}(\xi - \mathsf{M}\xi)^2$ мала, то мала и вероятность больших уклонений ξ от $\mathsf{M}\xi$ (большие уклонения ξ от $\mathsf{M}\xi$ встречаются изредка). Поэтому величину $\mathsf{M}(\xi - \mathsf{M}\xi)^2$ естественно рассматривать в качестве количественной меры разброса ξ относительно $\mathsf{M}\xi$, ее называют дисперсией ξ.

Определение. *Дисперсией* $\mathsf{D}\xi$ случайной величины ξ будем называть математическое ожидание квадрата отклонения случайной величины от своего среднего:

$$\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2.$$

Часто удобно пользоваться следующим представлением дисперсии:

$$\mathsf{D}\xi = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2,$$

которое получается из равенств

$$\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2 = \mathsf{M}(\xi^2 - 2\xi\mathsf{M}\xi + (\mathsf{M}\xi)^2) =$$

$$= \mathsf{M}\xi^2 - 2(\mathsf{M}\xi)^2 + (\mathsf{M}\xi)^2 = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2.$$

Свойства дисперсии. Приведенные далее свойства дисперсии непосредственно следуют из ее определения.

Свойство 1 (дисперсия $\boldsymbol{a\xi+b}$). *Пусть a, b — константы, тогда*

$$\mathsf{D}(a\xi + b) = a^2\mathsf{D}\xi.$$

Действительно,

$$\mathsf{D}(a\xi + b) = \mathsf{M}(a\xi + b - \mathsf{M}(a\xi + b))^2 =$$

$$= \mathsf{M}a^2(\xi - \mathsf{M}\xi)^2 = a^2\mathsf{M}(\xi - \mathsf{M}\xi)^2 = a^2\mathsf{D}\xi.$$

Следствие 1. *Дисперсия константы равна нулю:*

$$\mathsf{D}b = 0.$$

Достаточно в равенстве $\mathsf{D}(a\xi + b) = a^2\mathsf{D}\xi$ положить $a = 0$.

Следствие 2. *Константа выносится из-под знака дисперсии с квадратом:*

$$\mathsf{D}(a\xi) = a^2\mathsf{D}\xi.$$

Достаточно в равенстве $\mathsf{D}(a\xi + b) = a^2\mathsf{D}\xi$ положить $b = 0$.

Свойство 2 (аддитивность дисперсии). *Для независимых случайных величин ξ и η*

$$\mathsf{D}(\xi + \eta) = \mathsf{D}\xi + \mathsf{D}\eta$$

(*дисперсия суммы независимых случайных величин равна сумме их дисперсий*).

Доказательство.

$$\mathsf{D}(\xi+\eta) = \mathsf{M}\left((\xi+\eta) - \mathsf{M}(\xi+\eta)\right)^2 = \mathsf{M}\left((\xi - \mathsf{M}\xi) + (\eta - \mathsf{M}\eta)\right)^2 =$$

$$= \mathsf{M}\left((\xi - \mathsf{M}\xi)^2 + 2(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) + (\eta - \mathsf{M}\eta)^2\right) =$$

$$= \mathsf{M}(\xi - \mathsf{M}\xi)^2 + 2\mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) + \mathsf{M}(\eta - \mathsf{M}\eta)^2 =$$

$$= \mathsf{D}\xi + \mathsf{D}\eta + 2\mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) =$$

$$= \mathsf{D}\xi + \mathsf{D}\eta + 2\mathrm{cov}(\xi, \eta).$$

Для независимых случайных величин ξ и η в силу свойства мультипликативности математического ожидания

$$\mathrm{cov}(\xi,\eta) = \mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) = \mathsf{M}\xi\eta - \mathsf{M}\xi\mathsf{M}\eta = \mathsf{M}\xi\mathsf{M}\eta - \mathsf{M}\xi\mathsf{M}\eta = 0.$$

Так что

$$\mathsf{D}(\xi + \eta) = \mathsf{D}\xi + \mathsf{D}\eta.$$

Замечание. Очевидно, свойство аддитивности дисперсии имеет место не только для независимых случайных величин, но и для некоррелированных.

Закон больших чисел. Андрей Николаевич Колмогоров так сформулировал суть закона больших чисел: “Закон больших чисел — *общий принцип, в силу которого совокупное действие большого числа случайных факторов приводит, при некоторых весьма общих условиях, к результату, почти не зависящему от случая”.* В теории вероятностей закон больших чисел изучается как математический результат, отражающий этот принцип, лежащий в основе объективных внутренних закономерностей реального мира.

Теорема 5.4.1 (закон больших чисел в форме Чебышёва). *Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины с дисперсиями, ограниченными одной и той же константой C ($\mathsf{D}\xi_i \le C,\ i = 1, 2, \ldots$). Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathsf{M}\xi_i\right| > \varepsilon\right\} \to 0$$

при $n \to \infty$.

Доказательство. Воспользовавшись неравенством Чебышева и независимостью случайных величин $\xi_1, \xi_2, \ldots$, получим:

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathsf{M}\xi_i\right| > \varepsilon\right\} =$$

$$= \mathsf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mathsf{M}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)\right| > \varepsilon\right\} \le \frac{\mathsf{D}\left(\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i\right)}{\varepsilon^2} =$$

$$= \frac{\frac{1}{n^2}\sum\limits_{i=1}^{n}\mathsf{D}\xi_i}{\varepsilon^2} \le \frac{\frac{1}{n^2}nC}{\varepsilon^2} = \frac{C}{n\varepsilon^2},$$

что стремится к нулю при $n \to \infty$.

Следствие. *Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с математическим ожиданием $\mathsf{M}\xi_i = a$ и конечной дисперсией ($\mathsf{D}\xi_i = \sigma^2 < \infty$). Тогда для любого $\varepsilon > 0$ при $n \to \infty$*

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| > \varepsilon\right\} \to 0.$$

Доказательство. Достаточно заметить, что для случайных величин $\xi_1, \xi_2, \ldots$ выполняются условия теоремы и к тому же

$$\frac{1}{n}\sum_{i=1}^{n}\mathsf{M}\xi_i = \frac{1}{n}\sum_{i=1}^{n}a = a.$$

5.5 Примеры и задачи

Примеры

Пример 5.5.1 . *Подбрасывают две симметричные игральные кости. Вычислить математическое ожидание и дисперсию суммы выпавших очков.*

Решение. Пусть ξ — число очков, выпавших на первой игральной кости, η — на второй, тогда $\zeta = \xi + \eta$ — сумма очков, выпавших на двух костях. Так как кости симметричны, то распределение каждой из случайных величин ξ и η имеет следующий вид:

x_i	1	2	3	4	5	6
$\mathsf{P}(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Отсюда имеем:

$$\mathsf{M}\zeta = \mathsf{M}(\xi+\eta) = \mathsf{M}\xi + \mathsf{M}\eta = \frac{7}{2} + \frac{7}{2} = 7.$$

$$\mathsf{D}\xi = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2 = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Поскольку ξ и η — независимые случайные величины, то

$$\mathsf{D}\zeta = \mathsf{D}(\xi+\eta) = \mathsf{D}\xi + \mathsf{D}\eta = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}.$$

Отметим, что непосредственное вычисление математического ожидания и дисперсии суммы ζ очков было бы громоздким.

Пример 5.5.2. *Случайная величина ξ имеет геометрическое распределение с параметром p:*

$$\mathsf{P}(k) = (1-p)^k p,\ k = 0, 1, \ldots\ (0 < p < 1).$$

Вычислить: $\mathsf{M}\xi,\ \mathsf{M}\xi^2,\ \mathsf{D}\xi.$

Решение. В силу теоремы 5.3.1

$$\mathsf{M}\xi = \sum_{k=0}^{\infty} k(1-p)^k p,$$

$$\mathsf{M}\xi^2 = \sum_{k=0}^{\infty} k^2 (1-p)^k p.$$

Чтобы найти суммы этих рядов воспользуемся тем, что при $|x| < 1$

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

Продифференцировав последнее равенство и умножив на x, получим

$$\sum_{k=1}^{\infty} k x^k = \frac{x}{(1-x)^2}, \tag{5.5.1}$$

Продифференцировав равенство (5.5.1) и умножив его на x, получим

$$\sum_{k=0}^{\infty} k^2 x^k = \frac{x(x+1)}{(1-x)^3}. \tag{5.5.2}$$

Воспользовавшись равенствами (5.5.1) и (5.5.2), при $x = 1 - p$ получаем

$$\mathsf{M}\xi = \sum_{k=0}^{\infty} k(1-p)^k p = p \sum_{k=0}^{\infty} k(1-p)^k = p\frac{1-p}{p^2} = \frac{1-p}{p}.$$

$$\mathsf{M}\xi^2 = \sum_{k=0}^{\infty} k^2(1-p)^k p = p \sum_{k=0}^{\infty} k^2(1-p)^k = \frac{(1-p)(2-p)}{p^2}.$$

$$\mathsf{D}\xi = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2 = \frac{1-p}{p^2}.$$

Пример 5.5.3. *Теорема* 5.1.1 *гарантирует существование вероятностного пространства* $\{\Omega, \mathsf{P}\}$ *и случайной величины* $\xi = \xi(\omega)$ *на нем, распределение* P_ξ *которой совпадает с данным распределением* Q*:*

$$\mathsf{P}_\xi = \mathsf{Q},$$

но такое вероятностного пространства $\{\Omega, \mathsf{P}\}$ *и случайная величина* $\xi = \xi(\omega)$ *не единственны.*

Например, на вероятностном пространстве $\{\Omega, \mathsf{P}\}$:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\},\ \mathsf{P}(\omega_i) = 1/4,\ i = 1, 2, 3, 4,$$

случайные величины

$$\xi = \xi(\omega) = I_{\{\omega_1, \omega_2\}}(\omega),\ \omega \in \Omega,$$

$$\eta = \eta(\omega) = I_{\{\omega_3, \omega_4\}}(\omega)\ \omega \in \Omega,$$

не совпадают ни в одной точке, но их распределения P_ξ и P_η совпадают, каждое из них равно Q:

$$\mathsf{Q}(s) = 1/2,\ s = 0; 1.$$

Разумеется различными могут быть и сами пространства.

Пример 5.5.4. *Пусть $\xi_j = \xi_j(\omega)$, $j = 1, 2, \ldots, n$, — независимые случайные величины, для которых*

$$\mathsf{P}\{\xi_j > 0\} = p, \mathsf{P}\{\xi_j < 0\} = q, \mathsf{P}\{\xi_j = 0\} = f,\ \ p + q + f = 1,$$

$j = 1, 2, \ldots, n$; s — количество случайных величин среди ξ_j, $j = 1, 2, \ldots, n$, отличных от нуля, а μ — число положительных среди ξ_j, $j = 1, 2, \ldots, n$.

Найти совместное распределение случайных величин μ и s.

Решение. Вектор $(\mu, s) = (\mu(\zeta^*), s(\zeta^*))$ является функцией случайного вектора $\zeta^* = (\operatorname{sign}\xi_1, \operatorname{sign}\xi_2, \ldots, \operatorname{sign}\xi_n)$.

Сначала найдем распределение ζ^*:

$$\mathsf{P}\{\zeta^* = x\} = \mathsf{P}\{(\operatorname{sign}\xi_1, \operatorname{sign}\xi_2, \ldots, \operatorname{sign}\xi_n) = (x_1, x_2, \ldots, x_n)\} =$$

$$= \prod_{i=1}^{n} \mathsf{P}\{\operatorname{sign}\xi_i = x_i\} = p^{\mu(x)} q^{s(x)-\mu(x)} f^{n-s(x)},$$

где x_i, $i = 1, 2, \ldots, n$, принимает значения $-1, 0, 1$, $s(x)$ — количество компонент вектора x, отличных от нуля, $\mu(x)$ — равных $+1$.

$$\mathsf{P}\{\mu = k, s = j\} = \mathsf{P}\{\mu(\zeta^*) = k, s(\zeta^*) = j\} =$$

$$= \sum_{x:\mu(x)=k, s(x)=j} \mathsf{P}_{\zeta^*}(x) =$$

$$= \sum_{x:\mu(x)=k, s(x)=j} p^{\mu(x)} q^{s(x)-\mu(x)} f^{n-s(x)} = C_n^j C_j^k p^k q^{j-k} f^{n-j},$$

$k = 1, 2, \ldots, j$; $j = 1, 2, \ldots, n$ (см. теорему 5.1.2).

Пример 5.5.5. *Пусть ξ и η — независимые целочисленные случайные величины с распределениями*

$$P\{\xi = k\} = p_k,\ k = 0, 1, \ldots;\ P\{\eta = l\} = q_l,\ l = 0, 1, \ldots$$

Найти распределение случайной величины $\zeta = \xi + \eta$.

Решение. Поскольку случайные величины ξ и η независимы, то их совместным распределением является

$$P\{(\xi, \eta) = (k, l)\} = p_k q_l,\ \ k = 0, 1, \ldots; l = 0, 1, \ldots$$

Далее, согласно теореме 5.1.2 имеем

$$P\{\zeta = m\} = P\{\xi + \eta = m\} =$$

$$= \sum_{(k,l):k+l=m} p_k q_l = \sum_{k=0}^{m} p_k q_{m-k},\ m = 0, 1, \ldots$$

Распределение

$$f_m = \sum_{k=0}^{m} p_k q_{m-k} = p_0 q_m + p_1 q_{m-1} + \ldots + p_m q_0,\ m = 0, 1, \ldots$$

называют сверткой распределений $\{p_k\}$ и $\{q_l\}$.

Так что распределение суммы независимых целочисленных случайных величин равно свертке распределений слагаемых.

Задачи

5.1. Подбрасывают две симметричные игральные кости и рассматривают случайную величину $\zeta = (\xi, \eta)$, где ξ — число очков, выпавших на первой игральной кости, η — на второй.

1. Найти распределения случайных величин:

а) $\max\{\xi, \eta\}$;

б) $\min\{\xi, \eta\}$;

в) $\xi + \eta$.

2. Вычислить:

а) $\mathsf{P}\{\xi \le 2, \max\{\xi, \eta\} \ge 4\}$;

б) $\mathsf{P}\{\max\{\xi, \eta\} \ge 4\}$;

в) $\mathsf{P}\{|\eta - \xi| \ge 3\}$;

г) $\mathsf{P}\{4 \le \xi + \eta \le 6\}$;

д) $\mathsf{P}\{\xi \le 1, \max\{\xi, \eta\} \ge 3\}$;

е) $\mathsf{P}\{\max\{\xi, \eta\} \le 4\}$;

ж) $\mathsf{P}\{\max\{\xi, \eta\} \ge 3\}$;

з) $\mathsf{P}\{\xi \geq 2, \max\{\xi,\eta\} \geq 3\}$.

3. Найти совместное распределение случайных величин:

а) ξ и $\xi+\eta$; являются ли случайные величины ξ и $\xi+\eta$ независимыми?

б) ξ и $\max\{\xi,\eta\}$; являются ли случайные величины ξ и $\max\{\xi,\eta\}$ независимыми?

в) ξ и $\min\{\xi,\eta\}$; являются ли случайные величины ξ и $\min\{\xi,\eta\}$ независимыми?

Решение пункта 2(ж). Сначала найдем распределение $\zeta=(\xi,\eta)$.

Случайная величина $\zeta=\zeta(\omega)=(\xi(\omega),\eta(\omega))$ является функцией на вероятностном пространстве $\{\Omega,\mathsf{P}\}$, пространство элементарных событий Ω которого образуют пары (i,j), $i=1,2,\ldots,6$, $j=1,2,\ldots,6$, а $\mathsf{P}(i,j)=1/36$ для всех (i,j) (выбрана классическая модель, поскольку кости симметричны). Поэтому распределением $\zeta=(\xi,\eta)$ является

$$\mathsf{P}_\zeta(i,j)=\mathsf{P}\{\zeta=(i,j)\}=1/36,\ i=1,2,\ldots,6;\ j=1,2,\ldots,6,$$

его можно записать также в виде табл. 5.5.1.

Таблица 5.5.1. Распределение $\zeta=(\xi,\eta)$

Значения ξ	Значения η					
	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

По известному распределению случайной величины $\zeta=(\xi,\eta)$ (см. табл. 5.5.1), пользуясь теоремой 5.1.2 (здесь $g=g(s,t)= \max\{s,t\}$, $B=(-\infty,3)$), находим

$$\mathsf{P}\{\max\{\xi,\eta\}\geq 3\}=1-\mathsf{P}\{\max\{\xi,\eta\}<3\}=$$

$$= 1 - \sum_{(i,j):\max\{i,j\}<3} \mathsf{P}_\zeta(i,j) =$$

$$= 1 - (\mathsf{P}_\zeta(1,1) + \mathsf{P}_\zeta(1,2) + \mathsf{P}_\zeta(2,1) + \mathsf{P}_\zeta(2,2)) = 1 - 4/36 = 8/9.$$

Решение 3(в). Совместное распределение случайных величин ξ и $\min\{\xi,\eta\}$ (распределение вектора $(\xi,\min\{\xi,\eta\})$) находим как распределение функции $\theta = g(\xi,\eta) = (\xi,\min\{\xi,\eta\})$ от случайной величины $\zeta = (\xi,\eta)$ по ее распределению (см. табл. 5.5.1), пользуясь теоремой 5.1.2. Вычислим несколько значений $\mathsf{P}_\theta(i,j) = \mathsf{P}\{(\xi,\min\{\xi,\eta\}) = (i,j)\}$:

$$\mathsf{P}\{(\xi,\min\{\xi,\eta\}) = (1,1)\} = \sum_{(i,j):(i,\min\{i,j\})=(1,1)} \mathsf{P}_\zeta(i,j) =$$

$$= \mathsf{P}_\zeta(1,1) + \mathsf{P}_\zeta(1,2) + \ldots + \mathsf{P}_\zeta(1,6) = 6/36;$$

$$\mathsf{P}\{(\xi,\min\{\xi,\eta\}) = (1,2)\} = \sum_{(i,j):(i,\min\{i,j\})=(1,2)} \mathsf{P}_\zeta(i,j) = 0;$$

$$\mathsf{P}\{(\xi,\min\{\xi,\eta\}) = (1,k)\} = \sum_{(i,j):(i,\min\{i,j\})=(1,k)} \mathsf{P}_\zeta(i,j) = 0;$$

$k = 3,4,5,6$. И так далее.

Совместное распределение ξ и $\min\{\xi,\eta\}$ (распределение случайной величины $\theta = (\xi,\min\{\xi,\eta\})$) приведено в табл. 5.5.2.

Таблица 5.5.2. Распределение $\theta = (\xi,\min\{\xi,\eta\})$

Значения ξ	Значения $\min\{\xi,\eta\}$					
	1	2	3	4	5	6
1	6/36	0	0	0	0	0
2	1/36	5/36	0	0	0	0
3	1/36	1/36	4/36	0	0	0
4	1/36	1/36	1/36	3/36	0	0
5	1/36	1/36	1/36	1/36	2/36	0
6	1/36	1/36	1/36	1/36	1/36	1/36

.

Случайные величины ξ и $\min\{\xi, \eta\}$ не являются независимыми, поскольку их совместное распределение отлично от произведения распределения

1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

случайной величины ξ и распределения

1	2	3	4	5	6
11/36	9/36	7/36	5/36	3/36	1/36

случайной величины $\min\{\xi, \eta\}$ (см. теорему 5.2.1).

Ответ к пункту 1(а):

1	2	3	4	5	6
1/36	3/36	5/36	7/36	9/36	11/36

.

Ответ к пункту 1(б):

1	2	3	4	5	6
11/36	9/36	7/36	5/36	3/36	1/36

.

Ответы: 2(а) 1/6; 2(б) 3/4; 2(в) 1/3; 2(г) 1/3; 2(д) 1/9; 2(е) 4/9; 2(ж) 8/9; 2(з) 7/9.

Ответы к 3(б):

Распределение случайной величины $\theta = (\xi, \min\{\xi, \eta\})$ (совместное распределение ξ и $\min\{\xi, \eta\}$) приведено в табл. 5.5.2.

Случайные величины ξ и $\min\{\xi, \eta\}$ не являются независимыми, поскольку их совместное распределение (см. табл. 5.5.2) отлично от произведения распределений случайных величин ξ и $\min\{\xi, \eta\}$ (см. также замечание к теореме 5.2.1). Распределения $\min\{\xi, \eta\}$ приведены в ответе к 1(а).

5.2. Грани симметричной игральной кости занумерованы двумя единицами, двумя двойками и двумя тройками.

В квадратном уравнении

$$\xi x^2 + \eta x + \zeta = 0$$

ξ, η, ζ определяются как результат трех последовательных независимых подбрасываний игральной кости. Найти вероятность того, что: 1) уравнение имеет действительные корни; 2) уравнение имеет рациональные корни; 3) уравнение не имеет действительных корней.

Указание. Совместное распределение случайных величин ξ, η, ζ:

$$\mathsf{P}_{\varphi}(i, j, k) = \mathsf{P}\{\xi = i, \eta = j, \zeta = k\} = 1/27, \ i, j, k = 1, 2, 3.$$

Уравнение имеет действительные корни, если $\eta^2 - 4\xi\zeta \geq 0$;

$$\mathsf{P}\{\eta^2 - 4\xi\zeta \geq 0\} = \sum_{(i,j,k): j^2 - 4ik \geq 0} \mathsf{P}_{\varphi}(i, j, k) = \sum_{(i,j,k): j^2 - 4ik \geq 0} 1/27 = 4/27$$

(см. теорему 5.1.2).

5.3. Доказать, что если независимые случайные величины ξ_1 и ξ_2 имеют геометрическое распределение, то и случайная величина $\eta = \min\{\xi_1, \xi_2\}$ имеет геометрическое распределение. Найти параметр этого распределения, если параметры распределений случайных величин ξ_1 и ξ_2 равны соответственно p_1 и p_2.

Ответ: $p_1 + p_2 - p_1 p_2$.

5.4. Пусть ξ и η — независимые случайные величины, каждая из которых имеет распределение Пуассона с параметром λ:

$$\mathsf{P}_{\xi}(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ k = 0, 1, \ldots$$

Вычислить:

$$1)\, \mathsf{M}\frac{\xi}{1 + \eta}; \ 2)\, \mathsf{M}\xi\eta; \ 3)\, \mathsf{D}\xi\eta; \ 4)\, \mathsf{D}(\xi + \eta).$$

Ответы: 1) $1 - e^{-\lambda}$; 2) λ^2; 3) $\lambda^2(1 + 2\lambda)$; 4) 2λ.

5.5. Случайный вектор $\mu = (\mu_1, \mu_2, \ldots, \mu_r)$ имеет распределение:

$$\mathsf{P}\{(\mu_1, \mu_2, \ldots, \mu_k, \mu_{k+1}, \ldots, \mu_r) = (0, 0, \ldots, 0, 1, 0, \ldots, 0)\} = p_k,$$

в последовательности $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$ единица расположена на k-м месте, $k = 1, 2, \ldots, r$.

Вычислить

$$\mathsf{M}\exp\{i(t, \mu)\} = \mathsf{M}\exp\{i(t_1\mu_1 + t_2\mu_2 + \ldots + t_r\mu_r)\}.$$

Ответ: $\sum\limits_{k=1}^{r} p_k \exp\{it_k\}$.

5.6. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые одинаково распределенные случайные величины со значениями в $\mathbb{R}^1$, X_j, $j = 1, 2, \ldots, r$, — разбиение $\mathbb{R}^1$ на не пересекающиеся подмножества: $\bigcup\limits_{j=1}^{r} X_j = \mathbb{R}^1, X_s \cap X_l = \emptyset, s \neq l$, $\mathsf{P}\{\xi_k \in X_j\} = p_j$, $j = 1, 2, \ldots, r$, $\nu = (\nu_1, \nu_2, \ldots, \nu_j, \ldots, \nu_r)$ — случайный вектор, j-я компонента ν_j которого равна количеству случайных величин из $\xi_1, \xi_2, \ldots, \xi_n$, которые оказались в X_j, $j = 1, 2, \ldots, r$.

Вычислить

$$\mathsf{M}\exp\{i(t, \nu)\} = \mathsf{M}\exp\{i(t_1\nu_1 + t_2\nu_2 + \ldots + t_r\nu_r)\}.$$

Указание. Выразите вектор $\nu = (\nu_1, \nu_2, \ldots, \nu_r)$ через векторы

$$(\mu_1^{(k)}, \mu_2^{(k)}, \ldots, \mu_r^{(k)}) = (I_{X_1}(\xi_k), I_{X_2}(\xi_k), \ldots, I_{X_r}(\xi_k)),$$

$k = 1, 2, \ldots, r$, где I_A — индикатор множества A.

Ответ: $\left(\sum\limits_{k=1}^{r} p_k \exp\{it_k\}\right)^n$.

Глава 6

Испытания Бернулли

6.1 Испытания Бернулли. Биномиальное распределение

В теории вероятностей важное место занимает стохастический эксперимент, известный под называнием испытаний Бернулли. На неформальном языке его можно определить как последовательность независимых испытаний с двумя исходами (один из которых будем называть успехом, другой — неудачей) и вероятностью успеха, не изменяющейся от испытания к испытанию.

Примеры

Испытаниями Бернулли будут:

1) последовательные подбрасывания симметричных монет (герб — успех, решетка — неудача);

2) стрельба по мишени без корректировки (попадание — успех, промах — неудача);

3) наблюдения за погодой, проводимые в определенный день (скажем, 7 мая) каждого года (дождь — успех, нет дождя — неудача).

Испытаниями Бернулли не будут:

1) последовательные подбрасывания в разной степени изогнутых монет — от подбрасывания к подбрасыванию меняется вероятность успеха;

2) стрельба по мишени с корректировкой — нет независимости результатов отдельных выстрелов, изменяется вероятность успеха;

3) наблюдение за погодой в последовательные дни одного года — нет независимости наблюдений.

Вероятностное пространство испытаний Бернулли. В качестве пространства Ω элементарных событий n испытаний Бернулли естественно рассмотреть множество упорядоченных последовательностей длиной n, составленных из нулей и единиц. Каждая такая последовательность однозначно описывает исход эксперимента. Например, последовательность $\omega = (110\dots01)$ описывает исход: в первом испытании — успех, во втором — успех, в третьем — неудача, $\dots$, в n-м — успех (единицей мы обозначили успех, нулем — неудачу). Множество Ω конечно.

Вероятность на классе всех подмножеств дискретного Ω задается посредством неотрицательной функции $\mathsf{P}(\omega)$ точки $\omega \in \Omega$, такой что $\sum\limits_{\omega\in\Omega} \mathsf{P}(\omega) = 1$ (см. теорему 4.1.1). В испытаниях Бернулли в качестве $\mathsf{P}(\omega)$ естественно рассмотреть функцию

$$\mathsf{P} : \omega \to \mathsf{P}(\omega) = p^{\mu(\omega)}(1-p)^{n-\mu(\omega)} = p^{\mu(\omega)}q^{n-\mu(\omega)},\ \omega \in \Omega, \quad (6.1.1)$$

где $\mu(\omega)$ — число единиц в последовательности ω, $(n - \mu(\omega))$ — число нулей, p — параметр $(0 < p < 1)$, $q = 1 - p$.

Функция $\mathsf{P}(\omega) = p^{\mu(\omega)}q^{n-\mu(\omega)}$, во-первых, неотрицательна:

$$\mathsf{P}(\omega) \geq 0,\ \ \omega \in \Omega;$$

и, во вторых,

$$\sum_{\omega\in\Omega} \mathsf{P}(\omega) = 1.$$

Чтобы убедиться в справедливости последнего равенства, достаточно заметить, что

$$\bigcup_{k=0}^{n} \{\omega : \mu(\omega) = k\} = \Omega,\ \{\omega : \mu(\omega) = k\} \cap \{\omega : \mu(\omega) = j\} = \emptyset,\ k \neq j,$$

поэтому

$$\sum_{\omega\in\Omega} \mathsf{P}(\omega) = \sum_{\omega\in\Omega} p^{\mu(\omega)}q^{n-\mu(\omega)} =$$

$$= \sum_{\omega:\mu(\omega)=0} p^{\mu(\omega)}q^{n-\mu(\omega)} + \sum_{\omega:\mu(\omega)=1} p^{\mu(\omega)}q^{n-\mu(\omega)} + \dots$$

$$+ \sum_{\omega:\mu(\omega)=k} p^{\mu(\omega)}q^{n-\mu(\omega)} + \dots + \sum_{\omega:\mu(\omega)=n} p^{\mu(\omega)}q^{n-\mu(\omega)} =$$

$$= q^n + C_n^1 p^1 q^{n-1} + C_n^2 p^2 q^{n-2} + \ldots + C_n^k p^k q^{n-k} + \ldots + p^n = (p+q)^n = 1. \tag{6.1.2}$$

Каждая из сумм $\sum\limits_{\omega:\mu(\omega)=k} p^{\mu(\omega)} q^{n-\mu(\omega)}$, $k = 0, 1, \ldots, n$, вычисляется так:

$$\sum_{\omega:\mu(\omega)=k} p^{\mu(\omega)} q^{n-\mu(\omega)} = \sum_{\omega:\mu(\omega)=k} p^k q^{n-k} = C_n^k p^k q^{n-k} \tag{6.1.3}$$

— число одинаковых слагаемых $p^k q^{n-k}$ в сумме $\sum\limits_{\omega:\mu(\omega)=k} p^k q^{n-k}$ равно количеству последовательностей ω длиной n, составленных из k единиц и $n-k$ нулей, а число последних равно C_n^k.

Пару $\{\Omega, \mathsf{P}\}$, где Ω — множество последовательностей ω длиной n, составленных из нулей и единиц, $\mathsf{P}(\omega)$ — функция на Ω, определенная равенством (6.1.1), и будем рассматривать в качестве вероятностного пространства (математической модели) стохастического эксперимента, состоящего в проведении n испытаний Бернулли.

В наших интуитивных представлениях об испытаниях Бернулли события “успех” и “неудача”, связанные с различными испытаниями, независимы и вероятность успеха не меняется от испытания к испытанию. Мы покажем, что предложенная модель обладает свойствами, которые именно так и можно трактовать. Для этого определим на $\{\Omega, \mathsf{P}\}$ случайные величины μ_k:

$$\mu_k(\omega) = \begin{cases} 1, & \text{если в } \omega \text{ на } k\text{-м месте } 1; \\ 0, & \text{если в } \omega \text{ на } k\text{-м месте } 0, \end{cases}$$

$k = 1, 2, \ldots, n$. Другими словами, $\mu_k(\omega)$ — символ на k-м месте в последовательности ω, составленной из нулей и единиц.

Теорема 6.1.1. *Случайные величины $\mu_k, k = 1, 2, \ldots n$, независимы в совокупности и одинаково распределены:*

$$\mathsf{P}\{\mu_k = s\} = p^s q^{1-s},\ s = 0, 1;\ k = 1, 2, \ldots, n.$$

Доказательство. Сначала найдем распределения случайных величин μ_k, $k = 1, 2, \ldots, n$.

Случайная величина μ_k принимает два значения: 0 и 1. При этом

$$\mathsf{P}\{\mu_k = 1\} = \mathsf{P}\{\omega : \mu_k(\omega) = 1\} = \sum_{\omega:\mu_k(\omega)=1} \mathsf{P}(\omega) =$$

$$= pC_{n-1}^0 p^0 q^{n-1} + pC_{n-1}^1 p^1 q^{n-2} + \ldots + pC_{n-1}^{n-1} p^{n-1} q^0 =$$

$$= p(p+q)^{n-1} = p,$$

$$\mathsf{P}\{\mu_k = 0\} = 1 - p = q,\ k = 1, 2, \ldots, n.$$

Сумма $\sum\limits_{\omega:\mu_k(\omega)=1} \mathsf{P}(\omega)$ вычисляется аналогично сумме $\sum\limits_{\omega\in\Omega} \mathsf{P}(\omega)$, см. (6.1.2). Так что каждая из случайных величин $\mu_k, k = 1, 2, \ldots$ $\ldots, n$, имеет своим распределением

$$\mathsf{P}\{\mu_k = s\} = p^s q^{1-s},\ s = 0, 1.$$

Покажем, что случайные величины μ_k, $k = 1, 2, \ldots, n$, независимы в совокупности. Для этого выпишем их совместное распределение. Пусть x_k — возможное значение случайной величины μ_k (x_k принимает значения 0 или 1, $k = 1, 2, \ldots, n$),

$$\mathsf{P}\{\mu_1 = x_1, \mu_2 = x_2, \ldots, \mu_n = x_n\} =$$

$$= \mathsf{P}\{\omega : \mu_1(\omega) = x_1, \mu_2(\omega) = x_2, \ldots, \mu_n(\omega) = x_n\} =$$

$$= \mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\omega).$$

Множество

$$A = \{\omega : \mu_1(\omega) = x_1, \mu_2(\omega) = x_2, \ldots, \mu_n(\omega) = x_n\}$$

состоит из последовательностей длиной n, составленных из нулей и единиц, у которых на первом месте x_1, на втором x_2, $\ldots$, на n-м — x_n. Оно содержит только одну точку ω: $\omega =$ $= (x_1, x_2, \ldots, x_n)$. Поэтому

$$\mathsf{P}\{\mu_1 = x_1, \mu_2 = x_2, \ldots, \mu_n = x_n\} = \mathsf{P}(A) =$$

$$= \mathsf{P}(x_1, x_2, \ldots, x_n) = \mathsf{P}(\omega) = p^{\mu(\omega)} q^{n-\mu(\omega)} = p^{\sum\limits_{i=1}^n x_i} q^{n-\sum\limits_{i=1}^n x_i},$$

число единиц $\mu(\omega)$ в последовательности $\omega = (x_1, x_2, \ldots, x_n)$

$$\mu(\omega) = \sum_{i=1}^n x_i.$$

Так что

$$\mathsf{P}\{\mu_1 = x_1, \mu_2 = x_2, \ldots, \mu_n = x_n\} =$$

$$= p^{\sum\limits_{i=1}^{n} x_i} q^{n - \sum\limits_{i=1}^{n} x_i} = \prod_{i=1}^{n} p^{x_i} q^{1-x_i} = \prod_{i=1}^{n} \mathsf{P}\{\mu_i = x_i\}$$

— совместное распределение случайных величин $\mu_1, \mu_2, \ldots, \mu_n$ равно произведению их распределений. Поэтому случайные величины $\mu_1, \mu_2, \ldots, \mu_n$ независимы.

Замечание. Приведенная теорема показывает, что предложенная модель описывает (в терминах случайных величин μ_k) следующие свойства схемы испытаний Бернулли:

1) события, наблюдаемые в разных испытаниях, независимы — случайные величины $\mu_k, k = 1, 2, \ldots, n$, независимы;

2) вероятность успеха не меняется от испытания к испытанию — случайные величины $\mu_k, k = 1, 2, \ldots, n$, одинаково распределены:

$$\mathsf{P}\{\mu_k = s\} = p^s q^{1-s}, \; s = 0, 1.$$

Биномиальное распределение. В связи с испытаниями Бернулли нас часто будет интересовать ответ на вопрос: какова вероятность события, состоящего в том, что в n испытаниях Бернулли будет ровно k успехов.

Обозначим через μ число успехов в n испытаниях Бернулли. Случайная величина $\mu = \mu(\omega)$ — функция на Ω вероятностного пространства $\{\Omega, \mathsf{P}\}$ n испытаний Бернулли, принимающая в точке ω значение k, если в последовательности ω ровно k единиц. Вероятность $\mathsf{P}\{\mu = k\}$ обычно обозначают через $\mathsf{P}_n(k)$.

Теорема 6.1.2. *В последовательности n испытаний Бернулли с вероятностью успеха p в одном испытании вероятность $\mathsf{P}_n(k)$ появления k успехов равна $C_n^k p^k q^{n-k}$:*

$$\mathsf{P}_n(k) = C_n^k p^k q^{n-k}, \; k = 0, 1, \ldots, n. \tag{6.1.4}$$

Доказательство.

$$\mathsf{P}_n(k) = \mathsf{P}\{\mu = k\} = \mathsf{P}\{\omega : \mu(\omega) = k\} = \sum_{\omega:\mu(\omega)=k} \mathsf{P}(\omega) =$$

$$= \sum_{\omega:\mu(\omega)=k} p^{\mu(\omega)} q^{n-\mu(\omega)} = \sum_{\omega:\mu(\omega)=k} p^k q^{n-k} = C_n^k p^k q^{n-k},$$

относительно вычисления суммы $\sum\limits_{\omega:\mu(\omega)=k} p^{\mu(\omega)}q^{n-\mu(\omega)}$ см. равенство (6.1.3). Так что

$$\mathsf{P}\{\mu = k\} = \mathsf{P}_n(k) = C_n^k p^k q^{n-k}.$$

Эту формулу называют формулой Бернулли.

Определение. Распределение

$$B_{n;p}(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n,\ (0 < p < 1)$$

называется *биномиальным распределением* с параметрами $(n; p)$.

Так что число μ успехов в n испытаниях Бернулли с вероятностью успеха p в одном испытании имеет биномиальное распределение с параметрами $(n; p)$:

$$\mathsf{P}_n(k) = \mathsf{P}\{\mu = k\} = B_{n;p}(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

Определение. Распределение $B_{1;p}$ называют *распределением Бернулли* с параметром p.

Случайную величину, имеющую распределение Бернулли, называют *бернуллиевой* случайной величиной. Бернуллиевая случайная величина равна числу успехов в одном испытании Бернулли.

Если случайная величина μ имеет распределение Бернулли с параметром p, то

$$\mathsf{P}\{\mu = s\} = p^s (1-p)^{1-s},\ s = 0; 1;$$

$$\mathsf{M}\mu = p,\ \mathsf{D}\mu = p(1-p).$$

Теорема 6.1.3. *Пусть μ — число успехов в n испытаниях Бернулли с вероятностью успеха p в одном испытании, тогда*

$$\mathsf{M}\mu = np,\ \mathsf{D}\mu = npq.$$

Доказательство. Число μ успехов в n испытаниях складывается из числа успехов в каждом испытании, т. е.

$$\mu = \sum_{k=1}^{n} \mu_k, \tag{6.1.5}$$

где μ_k — число успехов в k-м испытании (равное 0 или 1). Случайные величины μ_k независимы и одинаково распределены:

$$\mathsf{P}\{\mu_k = s\} = p^s q^{1-s},\ s = 0, 1;\ k = 1, 2, \ldots, n,$$

(см. теорему 6.1.1). Отсюда, в частности, имеем, что

$$\mathsf{M}\mu_k = p,\ \mathsf{D}\mu_k = pq.$$

А из представления (6.1.5) получаем:

$$\mathsf{M}\mu = \mathsf{M}\sum_{k=1}^{n}\mu_k = \sum_{k=1}^{n}\mathsf{M}\mu_k = np,$$

$$\mathsf{D}\mu = \mathsf{D}\sum_{k=1}^{n}\mu_k = \sum_{k=1}^{n}\mathsf{D}\mu_k = npq$$

— мы воспользовались тем, что дисперсия суммы независимых случайных величин равна сумме их дисперсий.

Далее, как обычно, $[x]$ — целая часть числа x.

Теорема. *Если в последовательности n испытаний Бернулли с вероятностью успеха p в одном испытании число $(n+1)p$ — не целое и $(n+1)p > 1$, то*

$$\mathsf{P}_n(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n,$$

достигает наибольшего значения в точке $m = [(n+1)p]$, монотонно возрастая при $k < m$ и монотонно убывая при $k > m$; если $m = (n+1)p$ — целое, то $\mathsf{P}_n(k)$ достигает наибольшего значения в точках $m-1$ и m ($\mathsf{P}_n(m-1) = \mathsf{P}_n(m)$), монотонно возрастая при $k < m-1$ и монотонно убывая при $k > m$.

Если $(n+1)p < 1$, то $\mathsf{P}_n(k)$ монотонно убывает (наибольшее значение равно $\mathsf{P}_n(0)$), если $(n+1)p = 1$, то $\mathsf{P}_n(0) = \mathsf{P}_n(1)$ и $\mathsf{P}_n(k)$ монотонно убывает при $k > 1$.

Доказательство. Пусть $(n+1)p > 1$.

$$\frac{\mathsf{P}_n(k)}{\mathsf{P}_n(k-1)} = \frac{C_n^k p^k q^{n-k}}{C_n^{k-1} p^{k-1} q^{n-(k-1)}} = \frac{(n+1-k)p}{kq} =$$

$$= \frac{(n+1)p - k(1-q)}{kq} = 1 + \frac{(n+1)p - k}{kq}.$$

Поэтому если $k < (n+1)p$, то $\mathsf{P}_n(k)/\mathsf{P}_n(k-1) > 1$ (с возрастанием k вероятность $\mathsf{P}_n(k)$ возрастает); если же $k > (n+1)p$, то $\mathsf{P}_n(k)/\mathsf{P}_n(k-1) < 1$ (с возрастанием k вероятности $\mathsf{P}_n(k)$ убывает). Отсюда имеем, что если $(n+1)p$ не целое, то $\mathsf{P}_n(k)$ достигает наибольшего значения в точке $m = [(n+1)p]$. Если же $(n+1)p$

— целое число, то наибольших, равных между собой, значений $\mathsf{P}_n(k)$ два: $\mathsf{P}_n(m-1)$ и $\mathsf{P}_n(m)$ ($\mathsf{P}_n(m-1) = \mathsf{P}_n(m)$).

Случаи $(n+1)p < 1$ и $(n+1)p = 1$ рассматриваются аналогично.

Теорема 6.1.4. *Сумма n независимых бернуллиевых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ с параметром p:*

$$\mathsf{P}\{\xi_j = s\} = p^s(1-p)^{1-s},\ s = 0; 1;\ j = 1, 2, \ldots, n,$$

имеет биномиальное распределение с параметрами (n, p).

Доказательство. Поскольку случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы, их совместное распределение (распределение вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$) имеет вид

$$\mathsf{P}_\xi(x_1, x_2, \ldots, x_n) = \mathsf{P}\{\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n\} =$$

$$= \prod_{i=1}^{n} \mathsf{P}\{\xi_i = x_i\} = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum\limits_{i=1}^{n} x_i}(1-p)^{n-\sum\limits_{i=1}^{n} x_i}, \quad (6.1.6)$$

где x_j — возможное значение случайной величины ξ_j, равное 0 или 1, $j = 1, 2, \ldots, n$.

Распределение случайной величины $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$ вычислим как распределение функции

$$S_n = \xi_1 + \xi_2 + \ldots + \xi_n = g(\xi_1, \xi_2, \ldots, \xi_n)$$

от случайного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ по его распределению (6.1.6), см. теорему 5.1.2:

$$\mathsf{P}\{S_n = k\} = \mathsf{P}\{\xi_1 + \xi_2 + \ldots + \xi_n = k\} =$$

$$\sum_{(x_1, x_2, \ldots x_n): x_1 + x_2 + \ldots + x_n = k} \mathsf{P}_\xi(x_1, x_2, \ldots x_n) =$$

$$= \sum_{(x_1, x_2, \ldots x_n): x_1 + x_2 + \ldots + x_n = k} p^{\sum\limits_{i=1}^{n} x_i}(1-p)^{n-\sum\limits_{i=1}^{n} x_i} =$$

$$= \sum_{(x_1, x_2, \ldots x_n): x_1 + x_2 + \ldots + x_n = k} p^k(1-p)^{n-k} = C_n^k p^k (1-p)^{n-k}.$$

Так что

$$\mathsf{P}\{S_n = k\} = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

Мультиномиальное распределение. Естественным обобщением испытаний Бернулли является последовательность n независимых испытаний с r исходами, обозначим их через $a_1, a_2, \ldots$ $\ldots, a_r$, и вероятностями исходов, не меняющимися от испытания к испытанию.

В качестве пространства Ω элементарных событий n независимых испытаний с r исходами $a_1, a_2, \ldots, a_r$, будем рассматривать множество упорядоченных последовательностей длиной n, составленных из элементов $a_1, a_2, \ldots, a_r$, т. е. множество слов длиной n, составленных из букв $a_1, a_2, \ldots, a_r$ (о словах см. перестановки с повторениями в разд. 4.3 главы 4).

Множество Ω слов длиной n, составленных из букв $a_1, a_2, \ldots$ $\ldots, a_r$ конечно. Вероятность на классе подмножеств Ω зададим неотрицательной функцией точки

$$\mathsf{P} : \omega \to \mathsf{P}(\omega) = p_1^{\mu_1(\omega)} p_2^{\mu_2(\omega)} \ldots p_r^{\mu_r(\omega)}, \omega \in \Omega,$$

ставящей в соответствие слову ω длиной n, составленном из μ_1 букв a_1, μ_2 букв a_2 и т. д. μ_r букв a_r $(\mu_1 + \mu_2 + \ldots + \mu_r = n)$, число

$$p_1^{\mu_1(\omega)} p_2^{\mu_2(\omega)} \ldots p_r^{\mu_r(\omega)},$$

$p_1 > 0, p_2 > 0, \ldots, p_r > 0;$ $p_1 + p_2 + \ldots + p_r = 1.$

Легко видеть, что

$$\sum_{\omega\in\Omega} \mathsf{P}(\omega) = \sum_{\omega\in\Omega} p_1^{\mu_1(\omega)} p_2^{\mu_2(\omega)} \ldots p_r^{\mu_r(\omega)} =$$

$$= \sum_{k_1+k_2+\ldots+k_r=n} \quad \sum_{\omega:\mu_1(\omega)=k_1,\ldots,\mu_r(\omega)=k_r} p_1^{\mu_1(\omega)} p_2^{\mu_2(\omega)} \ldots p_r^{\mu_r(\omega)} =$$

$$= \sum_{k_1+k_2+\ldots+k_r=n} \quad \sum_{\omega:\mu_1(\omega)=k_1,\ldots,\mu_r(\omega)=k_r} p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r} =$$

$$= \sum_{k_1+k_2+\ldots+k_r=n} C_n(k_1, k_2, \ldots, k_r) p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r} =$$

$$= (p_1 + p_2 + \ldots + p_r)^n = 1,$$

во “внешней” сумме суммирование ведется по последовательностям $(k_1, k_2, \ldots, k_r)$ неотрицательных целых чисел $k_1, k_2, \ldots, k_r$, для которых $k_1 + k_2 + \ldots + k_r = n$ (см. также полиномиальную формулу в разд. 4.3 главы 4).

Пару $\{\Omega, \mathsf{P}\}$ и будем рассматривать в качестве математической модели последовательности n независимых испытаний с r исходами $a_1, a_2, \ldots, a_r$ и вероятностями исходов, не меняющимися от испытания к испытанию.

Аналогично тому, как было получено распределение числа успехов μ в n испытаниях Бернулли, получаем распределение числа исходов $a_1, a_2, \ldots, a_r$ — распределение вектора $(\mu_1, \mu_2, \ldots$ $\ldots, \mu_r)$ — в n независимых испытаний с r исходами:

$$\mathsf{P}\{\mu_1 = k_1, \mu_2 = k_2, \ldots, \mu_r = k_r\} = C_n(k_1, k_2, \ldots, k_r) p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r}, \tag{6.1.7}$$

$k_1 = 0, 1, \ldots, n; \ldots; k_r = 0, 1, \ldots, n;$ $k_1 + k_2 + \ldots + k_r = n$, p_i — вероятность исхода a_i $(i = 1, 2, \ldots, r)$.

Теорема Бернулли. Исторически первым теоретическим результатом, объясняющим давно известный факт устойчивости частоты, была знаменитая теорема Бернулли, впоследствии названная законом больших чисел. Этот результат положил начало теории вероятностей как самостоятельной науке.

Теорема 6.1.5 (Бернулли). *Пусть μ — число успехов в n испытаниях Бернулли с вероятностью успеха p в одном испытании. Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}\left\{\left|\frac{\mu}{n} - p\right| > \varepsilon\right\} \to 0$$

при $n \to \infty$.

Доказательство. Как известно (см. равенство (6.1.5)), число успехов μ можно представить в виде суммы

$$\mu = \sum_{k=1}^{n} \mu_k,$$

независимых случайных величин μ_k $(k = 1, 2, \ldots, n)$, каждая из которых имеет распределение

$$\mathsf{P}\{\mu_k = s\} = p^s (1-p)^{1-s},\ s = 0; 1.$$

Поскольку для случайных величин $\mu_1, \mu_2, \ldots$

$$\mathsf{M}\mu_k = p,\ \mathsf{D}\mu_k = p(1-p) < \infty,$$

то в силу следствия из закона больших чисел при $n \to \infty$

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} \mu_k - p\right| > \varepsilon\right\} \to 0.$$

Или, что то же: при $n \to \infty$

$$\mathsf{P}\left\{\left|\frac{\mu}{n} - p\right| > \varepsilon\right\} \to 0.$$

Тем самым теорема Бернулли доказана.

6.2 Предельные теоремы для биномиального распределения

Задача о телефонных линиях. Телефонная станция A должна связать 500 абонентов с соседней станцией B. Было бы дорого (и бессмысленно) прокладывать 500 линий связи. Разумно ограничиться таким минимальным числом линий, которое бы обеспечивало приемлемое качество связи, а именно, вероятность отказа в связи была бы мала, скажем, не превосходила 0,01 (в среднем на 100 требований связи приходится один отказ). Каким при этом должно быть число линий связи?

Попытаемся решить поставленную задачу. Обозначим через k минимальное число линий, которые обеспечивают описанное качество связи, через μ — суммарное число требований связи в наудачу выбранный момент наиболее напряженного часа дня. Если каждому абоненту станции A независимо от других необходима связь с B в среднем длительностью, скажем, одну минуту, то в наудачу выбранный момент данного часа каждый из 500 абонентов (независимо от остальных) потребует связь с вероятностью $1/60$ (момент связи приходится на данную минуту с вероятностью $1/60$). В этих предположениях случайную величину μ можно рассматривать как число успехов в 500 независимых испытаниях с вероятностью успеха $p = 1/60$ в одном испытании. Распределением μ является биномиальное распределение с параметрами $n = 500$ и $p = 1/60$, т. е.

$$\mathsf{P}\{\mu = i\} = C_{500}^{i} p^{i} (1-p)^{500-i},\ i = 0, 1, \ldots, 500.$$

Наступление события $\mu > k$ означает отказ в связи. И поскольку мы хотим, чтобы вероятность отказа $\mathsf{P}\{\mu > k\}$ не превосходила $0,01$, то k следует выбрать (естественно, минимальным) из условия

$$\mathsf{P}\{\mu > k\} = \sum_{i=k+1}^{500} C_{500}^{i} p^{i} (1-p)^{500-i} \le 0,01.$$

Для этого необходимо уметь вычислять значения выражений вида
$$B_{n;p}(m) = C_n^m p^m (1-p)^{n-m}$$
при больших n и m (порядка десятков и сотен), что делать непосредственно весьма затруднительно. Поэтому возникает необходимость в получении "хороших" приближенных формул для вычисления $B_{n;p}(m) = C_n^m p^m (1-p)^{n-m}$ при больших n и m. Далее рассматриваются два приближения для $B_{n;p}(m)$ — в различных предположениях относительно n и p. Эти приближения даются теоремой Пуассона и локальной предельной теоремой. Сначала рассмотрим теорему Пуассона.

Теорема Пуассона. Теорема Пуассона дает приближение для
$$B_{n;p}(m) = C_n^m p^m (1-p)^{n-m},$$
когда число испытаний n "велико", а вероятность успеха p "мала".

Чтобы на математическом языке корректно выразить утверждения "число n большое", "число p малое", величины n и p должны быть переменными, стремящимися соответственно к ∞ и к 0. Но в испытаниях Бернулли вероятность успеха постоянна. Поэтому теорема Пуассона формулируется в терминах серий испытаний Бернулли. Мы будем предполагать, что имеется бесконечная последовательность серий испытаний Бернулли: в первой серии одно испытание, вероятность успеха p_1, во второй серии два испытания, вероятность успеха p_2, и т. д., в n-й серии n испытаний, вероятность успеха p_n и т. д. В обозначении p_n индекс n будем опускать. Через μ будем обозначать число успехов.

Теорема 6.2.1 (Пуассона). *Если в испытаниях Бернулли при $n \to \infty$ среднее число успехов np стремится к конечному λ ($\lambda > 0$), то распределение числа успехов*
$$B_{n;p}(m) = C_n^m p^m (1-p)^{n-m},\ m = 0, 1, \ldots, n,$$
сходится к распределению Пуассона с параметром λ: для каждого фиксированного m, $m = 0, 1, 2, \ldots$,
$$\lim_{n\to\infty} C_n^m p^m (1-p)^{n-m} = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Доказательство. Согласно условию теоремы $np \to \lambda$ при $n \to \infty$, поэтому $np - \lambda = o(1)$ ($o(1)$ — величина, стремящаяся к нулю при $n \to \infty$). Отсюда
$$p = (\lambda + o(1))/n$$

и, следовательно, $C_n^m p^m (1-p)^{n-m}$ можно представить в виде:

$$C_n^m p^m (1-p)^{n-m} =$$

$$= \frac{n(n-1)\dots(n-(m-1))}{m!}\left(\frac{1}{n}(\lambda+o(1))\right)^m \times$$

$$\times \left(1-\frac{1}{n}(\lambda+o(1))\right)^{n-m} =$$

$$= \frac{1}{m!}\cdot\frac{n(n-1)\dots(n-(m-1))}{n^m}(\lambda+o(1))^m \times$$

$$\times \left(1-\frac{1}{n}(\lambda+o(1))\right)^{n-m}.$$

Далее,

$$\lim_{n\to\infty}\frac{n(n-1)\dots(n-(m-1))}{n^m}(\lambda+o(1))^m =$$

$$= \lim_{n\to\infty} 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{m-1}{n}\right)(\lambda+o(1))^m = \lambda^m.$$

Предел последовательности

$$\varphi_n = \left(1-\frac{1}{n}(\lambda+o(1))\right)^{n-m},\ n=1,2,\dots,$$

вычислим, воспользовавшись тем, что при $x\to 0$

$$\ln(1+x)\sim x.$$

Имеем:

$$\ln\varphi_n = (n-m)\ln\left(1-\frac{1}{n}(\lambda+o(1))\right)\sim(n-m)\left(-\frac{1}{n}(\lambda+o(1))\right).$$

Отсюда $\lim\limits_{n\to\infty}\ln\varphi_n = -\lambda$, а $\lim\limits_{n\to\infty}\varphi_n = e^{-\lambda}$. Так что в итоге получаем:

$$\lim_{n\to\infty} C_n^m p^m(1-p)^{n-m} = \frac{\lambda^m}{m!}e^{-\lambda}.$$

Замечание 1. В силу теоремы Пуассона при больших n и малых p можно считать, что $B_{n;p}(m)$ равно $\lambda^m e^{-\lambda}/m!$, где $\lambda = np$. Чтобы погрешность в результате замены $B_{n;p}(m)$ на $\lambda^m e^{-\lambda}/m!$ была незначительной, n должно быть не меньше нескольких десятков, а лучше сотен, а произведение np не должно превосходить 10. При других значениях np рекомендуется применять локальную предельную теорему.

В условиях теоремы Пуассона имеет место оценка

$$\sum_{m=0}^{\infty} \left| B_{n;p}(m) - \frac{\lambda^m}{m!} e^{-\lambda} \right| \le \frac{2\lambda \min\{2, \lambda\}}{n} \tag{6.2.1}$$

(Ю. В. Прохоров).

Замечание 2. Выражение $\dfrac{\lambda^m}{m!} e^{-\lambda}$ табулировано для различных значений m и λ (см., например, табл. 27.6.1).

Вернемся к задаче о телефонных линиях. Число μ требований связи в данный момент мы рассматриваем как число успехов в $n = 500$ испытаниях Бернулли с вероятностью успеха в одном испытании $p = 1/60$. Поэтому можно считать, что μ имеет пуассоновское распределение с параметром $\lambda = np$:

$$\mathsf{P}\{\mu = m\} = \frac{\lambda^m}{m!} e^{-\lambda},\ m = 0, 1, 2, \ldots$$

Тогда вероятность отказа

$$\mathsf{P}\{\mu > k\} = \sum_{m=k+1}^{500} \frac{\lambda^m}{m!} e^{-\lambda}.$$

А минимальное k, для которого $\mathsf{P}\{\mu > k\} \le 0{,}01$, находим из условия

$$\mathsf{P}\{\mu > k\} = \sum_{m=k+1}^{500} \frac{\lambda^m}{m!} e^{-\lambda} \le \sum_{m=k+1}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} \le 0{,}01,$$

где $\lambda = np = 500/60 = 8{,}33$. Это k оказалось равным 16.

Так что для обеспечения приемлемой работы телефонной станции A на 500 абонентов (вероятность отказа в связи не превосходит 0,01) достаточно 16 линий связи.

В силу теоремы Пуассона, решая задачу о телефонных линиях, можно было бы с самого начала исходить из пуассоновского распределения числа отказов μ.

Локальная предельная теорема. Напомним известные из анализа обозначения. Пусть функции $\alpha(n)$ и $\beta(n)$ со значениями $\mathbb{R}^1$, заданы на $\mathbb{N}_+$. Если $\alpha(n) \to 0$ и $\beta(n) \to 0$ при $n \to \infty$, причем так, что

$$\alpha(n)/\beta(n) \to 1,$$

то это записывается так:

$$\alpha(n) \sim \beta(n);$$

если

$$\alpha(n)/\beta(n) \to 0,$$

то это записывается так:

$$\alpha(n) = o(\beta(n)).$$

Будем предполагать, что имеется бесконечная последовательность серий испытаний Бернулли с вероятностью успеха p в каждой серии.

Теорема 6.2.2 (локальная предельная теорема). *В последовательности n испытаний Бернулли с вероятностью успеха p для числа успехов m, удовлетворяющих неравенству*

$$\left|\frac{m-np}{\sqrt{npq}}\right| \le C < \infty,$$

при $n \to \infty$

$$\frac{B_{n;p}(m)}{\dfrac{1}{\sqrt{2\pi npq}}\exp\left\{-\dfrac{1}{2}\left(\dfrac{m-np}{\sqrt{npq}}\right)^2\right\}} \to 1.$$

Доказательство. Обозначим $\dfrac{m-np}{\sqrt{npq}} = x_{n,m}$ (по условию $|x_{n,m}| \le C$). Тогда для m и $n-m$ получим представления

$$m = np + \sqrt{npq}x_{n,m}, \quad n - m = nq - \sqrt{npq}x_{n,m}. \tag{6.2.2}$$

Отсюда

$$\frac{m}{np} = 1 + \frac{q}{\sqrt{npq}}x_{n,m}, \quad \frac{n-m}{nq} = 1 - \frac{p}{\sqrt{npq}}x_{n,m}. \tag{6.2.3}$$

Далее, воспользовавшись формулой Стирлинга

$$n! \sim \sqrt{2\pi n} n^n e^{-n},$$

имеем

$$\sqrt{2\pi npq} B_{n;p}(m) = \sqrt{2\pi npq} \frac{n!}{m!(n-m)!} p^m q^{n-m} \sim$$

$$\sim \sqrt{2\pi npq} \frac{\sqrt{2\pi n} n^n e^{-n} p^m q^{n-m}}{\sqrt{2\pi m} m^m e^{-m} \sqrt{2\pi(n-m)}(n-m)^{n-m} e^{-(n-m)}} =$$

$$= \sqrt{\frac{np}{m} \cdot \frac{nq}{n-m}} \left(\frac{m}{np}\right)^{-m} \left(\frac{n-m}{nq}\right)^{-(n-m)}.$$

Из (6.2.3) следует, что радикал при $n \to \infty$ стремится к 1. Так что

$$\sqrt{2\pi npq} B_{n;p}(m) \sim \left(\frac{m}{np}\right)^{-m} \left(\frac{n-m}{nq}\right)^{-(n-m)},$$

или, воспользовавшись равенством (6.2.2),

$$\begin{aligned} \sqrt{2\pi npq} B_{n;p}(m) \sim \\ \sim \left(\frac{np + \sqrt{npq} x_{n,m}}{np}\right)^{-(np + \sqrt{npq} x_{n,m})} \times \\ \times \left(\frac{nq - \sqrt{npq} x_{n,m}}{nq}\right)^{-(nq - \sqrt{npq} x_{n,m})} = \\ = \left(1 + \frac{q}{\sqrt{npq}} x_{n,m}\right)^{-(np + \sqrt{npq} x_{n,m})} \times \\ \times \left(1 - \frac{p}{\sqrt{npq}} x_{n,m}\right)^{-(nq - \sqrt{npq} x_{n,m})} \end{aligned} \tag{6.2.4}$$

(заметим, что хотя $\sqrt{2\pi npq} B_{n;p}(m)$ зависит как от n так и от m, стремящихся к $+\infty$, но предельное поведение левой части, а вместе с тем и правой, определяется только n поскольку $|x_{n,m}| < C$). Для доказательства теоремы установим, что при $n \to \infty$

$$\varphi_n = \ln \frac{\sqrt{2\pi npq} B_{n;p}(m)}{\exp\left\{-\frac{1}{2} x_{n,m}^2\right\}} \to 0,$$

или, учитывая соотношение (6.2.4) для $\sqrt{2\pi npq}B_{n;p}(m)$, что

$$\varphi_n \sim -(np+\sqrt{npq}x_{n,m})\ln\left(1+\frac{q}{\sqrt{npq}}x_{n,m}\right)-$$

$$-(nq-\sqrt{npq}x_{n,m})\ln\left(1-\frac{p}{\sqrt{npq}}x_{n,m}\right)+\frac{1}{2}x_{n,m}^2 \to 0.$$

Воспользуемся тем, что

$$\ln(1+x) = x-\frac{1}{2}x^2+o(x^2) \text{ при } x\to 0.$$

Тогда, учитывая, что $npq\to\infty$ при $n\to\infty$, имеем:

$$\varphi_n \sim -(np+\sqrt{npq}x_{n,m})\ln\left(1+\frac{q}{\sqrt{npq}}x_{n,m}\right)-$$

$$-(nq-\sqrt{npq}x_{n,m})\ln\left(1-\frac{p}{\sqrt{npq}}x_{n,m}\right)+\frac{1}{2}x_{n,m}^2 =$$

$$= -\left(\frac{npq}{q}+\sqrt{npq}x_{n,m}\right)\times$$

$$\times\left(\frac{q}{\sqrt{npq}}x_{n,m}-\frac{1}{2}\left(\frac{q}{\sqrt{npq}}\right)^2x_{n,m}^2+o\left(\left(\frac{q}{\sqrt{npq}}\right)^2\right)\right)-$$

$$-\left(\frac{npq}{p}-\sqrt{npq}x_{n,m}\right)\times$$

$$\times\left(-\frac{p}{\sqrt{npq}}x_{n,m}-\frac{1}{2}\left(\frac{p}{\sqrt{npq}}\right)^2x_{n,m}^2+o\left(\left(\frac{p}{\sqrt{npq}}\right)^2\right)\right)+\frac{1}{2}x_{n,m}^2 =$$

$$= -\left(\sqrt{npq}x_{n,m}+qx_{n,m}^2-\frac{1}{2}qx_{n,m}^2+o(1)\right)-$$

$$-\left(-\sqrt{npq}x_{n,m}+px_{n,m}^2-\frac{1}{2}px_{n,m}^2+o(1)\right)+\frac{1}{2}x_{n,m}^2 =$$

$$= -\frac{1}{2}qx_{n,m}^2-\frac{1}{2}px_{n,m}^2+o(1)+\frac{1}{2}x_{n,m}^2 = o(1).$$

Так что при $n \to \infty$

$$\ln \frac{B_{n;p}(m)}{\frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}} \sim o\,(1)\,.$$

Отсюда

$$\frac{B_{n;p}(m)}{\frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}} \to 1 \qquad (6.2.5)$$

при $n \to \infty$.

Тем самым теорема доказана.

Следствие. *При $n \to \infty$ и m таких, что $|x_{n,m}| < C$*

$$B_{n;p}(m) = (1 + o(1))\frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}. \qquad (6.2.6)$$

Теорема 6.2.3 (интегральная предельная теорема Муавра–Лапласа). *Пусть μ — число успехов в n испытаниях Бернулли с вероятностью успеха в одном испытании p. Для любых $-\infty \le a < b \le +\infty$*

$$\lim_{n} \mathsf{P}\left\{a \le \frac{\mu - np}{\sqrt{npq}} < b\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2}dx.$$

Д о к а з а т е л ь с т в о. Сначала докажем теорему для конечных a, b.

Случайная величина μ имеет своим распределением $B_{n;p}$:

$$\mathsf{P}\{\mu = m\} = B_{n;p}(m),\ m = 0, 1, 2, , n,$$

поэтому согласно теореме 5.1.2 для функции $\dfrac{\mu - np}{\sqrt{npq}}$ от случайной величины μ

$$\mathsf{P}\left\{a \le \frac{\mu - np}{\sqrt{npq}} < b\right\} = \sum_{m: a \le x_{n,m} < b} B_{n;p}(m),$$

где, как и ранее, $x_{n,m} = (m - np)/\sqrt{npq}$.

В силу (6.2.6) для каждого фиксированного C при $|x_{n,m}| < C$ (у нас $a < x_{n,m} < b$) и $n \to \infty$

$$B_{n;p}(m) = (1 + o(1))\frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}.$$

Поэтому

$$\mathsf{P}\left\{a \le \frac{\mu - np}{\sqrt{npq}} < b\right\} =$$

$$= (1 + o(1))\frac{1}{\sqrt{2\pi}} \sum_{m: a \le x_{n,m} < b} \frac{1}{\sqrt{npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}.$$

Далее, для точек $x_{n,m}$

$$x_{n,m+1} - x_{n,m} = 1/\sqrt{npq},$$

поэтому сумма

$$\sum_{m: a \le x_{n,m} < b} \frac{1}{\sqrt{npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}$$

отличается от интегральной суммы

$$\sum\nolimits_{[a,b]}(x_{n,m+1} - x_{n,m}) \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}$$

функции $e^{-x^2/2}$ на отрезке $[a, b]$ не более чем на величину двух слагаемых интегральной суммы — “крайнего” левого и “крайнего” правого. Каждое из этих слагаемых не превосходит величины $1/\sqrt{npq}$ (поскольку $e^{-x^2/2} < 1$) и, следовательно, стремится к нулю при $n \to \infty$. Поэтому

$$\lim_n \sum_{m: a \le x_{n,m} < b} \frac{1}{\sqrt{npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\} =$$

$$= \lim_n \sum\nolimits_{[a,b]}(x_{n,m+1} - x_{n,m}) \exp\left\{-\frac{1}{2}x_{n,m}^2\right\} = \int\limits_a^b e^{-x^2/2} dx.$$

Так что
$$\lim_n \mathsf{P}\left\{a \le \frac{\mu - np}{\sqrt{npq}} < b\right\} =$$
$$= \lim_n (1+o(1))\frac{1}{\sqrt{2\pi}} \sum_{m: a \le x_{n,m} < b} \frac{1}{\sqrt{npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\} =$$
$$= \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2} dx.$$

Для конечных a, b теорема доказана.

Для бесконечных a и b теорема получается предельным переходом.

6.3 Неограниченные последовательности испытаний Бернулли

Испытания до первого успеха. Последовательность независимых испытаний с двумя исходами (успех, неудача) и вероятностью успеха p, не меняющейся от испытания к испытанию, определяет стохастический эксперимент, состоящий в проведении испытаний до первого успеха. Построим вероятностное пространство этого стохастического эксперимента.

В качестве пространства элементарных событий рассмотрим множество
$$\Omega = \{1, 01, 001, \ldots, \underbrace{00\ldots0}_{k}1, \ldots\}.$$

Точка $\omega = (\underbrace{00\ldots0}_{k}1)$ пространства Ω описывает исход эксперимента, состоящий в том, что первому успеху предшествует k неудач. Так как Ω дискретно, то вероятность на подмножествах пространства Ω задается посредством неотрицательной функции $\mathsf{P}(\omega)$ точки, $\omega \in \Omega$, для которой $\sum\limits_{\omega\in\Omega} \mathsf{P}(\omega) = 1$ (см. теорему 4.1.1). В качестве $\mathsf{P}(\omega)$ естественно рассмотреть функцию
$$\mathsf{P}(\omega) = (1-p)^{\nu(\omega)}p, \ \omega \in \Omega, \tag{6.3.1}$$

где $\nu(\omega)$ — число неудач до первого успеха (число нулей в последовательности $\omega = (00\ldots01)$). Ясно, что $\mathsf{P}(\omega) \geq 0$ и

$$\sum_{\omega\in\Omega}\mathsf{P}(\omega) = \sum_{\omega\in\Omega}(1-p)^{\nu(\omega)}p = p\sum_{k=0}^{\infty}(1-p)^k = p\frac{1}{1-(1-p)} = 1.$$

Поэтому функция $\mathsf{P}(\omega)$, определенная равенством (6.3.1), задает на подмножествах пространства Ω вероятность P. Пару $\{\Omega, \mathsf{P}\}$ и будем рассматривать в качестве математической модели стохастического эксперимента, состоящего в проведении независимых испытаний до первого успеха.

В наших интуитивных представлениях о последовательности независимых испытаний до первого успеха информация о том, что первые n испытаний закончились неудачей не меняет прогноз числа испытаний до первого успеха. Пусть, например, симметричную монету подбрасывают до первого выпадения герба (до первого успеха). Известно, что в первых десяти подбрасываниях монета легла решеткой (первые десять испытаний закончились неудачей). Эта информация не меняет прогноз числа испытаний до первого выпадения герба — наш прогноз такой же, как и до начала подбрасывания монеты. В противном случае нам пришлось бы согласиться с тем, что монета имеет память — помнит, что в первых десяти подбрасываниях герб не появился, а поэтому пора бы ему уже и выпасть.

Если предложенная выше модель $\{\Omega, \mathsf{P}\}$ адекватно описывает стохастический эксперимент, то она должна обладать свойством, соответствующим свойству отсутствия памяти у монеты. Для формального его описания введем необходимые определения.

Зададим на Ω вероятностного пространства $\{\Omega, \mathsf{P}\}$ случайную величину ν, равную числу неудач, предшествующих первому успеху, ее еще называют *временем ожидания* первого успеха. На исходе $\omega = (\underbrace{00\ldots0}_{k}1)$ значение $\nu(\omega)$ равно k. Распределение времени ожидания ν:

$$\mathsf{P}_\nu(k) = \mathsf{P}\{\omega : \nu(\omega) = k\} = \sum_{\omega:\nu(\omega)=k}\mathsf{P}(\omega) =$$

$$= \mathsf{P}((\underbrace{00\ldots0}_{k}1)) = (1-p)^k p,\ \ k = 0, 1, \ldots$$

Определение. *Геометрическим распределением* с параметром p $(0 < p < 1)$ называется распределение P_p, заданное на $\{0, 1, \ldots\}$ равенством

$$\mathsf{P}_p(k) = (1-p)^k p, \ k = 0, 1, \ldots$$

Ясно, что

$$\sum_{k=0}^{\infty} (1-p)^k p = 1.$$

Так что время ожидания первого успеха в последовательности независимых испытаний имеет геометрическое распределение.

Определение. Будем говорить, что неотрицательная целочисленная случайная величина ξ обладает свойством *отсутствия последействия,* если для каждого n, $n = 1, 2, \ldots,$

$$\mathsf{P}\{\xi = n + m / \xi \geq n\} = \mathsf{P}\{\xi = m\}, \ m = 0, 1, \ldots \tag{6.3.2}$$

Теорема 6.3.1 (об отсутствии последействия геометрического распределения). *Для того чтобы неотрицательная целочисленная случайная величина обладала свойством отсутствия последействия, необходимо и достаточно, чтобы она имела геометрическое распределение.*

Доказательство. Пусть случайная величина ξ имеет геометрическое распределение:

$$\mathsf{P}\{\xi = k\} = (1-p)^k p, \ k = 0, 1, \ldots, \ (p > 0).$$

Тогда

$$\mathsf{P}\{\xi = n + m \,/\, \xi \geq n\} = \frac{\mathsf{P}\{\xi = n + m, \xi \geq n\}}{\mathsf{P}\{\xi \geq n\}} =$$

$$= \frac{\mathsf{P}\{\xi = n + m\}}{\mathsf{P}\{\xi \geq n\}} = \frac{(1-p)^{n+m} p}{\sum\limits_{k=n}^{\infty} (1-p)^k p} = (1-p)^m p = \mathsf{P}\{\xi = m\}.$$

Так что

$$\mathsf{P}\{\xi = n + m \,/\, \xi \geq n\} = \mathsf{P}\{\xi = m\}.$$

Обратно. Пусть неотрицательная целочисленная случайная величина обладает свойством отсутствия последействия: для каждого n, $n = 1, 2, \ldots,$

$$\mathsf{P}\{\xi = n + m \,/\, \xi \geq n\} = \mathsf{P}\{\xi = m\}, \ m = 0, 1, \ldots.$$

Последнее равенство можно переписать так:

$$\frac{\mathsf{P}\{\xi = n + m\}}{\mathsf{P}\{\xi \geq n\}} = \mathsf{P}\{\xi = m\}$$

или

$$\mathsf{P}\{\xi = n + m\} = \mathsf{P}\{\xi = m\}\mathsf{P}\{\xi \geq n\}.$$

Положив в последнем равенстве $n = 1$, получим

$$\mathsf{P}\{\xi = m + 1\} = \mathsf{P}\{\xi = m\}\mathsf{P}\{\xi \geq 1\} = \mathsf{P}\{\xi = m\}(1 - \mathsf{P}\{\xi < 1\}) =$$

$$= \mathsf{P}\{\xi = m\}(1 - \mathsf{P}\{\xi = 0\})$$

или, обозначив $\mathsf{P}\{\xi = 0\} = p$,

$$\mathsf{P}\{\xi = m + 1\} = \mathsf{P}\{\xi = m\}(1 - p), \; m = 0, 1, \ldots$$

Полагая в последнем соотношении последовательно $m = 0, 1, \ldots,$ получаем:

$$\mathsf{P}\{\xi = 1\} = \mathsf{P}\{\xi = 0\}(1 - p) = p(1 - p),$$

$$\mathsf{P}\{\xi = 2\} = \mathsf{P}\{\xi = 1\}(1 - p) = p(1 - p)^2,$$

$$\vdots$$

$$\mathsf{P}\{\xi = m\} = \mathsf{P}\{\xi = m - 1\}(1 - p) = p(1 - p)^m.$$

Так что для каждого $m = 0, 1, \ldots$

$$\mathsf{P}\{\xi = m\} = p(1 - p)^m$$

— случайная величина ξ имеет геометрическое распределение.

Время ожидания ν первого успеха в последовательности независимых испытаний имеет геометрическое распределение и, следовательно, обладает свойством отсутствия последействия:

$$\mathsf{P}\{\nu = n + m \,|\, \nu \geq n\} = \mathsf{P}\{\nu = m\}, \; m = 0, 1, 2, \ldots,$$

последнее как раз и означает, что информация о первых n испытаниях, закончившихся неудачей, не меняет прогноз о числе испытаний до первого успеха.

Испытания до r-го успеха. Последовательность независимых испытаний с двумя исходами (успех, неудача) и вероятностью успеха p, не меняющейся от испытания к испытанию, определяет стохастический эксперимент, состоящий в проведении испытаний до r-го успеха. Построим вероятностное пространство этого стохастического эксперимента.

В качестве пространства элементарных событий Ω рассмотрим множество последовательностей, составленных из нулей и r единиц и оканчивающихся единицей. Например, точка $\omega = (100\ldots 11)$ описывает исход, состоящий в том, что в первом испытании наблюдался успех, в двух последующих испытаниях наблюдались неудачи и т. д., в двух последних испытаниях наблюдались успехи. Пространство Ω дискретно, поэтому вероятность на классе всех подмножеств пространства Ω задается неотрицательной функцией точки $\mathsf{P}(\omega)$, $\omega \in \Omega$, такой что $\sum\limits_{\omega\in\Omega} \mathsf{P}(\omega) = 1$ (см. теорему 4.1.1). В качестве $\mathsf{P}(\omega)$ естественно рассмотреть

$$\mathsf{P}(\omega) = (1-p)^{\nu(\omega)} p^r,\ \omega \in \Omega,$$

где $\nu(\omega)$ — число нулей в последовательности $\omega = (100\ldots 11)$ (число неудач до r-го успеха). Для функции

$$\mathsf{P}(\omega) = (1-p)^{\nu(\omega)} p^r,\ \omega \in \Omega,$$

имеет место равенство $\sum\limits_{\omega\in\Omega} \mathsf{P}(\omega) = 1$. Убедимся в этом. Ясно, что

$$\sum_{\omega\in\Omega} \mathsf{P}(\omega) = \sum_{\omega\in\Omega} (1-p)^{\nu(\omega)} p^r = \sum_{\omega:\nu(\omega)=0} (1-p)^{\nu(\omega)} p^r +$$

$$+ \sum_{\omega:\nu(\omega)=1} (1-p)^{\nu(\omega)} p^r + \ldots + \sum_{\omega:\nu(\omega)=k} (1-p)^{\nu(\omega)} p^r + \ldots =$$

$$= \sum_{k=0}^{\infty} \sum_{\omega:\nu(\omega)=k} (1-p)^{\nu(\omega)} p^r.$$

Для каждого $k = 0, 1, \ldots$

$$\sum_{\omega:\nu(\omega)=k} (1-p)^{\nu(\omega)} p^r = \sum_{\omega:\nu(\omega)=k} (1-p)^k p^r = C_{k+r-1}^k (1-p)^k p^r,$$

поскольку число одинаковых слагаемых в сумме

$$\sum_{\omega:\nu(\omega)=k} (1-p)^k p^r$$

равно числу $C^k_{k+(r-1)}$ последовательностей длиной $k+(r-1)$, составленных из k нулей и $(r-1)$ единиц. Поэтому

$$\sum_{\omega\in\Omega} \mathsf{P}(\omega) = \sum_{\omega\in\Omega} (1-p)^{\nu(\omega)} p^r = \sum_{k=0}^{\infty} C^k_{k+r-1}(1-p)^k p^r.$$

Покажем, что

$$\sum_{k=0}^{\infty} C^k_{k+r-1}(1-p)^k p^r = 1.$$

Для этого разложим функцию $f(x) = (1-x)^{-r}$ в ряд Тейлора при $|x| < 1$. Имеем:

$$f^{(k)}(x) = r(r+1)\ldots(r+(k-1))(1-x)^{-(r+k)},\ k = 1, 2, \ldots,$$

$$\frac{f^{(k)}(0)}{k!} = \frac{r(r+1)\ldots(r+k-1)}{k!} = \frac{(k+r-1)!}{k!(r-1)!} = C^k_{k+r-1},$$

$$\frac{1}{(1-x)^r} = f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^{\infty} C^k_{k+r-1} x^k.$$

Положив в последнем равенстве $x = 1-p$, получим:

$$\frac{1}{p^r} = \sum_{k=0}^{\infty} C^k_{k+r-1}(1-p)^k$$

или

$$1 = \sum_{k=0}^{\infty} C^k_{k+r-1}(1-p)^k p^r. \tag{6.3.3}$$

Пару $\{\Omega, \mathsf{P}\}$, где $\mathsf{P}(\omega) = (1-p)^{\nu(\omega)} p^r$, $\nu(\omega)$ — число неудач до r-го успеха, и будем рассматривать в качестве математической модели стохастического эксперимента, состоящего в проведении независимых испытаний до r-го успеха.

Определение. *Отрицательным биномиальным распределением* с параметрами $(r; p)$ будем называть вероятностное распределение $\mathsf{P}_{r;p}$, заданное на $\{0, 1, \ldots\}$ равенством

$$\mathsf{P}_{r;p}(k) = C^k_{k+r-1}(1-p)^k p^r,\ k = 0, 1, \ldots$$

Будем говорить, что случайная величина ξ имеет отрицательное биномиальное распределение с параметрами $(r;p)$, если

$$\mathsf{P}_\xi(k) = C_{k+r-1}^k (1-p)^k p^r,\ k = 0, 1, \ldots$$

Случайную величину $\nu = \nu(\omega)$ на $\{\Omega, \mathsf{P}\}$, равную числу неудач до r-го успеха, называют *временем ожидания r-го успеха.*

Время ν ожидания r-го успеха имеет отрицательное биномиальное распределение:

$$\mathsf{P}_\nu(k) = \mathsf{P}\{\nu = k\} = \mathsf{P}\{\omega : \nu(\omega) = k\} = \sum_{\omega:\nu(\omega)=k} \mathsf{P}(\omega) =$$

$$= \sum_{\omega:\nu(\omega)=k} (1-p)^{\nu(\omega)} p^r = \sum_{\omega:\nu(\omega)=k} (1-p)^k p^r =$$

$$= C_{k+r-1}^k (1-p)^k p^r,\ k = 0, 1, \ldots$$

Теорема 6.3.2. *Сумма r независимых геометрически распределенных с параметром p случайных величин имеет отрицательное биномиальное распределение с параметрами $(r;p)$.*

Доказательство сводится к вычислению распределения функции

$$\eta = \xi_1 + \xi_2 + \ldots + \xi_r$$

от независимых геометрически распределенных случайных величин $\xi_1, \xi_2, \ldots, \xi_r$ по их совместному распределению. Совместное распределение $\xi_1, \xi_2, \ldots, \xi_r$

$$\mathsf{P}\{\xi_1 = k_1, \xi_2 = k_2, \ldots, \xi_r = k_r\} =$$

$$= \prod_{i=1}^r \mathsf{P}\{\xi_i = k_i\} = \prod_{i=1}^r p(1-p)^{k_i} = p^r (1-p)^{k_1+k_2+\cdots+k_r},$$

$$k_1 = 0, 1, \ldots;\ k_2 = 0, 1, \ldots;\ \ldots; k_r = 0, 1, \ldots$$

Согласно теореме 5.1.2 о вычислении распределения функции от случайной величины (здесь $g(x) = g(x_1, x_2, \ldots, x_r) = x_1 + x_2 + \ldots + x_r$, $B = \{k\}$) для каждого k, $k = 0, 1, \ldots$, имеем:

$$\mathsf{P}\{\eta = k\} = \mathsf{P}\{\xi_1 + \xi_2 + \ldots + \xi_r = k\} =$$

$$= \sum_{(k_1,k_2,\ldots,k_r):k_1+k_2+\cdots+k_r=k} p^r (1-p)^{k_1+k_2+\cdots+k_r} =$$

$$= \sum_{(k_1,k_2,\ldots,k_r):k_1+k_2+\cdots+k_r=k} (1-p)^k p^r = C_{k+r-1}^k (1-p)^k p^r.$$

Суммирование ведется по всем неотрицательным целочисленным последовательностям $(k_1, k_2, \ldots, k_r)$, для которых $k_1 + k_2 + \cdots + k_r = k$, т. е. по всем неотрицательным целочисленным решениям уравнения $k_1 + k_2 + \cdots + k_r = k$, а число таких решений равно C_{k+r-1}^k (см. теорему 4.3.3).

Следствие. *Математическое ожидание и дисперсия случайной величины ζ, имеющей отрицательное биномиальное распределение с параметрами (r, p), соответственно равны*

$$\mathsf{M}\zeta = r\frac{1-p}{p},\ \mathsf{D}\zeta = r\frac{1-p}{p^2}.$$

Доказательство. Согласно теореме распределение случайной величины ζ совпадает с распределением суммы $\xi_1 + \xi_2 + \ldots + \xi_r$ независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_r$, имеющих геометрическое распределение с параметром p. Поэтому совпадают их математические ожидания и дисперсии:

$$\mathsf{M}\zeta = \mathsf{M}(\xi_1 + \xi_2 + \ldots + \xi_r) = r\mathsf{M}\xi_1 = r\frac{1-p}{p},$$

$$\mathsf{D}\zeta = \mathsf{D}(\xi_1 + \xi_2 + \ldots + \xi_r) = r\mathsf{D}\xi_1 = r\frac{1-p}{p^2}.$$

Математическое ожидание и дисперсия геометрически распределенной с параметром p случайной величины ξ

$$\mathsf{M}\xi = \frac{1-p}{p},\ \mathsf{D}\xi = \frac{1-p}{p^2},$$

см. пример 5.5.2.

6.4 Дискретные распределения на $\mathbb{R}^1$

Ниже рассматриваются наиболее важные и часто встречающиеся в теории вероятностей дискретные распределения на $\mathbb{R}^1$.

Напомним, что дискретным вероятностным распределением на $\mathbb{R}^n$ мы называем неотрицательную функцию точки:

$$\mathsf{P} : x_i \to \mathsf{P}(x_i),\ x_i \in X \subset \mathbb{R}^n.$$

со значениями в $\mathbb{R}^1$ на дискретном $X \subset \mathbb{R}^n$ такую, что

$$\sum_{x_i \in X} \mathsf{P}(x_i) = 1$$

(см. также (4.1.4)).

Функция множества

$$\mathsf{P}(A) = \sum_{x_i \in A} \mathsf{P}(x_i)$$

является неотрицательной счетно-аддитивной нормированной на классе всех подмножеств дискретного $X \subset \mathbb{R}^n$.

Биномиальное распределение. Биномиальным распределением с параметрами $(n; p)$ называется вероятностное распределение $B_{n;p}$, заданное на $\{0, 1, \dots, n\}$ равенством

$$B_{n;p}(k) = C_n^k p^k (1-p)^{n-k}, \ k = 0, 1, \dots, n.$$

Определение корректно, поскольку

$$\sum_{k=0}^{n} B_{n;p}(k) = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Мультиномиальное распределение. Мультиномиальным распределением с параметрами $(n; p_1, p_2, \dots, p_r)$, $p_1 + p_2 + \dots + p_r = 1$, $0 < p_i < 1$, $i = 1, 2, \dots, r$, называется вероятностное распределение P, заданное на множестве последовательностей $(k_1, k_2, \dots, k_r)$ целых неотрицательных чисел $k_1, k_2, \dots, k_r$ таких, что $k_1 + k_2 + \dots + k_r = n$, равенством

$$\mathsf{P}(k_1, k_2, \dots, k_r) = C_n(k_1, k_2, \dots, k_r) p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$$

(см. (6.1.7)). Определение корректно, поскольку

$$\sum_{k_1 + k_2 + \dots + k_r = n} C_n(k_1, k_2, \dots, k_r) p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} =$$

$$= (p_1 + p_2 + \dots + p_r)^n = 1,$$

суммирование ведется по всем последовательностям $k_1, k_2, \dots, k_r$, для которых $k_1 + k_2 + \dots + k_r = n$ (см. полиномиальную формулу в разд. 4.3 главы 4).

Распределение Пуассона. *Распределением Пуассона* (*пуассоновским распределением*) с параметром λ $(\lambda > 0)$ называется вероятностное распределение P_λ, заданное на $\{0, 1, \ldots\}$ равенством

$$\mathsf{P}_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, \ldots$$

Определение корректно поскольку

$$\sum_{k=0}^{\infty} \mathsf{P}_\lambda(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Распределение Пуассона играет важную роль в теории массового обслуживания, теории надежности, теории случайных процессов. Например, пуассоновское распределение имеет число частиц, достигающих заданной части пространства при радиоактивном распаде вещества; число несчастных случаев, происходящих за единицу времени в данной совокупности индивидуумов.

Замечание. Случайные величины часто называют по их распределениям. Например, случайную величину ξ называют биномиально распределенной с параметрами $(n; p)$, если ее распределение биномиальное:

$$\mathsf{P}_\xi(k) = B_{n;p}(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n,$$

пуассоновской — если ее распределение пуассоновское:

$$\mathsf{P}_\xi(k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, \ldots$$

Геометрическое распределение. *Геометрическим распределением* с параметром p $(0 < p < 1)$ называется распределение P_p, заданное на $\{0, 1, \ldots\}$ равенством

$$\mathsf{P}_p(k) = (1-p)^k p,\ k = 0, 1, \ldots$$

Ясно, что

$$\sum_{k=0}^{\infty} (1-p)^k p = 1.$$

Отрицательное биномиальное распределение. *Отрицательным биномиальным распределением* с параметрами $(r; p)$

будем называть вероятностное распределение $\mathsf{P}_{r;p}$, заданное на $\{0, 1, \ldots\}$ равенством

$$\mathsf{P}_{r;p}(k) = C_{k+r-1}^{k}(1-p)^k p^r,\ k = 0, 1, \ldots$$

Определение корректно, см. равенство (6.3.3).

Гипергеометрическое распределение. В урне имеется N различимых шаров (например, занумерованных числами от 1 до N), среди них M белых, остальные $N-M$ черные. Рассмотрим стохастический эксперимент, состоящий в выборе наудачу из урны n шаров, число n меньше числа шаров каждого цвета: $n \le M$, $n \le N-M$. Исход ω эксперимента — n-элементное подмножество N-элементного множества — сочетание из N элементов по n, Ω — совокупность всех таких ω. Поскольку выбор производится наудачу, то каждому исходу ω следует приписать вероятность $\mathsf{P}(\omega) = 1/C_N^n$. Тем самым вероятностное пространство стохастического эксперимента построено. Вычислим вероятность события “среди n выбранных шаров m белых”. Поскольку модель классическая, то искомая вероятность

$$\mathsf{P}_{N;M;n}(m) = \mathsf{P}\{\xi = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n},\ m = 0, 1, \ldots, n.$$

Определение. *Гипергеометрическим распределением* с параметрами $(N; M; n)$, $n \le M$, $n \le N-M$, будем называть вероятностное распределение $\mathsf{P}_{N;M;n}$, заданное на $\{0, 1, \ldots, n\}$ равенством:

$$\mathsf{P}_{N;M;n}(m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n},\ m = 0, 1, \ldots, n.$$

Определение корректно. Убедимся в этом. Пусть в урне содержится N шаров, среди них M белых и $N-M$ черных. Число n-элементных подмножеств N-элементного множества ($n \le M$, $n \le N-M$), среди элементов которого ровно m белых шаров, очевидно, равно $C_M^m C_{N-M}^{n-m}$. Тогда с одной стороны число всех n-элементных подмножеств N-элементного множества равно $\sum\limits_{m=0}^{n} C_M^m C_{N-M}^{n-m}$, а с другой C_N^n, поэтому

$$\sum_{m=0}^{n} C_M^m C_{N-M}^{n-m} = C_N^n.$$

Отсюда

$$\sum_{m=0}^{n} \mathsf{P}_{N;M;n}(m) = 1.$$

Теорема 6.4.1 (о предельном поведении $\mathsf{P}_{N;M;n}(m)$). *Пусть $N \to \infty$ и $M \to \infty$, причем так, что $M/N \to p$ $(0 < p < 1)$, тогда*

$$\mathsf{P}_{N;M;n}(m) \to B_{n;p}(m) = C_n^m p^m (1-p)^{n-m},\ m = 0, 1, 2, \ldots, n.$$

Доказательство.

$$\mathsf{P}_{N;M;n}(m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} =$$

$$= \frac{n!(N-n)!}{N!} \cdot \frac{M!}{m!(M-m)!} \cdot \frac{(N-M)!}{(n-m)!((N-M)-(n-m))!} =$$

$$= \frac{n!}{m!(n-m)!} \cdot \frac{(N-n)!}{N!} \cdot \frac{M!}{(M-m)!} \cdot \frac{(N-M)!}{((N-M)-(n-m))!} =$$

$$= \frac{n!}{m!(n-m)!} \cdot \frac{1}{N(N-1)\ldots(N-(n-1))} \times$$

$$\times \frac{M(M-1)\ldots(M-(m-1))}{1} \times$$

$$\times \frac{(N-M)(N-M-1)\ldots(N-M-(n-m-1))}{1}.$$

Разделив числитель и знаменатель правой части на N^n, получаем:

$$\mathsf{P}_{N;M;n}(m) = \frac{n!}{m!(n-m)!} \cdot \frac{1}{1\left(1-\frac{1}{N}\right)\ldots\left(1-\frac{n-1}{N}\right)} \times$$

$$\times \frac{M}{N}\left(\frac{M}{N} - \frac{1}{N}\right)\ldots\left(\frac{M}{N} - \frac{m-1}{N}\right) \times$$

$$\times \left(\frac{N-M}{N}\right)\left(\frac{N-M}{N} - \frac{1}{N}\right)\ldots\left(\frac{N-M}{N} - \frac{n-m-1}{N}\right).$$

Устремив N и M к бесконечности так, чтобы $M/N \to p$, получим

$$\mathsf{P}_{N;M;n}(m) \to \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} = B_{n;p}(m).$$

Утверждение теоремы согласуется с нашей интуицией: если N и M большие, то результат однократного выбора n шаров из N можно считать совпадающим с результатом их последовательного выбора. При больших N и M последовательный выбор шаров можно считать независимыми событиями, а вероятность вынуть белый шар — постоянной и равной $p = M/N$; это предположение нельзя принять, если N и M малые (рассмотрите, например, случай $N = 4$, $M = 2$, $n = 2$).

Многомерное гипергеометрическое распределение. В урне имеется N различимых шаров (например, занумерованных числами от 1 до N), среди них M_1 шаров цвета b_1, M_2 шаров цвета b_2, ..., M_r шаров цвета b_r $(M_1 + M_2 + \ldots + M_r = N)$. Из урны наудачу выбирают n шаров, число n меньше числа шаров каждого цвета: $n \leq M_1, n \leq M_2, \ldots, n \leq M_r$. Исход ω эксперимента — n-элементное подмножество N-элементного множества (сочетание из N элементов по n), Ω — совокупность всех таких ω. Поскольку выбор производится наудачу, то каждому исходу ω следует приписать вероятность $\mathsf{P}(\omega) = 1/C_N^n$. Тем самым вероятностное пространство стохастического эксперимента построено.

Вычислим вероятность $\mathsf{P}(m_1, m_2, \ldots, m_r)$ того, что среди n выбранных шаров m_1 шаров цвета b_1, m_2 шаров цвета b_2, и т. д., m_r шаров цвета b_r $(m_1 \leq M_1, m_2 \leq M_2, \ldots, m_r \leq M_r)$. Поскольку $\{\Omega, \mathsf{P}\}$ — классическая модель, то искомая вероятность

$$\mathsf{P}(m_1, m_2, \ldots, m_r) = \frac{C_{M_1}^{m_1} C_{M_2}^{m_2} \ldots C_{M_r}^{m_r}}{C_N^n}.$$

Определение. *Многомерным гипергеометрическим распределением* с параметрами $N, M_1, M_2, \ldots, M_r, n;$ $M_1 + M_2 + \ldots$ $\ldots + M_r = N; n \leq M_1, n \leq M_2, \ldots, n \leq M_r$, будем называть вероятностное распределение P, заданное на множестве последовательностей $(m_1, m_2, \ldots, m_r)$ целых неотрицательных чисел таких, что $m_1 + m_2 + \ldots + m_r = n$, равенством

$$\mathsf{P}(m_1, m_2, \ldots, m_r) = \frac{C_{M_1}^{m_1} C_{M_2}^{m_2} \ldots C_{M_r}^{m_r}}{C_N^n}.$$

В том, что

$$\sum_{m_1+m_2+\ldots+m_r=n} \mathsf{P}(m_1, m_2, \ldots, m_r) = 1,$$

убеждаемся аналогично тому, как это делалось для гипергеометрического распределения с параметрами (N, M, n).

Как и для гипергеометрического распределения, для многомерного гипергеометрического распределения

$$\mathsf{P}(m_1, m_2, \ldots, m_r) = \frac{C_{M_1}^{m_1} C_{M_2}^{m_2} \ldots C_{M_r}^{m_r}}{C_N^n},$$

$(m_1 + m_2 + \ldots + m_r = n,\ m_i$ — целые неотрицательные) имеет место утверждение о его предельном поведении.

Теорема 6.4.2 *Пусть $N, M_1, M_2, \ldots, M_r$, стремятся к $+\infty$, причем так, что*

$$M_i/N \to p_i,\ \ i = 1, 2, \ldots, r,$$

тогда многомерное гипергеометрическое распределение

$$\mathsf{P}(m_1, m_2, \ldots, m_r) = \frac{C_{M_1}^{m_1} C_{M_2}^{m_2} \ldots C_{M_r}^{m_r}}{C_N^n}$$

сходится к мультиномиальному распределению с параметрами $(n, p_1, p_2, \ldots, p_r)$:

$$\mathsf{P}(m_1, m_2, \ldots, m_r) \to C_n(m_1, m_2, \ldots, m_r) p_1^{m_1} p_2^{m_2} \ldots p_r^{m_r}$$

$(m_1 + m_2 + \ldots + m_r = n,\ m_i$ — целые неотрицательные).

Доказательство аналогично доказательству теоремы 6.4.1.

Теорема 6.4.3 (о сохранении вида распределения).

1° *Сумма независимых пуассоновских случайных величин является пуассоновской случайной величиной с параметром, равным сумме параметров слагаемых.*

2° *Сумма независимых случайных величин, имеющих отрицательное биномиальное распределение с параметрами (r_1, θ) и (r_2, θ), имеет отрицательное биномиальное распределение с параметрами $(r_1 + r_2, \theta)$.*

3° *Сумма независимых биномиально распределенных случайных величин соответственно с параметрами $(n; \theta)$ и $(m; \theta)$ является биномиально распределенной случайной величиной с параметрами $(n + m; \theta)$.*

Доказательство.

1° Доказательства пункта 1° приведено в примере 5.3.5.

2° Пусть случайная величина ξ имеет отрицательное биномиальное распределение с параметрами (r, θ):

$$\mathsf{P}\{\xi = k\} = C_{k+r-1}^{k}(1-\theta)^{k}\theta^{r},\ k = 0, 1, \ldots;\ \theta \in (0; 1).$$

Для её производящей функции

$$P_{\xi}(t) = \mathsf{M}t^{\xi} = \sum_{k=0}^{\infty} t^{k} C_{k+r-1}^{k}(1-\theta)^{k}\theta^{r} =$$

$$= \theta^{r} \sum_{k=0}^{\infty} C_{k+r-1}^{k}((1-\theta)t)^{k},$$

учитывая разложение функции $1/(1-x)^{r}$ в ряд Тейлора

$$\frac{1}{(1-x)^{r}} = \sum_{k=0}^{\infty} C_{k+r-1}^{k} x^{k},\ |x| < 1,$$

имеем следующее представление:

$$P_{\xi}(t) = \frac{\theta^{r}}{(1-t(1-\theta))^{r}},\ |t| < 1. \tag{6.4.1}$$

Производящая функция суммы $\zeta = \xi_1 + \xi_2$ независимых случайных величин ξ_1 и ξ_2, имеющих отрицательные биномиальные распределения соответственно с параметрами (r_1, θ) и (r_2, θ)

$$P_{\zeta}(t) = P_{\xi_1}(t)P_{\xi_2}(t) = \frac{\theta^{r_1}}{(1-t(1-\theta))^{r_1}} \frac{\theta^{r_2}}{(1-t(1-\theta))^{r_2}} =$$

$$= \frac{\theta^{r_1+r_2}}{(1-t(1-\theta))^{r_1+r_2}},\ |t| < 1,$$

совпадает с производящей функцией отрицательного биномиального распределения с параметрами $(r_1 + r_2, \theta)$. Поэтому в силу теоремы единственности (теорема 5.3.5) сумма независимых случайных величин, имеющих отрицательные биномиальные распределения соответственно с параметрами (r_1, θ) и (r_2, θ), имеет отрицательное биномиальное распределение с параметрами $(r_1 + r_2, \theta)$.

3° Пусть случайная величина ξ имеет биномиальное распределение с параметрами (n, θ):

$$\mathsf{P}\{\xi = k\} = C_n^k \theta^k (1-\theta)^{n-k},\ k = 0, 1, \ldots, n;\ \theta \in (0; 1).$$

Её производящая функция

$$P_\xi(t) = \mathsf{M} t^\xi = \sum_{k=0}^{n} t^k C_n^k \theta^k (1-\theta)^{n-k} =$$

$$= \sum_{k=0}^{n} C_n^k (t\theta)^k (1-\theta)^{n-k} = (t\theta + (1-\theta))^n = (1 + \theta(t-1))^n. \quad (6.4.2)$$

Производящая функция суммы $\zeta = \xi + \eta$ независимых биномиально распределенных случайных величин с параметрами (n, θ) и (m, θ)

$$P_\zeta(t) = P_\xi(t) P_\eta(t) =$$

$$= (1 + \theta(t-1))^n (1 + \theta(t-1))^m = (1 + \theta(t-1))^{n+m}$$

совпадает с производящей функцией биномиально распределенной случайной величины с параметрами $(n + m, \theta)$. Поэтому, согласно теореме единственности (см. теорему 5.3.5), распределение суммы $\xi + \eta$ независимых биномиально распределенных случайных величин с параметрами (n, θ) и (m, θ) совпадает с биномиальным распределением с параметрами $(n + m, \theta)$.

Следствие. *Сумма r независимых геометрически распределенных случайных величин с параметром θ имеет отрицательное биномиальное распределение с параметрами (r, θ).*

Достаточно заметить, что геометрическое распределение:

$$\mathsf{P}_\theta(k) = (1-\theta)^k \theta,\ k = 0, 1, 2, \ldots$$

является отрицательным биномиальным распределением с параметрами $(1, \theta)$.

Моменты дискретной случайной величины (дискретного вероятностного распределения). Пусть ξ — случайная величина со значениями в $\mathbb{R}^1$ на $\{\Omega, \mathsf{P}\}$, X — множество ее возможных значений,

$$\mathsf{Q} : x \to \mathsf{Q}(x),\ x \in X \subset \mathbb{R}^1$$

— распределение ξ. Будем называть

$$\mathsf{M}\xi^r = \sum_{x \in X} x^r \mathsf{Q}(x) = m_r,$$

$r = 1, 2, \ldots,$ *моментом порядка* r случайной величины ξ (моментом порядка r распределения Q);

$$\mathsf{M}|\xi|^r = \sum_{x \in X} |x|^r \mathsf{Q}(x) = \mathsf{M}_r$$

— *абсолютным моментом порядка* r случайной величины ξ (абсолютным моментом порядка r распределения Q);

$$\mathsf{M}(\xi - m_1)^r = \sum_{x \in X} (x - m_1)^r \mathsf{Q}(x)$$

— *центральным моментом порядка* r случайной величины ξ (центральным моментом порядка r распределения Q);

$$\mathsf{M}|\xi - m_1|^r = \sum_{x \in X} |x - m_1|^r \mathsf{Q}(x)$$

— *абсолютным центральным моментом порядка* r случайной величины ξ (абсолютным центральным моментом порядка r распределения Q).

Часто используемой характеристикой случайной величины (распределения) является центральный момент второго порядка, называемый дисперсией случайной величины (дисперсией распределения).

Определение. Пусть ξ — случайная величина с распределением Q. *Дисперсией* случайной величины ξ (дисперсией распределения Q) будем называть

$$\mathsf{M}(\xi - \mathsf{M}\xi)^2 = \sum_{x \in X} (x - m_1)^2 \mathsf{Q}(x).$$

Дисперсия обозначается через $\mathsf{D}\xi$ или σ^2, т. е. по определению

$$\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2 = \sum_{x \in X} (x - m_1)^2 \mathsf{Q}(x) = \sigma^2.$$

Величина

$$\sigma = \sqrt{\mathsf{D}\xi}$$

называется *средним квадратическим уклонением* или *стандартным уклонением* случайной величины ξ (стандартным уклонением распределения Q).

Моменты некоторых дискретных случайных величин (дискретных вероятностных распределений). *Моменты биномиального распределения $B_{n,p}$ (моменты биномиально распределенной с параметрами (n, p) случайной величины ξ):*

$$m_1 = np, \ \sigma^2 = npq$$

(см. теорему 6.1.3).

Моменты пуассоновского распределения с параметром λ (моменты пуассоновской с параметром λ случайной величины ξ):

$$m_1 = \lambda, \ \sigma^2 = \lambda.$$

Моменты геометрического распределения с параметром p (моменты геометрически распределенной с параметром p случайной величины ξ):

$$m_1 = \frac{1-p}{p}, \ m_2 = \frac{(1-p)(2-p)}{p^2}, \ \sigma^2 = \frac{1-p}{p^2}$$

(см. пример 5.5.2).

Моменты отрицательного биномиального распределения с параметрами (r, p) (моменты отрицательно биномиально распределенной с параметрами (r, p) случайной величины ξ):

$$m_1 = r\frac{1-p}{p}, \ \sigma^2 = r\frac{1-p}{p^2}$$

(см. пример 5.5.2 и теорему 6.3.2).

6.5 Примеры и задачи

Примеры

Пример 6.5.1. *У большого числа k людей исследуется кровь. Это исследование можно организовать двумя способами.*

Способ 1. *Кровь каждого человека исследуется отдельно. В этом случае потребуется k анализов.*

Способ 2. *Кровь k людей смешивается и анализируется полученная смесь. Если результат анализа отрицателен (диагноз не подтверждается), то этого одного анализа достаточно для k человек. Если анализ положителен, то кровь каждого дополнительно приходится анализировать отдельно, и в целом на k человек потребуется $k+1$ анализ. Предположим, что вероятность p положительного анализа одна и та же для всех людей и что результаты анализов независимы (в теоретико-вероятностном смысле).*

1. *Найти вероятность того, что анализ смешанной крови k людей положителен.*

2. *Вычислить математическое ожидание числа ξ анализов, необходимых при втором методе обследования.*

Решение. Найдем распределение случайной величины ξ — количества анализов, необходимых для исследования крови k человек, если исследование проводится способом 2.

Случайная величина ξ принимает значения: 1 и $k+1$. Обозначим через η количество положительных анализов на k проведенных, если кровь каждого из k человек анализируется отдельно. Случайная величина η имеет биномиальное распределение с параметрами $(k;p)$, вероятность хотя бы одного положительного анализа на k равна

$$\mathsf{P}\{\eta \geq 1\} = 1 - \mathsf{P}\{\eta = 0\} = 1 - C_k^0 p^0 (1-p)^k = 1 - (1-p)^k,$$

а вероятность того, что все анализы отрицательны, равна $(1-p)^k$. Поэтому случайная величина ξ имеет такое распределение:

$$\mathsf{P}\{\xi = 1\} = \mathsf{P}\{\eta = 0\} = (1-p)^k,$$

$$\mathsf{P}\{\xi = k+1\} = \mathsf{P}\{\eta \geq 1\} = 1 - (1-p)^k.$$

По распределению ξ находим:

$$\mathsf{M}\xi = 1(1-p)^k + (k+1)(1-(1-p)^k) = k(1-(1-p)^k) + 1.$$

Пример 6.5.2 (задача Банаха). *Математик носит с собой две коробки спичек. Каждый раз, когда ему нужна спичка, он наудачу берет одну из коробок, вынимает спичку и возвращает коробку обратно. Рано или поздно наступит момент, когда выбранная коробка окажется пустой. Найти вероятность того, что другая коробка содержит r спичек, в предположении, что сначала в каждой коробке содержится N $(r \le N)$ спичек.*

Решение. Пометим одну из коробок, например, звездочкой. Обозначим через B^* событие "помеченная коробка пустая", через B — событие "непомеченная коробка пустая", через A_r обозначим событие "непустая коробка содержит r спичек". Тогда

$$\mathsf{P}(A_r) = \mathsf{P}(A_r|B^*)\mathsf{P}(B^*) + \mathsf{P}(A_r|B)\mathsf{P}(B).$$

Коробки берут наудачу, поэтому естественно положить $\mathsf{P}(B^*) = 1/2$, $\mathsf{P}(B) = 1/2$. Вычислим $\mathsf{P}(A_r|B^*)$. Коль скоро произошло событие B^* — "помеченная коробка пустая", то событие A_r — "непустая коробка содержится r спичек" происходит тогда, когда коробки брали $(2N - r)$ раз, причем помеченную брали N раз. Поэтому $\mathsf{P}(A_r|B^*)$ равна вероятности N успехов (успех — выбрана помеченная коробка) в $(2N-r)$ испытаниях Бернулли с вероятностью успеха $1/2$ в одном испытании, т. е.

$$\mathsf{P}(A_r|B^*) = C_{2N-r}^{N}(1/2)^{2N-r}.$$

Аналогично получаем

$$\mathsf{P}(A_r|B) = C_{2N-r}^{N-r}(1/2)^{2N-r}.$$

Так что искомая вероятность

$$\mathsf{P}(A_r) = \frac{1}{2}C_{2N-r}^{N}(1/2)^{2N-r} + \frac{1}{2}C_{2N-r}^{N-r}(1/2)^{2N-r} =$$

$$= C_{2N-r}^{N}(1/2)^{2N-r}.$$

Задачи

6.1. Пусть ξ и η — независимые случайные величины, имеющие распределение Пуассона соответственно с параметрами λ и μ.

Доказать, что условным распределением случайной величины ξ при условии, что $\xi+\eta=n$, является биномиальное распределение с параметрами $(n;\lambda/(\lambda+\mu))$, т. е.

$$\mathsf{P}\{\xi=k/\xi+\eta=n\}=C_n^k\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(1-\frac{\lambda}{\lambda+\mu}\right)^{n-k},$$

$k=0,1,\ldots,n$.

6.2. Технический контроль проверяет изделия, каждое из которых, независимо от остальных, может с вероятностью p оказаться дефектным.

1) Какова вероятность того, что из 10 проверенных изделий только одно окажется дефектным?

2) Найти вероятность того, что первым дефектным изделием окажется k-е проверенное изделие.

3) Найти вероятность того, что последующие 10 изделий окажутся стандартными, при условии, что предыдущие l изделий были стандартными. Зависит ли эта вероятность от l?

Ответы: 1) $C_{10}^1p(1-p)^9$; 2) $p(1-p)^{k-1}$; 3) $(1-p)^{10}$.

6.3. Найти вероятность выпадения по крайней мере:

а) одной шестерки при подбрасывании 6 игральных костей;

б) двух шестерок при подбрасывании 12 игральных костей;

в) трех шестерок при подбрасывании 18 игральных костей.

Ответы: а) $1-\dfrac{5^6}{6^6}$; б) $1-\dfrac{17\cdot 5^{11}}{6^{12}}$; в) $1-\dfrac{196\cdot 5^{16}}{6^{18}}$.

6.4. Что вероятнее: выиграть у игрока, равного по силе (игра ведется без ничьих),

1) 4 партии из 8 или 3 из 5;

2) 3 партии из 6 или 2 из 4.

Ответы: вероятнее выиграть

1) 3 партии из 5;

2) 2 партии из 4.

6.5. В экспериментах на встречных электрон-позитронных пучках вероятность того, что за единицу времени произойдет j столкновений, сопровождающихся рождением новых элементарных частиц, равна

$$p_j=\frac{\lambda^j}{j!}e^{-\lambda},\ j=0,1,\ldots,$$

λ — положительный параметр. При каждом столкновении в результате взаимодействия могут образоваться разные группы элементарных частиц, при этом вероятность появления каждой группы постоянна и не зависит от результатов взаимодействия при других столкновениях. В качестве одной из таких групп рассмотрим пару μ-мезонов и обозначим через p вероятность их образования при одном столкновении. Найти распределение случайной величины — количества пар μ-мезонов, рождающихся за единицу времени.

Ответ: пуассоновское распределение с параметром λp.

6.6. Двое играют в игру, подбрасывая по очереди монету. Выигрывает тот, у кого впервые выпадет герб. Найти вероятность p_k того, что игра закончится на k-м подбрасывании. Во сколько раз больше вероятность выигрыша для игрока, который начинает первым?

Ответ: $p_k = (1/2)^k$; вероятность выигрыша для того, кто начинает первым, в два раза больше.

6.7. Театр, вмещающий 1000 человек, имеет два разных входа. Возле каждого из них имеется гардероб. Количество мест в гардеробах одинаковое.

Какое минимальное количество мест должно быть в гардеробах, чтобы в среднем в 9 случаях из 10 все зрители могли раздеться в гардеробе того входа, через который они вошли? Рассмотрите два случая: а) зрители приходят парами; б) зрители приходят по одному. Предположите, что входы зрители выбирают с одинаковыми вероятностями.

Р е ш е н и е. Обозначим через k количество мест в каждом гардеробе. Пусть μ_i — случайная величина, принимающая значение 1, если i-й зритель выбрал вход 1, и 0, если выбрал вход 2, $i = 1, 2, \ldots, 1000$; μ_i — независимые случайные величины, каждая с распределением

$$\mathsf{P}\{\mu_i = x\} = (1/2)^x (1/2)^{1-x},\ x = 0, 1.$$

Тогда $\mu = \sum\limits_{i=1}^{1000} \mu_i$ — количество зрителей, которые вошли через вход 1, а $1000 - \mu$ — через вход 2.

Событие “все зрители имеют возможность раздеться” в терминах случайной величины μ запишется так:

$$\{\mu < k, 1000 - \mu < k\} = \{1000 - k < \mu < k\}.$$

Количество мест k (у каждом гардеробе) находится как минимальное натуральное число, удовлетворяющее условию

$$\mathsf{P}\{1000 - k < \mu < k\} \geq 0,90.$$

Чтобы найти k, воспользуемся интегральной теоремой Муавра—Лапласа.

Ответ: $k = 527$.

6.8. Симметричную монету подбрасывают 10 раз. Как известно, вероятность того, что герб выпадет 4 раза, равна

$$B_{10;1/2}(4) = C_{10}^4(1/2)^{10}.$$

А чему равна вероятность того, что при последовательном подбрасывании монеты для появления 4-х гербов необходимо сделать 10 подбрасываний?

Решение. Число неудач ξ до r-го успеха в последовательности независимых испытаний с вероятностью успеха p в каждом испытании имеет отрицательное биномиальное распределение с параметрами $(r;p)$:

$$\mathsf{P}_\xi(k) = C_{k+r-1}^{r-1} p^r (1-p)^k,\ k = 0, 1, \ldots$$

Решение задачи сводится к вычислению вероятности того, что число неудач до 4-го успеха равно 6 (вероятность успеха в каждом испытании 1/2). Искомая вероятность

$$C_{6+4-1}^{4-1}(1/2)^4(1-1/2)^6 = C_9^3(1/2)^{10}.$$

6.9. Вероятность того, что в справочное бюро на протяжении часа обратилось k человек, равна

$$\frac{\lambda^k}{k!}e^{-\lambda},\ k = 0, 1, \ldots,$$

(λ — положительный параметр). Для каждого человека вероятность отказа равна p $(0 < p < 1)$. Найти вероятность того, что на протяжении часа s человек не получат ответ на свой вопрос.

Ответ: $e^{-\lambda p}(\lambda p)^s/s!$

Глава 7

Построение вероятностных пространств

7.1 Продолжение вероятности. Теорема Каратеодори

Построить математическую модель стохастического эксперимента (вероятностное пространство) — это значит определить пространство элементарных событий Ω, σ-алгебру $\mathfrak{F}$ подмножеств Ω и задать вероятность на $\mathfrak{F}$. Исходными данными для построения математической модели стохастического эксперимента, как правило, являются частоты событий на некотором классе $\mathfrak{K}$ наблюдаемых событий.

Примеры

1. Бросают симметричную игральную кость. Построить математическую модель данного стохастического эксперимента.

События “выпало одно очко”, “выпало два очка”, ..., “выпало шесть очков” являются наблюдаемыми (обозначим их $1, 2, \ldots, 6$), класс этих событий обозначим через $\mathfrak{K}$. События из класса $\mathfrak{K}$ в последовательности экспериментов появляются одинаково часто. Это опытные данные. Они являются исходными для построения математической модели стохастического эксперимента. А именно, событиям из класса $\mathfrak{K}$ естественно приписать вероятности так:

$$\mathsf{P}(i) = 1/6,\ i = 1, 2, \ldots, 6.$$

Вероятность P, заданная на классе $\mathfrak{K}$, продолжается на класс всех подмножеств пространства Ω: для каждого $A \subset \Omega$

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega).$$

2. На отрезок $[a; b]$ наудачу бросают точку. Предложить математическую модель этого стохастического эксперимента.

Класс $\mathfrak{K}$ наблюдаемых событий образуют события $A_{a',b'} =$ = "брошенная точка попала в промежуток с концами a', b'; $a', b' \in [a; b]$". Частота $\nu(A_{a',b'})$ события $A_{a',b'}$ в последовательности экспериментов близка к $(b'-a')/(b-a)$. Это опытные данные. Они являются исходными для построения математической модели стохастического эксперимента. А именно, событию $A_{a',b'}$ естественно приписать вероятность

$$\mathsf{P}(A_{a',b'}) = (b' - a')/(b - a).$$

Далее, вероятность P, заданная на классе $\mathfrak{K}$, продолжается на наименьшую σ-алгебру, содержащую класс $\mathfrak{K}$.

О задании вероятности на несчетном пространстве элементарных событий. Если пространство элементарных событий Ω стохастического эксперимента дискретно, то вероятности элементарных событий $\mathsf{P}(\omega)$ однозначно задают вероятности событий из σ-алгебры всех подмножеств Ω — самого "богатого" класса подмножеств Ω, а именно, для каждого $A \subset \Omega$

$$\mathsf{P}(A) = \sum_{\omega \in A} \mathsf{P}(\omega)$$

(класс $\mathfrak{K}$, на котором первоначально определена вероятность, образуют элементарные события).

В общей ситуации определить вероятность на достаточно "богатом" классе событий по вероятностям элементарных событий заведомо не удастся. Это невозможно сделать даже при построении вероятностного пространства такого стохастического эксперимента как бросание случайным образом точки на отрезок $[0; 1]$. Действительно, вероятности элементарных событий следует считать равными нулю: $\mathsf{P}(\omega) = 0$, $\omega \in [0; 1]$, поскольку нулю равна частота попадания бросаемой точки в заданную точку ω промежутка $[0; 1]$. Но по вероятностям элементарных событий, равным нулю, невозможно корректно задать вероятность попадания точки даже в промежуток, например, $[0; 1/2]$. На подмножествах промежутка $[0; 1]$ можно задать много вероятностей P, значения каждой из которых на элементарных событиях ω равно нулю, а значение на промежутке $[0;1/2]$ свое, вообще говоря, любое число (из $[0;1]$). Например, для любой абсолютно непрерывной вероятности P на $[0;1]$, т. е. имеющей плотность $f(t)$ (см. параграф 9.6 в гл. 9), вероятность $\mathsf{P}(\{\omega\})$ точки $\omega \in [0;1]$ равна нулю, как значение

$$\mathsf{P}(\{\omega\}) = \int\limits_{\{\omega\}} f(t)dt$$

интеграла по множеству меры нуль, но при этом за счет выбора плотности f, скажем так: $f(t) = a$, когда $t \in [0, 1/2]$ и $f(t) = 2-a$, когда $t \in (1/2, 1]$ $(0 \le a \le 2)$, значение вероятности

$$\mathsf{P}([0;1/2]) = \int\limits_{[0;1/2]} f(t)dt = a/2$$

может быть сделано любым числом из $[0;1]$. С другой стороны, опыт показывает, что частота попадания точки в промежуток $[0;1/2]$ при ее бросании на отрезок $[0;1]$ близка к $1/2$, а поэтому вероятность попадания точки на отрезок $[0;1/2]$ следует считать равной $1/2$.

Среди различных способов случайного бросания точки на отрезок $[0;1]$ выделяется бросание, когда вероятность попадания точки в промежуток $[a;b] \subset [0;1]$ пропорциональна длине промежутка $[a;b]$. Такое бросание точки будем называть бросанием наудачу.

Так что при построении вероятностного пространства стохастического эксперимента с несчетным множеством исходов Ω класс $\mathfrak{K}$ наблюдаемых событий, на котором определены вероятности, должен быть достаточно богатым и уж заведомо богаче класса элементарных событий. Разумеется, для каждого стохастического эксперимента класс $\mathfrak{K}$ свой.

Продолжение вероятности на $\boldsymbol{\sigma}$-алгебру. “Исходным материалом” для построения вероятностного пространства является вероятность, заданная из тех или иных соображений на некотором классе $\mathfrak{A}$ подмножеств пространства Ω. При этом можно считать, что класс $\mathfrak{A}$ вместе с $A, B \in \mathfrak{A}$ содержит и $A \cap B$, вместе с $A \in \mathfrak{A}$ содержит и $\overline{A}$, т. е. является алгеброй. Оказывается, что вероятность, заданную на алгебре $\mathfrak{A}$ подмножеств Ω, всегда (без дополнительных предположений) можно продолжить на так называемую σ-алгебру $\mathfrak{F} = \sigma(\mathfrak{A})$, порожденную алгеброй $\mathfrak{A}$.

Определение. *σ-Алгеброй, порожденной классом* $\mathfrak{A}$, будем называть σ-алгебру $\sigma(\mathfrak{A})$, которая, во-первых, содержит класс $\mathfrak{A}$:

$$\mathfrak{A} \subset \sigma(\mathfrak{A}),$$

и, во-вторых, содержится в любой другой σ-алгебре $\mathfrak{F}$, содержащей класс $\mathfrak{A}$:

$$\sigma(\mathfrak{A}) \subset \mathfrak{F}.$$

σ-Алгебру $\sigma(\mathfrak{A})$ еще называют *наименьшей σ-алгеброй, содержащей класс* $\mathfrak{A}$.

Теорема 7.1.1. *Для любого класса $\mathfrak{K}$ подмножеств Ω существует наименьшая σ-алгебра $\sigma(\mathfrak{K})$, его содержащая.*

Доказательство. Заметим, что семейство σ-алгебр, содержащих класс $\mathfrak{K}$, не пусто, например, σ-алгебра всех подмножеств Ω содержит $\mathfrak{K}$.

Пересечение любого семейства σ-алгебр $\mathfrak{F}_\alpha, \alpha \in J$, есть σ-алгебра. Убедимся в этом.

Пусть $A \in \bigcap\limits_{\alpha\in J} \mathfrak{F}_\alpha$, т. е. $A \in \mathfrak{F}_\alpha$ для каждого $\alpha \in J$. Так как $\mathfrak{F}_\alpha$ — σ-алгебра, то и $\overline{A} \in \mathfrak{F}_\alpha$ для каждого $\alpha \in J$, т. е. $\overline{A} \in \bigcap\limits_{\alpha\in J} \mathfrak{F}_\alpha$.

Пусть $A_i \in \bigcap\limits_{\alpha\in J} \mathfrak{F}_\alpha$, $i = 1, 2, \ldots$, т. е. $A_i \in \mathfrak{F}_\alpha$, $i = 1, 2, \ldots$, для каждого $\alpha \in J$. Так как $\mathfrak{F}_\alpha$ — σ-алгебра, то и $\bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{F}_\alpha$ для каждого $\alpha \in J$, т. е. $\bigcup\limits_{i=1}^{\infty} A_i \in \bigcap\limits_{\alpha\in J} \mathfrak{F}_\alpha$.

Поэтому по определению σ-алгебры класс $\bigcap\limits_{\alpha\in J} \mathfrak{F}_\alpha$ является σ-алгеброй.

Пересечение $\bigcap\limits_{\alpha\in I} \mathfrak{F}_\alpha$ всех σ-алгебр, содержащих класс $\mathfrak{K}$, и является σ-алгеброй, порожденной $\mathfrak{K}$.

В самом деле, во-первых, $\bigcap\limits_{\alpha\in I} \mathfrak{F}_\alpha$ является σ-алгеброй; во-вторых, $\bigcap\limits_{\alpha\in I} \mathfrak{F}_\alpha$ содержит класс $\mathfrak{K}$; и, наконец, σ-алгебра $\bigcap\limits_{\alpha\in I} \mathfrak{F}_\alpha$ как пересечение всех σ-алгебр, содержащих класс $\mathfrak{K}$, содержится в любой σ-алгебре, содержащей $\mathfrak{K}$.

Пример 7.1.1. *Пусть $\mathfrak{K}_1$ и $\mathfrak{K}_2$ — два класса подмножеств Ω, причем $\mathfrak{K}_1 \subset \mathfrak{K}_2$. Доказать, что $\sigma(\mathfrak{K}_1) \subset \sigma(\mathfrak{K}_2)$.*

Решение. Из условия и определения σ-алгебры, порожденной данным классом, имеем: $\mathfrak{K}_1 \subset \mathfrak{K}_2 \subset \sigma(\mathfrak{K}_2)$. Поэтому

$\mathfrak{K}_1 \subset \sigma(\mathfrak{K}_2)$. А поскольку $\sigma(\mathfrak{K}_1)$ — наименьшая σ-алгебра, содержащая класс $\mathfrak{K}_1$, а $\sigma(\mathfrak{K}_2)$ — одна из σ-алгебр, которая содержит класс $\mathfrak{K}_1$, то $\sigma(\mathfrak{K}_1) \subset \sigma(\mathfrak{K}_2)$.

Теорема 7.1.2 (Каратеодори о продолжении вероятности). *Пусть* $\mathsf{P}(\cdot)$ *— вероятность, заданная на алгебре* $\mathfrak{A}$ *подмножеств пространства* Ω. *Существует, и притом единственная, вероятность* $\mathsf{Q}(\cdot)$, *заданная на* σ*-алгебре* $\sigma(\mathfrak{A})$, *порожденной алгеброй* $\mathfrak{A}$, *такая, что*

$$\mathsf{Q}(A) = \mathsf{P}(A)$$

для каждого $A \in \mathfrak{A}$.

Вероятность Q называют *продолжением вероятности* P с алгебры $\mathfrak{A}$ на σ-алгебру $\sigma(\mathfrak{A})$.

Единственность продолжения вероятности P с алгебры $\mathfrak{A}$ на σ-алгебру $\sigma(\mathfrak{A})$ означает следующее. Если Q и Q' — два продолжения вероятности P на σ-алгебру $\sigma(\mathfrak{A})$, т. е. Q — вероятность, заданная на $\sigma(\mathfrak{A})$, и такая, что

$$\mathsf{Q}(A) = \mathsf{P}(A)$$

для каждого $A \in \mathfrak{A}$, и Q' — вероятность, заданная на $\sigma(\mathfrak{A})$, и такая, что

$$\mathsf{Q}'(A) = \mathsf{P}(A)$$

для каждого A из $\mathfrak{A}$, то

$$\mathsf{Q}'(A) = \mathsf{Q}(A)$$

для всех A из $\sigma(\mathfrak{A})$.

Следствие (из теоремы Каратеодори). *Вероятности* $\mathsf{P}(\cdot)$ *и* $\mathsf{Q}(\cdot)$, *заданные на* σ*-алгебре* $\mathfrak{F}$, *совпадают, если они совпадают на алгебре* $\mathfrak{A}$, *порождающей* σ*-алгебру* $\mathfrak{F}$ ($\mathfrak{F} = \sigma(\mathfrak{A})$).

Чтобы убедиться, что вероятности $\mathsf{P}(\cdot)$ *и* $\mathsf{Q}(\cdot)$, *заданные на* σ*-алгебре* $\mathfrak{F}$, *совпадают, достаточно убедиться, что они совпадают на алгебре* $\mathfrak{A}$, *порождающей* σ*-алгебру* $\mathfrak{F}$ ($\mathfrak{F} = \sigma(\mathfrak{A})$).

В силу теоремы Каратеодори вероятность P, заданную на алгебре $\mathfrak{A}$, можно продолжить на $\sigma(\mathfrak{A})$, причем без каких-либо дополнительных ограничений на $\mathfrak{A}$ и P. Поэтому в дальнейшем всегда будем считать вероятность заданной на σ-алгебре.

Меры. Далее приводятся необходимые определения и факты из теории меры.

Определение. *Измеримым пространством* будем называть пару $\{\Omega, \mathfrak{F}\}$, где $\mathfrak{F}$ — σ-алгебра подмножеств пространства Ω.

Определение. *Мерой* будем называть неотрицательную счетно-аддитивную функцию множества $\mu(\cdot)$, заданную на σ-алгебре $\mathfrak{F}$ измеримого пространства $\{\Omega, \mathfrak{F}\}$ и такую, что $\mu(\emptyset) = 0$.

З а м е ч а н и е. Условие $\mu(\emptyset) = 0$ эквивалентно тому, что мера μ не равна тождественно $+\infty$ на $\mathfrak{F}$ (проверить).

Для мер имеет место свойство непрерывности, аналогичное сформулированному ранее свойству непрерывности для вероятностей (см. теорему 3.2.1).

Определение. *Пространством с мерой* будем называть тройку $\{\Omega, \mathfrak{F}, \mu\}$, где $\{\Omega, \mathfrak{F}\}$ — измеримое пространство, а μ — мера на $\mathfrak{F}$.

Определение. Меру μ, для которой $\mu(\Omega) < +\infty$, будем называть *конечной мерой.*

Меру μ, для которой $\mu(\Omega) = 1$, будем называть *вероятностной мерой.*

Вероятностная мера является вероятностью.

Определение. Меру μ на $\{\Omega, \mathfrak{F}\}$ будем называть *σ-конечной*, если Ω можно представить в виде объединения счетного числа непересекающихся множеств Ω_i, таких, что $\mu(\Omega_i)$ конечно для каждого $i = 1, 2, \ldots$

В дальнейшем мы будем иметь дело только с конечными и σ-конечными мерами.

Определение. Меру μ, заданную на σ-алгебре $\mathfrak{F}$, будем называть *полной*, если для каждого множества $A \in \mathfrak{F}$ такого, что $\mu(A) = 0$, из $A' \subset A$ следует $A' \in \mathfrak{F}$.

Меру всегда можно считать полной, что следует из приводимого далее утверждения.

Пусть ν — мера на σ-алгебре $\mathfrak{S}$ подмножеств пространства Ω. Определим класс $\mathfrak{N}$ подмножеств Ω следующим образом:

$$\mathfrak{N} = \{N : N \subset S \in \mathfrak{S}, \nu(S) = 0\}$$

(класс $\mathfrak{N}$ образуют подмножества множеств ν-меры нуль). Через $\mathfrak{F}$ обозначим класс множеств вида

$$A = \left(E \cup N'\right) \setminus N'', \text{ где } E \in \mathfrak{S}, N' \in \mathfrak{N}, N'' \in \mathfrak{N}.$$

Класс множеств $\mathfrak{F}$ является σ-алгеброй, а функция множества μ, задаваемая на $\mathfrak{F}$ равенством

$$\mu(E) = \mu\left(\left(E \cup N'\right) \setminus N''\right) = \nu(E),$$

является полной мерой на σ-алгебре $\mathfrak{F}$.

Меру μ можно рассматривать как продолжение меры ν в том смысле, что

$$\mu(E) = \nu(E),\ E \in \mathfrak{S}.$$

Мера может задаваться не только на σ-алгебре, но и на алгебре. При этом имеет место

Теорема 7.1.3 (Каратеодори о продолжении меры). *Пусть $\mu(\cdot)$ — мера, заданная на алгебре $\mathfrak{A}$ подмножеств пространства Ω. Существует, и притом единственная, мера $M(\cdot)$, заданная на σ-алгебре $\sigma(\mathfrak{A})$, порожденной алгеброй $\mathfrak{A}$, такая, что*

$$M(A) = \mu(A)$$

для каждого $A \in \mathfrak{A}$.

Из единственности продолжения меры с алгебры на σ-алгебру $\sigma(\mathfrak{A})$ получаем следующее утверждение.

Следствие (из теоремы Каратеодори). *Меры $\mu(\cdot)$ и $\nu(\cdot)$, заданные на σ-алгебре $\mathfrak{F}$, совпадают, если они совпадают на алгебре $\mathfrak{A}$, порождающей σ-алгебру $\mathfrak{F}$ $(\mathfrak{F} = \sigma(\mathfrak{A}))$.*

Чтобы убедиться, что меры $\mu(\cdot)$ и $\nu(\cdot)$, заданные на σ-алгебре $\mathfrak{F}$, совпадают, достаточно убедиться, что они совпадают на алгебре $\mathfrak{A}$, порождающей σ-алгебру $\mathfrak{F}$ $(\mathfrak{F} = \sigma(\mathfrak{A}))$.

7.2 Задание распределения на $\mathbb{R}^n$

Стохастические эксперименты с исходами в $\mathbb{R}^n$ (случайные величины) представляют собой важный класс стохастических экспериментов. Их математическими моделями являются вероятностные пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ с $\mathbb{R}^n$ (или частью $\mathbb{R}^n$) в качестве Ω, σ-алгеброй борелевских множеств $\mathfrak{B}^n$ в качестве $\mathfrak{F}$ (о ней далее) и вероятностью (вероятностным распределением) на σ-алгебре $\mathfrak{B}^n$.

Борелевские множества. *σ-Алгеброй борелевских множеств на* $\mathbb{R}^1$ называется σ-алгебра, порожденная классом $\mathfrak{K}$ промежутков вида $[a, b), a, b \in \mathbb{R}^1$. Обозначать σ-алгебру борелевских множеств на $\mathbb{R}^1$ будем через $\mathfrak{B}^1$.

σ-Алгебра борелевских множеств является достаточно "богатым" классом:

$\mathfrak{B}^1$ содержит все одноточечные множества, так как

$$\{a\} = \bigcap_{n=1}^{\infty} [a, a + 1/n);$$

$\mathfrak{B}^1$ содержит открытые интервалы (a, b), так как любой открытый интервал (a, b) можно представить в виде

$$(a, b) = [a, b) \cap \overline{\{a\}};$$

$\mathfrak{B}^1$ содержит все открытые множества, так как каждое открытое множество на $\mathbb{R}^1$ представимо в виде объединения конечного или счетного числа открытых интервалов;

$\mathfrak{B}^1$ содержит все замкнутые множества, так как замкнутое множество является дополнением открытого.

Пример 7.2.1. *Доказать, что класс $\mathfrak{K}$ открытых интервалов (a, b) на числовой прямой порождает σ-алгебру борелевских множеств на $\mathbb{R}^1$.*

Р е ш е н и е. Имеет место включение

$$\{[a, b) : a, b \in \mathbb{R}^1\} = \mathfrak{K}_1 \subset \sigma(\mathfrak{K}),$$

поскольку $\sigma(\mathfrak{K})$ содержит одноточечные множества $\{a\}$, это следует из равенства $\{a\} = \bigcap\limits_{i=1}^{n} (a-1/n,\ a+1/n)$, и $[a, b) = (a, b) \cup \{a\}$. Поэтому

$$\mathfrak{B}^1 = \sigma(\mathfrak{K}_1) \subset \sigma(\mathfrak{K}),$$

поскольку $\sigma(\mathfrak{K}_1)$ — наименьшая σ-алгебра, содержащая класс $\mathfrak{K}_1$, а $\sigma(\mathfrak{K})$ — одна из σ-алгебр, содержащих класс $\mathfrak{K}_1$. Далее, $\mathfrak{K} \subset \mathfrak{B}^1$, поэтому

$$\sigma(\mathfrak{K}) \subset \mathfrak{B}^1,$$

так как $\sigma(\mathfrak{K})$ — наименьшая σ-алгебра, содержащая класс $\mathfrak{K}$, а $\mathfrak{B}^1$ — одна из таких σ-алгебр.

Следовательно, $\sigma(\mathfrak{K}) = \mathfrak{B}^1$.

σ-Алгебру $\mathfrak{B}^1$ порождают, в частности, классы: $\mathfrak{K}_1 = \{[a, b];\ a, b \in \mathbb{R}^1\}$; $\mathfrak{K}_2 = \{(-\infty, x]; x \in \mathbb{R}^1\}$, $\mathfrak{K}_3 = \{(-\infty, x); x \in \mathbb{R}^1\}$, $\mathfrak{K}_4 = \{(x, +\infty); x \in \mathbb{R}^1\}$, $\mathfrak{K}_5 = \{[x, +\infty); x \in \mathbb{R}^1\}$, класс открытых множеств, класс замкнутых множеств.

Определение. *σ-Алгеброй борелевских множеств на* $\mathbb{R}^n$ называется σ-алгебра, порожденная классом параллелепипедов вида $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$.

Обозначать σ-алгебру борелевских множеств на $\mathbb{R}^n$ будем через $\mathfrak{B}^n$.

Определение. Меру на σ-алгебре $\mathfrak{B}^n$ борелевских множеств измеримого пространства $\{\mathbb{R}^n, \mathfrak{B}^n\}$ называют *распределением* на $\mathbb{R}^n$.

Распределение F на $\mathbb{R}^n$ такое, что $\mathsf{F}(\mathbb{R}^n) = 1$, называется *вероятностным* или *распределением вероятностей*, или *вероятностью.*

Теорема Каратеодори (см. теорему 7.1.2), в частности, имеет место для распределений на $\mathbb{R}^n$.

Теорема 7.2.1 (Каратеодори о продолжении распределений). *Пусть F — распределение на алгебре $\mathfrak{A}$ конечных объединений непересекающихся параллелепипедов $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$ пространства $\mathbb{R}^n$.*

Существует, и притом единственное, распределение Q, заданное на σ-алгебре $\mathfrak{B}^n = \sigma(\mathfrak{A})$, порожденной алгеброй $\mathfrak{A}$, такая, что

$$\mathsf{Q}(A) = \mathsf{F}(A)$$

для каждого $A \in \mathfrak{A}$.

Распределение Q называют *продолжением распределения* F с алгебры $\mathfrak{A}$ на σ-алгебру $\sigma(\mathfrak{A})$.

Единственность продолжения распределения F с алгебры $\mathfrak{A}$ на σ-алгебру $\sigma(\mathfrak{A})$ означает следующее. Если Q и Q' — два продолжения распределения F на σ-алгебру $\sigma(\mathfrak{A})$, т. е. Q — распределение, заданное на $\sigma(\mathfrak{A})$, такое, что

$$\mathsf{Q}(A) = \mathsf{F}(A)$$

для каждого $A \in \mathfrak{A}$, Q' — распределение, заданное на $\sigma(\mathfrak{A})$, такое, что

$$\mathsf{Q}'(A) = \mathsf{F}(A)$$

для каждого A из $\mathfrak{A}$, то

$$\mathsf{Q}(A) = \mathsf{Q}'(A)$$

для всех $\sigma(\mathfrak{A})$.

Следствие 1. *Распределения F и Q на $\mathfrak{B}^n$ совпадают, если они совпадают на параллелепипедах $[a_1, b_1) \times [a_2, b_2) \times \ldots \ldots \times [a_n, b_n)$.*

В самом деле, если распределения F и Q, заданные на $\mathfrak{B}^n$, совпадают параллелепипедах $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$, то они совпадают и на алгебре $\mathfrak{A}$ конечных объединений непересекающихся параллелепипедов, а в силу теоремы Каратеодори и на σ-алгебре $\mathfrak{B}^n = \sigma(\mathfrak{A})$, порожденной алгеброй $\mathfrak{A}$.

Поэтому чтобы убедиться, что два распределения на $\mathbb{R}^n$ (на $\mathfrak{B}^n$) совпадают, достаточно убедиться, что они совпадают на параллелепипедах $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$.

В частности для $n = 1$ имеем.

Следствие 2. *Распределения* F *и* Q *на* $\mathfrak{B}^1$ *совпадают, если они совпадают* 1) *на промежутках вида* $[a, b)$; 2) *на промежутках вида* $(-\infty, c)$.

Действительно, если распределения F и Q, совпадают на промежутках $[a, b)$, то они совпадают и на алгебре $\mathfrak{A}$ конечных объединений таких непересекающихся промежутков, а в силу теоремы Каратеодори и на σ-алгебре $\mathfrak{B}^1 = \sigma(\mathfrak{A})$, порожденной алгеброй $\mathfrak{A}$.

Поэтому чтобы убедиться, что два распределения на $\mathbb{R}^1$ (*на* $\mathfrak{B}^1$) *совпадают, достаточно убедиться, что они совпадают на промежутках вида* $[a, b)$ (*на промежутках вида* $(-\infty, c)$).

Построение распределения на $\mathbb{R}^1$ по функции распределения. Как в случае дискретного пространства элементарных событий Ω исходным “материалом” для задания вероятности на подмножествах Ω являются вероятности $\mathsf{P}(\omega)$ элементарных событий (см. теорему 4.1.1), так при построении меры (распределения) на $\mathbb{R}^1$ (на σ-алгебре борелевских множеств $\mathfrak{B}^1$) исходным “материалом” являются значения $\mu([a, b))$ распределения μ на классе промежутков $\{[a, b), a, b \in \mathbb{R}^1\}$, последние, обычно, задают с помощью, так называемой, функции распределения $F(x)$ на $\mathbb{R}^1$ равенством

$$\mu([a, b)) = F(b) - F(a).$$

Определение. Неотрицательную ограниченную монотонно неубывающую непрерывную слева функцию $F(x)$ на $\mathbb{R}^1$ будем называть *функцией распределения* на $\mathbb{R}^1$.

Функцию распределения $F(x)$, удовлетворяющую условиям

$$F(+\infty) = 1, \quad F(-\infty) = 0,$$

где $F(-\infty) = \lim_{x \to -\infty} F(x)$, $F(+\infty) = \lim_{x \to +\infty} F(x)$, будем называть *вероятностной функцией распределения* (собственной).

Теорема 7.2.2 (о построении распределения на $\mathbb{R}^1$). *Пусть* $F(x)$ *— функция распределения на* $\mathbb{R}^1$. *Существует, и притом единственное, распределение* μ *на* σ*-алгебре* $\mathfrak{B}^1$ *такое, что для любых* $a, b \in \mathbb{R}^1$ $(a \le b)$

$$\mu([a, b)) = F(b) - F(a).$$

Если к тому же F *— вероятностная функция распределения, то* μ *— вероятностное распределение* (*вероятность*).

Доказательство. Обозначим через $\mathfrak{A}$ алгебру подмножеств $\mathbb{R}^1$, состоящую из конечных объединений непересекающихся промежутков вида $[a, b)$ (a и b могут принимать значения $\pm\infty$). По функции распределения F определим на $\mathfrak{A}$ функцию множества μ. Сначала определим μ на промежутках $[a, b)$:

$$\mu([a, b)) = F(b) - F(a), \tag{7.2.1}$$

затем на конечных объединениях

$$A = \bigcup_{i=1}^{n} [a_i, b_i) \tag{7.2.2}$$

непересекающихся промежутков ($[a_i, b_i) \cap [a_j, b_j) = \emptyset$, $i \neq j$), естественно, так:

$$\mu(A) = \mu\left(\bigcup_{i=1}^{n} [a_i, b_i)\right) = \sum_{i=1}^{n} \mu\left([a_i, b_i)\right). \tag{7.2.3}$$

Несложно убедиться, что определение $\mu(A)$ корректно, а именно, не зависит от представления A в виде объединения непересекающихся промежутков, хотя такое представление и не единственно.

Так определенная на $\mathfrak{A}$ неотрицательная функция множества μ является конечно-аддитивной. Покажем, что она является и счетно-аддитивной.

Для доказательства счетной аддитивности μ на алгебре $\mathfrak{A}$ достаточно установить, что если

$$[a, b) = \bigcup_{k=1}^{\infty} [a_k, b_k), \ [a_k, b_k) \cap [a_l, b_l) = \emptyset, \ k \neq l,$$

то

$$\mu([a, b)) = \sum_{k=1}^{\infty} \mu([a_k, b_k)).$$

Очевидно, для любого конечного n имеет место включение

$$\bigcup_{k=1}^{n} [a_k, b_k) \subset [a, b),$$

поэтому
$$\sum_{k=1}^{n} \mu([a_k, b_k)) \le \mu([a, b)),$$
а, следовательно, и
$$\sum_{k=1}^{\infty} \mu([a_k, b_k)) \le \mu([a, b)). \tag{7.2.4}$$

Установим неравенство противоположного знака. Пусть $\varepsilon > 0$ (ε произвольно, но фиксировано). Выберем для точки b точку b' ($b' < b$) так, чтобы
$$F(b) - F(b') \le \frac{\varepsilon}{2}, \tag{7.2.5}$$
и для каждой точки a_k точку a'_k ($a'_k < a_k$) так, чтобы
$$F(a_k) - F(a'_k) \le \frac{\varepsilon}{2^{k+1}} \tag{7.2.6}$$
(это всегда можно сделать, поскольку функция распределения $F(x)$ непрерывна слева). Из равенства
$$[a, b) = \bigcup_{k=1}^{\infty} [a_k, b_k)$$
и включений $[a, b'] \subset [a, b)$, $[a_k, b_k) \subset (a'_k, b_k)$, $k = 1, 2, \ldots$, имеем:
$$[a, b'] \subset \bigcup_{k=1}^{\infty} (a'_k, b_k).$$

В силу леммы Гейне—Бореля из бесконечного покрытия замкнутого промежутка $[a, b']$ открытыми (a'_k, b_k), $k = 1, 2, \ldots$, можно извлечь конечное подпокрытие (a'_{k_s}, b_{k_s}), $s = 1, 2, \ldots, p$:
$$[a, b'] \subset \bigcup_{s=1}^{p} (a'_{k_s}, b_{k_s}).$$

Тогда и
$$[a, b') \subset \bigcup_{s=1}^{p} [a'_{k_s}, b_{k_s}).$$

Отсюда, учитывая определение $\mu([a,b))$ (см. (7.2.3)), получаем:

$$\mu([a,b')) \le \sum_{s=1}^{p} \mu([a'_{k_s}, b_{k_s})) \le \sum_{k=1}^{\infty} \mu([a'_k, b_k))$$

или (в терминах функции распределения $F(x)$)

$$F(b') - F(a) \le \sum_{k=1}^{\infty} (F(b_k) - F(a'_k)).$$

Из последнего неравенства, с учетом выбора точек b' и a'_k (см. (7.2.5) и (7.2.6)), получаем:

$$F(b) - F(a) - \frac{\varepsilon}{2} \le \sum_{k=1}^{\infty} \left(F(b_k) - F(a_k) + \frac{\varepsilon}{2^{k+1}} \right)$$

или

$$F(b) - F(a) \le \sum_{k=1}^{\infty} (F(b_k) - F(a_k)) + \varepsilon.$$

А поскольку ε произвольно, то

$$F(b) - F(a) \le \sum_{k=1}^{\infty} (F(b_k) - F(a_k)),$$

или

$$\mu([a,b)) \le \sum_{k=1}^{\infty} \mu([a_k, b_k)),$$

что вместе с неравенством противоположного смысла (см. (7.2.4)) дает

$$\mu([a,b)) = \sum_{k=1}^{\infty} \mu([a_k, b_k)).$$

Итак, конечно-аддитивная функция μ, определенная на алгебре $\mathfrak{A}$, является счетно-аддитивной (на $\mathfrak{A}$), поэтому в силу теоремы Каратеодори она может быть продолжена с алгебры $\mathfrak{A}$ на σ-алгебру $\sigma(\mathfrak{A}) = \mathfrak{B}^1$ и притом единственным образом.

Функция распределения $F(x)$ и построенное по ней распределение (мера) μ связаны соотношением

$$\mu([a,b)) = F(b) - F(a).$$

Далее, если F — вероятностная функция распределения ($F(+\infty) = 1$, $F(-\infty) = 0$), то соответствующее распределение μ является вероятностным ($\mu(\mathbb{R}^1) = 1$). Действительно, пусть $a_n \downarrow -\infty$, $b_n \uparrow +\infty$ при $n \to \infty$, тогда

$$\mathbb{R}^1 = \bigcup_{n=1}^{\infty} [a_n, b_n),$$

причем $[a_n, b_n) \subset [a_{n+1}, b_{n+1})$. Воспользовавшись свойством непрерывности меры, имеем:

$$\mu(\mathbb{R}^1) = \mu\left(\bigcup_{n=1}^{\infty} [a_n, b_n)\right) = \lim_n \mu([a_n, b_n)) =$$
$$= \lim_n (F(b_n) - F(a_n)) = F(+\infty) - F(-\infty) = 1.$$

Тем самым теорема полностью доказана.

Построенное в теореме по $F(x)$ распределение будем называть распределением, соответствующим функции распределения $F(x)$.

Построение распределения на $\mathbb{R}^n$ по функции распределения. Как на $\mathbb{R}^1$ исходными данными для задания распределения являются значения распределения F на промежутках $[a, b)$, задаваемые с помощью функции распределения $F(x)$ на $\mathbb{R}^1$, так при построении распределения F на $\mathbb{R}^n$, исходными данными являются значения распределения на параллелепипедах $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$, задаваемые функцией распределения $F(x_1, x_2, \ldots, x_n)$ на $\mathbb{R}^n$.

Определение. *Функцией распределения на* $\mathbb{R}^n$ будем называть неотрицательную ограниченную функцию $F(x_1, x_2, \ldots, x_n)$, удовлетворяющую условиям:

1) $F(x_1, x_2, \ldots, x_n)$ неубывающая по каждой переменной;

2) $F(x_1, x_2, \ldots, x_n)$ непрерывна слева по каждой переменной;

3) $\Delta^1_{I_1}\ \Delta^2_{I_2} \ldots \Delta^n_{I_n} F(x_1, x_2, \ldots, x_n) \geq 0$,

где $I_k = [a_k, b_k)$, $a_k \leq b_k$, а $\Delta^k_{I_k} g(x_1, x_2, \ldots, x_k, \ldots, x_n)$ для функции g определяется как разность

$$g(x_1, \ldots, b_k, \ldots, x_n) - g(x_1, \ldots, a_k, \ldots, x_n).$$

Теорема (о построении распределения на $\mathbb{R}^n$). *Функция распределения $F(x_1, x_2, \ldots, x_n)$ на $\mathbb{R}^n$ единственным образом задает на σ-алгебре борелевских множеств $\mathfrak{B}^n$ распределение F, причем так, что*

$$\mathsf{F}(I_1 \times I_2 \times \ldots \times I_n) = \Delta^1_{I_1}\ \Delta^2_{I_2} \ldots \Delta^n_{I_n} F(x_1, x_2, \ldots, x_n).$$

Распределение F на $\mathfrak{B}^n$ строится следующим образом. Значения F на параллелепипедах

$$A = I_1 \times I_2 \times \ldots \times I_n \ \ (I_k = [a_k, b_k),\ k = 1, 2, \ldots, n)$$

определяются равенством

$$\mathsf{F}(A) = \Delta^1_{I_1}\ \Delta^2_{I_2} \ldots \Delta^n_{I_n} F(x_1, x_2, \ldots, x_n),$$

на конечных объединениях $\bigcup\limits_{i=1}^{m} A_i$ непересекающихся параллелепипедов $A_1, A_2, \ldots, A_m$ (образующих алгебру $\mathfrak{A}$) равенством

$$\mathsf{F}\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} \mathsf{F}(A_i).$$

Так определенная на алгебре $\mathfrak{A}$ функция множества F является распределением, причем

$$\mathsf{F}(I_1 \times I_2 \times \ldots \times I_n) = \Delta^1_{I_1}\ \Delta^2_{I_2} \ldots \Delta^n_{I_n} F(x_1, x_2, \ldots, x_n).$$

С алгебры $\mathfrak{A}$ на σ-алгебру $\sigma(\mathfrak{A}) = \mathfrak{B}^n$ распределение F продолжается с помощью теоремы Каратеодори.

Пример. *Пусть $F(x)$ — функция распределения, тождественно равная постоянной c. Построить по $F(x)$ меру (распределение) μ.*

Решение. Согласно теореме 7.2.2 на промежутках $[a, b)$

$$\mu([a, b)) = F(b) - F(a) = c - c = 0.$$

Поэтому μ равна нулю и для каждого множества из алгебры $\mathfrak{A}$, образованной конечными объединениями непересекающихся промежутков вида $[a_i, b_i)$. И поскольку $\mathfrak{A}$ порождает σ-алгебру $\mathfrak{B}^1$, то в силу теоремы Каратеодори μ можно продолжить с $\mathfrak{A}$ на $\mathfrak{B}^1$.

С другой стороны, пусть μ' — мера на $\mathfrak{B}^1$, тождественно равная нулю, т. е. $\mu'(B) = 0$ на каждом $B \in \mathfrak{B}^1$, в частности, и на каждом B' из алгебры $\mathfrak{A}$, образованной конечными объединениями непересекающихся промежутков вида $[a_i, b_i)$. Тогда в силу единственности продолжения меры с алгебры на σ-алгебру (см. теорему 7.1.2), $\mu = \mu'$. Поэтому мера, соответствующая функции распределения $F(x) \equiv c$, тождественно равна нулю.

Функция распределения данного распределения. По функции распределения всегда можно построить распределение. Справедливо и обратное: каждому распределению соответствует функция распределения.

Теорема 7.2.3. *Пусть* Q — *вероятностное распределение на* $\mathbb{R}^1$. *Функция точки* $Q(x)$, *определенная равенством*

$$Q(x) = \mathsf{Q}((-\infty, x)),$$

является вероятностной функцией распределения, при этом

$$\mathsf{Q}([a, b)) = Q(b) - Q(a),$$

$$\mathsf{Q}(\{x_0\}) = Q(x_0 + 0) - Q(x_0 - 0).$$

Доказательство. Сначала убедимся, что $Q(x)$ обладает всеми свойствами вероятностной функции распределения.

Функция $Q(x)$ неотрицательна по определению:

$$Q(x) = \mathsf{Q}((-\infty, x)).$$

Далее, если $x_1 \leq x_2$, то $(-\infty, x_1) \subset (-\infty, x_2)$. Поэтому

$$Q(x_1) = \mathsf{Q}((-\infty, x_1)) \leq \mathsf{Q}((-\infty, x_2)) = Q(x_2),$$

т. е. $Q(x)$ — монотонно неубывающая функция.

Непрерывность $Q(x)$ слева следует из непрерывности распределения на монотонно возрастающей последовательности множеств: если $x_n \uparrow x$, то $\bigcup\limits_{n=1}^{\infty} (-\infty, x_n) = (-\infty, x)$, поэтому

$$Q(x) = \mathsf{Q}((-\infty, x)) = \mathsf{Q}\left(\bigcup_{n=1}^{\infty} (-\infty, x_n)\right) =$$

$$= \lim_{n\to\infty} \mathsf{Q}((-\infty, x_n)) = \lim_{n\to\infty} Q(x_n),$$

т. е.

$$Q(x) = \lim_{x_n \uparrow x} Q(x_n).$$

Равенства $Q(-\infty) = 0,\ Q(+\infty) = 1$ следуют из свойств непрерывности распределения Q на монотонных последовательностях множеств и соотношений $\bigcap\limits_{n=1}^{\infty} (-\infty, x_n) = \emptyset$ при $x_n \downarrow -\infty$, $\bigcup\limits_{n=1}^{\infty} (-\infty, x_n) = (-\infty, +\infty)$ при $x_n \uparrow +\infty$.

Так что $Q(x) = \mathsf{Q}((-\infty, x))$ является функцией распределения.

Далее, если $a < b$, то $(-\infty, a) \subset (-\infty, b)$ и

$$[a, b) = (-\infty, b) \setminus (-\infty, a),$$

поэтому

$$\mathsf{Q}([a, b)) = \mathsf{Q}((-\infty, b)) - \mathsf{Q}((-\infty, a)) = Q(b) - Q(a).$$

Осталось доказать, что

$$\mathsf{Q}(\{x_0\}) = Q(x_0 + 0) - Q(x_0 - 0).$$

Очевидно,

$$\{x_0\} = \bigcap_{n=1}^{\infty} \left[x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right),$$

причем множества $[x_0 - 1/n, x_0 + 1/n), n = 1, 2, \ldots,$ образуют монотонно убывающую последовательность. Поэтому в силу свойства непрерывности распределения

$$\mathsf{Q}(\{x_0\}) = \mathsf{Q}\left(\bigcap_{n=1}^{\infty} \left[x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right)\right) =$$

$$= \lim_{n} \mathsf{Q}\left(\left[x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right)\right) =$$

$$= \lim_{n} \left(Q\left(x_0 + \frac{1}{n}\right) - Q\left(x_0 - \frac{1}{n}\right)\right) =$$

$$= Q(x_0 + 0) - Q(x_0 - 0).$$

Определение. *Функцией распределения данного вероятностного распределения* (вероятностной меры) Q на $\mathbb{R}^1$ будем называть функцию точки $Q(x)$, определенную на $\mathbb{R}^1$ равенством

$$Q(x) = \mathsf{Q}((-\infty, x)).$$

Распределение и его функцию распределения, как правило, обозначают одной и той же буквой.

Теорема. *Между семейством вероятностных функций распределений $F(x)$ на $\mathbb{R}^1$ и классом вероятностных распределений μ существует взаимно однозначное соответствие, причем такое, что для функции распределения $F(x)$ и соответствующего ей распределения μ имеет место равенство*

$$\mu([a, b)) = F(b) - F(a).$$

Доказательство. Каждой вероятностной функции распределения $F(x)$ поставим в соответствие вероятностное распределение μ, которое строится по $F(x)$ согласно теореме 7.2.2, при этом

$$\mu([a, b)) = F(b) - F(a).$$

Убедимся, что это соответствие между вероятностными функциями распределения и вероятностными распределениями взаимно однозначное.

Различным вероятностным функциям распределения ставятся в соответствие различные вероятностные распределения. Если функциям распределения $F(x)$ и $G(x)$ поставлено в соответствие одно и то же распределение μ, то $F(x) = G(x)$. В самом деле, для любого промежутка $[x_n, x)$

$$G(x) - G(x_n) = \mu([x_n, x)) = F(x) - F(x_n).$$

Отсюда, переходя к пределу при $x_n \downarrow -\infty$, получаем

$$G(x) = \mu((-\infty, x)) = F(x).$$

Осталось заметить, что у каждого вероятностного распределения μ существует функция распределения $G(x)$, которой распределение μ ставится в соответствие; эта функция распределения определяется равенством

$$G(x) = \mu((-\infty, x)).$$

Мера Лебега. Пусть $[a, b)$ — произвольный промежуток на $\mathbb{R}^1$. Через $\mathfrak{B}^1_{[a,b)}$ обозначим σ-алгебру борелевских множеств вида $B \cap [a, b), B \in \mathfrak{B}^1$ ($\mathfrak{B}^1_{[a,b)}$ — σ-алгебра борелевских множеств промежутка $[a, b)$).

Определим на $\mathbb{R}^1$ функцию распределения $F(x)$:

$$F(x) = \begin{cases} 0, & \text{если } x \le a; \\ x - a, & \text{если } a < x \le b; \\ b - a, & \text{если } x > b \end{cases} \tag{7.2.7}$$

(см. также рис. 7.2.1). Как и каждая функция распределения, $F(x)$ задает на $\mathfrak{B}^1$ некоторое распределение (меру) μ. В частности, эта мера оказывается определенной на σ-алгебре $\mathfrak{B}^1_{[a,b)}$, поскольку $\mathfrak{B}^1_{[a,b)} \subset \mathfrak{B}^1$. Функция распределения $F(x)$ и построенная по ней мера μ связаны соотношением

$$\mu([a', b')) = F(b') - F(a').$$

На $[a, b)$ функция распределения $F(x)$ равна $x - a$, поэтому для промежутков $[a', b') \subset [a, b)$

$$\mu([a', b')) = F(b') - F(a') = b' - a'.$$

Так что для меры, построенной по функции распределения $F(x)$, заданной равенством (7.2.7), ее значение на каждом промежутке $[a', b')$ из $[a, b)$ совпадает с длиной $b' - a'$ промежутка $[a', b')$.

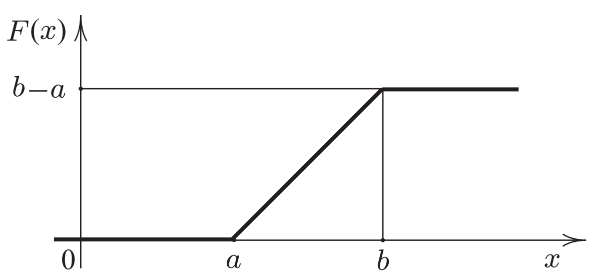

Рис. 7.2.1: График функции $F(x)$

Определение. Меру, заданную на σ-алгебре $\mathfrak{B}^1_{[a,b)}$, значение которой на каждом промежутке $[a', b') \subset [a, b)$ совпадает с его длиной, будем называть *мерой Лебега* на $[a, b)$.

Определение. Меру L, заданную на σ-алгебре $\mathfrak{B}^1$, значение которой на каждом промежутке $[a,b)$ совпадает с его длиной:

$$\mathsf{L}([a,b)) = b - a,$$

будем называть *мерой Лебега* на $\mathbb{R}^1$.

Мера Лебега на прямой строится так. Представим прямую $\mathbb{R}^1$ в виде $\mathbb{R}^1 = \bigcup\limits_{n=-\infty}^{+\infty} [n, n+1)$. Через μ_n обозначим меру Лебега на промежутке $[n, n+1)$, $n \in \mathbb{Z}$. Далее, определим на $\mathfrak{B}^1$ функцию множества следующим равенством:

$$\mathsf{L}(B) = \sum_{n=-\infty}^{+\infty} \mu_n \left(B \cap [n, n+1)\right), \; B \in \mathfrak{B}^1.$$

Функция L неотрицательна и счетно-аддитивна (счетная аддитивность L следует из счетной аддитивности μ_n на $\mathfrak{B}^1_{[n,n+1)}$). Тем самым на $\mathfrak{B}^1$ задана мера L, причем по построению для любого промежутка $[a,b)$

$$\mathsf{L}([a,b)) = b - a.$$

Мера Лебега (функция множества, заданная на $\mathfrak{B}^1$) является обобщением длины, в том смысле, что значение меры на каждом промежутке совпадает с его длиной.

Пример. *Найти меру Лебега множества иррациональных чисел промежутка* $[0;1)$.

Решение. Сначала найдем меру Лебега множества $Q_{[0;1)}$ рациональных чисел промежутка $[0;1)$.

$Q_{[0;1)}$ можно представить так: $Q_{[0;1)} = \bigcup\limits_{r} \{r\}$, где $\{r\}$ — одноточечное множество, состоящее из одного рационального числа r. Мера Лебега множества $\{r\}$ равна нулю:

$$\mathsf{L}(\{r\}) = \mathsf{L}\left(\bigcap_{n=1}^{\infty}\left[r, r+\frac{1}{n}\right)\right) = \lim_{n} \mathsf{L}\left(\left[r, r+\frac{1}{n}\right)\right) = \lim_{n} \frac{1}{n} = 0.$$

Поэтому

$$\mathsf{L}(Q_{[0;1)}) = \mathsf{L}\left(\bigcup_{r\in[0;1)} \{r\}\right) = \sum_{r\in[0;1)} \mathsf{L}(\{r\}) = \sum_{r\in[0;1)} 0 = 0,$$

$$\mathsf{L}(\overline{Q_{[0;1)}}) = \mathsf{L}([0;1)) - \mathsf{L}(Q_{[0;1)}) = 1.$$

Определение. Мера L, заданная на σ-алгебре $\mathfrak{B}^2$, значение которой на каждом прямоугольнике $[a_1, b_1) \times [a_2, b_2)$ совпадает с его площадью:

$$\mathsf{L}([a_1, b_1) \times [a_2, b_2)) = (b_2 - a_2)(b_1 - a_1),$$

называется *мерой Лебега* на $\mathbb{R}^2$.

Определение. Мера L, заданная на σ-алгебре $\mathfrak{B}^n$, значение которой на каждом параллелепипеде

$$[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$$

совпадает с его объемом:

$$\mathsf{L}([a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)) = \prod_{i=1}^{n} (b_i - a_i),$$

называется *мерой Лебега* на $\mathbb{R}^n$.

Построение меры Лебега на плоскости $\mathbb{R}^2$, в пространстве $\mathbb{R}^3$ и любом конечномерном пространстве $\mathbb{R}^n$ аналогично построению меры Лебега на прямой.

Мера Лебега на плоскости — обобщение площади, мера Лебега в пространстве — обобщение объема.

Меру Лебега, как правило, будем обозначать через L.

7.3 Вероятностные распределения на $\mathbb{R}^1$

Определение. Неотрицательную интегрируемую функцию[1] $p(t)$ на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$ такую, что

$$\int\limits_{\mathbb{R}^1} p(t)dt = 1,$$

будем называть *плотностью на* $\mathbb{R}^1$.

[1]Здесь под интегралом понимается интеграл в смысле Римана, в общем случае — в смысле Лебега (см. гл. 9).

Определение. Неотрицательную интегрируемую функцию $p(t_1, t_2, \ldots, t_n)$ на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$ такую, что

$$\int\limits_{\mathbb{R}^n} p(t_1, t_2, \ldots, t_n) d(t_1, t_2, \ldots, t_n) = 1,$$

будем называть *плотностью на* $\mathbb{R}^n$.

Плотности являются удобным средством для задания вероятностных распределений, поскольку каждая плотность $p(t)$ задает функцию распределения:

$$F(x) = \int\limits_{-\infty}^{x} p(t) dt, \ x \in \mathbb{R}^1,$$

а последняя — распределение.

Далее приводятся наиболее часто встречающиеся вероятностные распределения на $\mathbb{R}^1$, задаваемые плотностями.

1. Нормальное распределение. *Нормальным распределением с параметрами* $(a; \sigma^2)$ ($\mathsf{N}_{a;\sigma^2}$*-распределением, гауссовским распределением*) называют вероятностное распределение, задаваемое плотностью

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

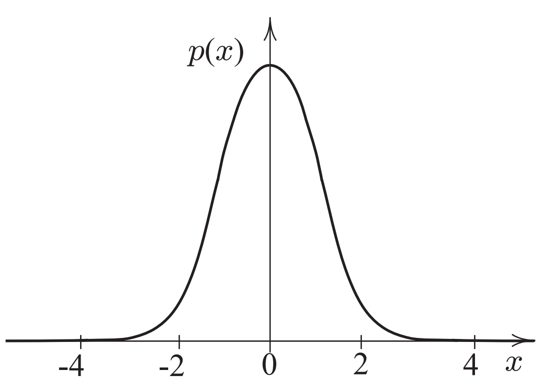

Рис. 7.3.1: График плотности нормального с параметрами $(0; 1)$ распределения

Убедимся в корректности определения. Для этого проверим, что

$$\int_{-\infty}^{\infty} p(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = 1.$$

Положив $\dfrac{x-a}{\sigma} = t$, получим:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

Обозначим

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{t^2}{2}\right\} dt = 2A,$$

тогда

$$\int_{0}^{\infty} \exp\left\{-\frac{u^2}{2}\right\} du = A, \ \int_{0}^{\infty} \exp\left\{-\frac{v^2}{2}\right\} dv = A,$$

$$A^2 = \int_{0}^{\infty} \exp\left\{-\frac{u^2}{2}\right\} du \int_{0}^{\infty} \exp\left\{-\frac{v^2}{2}\right\} dv =$$

$$= \int_{B} \exp\left\{-\frac{u^2+v^2}{2}\right\} d(u,v),$$

где $B = \{(u,v) : u \in [0,\infty), v \in [0,\infty)\}$.

Для вычисления последнего интеграла воспользуемся заменой $u = r\cos\varphi$, $v = r\sin\varphi$. Получим:

$$A^2 = \int_{C} \exp\left\{-\frac{r^2}{2}\right\} r d(\varphi, r) = \int_{0}^{\pi/2} d\varphi \int_{0}^{\infty} \exp\left\{-\frac{r^2}{2}\right\} r dr =$$

$$= \frac{\pi}{2} \int_{0}^{\infty} r \exp\left\{-\frac{r^2}{2}\right\} dr = \frac{\pi}{2} \int_{0}^{\infty} \exp\left\{-\frac{r^2}{2}\right\} d\left(\frac{r^2}{2}\right) = \frac{\pi}{2},$$

где $C = \{(\varphi, r) : \varphi \in [0, \pi/2],\ r \in [0, \infty)\}$. Отсюда

$$2A = \int\limits_{-\infty}^{\infty} \exp\left\{-\frac{t^2}{2}\right\} dt = \sqrt{2\pi}$$

(последний интеграл известен под названием интеграла Эйлера–Пуассона).

2. Гамма-распределение. Гамма-функция $\Gamma(\nu), \nu > 0$, определяется равенством

$$\Gamma(\nu) = \int\limits_0^{\infty} x^{\nu-1} e^{-x} dx.$$

Интегрируя по частям, получаем:

$$\Gamma(\nu) = (\nu - 1)\Gamma(\nu - 1),$$

в частности, для $n = 1, 2, \ldots$ имеем $\Gamma(n) = (n-1)!$

Определение. *Гамма-распределением* с параметрами $(\nu; \theta)$ $\nu > 0$, $\theta > 0$ (распределением $\mathsf{G}_{\nu;\theta}$) называется вероятностное распределение, задаваемое плотностью

$$p(x) = \begin{cases} \dfrac{\theta^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0. \end{cases}$$

Определение корректно, так как

$$\int\limits_{-\infty}^{\infty} p(x) dx = \int\limits_0^{\infty} \frac{\theta^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\theta x} dx = 1.$$

Определяемые далее показательное, Эрланга, χ^2 распределения являются частными случаями гамма-распределения.

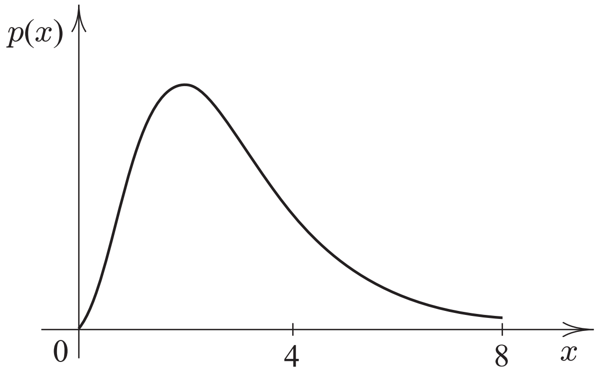

Рис. 7.3.2: График плотности гамма-распределения, $\nu > 1$

Показательное распределение. *Показательным распределением* с параметром θ, $\theta > 0$, называется вероятностное распределение, задаваемое плотностью

$$p(x) = \begin{cases} \theta \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Показательное распределение является гамма-распределением с параметрами $(1; \theta)$ (распределением $\mathsf{G}_{1;\theta}$).

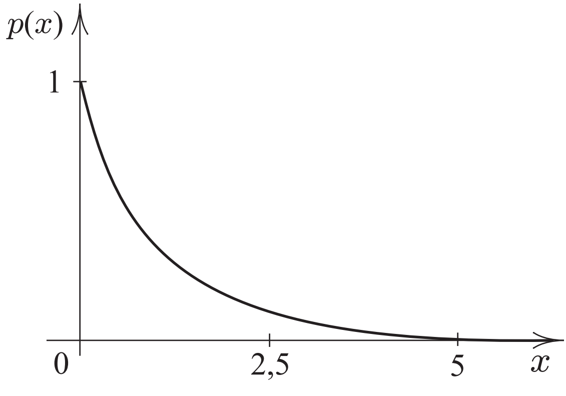

Рис. 7.3.3: График плотности показательного распределения,$\theta = 1$

Распределение Эрланга. *Распределением Эрланга* с параметрами $(m;\theta)$, $\theta > 0, m \in \mathbb{N}$, называется распределение, задаваемое плотностью

$$p(x) = \begin{cases} \dfrac{\theta^m}{(m-1)!} x^{m-1} \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Распределение Эрланга является гамма-распределением с параметрами $(m;\theta)$ (распределением $\mathsf{G}_{m;\theta}$).

χ^2-распределение. *χ^2-распределением* с n степенями свободы называется распределение, задаваемое плотностью

$$p(x) = \begin{cases} \dfrac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} \exp\left\{-\frac{1}{2}x\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

χ^2-распределение является гамма-распределением с параметрами $(n/2; 1/2)$ (распределением $\mathsf{G}_{n/2;1/2}$).

3. Двустороннее показательное распределение. *Двусторонним показательным распределением* с параметром θ, $\theta > 0$, будем называть распределение, задаваемое плотностью

$$p(x) = \frac{\theta}{2} \exp\{-\theta|x|\}.$$

Распределение Лапласа. *Распределением Лапласа* с параметрами (θ, b), $\theta > 0$, будем называть распределение, задаваемое плотностью

$$p(x) = \frac{\theta}{2} \exp\{-\theta|x-b|\}.$$

Распределение Лапласа — это сдвинутое двустороннее показательное распределение.

4. Равномерное распределение. *Равномерным распределением* на промежутке $[a,b]$ (распределением $\mathsf{U}_{a;b}$) называется вероятностное распределение, задаваемое плотностью

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & \text{если } x \in [a,b]; \\ 0, & \text{если } x \notin [a,b]. \end{cases}$$

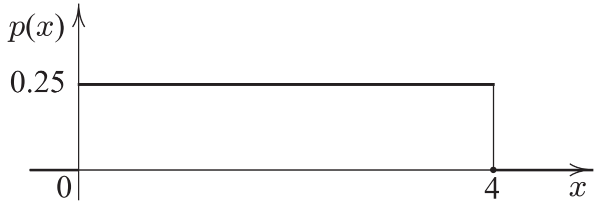

Рис. 7.3.4: График плотности равномерного на промежутке $[0; 4]$ распределения

Плотность равномерного на промежутке $[a, b]$ распределения удобно записывать в виде

$$p(x) = \frac{1}{b-a} I_{[a,b]}(x).$$

6. Распределение Коши. *Распределением Коши* с параметрами $(a; b)$, $a > 0$ (распределением $\mathsf{C}_{a;b}$) называется распределение, задаваемое плотностью

$$p(x) = \frac{1}{\pi} \frac{a}{a^2 + (x-b)^2}.$$

Распределением Коши с параметром a, $a > 0$, называется распределение $\mathsf{C}_{a;0}$, его плотность

$$p(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}.$$

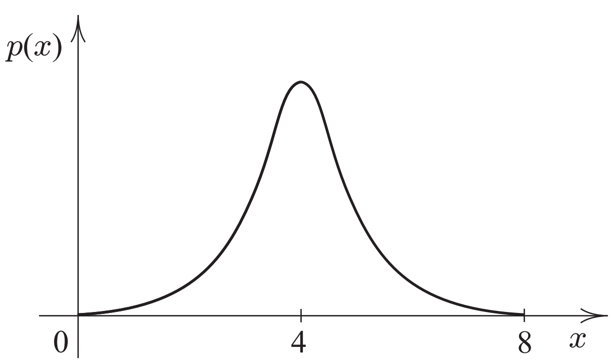

Рис. 7.3.5: График плотности распределения Коши

5. Бета-распределение. *Бета-распределением* с параметрами $(\mu;\nu)$, $\mu > 0$, $\nu > 0$, будем называть абсолютно непрерывное распределение, задаваемое плотностью

$$\beta_{\mu;\nu}(x) = \begin{cases} \dfrac{1}{B(\mu;\nu)} x^{\mu-1}(1-x)^{\nu-1}, & \text{если } x \in (0,1); \\ 0, & \text{если } x \overline{\in} (0,1), \end{cases}$$

где

$$B(\mu;\nu) = \int\limits_0^1 t^{\mu-1}(1-t)^{\nu-1} dt.$$

7. Логарифмически нормальное распределение. *Логарифмически нормальным распределением* с параметрами $(\mu;\sigma^2)$ называется вероятностное распределение, задаваемое плотностью

$$p(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma^2}x} \exp\left\{-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0. \end{cases}$$

Заметим, что

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_0^\infty \frac{1}{x} \exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\} dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^\infty \exp\left\{-\frac{t^2}{2}\right\} dt = 1$$

(вычисляя интеграл, воспользовались заменой $(\ln x-\mu)/\sigma = t$).

8. Распределение Парето. *Распределением Парето* с параметрами $(\lambda;\theta)$, $\lambda > 0$, $\theta > 2$, называется вероятностное распределение, задаваемое плотностью

$$p(x) = \begin{cases} \dfrac{\theta\lambda^\theta}{x^{\theta+1}}, & \text{если } x > \lambda; \\ 0, & \text{если } x \le \lambda. \end{cases}$$

7.4 Геометрические вероятности

Здесь мы построим математическую модель стохастического эксперимента, состоящего в бросании наудачу точки в множество B из $\mathbb{R}^n$, например, на отрезок $[a, b] \subset \mathbb{R}^1$, в прямоугольник $[a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$, в параллелепипед $[a_1, b_1] \times [a_2, b_2] \times \times [a_3, b_3] \subset \mathbb{R}^3$ и т. д. Наблюдаемыми событиями таких стохастических экспериментов являются события вида “точка попала в промежуток $[a', b'] \subset [a, b]$”, “точка попала в прямоугольник $[a'_1, b'_1] \times [a'_2, b'_2] \subset [a_1, b_1] \times [a_2, b_2]$” и т. д. Из опыта известно, что при бросании наудачу[2] точки на отрезок $[a, b]$ частота её попадания в промежуток $[a', b'] \subset [a, b]$ близка к

$$(b' - a')/(b - a) = \mathsf{L}([a', b'])/\mathsf{L}([a, b]),$$

при бросании наудачу точки в множество $B = [a_1, b_1] \times [a_2, b_2]$ частота её попадания в прямоугольник $B' = [a'_1, b'_1] \times [a'_2, b'_2]$ $(B' \subset B)$ близка к

$$(b'_1 - a'_1)(b'_2 - a'_1)/(b_1 - a_1)(b_2 - a_1) = \mathsf{L}(B')/\mathsf{L}(B)$$

и т. д., где, как обычно, $\mathsf{L}(\cdot)$ — мера Лебега. Поэтому для стохастического эксперимента, состоящего в бросании наудачу точки в множество B из $\mathbb{R}^n$ $(0 < \mathsf{L}(B) < \infty)$, в качестве пространства элементарных событий естественно рассмотреть множество $B \in \mathfrak{B}^n$, в качестве σ-алгебры событий — σ-алгебру $\mathfrak{B}^n_B$ борелевских подмножеств B (она образована множествами вида $B \cap A$,

---

[2] Стохастический эксперимент, состоящий в бросании наудачу точки, скажем, на отрезок $[0; 1]$ можно реализовать, например, так. Из таблицы случайных чисел выбираем число $\overline{abcd}$ (оно может быть и “длиннее”) и точку на отрезке $[0; 1]$ с координатой $a \cdot 10^{-1} + b \cdot 10^{-2} + c \cdot 10^{-3} + d \cdot 10^{-4}$ рассматриваем в качестве исхода стохастического эксперимента.

Этот стохастический эксперимент можно реализовать и так. Последовательно независимым образом подбрасываем симметричную монету. Выпадение герба обозначаем через 1, решетки через — 0. Пусть к примеру, монету подбрасываем четыре раза (можно подбрасывать и большее число раз) и 1101 — результат подбрасываний монеты. Точку на отрезке $[0; 1]$ с координатой $1 \cdot 2^{-1} + 1 \cdot 2^{-2} + 0 \cdot 2^{-3} + 1 \cdot 2^{-4}$ можно рассматривать как исход стохастического эксперимента, состоящего в бросании наудачу точки на отрезок $[0; 1]$.

Следует отметить, что фактически мы моделируем дискретное приближение равномерного распределения.

$A \in \mathfrak{B}^n$), а вероятность на $\mathfrak{B}^n_B$ естественно задать равенством

$$\mathsf{P}(\cdot) = \frac{\mathsf{L}(\cdot)}{\mathsf{L}(B)}. \qquad (7.4.1)$$

Вероятность, определенную равенством (7.4.1), называют *геометрической вероятностью.*

Далее в слова "точку наудачу бросают в множество B" мы будем вкладывать вполне конкретное содержание: брошенная точка может попасть в любое подмножество B' (борелевское) множества B и вероятность попадания в $B' \subset B$ пропорциональна его мере Лебега $\mathsf{L}(B')$. Формально это означает, что в качестве математической модели стохастического эксперимента, состоящего в бросании наудачу точки в множество B из $\mathbb{R}^n$, мы будем рассматривать вероятностное пространство $\{B, \mathfrak{B}^n_B, \mathsf{P}\}$, где P — геометрическая вероятность на $\mathfrak{B}^n_B$.

Пример 7.4.1 (задача Бюффона). *Плоскость разграфлена параллельными прямыми, отстоящими друг от друга на расстоянии* $2a$. *На плоскость наудачу*[3] *бросают иглу длиной* $2l$ ($2l < 2a$). *Найти вероятность того, что игла пересечет какую-нибудь прямую.*

Решение. Обозначим через x расстояние от середины иглы до ближайшей прямой, через φ — угол, составленный иглой с прямой (φ будем откладывать против часовой стрелки от направления иглы до направления прямой, см. рис. 7.4.1).

Тогда игла на плоскости, разграфленной параллельными прямыми, задает упорядоченную пару чисел (φ, x), $\varphi \in [0, \pi]$, $x \in [0, a]$. Та же пара чисел (φ, x) задает на координатной плоскости φ, x точку с координатами (φ, x), принадлежащую прямоугольнику $B = [0, \pi] \times [0, a]$ (см. рис. 7.4.2). И обратно: каждая точка из прямоугольника B определяет пару чисел (φ, x), а вместе с этой парой и положение иглы по отношению к параллельным прямым на плоскости. Поэтому бросание наудачу иглы длиной $2l$ на плоскость, разграфленную параллельными прямыми, отстоящими друг от друга на расстоянии $2a$, и регистрация ее положения по отношению к прямым равносильны бросанию наудачу точки в прямоугольник $B = [0, \pi] \times [0, a]$ (см. рис. 7.4.2).

[3] В задаче оборот "наудачу бросают иглу" означает следующее: 1° середина иглы (точка) равномерно распределена на отрезке длиной $2a$, перпендикулярном параллельным прямым; 2° угол φ, образованный иглой с параллельными прямыми, распределен равномерно на промежутке $[0, \pi]$; 3° случайные величины x — расстояние от центра иглы до ближайшей прямой — и φ являются независимыми (см. гл. 8).

Математической же моделью стохастического эксперимента, состоящего в бросании наудачу точки в прямоугольник B, является вероятностное пространство $\{B, \mathfrak{B}_B^2, \mathsf{P}\}$, где $B = [0, \pi] \times [0, a]$, $\mathfrak{B}_B^2$ — σ-алгебра борелевских множеств пространства B, P — геометрическая вероятность на $\mathfrak{B}_B^2$.

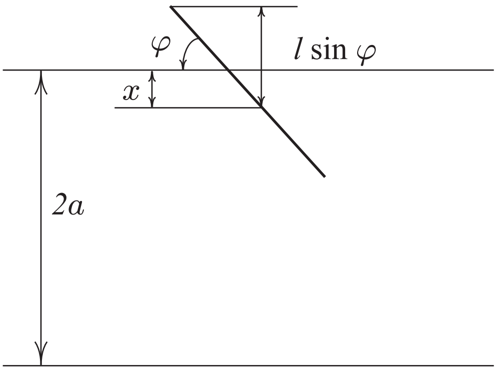

Рис. 7.4.1: Игла на разграфленной плоскости

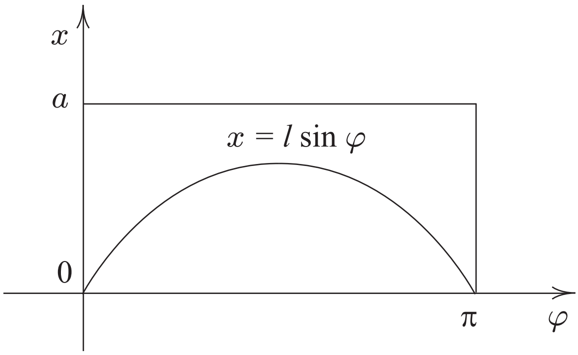

Рис. 7.4.2: К задаче Бюффона

Игла пересекает прямую тогда и только тогда, когда φ и x связаны соотношением $x \leq l \sin \varphi$ (см. рис. 7.4.1) или, что то же, брошенная в прямоугольник $B = [0, \pi] \times [0, a]$ точка попадает в область, ограниченную кривой $x = l \sin \varphi$ и осью $O\varphi$

(см. рис. 7.4.2). И поскольку точка бросается наудачу, то искомая вероятность вычисляется как геометрическая вероятность, т. е.

$$p = \frac{1}{a\pi}\int\limits_0^\pi l\sin\varphi d\varphi = \frac{2l}{a\pi}.$$

Интересно, что полученное соотношение можно использовать для экспериментального определения числа π.

Предположим, что игла брошена на плоскость n раз и пусть m — число пересечений иглы с прямой. При большом n частота m/n события "игла пересечет прямую" близка к его вероятности p, т. е. выполняется соотношение

$$\frac{2l}{a\pi} \approx \frac{m}{n},$$

откуда получаем:

$$\pi \approx \frac{2l}{a}\frac{n}{m}.$$

7.5 Прямое произведение вероятностных пространств

Рассмотрим пару независимых стохастических экспериментов — стохастический эксперимент 1° и стохастический эксперимент 2°, описываемых соответственно вероятностными пространствами $\left\{\Omega^{(1)}, \mathfrak{F}^{(1)}, \mathsf{P}^{(1)}\right\}$ и $\left\{\Omega^{(2)}, \mathfrak{F}^{(2)}, \mathsf{P}^{(2)}\right\}$, а вместе с ними стохастический эксперимент, состоящий в проведении пары независимых стохастических экспериментов 1° и 2°. Построим вероятностное пространство этого стохастического эксперимента

В качестве пространства элементарных событий пары стохастических экспериментов естественно рассмотреть декартово произведение $\Omega = \Omega^{(1)} \times \Omega^{(2)}$ — множество упорядоченных пар $\omega = \left(\omega^{(1)}, \omega^{(2)}\right)$, $\omega^{(1)} \in \Omega^{(1)}$, $\omega^{(2)} \in \Omega^{(2)}$, в качестве σ-алгебры событий — σ-алгебру, порожденную "прямоугольниками" $A^{(1)} \times A^{(2)}$ $(A^{(1)} \in \mathfrak{F}^{(1)}$, $A^{(2)} \in \mathfrak{F}^{(2)})$. Эту σ-алгебру будем обозначать $\mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}$ и называть произведением σ-алгебр $\mathfrak{F}^{(1)}$ и $\mathfrak{F}^{(2)}$.

Далее, коль скоро стохастические эксперименты 1° и 2° независимы, вероятность на σ-алгебре $\mathfrak{F}=\mathfrak{F}^{(1)}\times\mathfrak{F}^{(2)}$ необходимо задать так, чтобы события стохастического эксперимента 1° и стохастического эксперимента 2° были независимы. Оказывается, что так определить вероятность на σ-алгебре $\mathfrak{F}^{(1)}\times\mathfrak{F}^{(2)}$ можно и, более того, единственным образом.

Теорема (о произведении вероятностей). *Пусть* $\left\{\Omega^{(1)},\mathfrak{F}^{(1)},\mathsf{P}^{(1)}\right\}$ *и* $\left\{\Omega^{(2)},\mathfrak{F}^{(2)},\mathsf{P}^{(2)}\right\}$ *— вероятностные пространства. Существует, и притом единственная, вероятность* P, *заданная на* σ*-алгебре* $\mathfrak{F}=\mathfrak{F}^{(1)}\times\mathfrak{F}^{(2)}$ *подмножеств* $\Omega=\Omega^{(1)}\times\Omega^{(2)}$ *такая, что*

$$\mathsf{P}\left(A^{(1)}\times A^{(2)}\right)=\mathsf{P}^{(1)}\left(A^{(1)}\right)\cdot\mathsf{P}^{(2)}\left(A^{(2)}\right) \qquad (7.5.1)$$

для любых $A^{(1)}\in\mathfrak{F}^{(1)}$, $A^{(2)}\in\mathfrak{F}^{(2)}$.

(Без доказательства.)

Вероятность $\mathsf{P}(\cdot)$, заданную на σ-алгебре $\mathfrak{F}=\mathfrak{F}^{(1)}\times\mathfrak{F}^{(2)}$ равенством (7.5.1) будем обозначать $\mathsf{P}^{(1)}\times\mathsf{P}^{(2)}$ и называть *произведением вероятностей* $\mathsf{P}^{(1)}$ и $\mathsf{P}^{(2)}$.

Вероятностное пространство $\{\Omega,\mathfrak{F},\mathsf{P}\}$, построенное по $\left\{\Omega^{(1)},\mathfrak{F}^{(1)},\mathsf{P}^{(1)}\right\}$ и $\left\{\Omega^{(2)},\mathfrak{F}^{(2)},\mathsf{P}^{(2)}\right\}$ так, как это описано в теореме, называется *прямым произведением* вероятностных пространств $\left\{\Omega^{(1)},\mathfrak{F}^{(1)},\mathsf{P}^{(1)}\right\}$ и $\left\{\Omega^{(2)},\mathfrak{F}^{(2)},\mathsf{P}^{(2)}\right\}$.

Аналогичным образом определяется прямое произведение n вероятностных пространств.

Независимость событий и прямое произведение вероятностных пространств. События стохастического эксперимента 1° можно описывать не только множествами из σ-алгебры $\mathfrak{F}^{(1)}$, но и множествами из σ-алгебры $\mathfrak{F}^{(1)}\times\mathfrak{F}^{(2)}$. Действительно, обозначим через $\tilde{\mathfrak{F}}^{(1)}$ σ-алгебру цилиндрических множеств в $\Omega^{(1)}\times\Omega^{(2)}$ с основаниями в $\Omega^{(1)}$. В $\tilde{\mathfrak{F}}^{(1)}$ входят подмножества из $\Omega^{(1)}\times\Omega^{(2)}$ вида $A^{(1)}\times\Omega^{(2)}$, где $A^{(1)}\in\mathfrak{F}^{(1)}$. Ясно, что $\tilde{\mathfrak{F}}^{(1)}\subset\mathfrak{F}^{(1)}\times\mathfrak{F}^{(2)}$. Между σ-алгебрами $\mathfrak{F}^{(1)}$ и $\tilde{\mathfrak{F}}^{(1)}$ устанавливается естественное взаимно однозначное соответствие: каждому $A^{(1)}\in\mathfrak{F}^{(1)}$ ставится в соответствие цилиндрическое множество $\tilde{A}^{(1)}=A^{(1)}\times\Omega^{(2)}$ из $\tilde{\mathfrak{F}}^{(1)}$ с основанием $A^{(1)}$. Поэтому цилиндрическими множествами $\tilde{A}^{(1)}=A^{(1)}\times\Omega^{(2)}$ можно описывать события стохастического эксперимента 1°, а именно $\tilde{A}^{(1)}=A^{(1)}\times\Omega^{(2)}\in\mathfrak{F}^{(1)}\times\mathfrak{F}^{(2)}$ описывает событие $A^{(1)}\in\mathfrak{F}^{(1)}$.

Аналогично определяется σ-алгебра $\tilde{\mathfrak{F}}^{(2)}$ и устанавливается взаимно однозначное соответствие между $\mathfrak{F}^{(2)}$ и $\tilde{\mathfrak{F}}^{(2)}$.

Вероятность P на σ-алгебре $\mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}$ определена как произведение вероятностей $\mathsf{P}^{(1)}$ и $\mathsf{P}^{(2)}$ (см. (7.5.1)) поэтому

$$\mathsf{P}\left(\tilde{A}^{(1)}\right) = \mathsf{P}\left(A^{(1)} \times \Omega^{(2)}\right) = \mathsf{P}^{(1)}\left(A^{(1)}\right) \cdot \mathsf{P}^{(2)}\left(\Omega^{(2)}\right) =$$

$$= \mathsf{P}^{(1)}\left(A^{(1)}\right) \cdot 1 = \mathsf{P}^{(1)}\left(A^{(1)}\right).$$

Аналогично

$$\mathsf{P}\left(\tilde{A}^{(2)}\right) = \mathsf{P}^{(2)}\left(A^{(2)}\right).$$

Теорема. *Пусть* $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ *— прямое произведение вероятностных пространств* $\left\{\Omega^{(1)}, \mathfrak{F}^{(1)}, \mathsf{P}^{(1)}\right\}$ *и* $\left\{\Omega^{(2)}, \mathfrak{F}^{(2)}, \mathsf{P}^{(2)}\right\}$*, тогда для любых* $A^{(1)} \in \mathfrak{F}^{(1)}$ *и* $A^{(2)} \in \mathfrak{F}^{(2)}$ *события* $\tilde{A}^{(1)}$ *и* $\tilde{A}^{(2)}$ *из* $\mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}$ *независимы в* $\{\Omega, \mathfrak{F}, \mathsf{P}\}$.

Действительно,

$$\mathsf{P}\left(\tilde{A}^{(1)} \cap \tilde{A}^{(2)}\right) = \mathsf{P}\left(\left(A^{(1)} \times \Omega^{(2)}\right) \cap \left(\Omega^{(1)} \times A^{(2)}\right)\right) =$$

$$= \mathsf{P}\left(A^{(1)} \times A^{(2)}\right) = \mathsf{P}^{(1)}\left(A^{(1)}\right) \cdot \mathsf{P}^{(2)}\left(A^{(2)}\right) = \mathsf{P}\left(\tilde{A}^{(1)}\right) \cdot \mathsf{P}\left(\tilde{A}^{(2)}\right).$$

Вероятностное пространство $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ является математической моделью пары независимых стохастических экспериментов. В таком вероятностном пространстве события, связанные с первым стохастическим экспериментом и со вторым, всегда независимы.

Пример. Стохастический эксперимент 1° состоит в подбрасывании симметричной монеты и регистрации числа выпавших гербов, описывается вероятностным пространством $\left\{\Omega^{(1)}, \mathsf{P}^{(1)}\right\}$, где $\Omega^{(1)} = \{0, 1\}$, $\mathsf{P}^{(1)}(i) = 1/2$, $i = 0, 1$. Стохастический эксперимент 2° состоит в подбрасывании симметричной игральной кости и регистрации числа выпавших очков, описывается вероятностным пространством $\left\{\Omega^{(2)}, \mathsf{P}^{(2)}\right\}$, где $\Omega^{(2)} = \{1, 2, \ldots, 6\}$, $\mathsf{P}^{(2)}(j) = 1/6$, $j = 1, 2, \ldots, 6$. Пара независимых стохастических экспериментов 1° и 2° описывается вероятностным пространством $\{\Omega, \mathsf{P}\}$, где $\Omega = \Omega^{(1)} \times \Omega^{(2)}$, $\mathsf{P} = \mathsf{P}^{(1)} \times \mathsf{P}^{(2)}$, подробнее, Ω состоит из пар (i, j), $i = 0, 1; j = 1, 2, \ldots, 6$; $\mathsf{P}(i, j) = \mathsf{P}^{(1)}(i) \times \mathsf{P}^{(2)}(j) = (1/2) \cdot (1/6) = 1/12$, $i = 0, 1; j = 1, 2, \ldots, 6$

7.6 Примеры и задачи

Примеры

Пример 7.6.1. *На отрезок* $[0;1]$ *наудачу бросают пару точек. Вычислить вероятность того, что расстояние между ними меньше* $1/2$.

Решение. Точки будем считать различимыми (например, они окрашены в разные цвета). Пусть x и y — координаты этих точек.

На отрезке $[0;1]$ упорядоченной паре точек с координатами x и y соответствует одна точка в квадрате $[0;1]\times[0;1]$ с координатами (x,y) и наоборот. Поэтому бросание наудачу пары точек на отрезок $[0;1]$ равносильно бросанию наудачу одной точки в квадрат $[0;1]\times[0;1]$. При этом парам точек на отрезке $[0;1]$, для которых расстояние меньше $1/2$ (т. е. $|y-x|<1/2$), соответствует множество точек (x,y) квадрата $[0;1]\times[0;1]$, координаты которых удовлетворяют соотношению $|y-x|<1/2$, и наоборот. Обозначим это множество через A:

$$A=\{(x,y)\in[0;1]\times[0;1]:|y-x|<1/2\}=$$

$$=\{(x,y)\in[0,1]\times[0,1]:x-1/2<y<x+1/2\}$$

(см. рис. 7.6.1, множество A заштриховано).

Вероятность того, что расстояние между брошенными наудачу на отрезок $[0;1]$ точками будет меньше $1/2$, равна вероятности попадания точки брошенной наудачу в квадрат $[0;1]\times[0;1]$ в множество A.

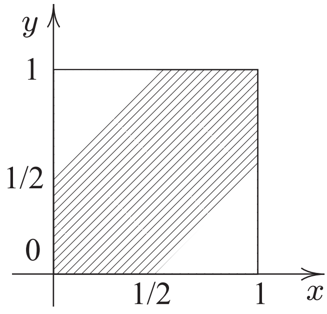

Рис. 7.6.1: К примеру 7.6.1

Математической моделью стохастического эксперимента, состоящего в бросании наудачу точки в квадрат $[0;1] \times [0;1]$, является вероятностное пространство $\{B, \mathfrak{B}_B^2, \mathsf{P}\}$, где $B = [0;1]\times$ $\times[0;1]$, $\mathfrak{B}_B^2$ — класс (σ-алгебра) борелевских подмножеств квадрата $[0;1] \times [0;1]$, P — геометрическая вероятность на $\mathfrak{B}_B^2$. Поэтому

$$\mathsf{P}(A) = \frac{\mathsf{L}(A)}{\mathsf{L}(B)} = \frac{3/4}{1} = \frac{3}{4}.$$

Пример 7.6.2. *Стержень длиной* 1 *наудачу разламывают на три части. Вычислить вероятность того, что из образовавшихся частей можно построить треугольник.*

Решение. Стержень длиной 1 удобно интерпретировать как отрезок $[0;1]$ на координатной прямой. Разламывать отрезок $[0;1]$ на три части будем так: бросим наудачу на отрезок две точки и разломаем его в этих точках (точки будем считать различимыми, их координаты обозначим x, y). При этом образовалось три отрезка: крайний левый (длиной $\min\{x,y\}$), средний (длиной $|x-y|$), крайний правый (длиной $1-\max\{x,y\}$). Из полученных отрезков можно построить треугольник, если длина каждого из них меньше суммы длин других:

$$\begin{cases} \min\{x,y\} < |x-y| + (1-\max\{x,y\}); \\ |x-y| < \min\{x,y\} + (1-\max\{x,y\}); \\ 1-\max\{x,y\} < \min\{x,y\} + |x-y|. \end{cases} \tag{7.6.1}$$

Бросание пары точек на отрезок $[0;1]$ равносильно бросанию одной точки в квадрат $[0;1]\times[0;1]$ (см. также пример 7.6.1). При этом парам точек на отрезке $[0;1]$, координаты которых удовлетворяют неравенствам (7.6.1), соответствуют точки (x,y) квадрата $[0;1]\times[0;1]$, координаты которых удовлетворяют тем же соотношениям. Обозначим это множество точек через A (см. также рис. 7.6.2, множество A заштриховано). Его удобно представить в виде:

$$A = A \cap (\{(x,y) : x \le y\} \cup \{(x,y) : x > y\}) =$$

$$= (A \cap \{(x,y) : x \le y\}) \cup (A \cap \{(x,y) : x > y\}) =$$

$$= \{(x,y) : x < 1/2,\ x + 1/2 > y,\ y > 1/2\} \cup$$

$$\cup \{(x,y) : y < 1/2,\ x - 1/2 < y,\ x > 1/2\}.$$

Поэтому вычисление вероятности построения треугольника из частей отрезка сводится к вычислению вероятности попадания точки, брошенной в квадрат $[0;1] \times [0;1]$, в множество A (см. рис. 7.6.2). И поскольку точку бросают наудачу, то искомая вероятность вычисляется как геометрическая вероятность:

$$\mathsf{P}(A) = \frac{\mathsf{L}(A)}{\mathsf{L}([0;1] \times [0;1])} = \frac{1/4}{1} = \frac{1}{4}.$$

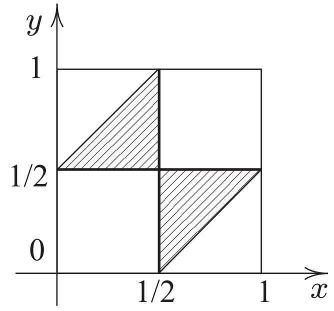

Рис. 7.6.2: К примеру 7.6.2

Пример 7.6.3 . *На окружности наудачу выбирают три точки. Какова вероятность того, что треугольник с вершинами в этих точках остроугольный?*

Решение. При любом повороте окружности вероятность события “треугольник остроугольный” остается неизменной, поэтому можно считать, что из трех вершин A, B, C одна, например C, фиксирована. Две другие выбирают наудачу. Будем задавать положения точек A и B относительно точки C величинами дуг $CA = \alpha$ и $CB = \beta$ (см. рис. 7.6.3 откладывая их против движения часовой стрелки, дуги будем измерять в радианах. Треугольник ABC остроугольный (при $\alpha < \beta$), если каждая из дуг $CA = \alpha, AB = \beta - \alpha, BC = 2\pi - \beta$ меньше π, т. е.

$$\beta > \alpha,\ \alpha < \pi,\ \beta - \alpha < \pi,\ \beta > \pi, \tag{7.6.2}$$

или (при $\beta < \alpha$) каждая из дуг $CB = \beta, BA = \alpha - \beta, CA = 2\pi - \alpha$ меньше π, т. е.

$$\beta < \alpha,\ \beta < \pi,\ \alpha - \beta < \pi,\ \alpha > \pi. \tag{7.6.3}$$

Далее, упорядоченной паре чисел (α, β), $0 < \alpha < 2\pi$, $0 < \beta < 2\pi$ соответствует одна точка квадрата $[0, 2\pi] \times [0, 2\pi]$ на плоскости с осями $O\alpha$ и $O\beta$ и наоборот. А следовательно, множеству дуг α, β, которые удовлетворяют одному из условий: (7.6.2) или (7.6.3), соответствует множество точек квадрата $[0, 2\pi] \times [0, 2\pi]$, координаты которых удовлетворяют одному из условий: (7.6.2) или (7.6.3). Это множество точек квадрата имеет вид, аналогичный изображенному на рис. 7.6.2.

Искомая вероятность равна $\dfrac{2\pi^2/2}{(2\pi)^2} = \dfrac{1}{4}$.

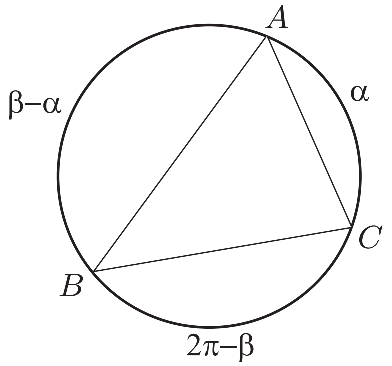

Рис. 7.6.3: К примеру 7.6.3

Задачи

7.1. Пусть $\mathfrak{A}$ — алгебра множеств на $\mathbb{R}^1$, элементами которой являются конечные объединения непересекающихся промежутков вида $[a, b)$; $a, b \in \mathbb{R}^1$, a и b могут принимать значения $+\infty$, $-\infty$. Доказать, что σ-алгебра, порожденная алгеброй $\mathfrak{A}$, является σ-алгеброй борелевских множеств на $\mathbb{R}^1$.

Указание. См. пример 7.2.1.

7.2. Пусть $\mathfrak{B}^1$ — σ-алгебра борелевских множеств на $\mathbb{R}^1$, B — произвольное, но фиксированное множество из $\mathfrak{B}^1$. Обозначим через $\mathfrak{B}^1_B$ класс множеств вида $B \cap A$, где $A \in \mathfrak{B}^1$. Доказать, что $\mathfrak{B}^1_B$ является σ-алгеброй.

Примечание. Дополнение берется до множества B.

7.3. На отрезок $[-2; 2]$ наудачу бросают пару точек. Пусть x — координата одной, y — другой. Найти вероятности таких событий:

1) $x+|x|=y+|y|$; 8) $(y-2x)(y+2x)\geq 0$;
2) $x-|x|=y-|y|$; 9) $(|x|+|y|-1)(|x|+|y|-2)\leq 0$;
3) $[y]=[x]$; 10) $|x-1|+|y-1|\leq 1$;
4) $[y]=-[x]$; 11) $(y-x)(y+x-2)\geq 0$;
5) $[y]=[x-1]$; 12) $([y]-[x])(\{y\}-\{x\})\geq 0$;
6) $\{y\}\leq\{x\}$; 13) $(x-\operatorname{sign}x)^2+(y-\operatorname{sign}y)^2\leq 1$;
7) $|x|+|y|\leq 1$; 14) $|x-\operatorname{sign}x|+|y-\operatorname{sign}y|\leq 1$;
15) $(|x-1|+|y-1|-1)(|x-1|+|y-1|-2)\leq 0$;
16) $((x-\operatorname{sign}x)^2+(y-\operatorname{sign}y)^2-1)(|x-\operatorname{sign}x|+$
$+|y-\operatorname{sign}y|-1)\leq 0$.

Здесь $[a]$ — целая часть числа a, $\{a\}=a-[a]$ — дробная часть числа a.

Ответы: 1) 1/4; 2) 1/4; 3) 1/4; 4) 3/16; 5) 3/16; 6) 1/2; 7) 1/8; 8) 1/4; 9) 3/8; 10) 1/8; 11) 1/2; 12) 5/8; 13) $\pi/4$; 14) 1/2; 15) 1/4; 16) $(\pi-2)/4$.

7.4. На отрезке $[-1;1]$ наудачу выбирают две точки. Пусть p и q — координаты этих точек. Найти вероятность того, что квадратное уравнение $x^2+px+q=0$: 1) имеет действительные корни; 2) не имеет действительных корней.

Ответ: 13/24, 11/24.

7.5. На отрезок наудачу бросают три точки одну за другой. Какова вероятность того, что третья по счету точка окажется между двумя первыми?

Ответ: 1/3.

7.6. На отрезке $[0;1]$ последовательно случайным образом выбирают три числа. Какова вероятность того, что:

1) выбранное последним число наибольшее;

2) числа идут в порядке возрастания?

Ответы: 1) 1/3; 2)1/6.

7.7. Отрезок разделен на три равные части. Какова вероятность того, что три точки, наудачу брошенные на отрезок, попадут в три разные части?

7.8. На окружности наудачу выбраны три точки. Вычислить вероятность того, что в треугольнике с вершинами в этих точках:

1) имеется угол, меньший 30°;

2) все углы больше 30°;

3) имеется угол 90°.

Ответы: 1) 3/4; 2) 1/4.

7.9. На отрезок $[0,nd]$, разбитый точками kd, $k=0,1,\ldots,n$, на части, наудачу бросают отрезок случайной длины. Вычислить вероятность того, что отрезок “накроет” одну из точек деления, если его длина 1° распределена равномерно на $[0,d]$, 2° распределена равномерно на $[0,l]$ $(0<l<d)$.

7.9.[4] На окружности наудачу выбирают точки: A,B,C,D. Вычислить вероятность того, что отрезки AC и BD пересекаются.

Указание. Зафиксируем одну из точек, например, A. Тогда положение других можно описать дугами AB, AC, AD, отложенными против движения часовой стрелки.

Ответ: $1/3$.

7.10. Чтобы собрать шарикоподшипник, радиус R внешнего кольца, радиус r внутреннего кольца и диаметр d шарика должны удовлетворять условиям

$$0 \le R-r-d \le \delta.$$

Предположим, что R,r,d — независимые и равномерно распределенные соответственно на отрезках [50,0; 51,0], [40,0; 41,0], [9,5; 10,0] случайные величины.

Найти вероятность того, что шарикоподшипник будет собран, если $\delta = 0{,}5$.

7.11. Два судна должны подойти к одному причалу. Моменты прихода суден к причалу — независимые события, равновозможные в течение суток. Найти вероятность того, что одному из суден придется ждать освобождения причала, если время стоянки первого судна (судна номер 1) — один час, а второго (судна номер 2) — два часа.

[4]Васильев. Н. Б. Геометрические вероятности // Квант. — 1991. — N 1, с. 47–53.

Глава 8

Случайная величина и её распределение

8.1 Случайная величина

Прообразы множеств. Напомним определение прообраза множества. Пусть X и Y — произвольные множества; f — отображение X в Y, ставящее в соответствие каждому элементу $x \in X$ элемент $y \in Y$, который будем обозначать $f(x)$ $(f : x \to y = f(x))$.

Прообразом множества B, $B \subset Y$, при отображении f будем называть совокупность точек x множества X, для которых $f(x) \in B$. Обозначается прообраз множества B при отображении f символом $f^{-1}(B)$.

Пусть B_λ, $\lambda \in \Lambda$, — семейство подмножеств множества Y Имеют место следующие легко проверяемые соотношения:

$$f^{-1}\Big(\bigcup_\lambda B_\lambda\Big) = \bigcup_\lambda f^{-1}(B_\lambda),$$

$$f^{-1}\Big(\bigcap_\lambda B_\lambda\Big) = \bigcap_\lambda f^{-1}(B_\lambda),$$

$$f^{-1}(\overline{B}) = \overline{f^{-1}(B)}.$$

Случайная величина как функция на пространстве элементарных событий. Рассматривая стохастические эксперименты с исходами в $\mathbb{R}^n$ (случайные величины), мы пришли к двум их математическим моделям (см. параграф 5.1 гл. 5):

1° вероятностному распределению на $\mathbb{R}^n$;

2° функции со значениями в $\mathbb{R}^n$, заданной на пространстве элементарных событий Ω некоторого вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$.

Вообще говоря, не любая функция $\xi = \xi(\omega)$ на Ω может рассматриваться как модель случайной величины. Приведем соображения, которые мотивируют ограничения на функцию $\xi = \xi(\omega)$, заданную на Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, позволяющие рассматривать $\xi = \xi(\omega)$ в качестве математической модели стохастического эксперимента с исходами в $\mathbb{R}^n$.

Первый вопрос, возникающий при изучении стохастических экспериментов с исходами ξ в $\mathbb{R}^n$: "Какова вероятность попадания ξ в данное множество B, например, в параллелепипед $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$ (в одномерном случае — в промежуток $[a, b]$, в двумерном в прямоугольник $[a_1, b_1] \times [a_2, b_2]$)?" Если мы рассматриваем ξ как функцию от ω, т. е. $\xi = \xi(\omega)$, $\omega \in \Omega$, то ξ попадает в B тогда и только тогда, когда ω попадает в множество $\{\omega : \xi(\omega) \in B\} = \xi^{-1}(B)$ из Ω. И, следовательно, чтобы можно было говорить о вероятности попадания ξ в B, множество $\{\omega : \xi(\omega) \in B\} = \xi^{-1}(B)$ должно принадлежать классу тех подмножеств из Ω, на котором задана вероятность, т. е. должно принадлежать σ-алгебре $\mathfrak{F}$:

$$\xi^{-1}(B) \in \mathfrak{F}.$$

Так что в качестве математической модели стохастического эксперимента с исходами в $\mathbb{R}^n$ естественно рассматривать такую функцию $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$ на пространстве элементарных событий Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, для которой

$$\xi^{-1}(B) \in \mathfrak{F},$$

на достаточно широком классе подмножеств B из $\mathbb{R}^n$, содержащем, в частности, параллелепипеды $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$ (в одномерном случае — промежутки $[a, b]$, в двумерном — прямоугольники $[a_1, b_1] \times [a_2, b_2]$). Такие функции мы будем называть случайными величинами.

Определение. *Случайной величиной* со значениями в $\mathbb{R}^n$ будем называть функцию $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ со значениями в $\mathbb{R}^n$, заданную на пространстве элементарных событий Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, такую, что для каждого $B \in \mathfrak{B}^n$

$$\xi^{-1}(B) \in \mathfrak{F}. \tag{8.1.1}$$

Если $n > 1$, то случайную величину $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ называют многомерной, если $n = 2$ — двумерной, если $n = 1$ — одномерной (скалярной) или просто случайной величиной.

Для часто встречающегося одномерного случая имеем.

Определение. *Случайной величиной* со значениями в $\mathbb{R}^1$ будем называть функцию $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^1$, заданную на пространстве элементарных событий Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, такую, что для каждого множества $B \in \mathfrak{B}^1$

$$\xi^{-1}(B) \in \mathfrak{F}.$$

Рисунок 8.1.1 иллюстрирует определение случайной величины со значениями в $\mathbb{R}^1$.

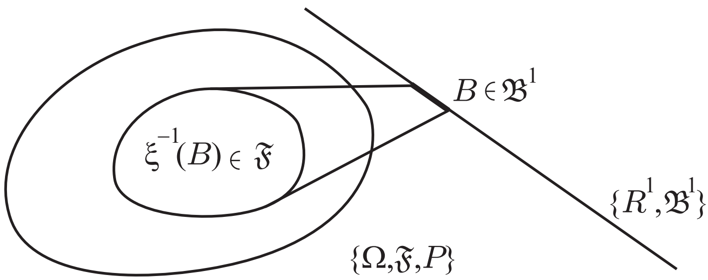

Рис. 8.1.1: К определению случайной величины

В качестве $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, в частности, может рассматриваться и $\left\{\mathbb{R}^l, \mathfrak{B}^l, \mathsf{P}\right\}$. Для этого частного, но важного случая имеем.

Определение. *Борелевской случайной величиной* (борелевской функцией) со значениями в $\mathbb{R}^n$ будем называть функцию $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ со значениями в $\mathbb{R}^n$, заданную на пространстве элементарных событий $\mathbb{R}^l$ вероятностного пространства $\left\{\mathbb{R}^l, \mathfrak{B}^l, \mathsf{P}\right\}$, такую, что для каждого $B \in \mathfrak{B}^n$

$$\xi^{-1}(B) \in \mathfrak{B}^l.$$

Рисунок 8.1.2 иллюстрирует определение борелевской случайной величины (борелевской функции) со значениями в $\mathbb{R}^1$, заданной на $\mathbb{R}^1$.

Замечание 1. Случайная величина $\xi = \xi(\omega)$ — это функция со значениями в $\mathbb{R}^n$ на пространстве элементарных событий Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, для которой выполняется соотношение (8.1.1). В анализе такие функции называют $\mathfrak{F}$-измеримыми или измеримыми относительно σ-алгебры $\mathfrak{F}$.

Замечание 2. В определении случайной величины $\xi = \xi(\omega)$, заданной на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, вероятность P никак не участвует, поэтому можно говорить о случайной величине, заданной на измеримом пространстве $\{\Omega, \mathfrak{F}\}$.

Замечание 3. В дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\} = \{\Omega, \mathfrak{F}, \mathsf{P}\}$ σ-алгебру событий образует σ-алгебра $\mathfrak{F}$ всех подмножеств пространства Ω. Поэтому для любой функции $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$, заданной на дискретном Ω,

$$\xi^{-1}(B) \in \mathfrak{F}, B \in \mathfrak{B}^n.$$

Так что любая функция со значениями в $\mathbb{R}^n$ на дискретном вероятностном пространстве $\{\Omega, \mathsf{P}\}$ является случайной величиной.

Напомним, что случайную величину, принимающую не более чем счетное число значений, называют *дискретной.*

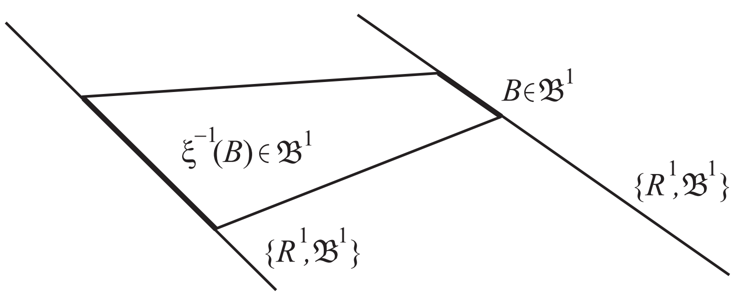

Рис. 8.1.2: К определению борелевской случайной величины ($n = 1,\ l = 1$)

Замечание о терминологии. Обычно вместо "случайная величина ξ задана на пространстве элементарных событий Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$", говорят "$\xi$ задана на вероятностном пространстве $\{\Omega, \mathfrak{F}, \mathsf{P}\}$".

Эквивалентные определения случайной величины. Функция $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$, заданная на пространстве элементарных событий Ω вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, является случайной величиной, если для каждого $B \in \mathfrak{B}^n$ его прообраз $\xi^{-1}(B)$ принадлежит σ-алгебре $\mathfrak{F}$. Но оказывается, для того чтобы убедиться, что функция $\xi = \xi(\omega)$ является случайной величиной, не обязательно проверять принадлежность $\xi^{-1}(B)$ σ-алгебре $\mathfrak{F}$ для каждого B из $\mathfrak{B}^n$, можно обойтись "более узким" классом подмножеств из $\mathbb{R}^n$ — классом $\mathfrak{M}$ порождающим σ-алгебру $\mathfrak{B}^n$.

Лемма 8.1.1 (о $\boldsymbol{\sigma}$-алгебре $\boldsymbol{\xi(\mathfrak{F})}$). *Пусть $\xi = \xi(\omega)$ — функция со значениями в $\mathbb{R}^n$, заданная на Ω измеримого пространства $\{\Omega, \mathfrak{F}\}$. Класс $\mathfrak{S}$ подмножеств B из $\mathbb{R}^n$, для которых*

$$\xi^{-1}(B) \in \mathfrak{F},$$

образует σ-алгебру.

Доказательство. Проверим, что $\mathfrak{S}$ является σ-алгеброй.

1. Пусть $B \in \mathfrak{S}$, это означает, что $\xi^{-1}(B) \in \mathfrak{F}$. Тогда и

$$\overline{B} \in \mathfrak{S},$$

поскольку

$$\xi^{-1}(\overline{B}) = \overline{\xi^{-1}(B)} \in \mathfrak{F}.$$

2. Пусть $B_k \in \mathfrak{S}, k = 1, 2, \ldots$, это равносильно тому, что $\xi^{-1}(B_k) \in \mathfrak{F},\ k = 1, 2, \ldots$ Тогда и

$$\bigcup_{k=1}^{\infty} B_k \in \mathfrak{S},$$

поскольку

$$\xi^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right) = \bigcup_{k=1}^{\infty} \xi^{-1}(B_k) \in \mathfrak{F}.$$

Так что $\mathfrak{S}$ является σ-алгеброй. σ-Алгебру $\mathfrak{S}$ еще обозначают $\xi(\mathfrak{F})$.

Теорема 8.1.1 *Для того чтобы функция $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$, заданная на измеримом пространстве $\{\Omega, \mathfrak{F}\}$, была случайной величиной, достаточно, чтобы для множеств $B \subset \mathbb{R}^n$ из класса $\mathfrak{M}$, порождающего σ-алгебру $\mathfrak{B}^n$,*

$$\xi^{-1}(B) \in \mathfrak{F}.$$

Доказательство. Пусть $\xi = \xi(\omega)$ — функция на $\{\Omega, \mathfrak{F}\}$ со значениями в $\mathbb{R}^n$ такая, что для подмножеств $B \subset \mathbb{R}^n$ из класса $\mathfrak{M}$, порождающего $\mathfrak{B}^n$, их прообразы $\xi^{-1}(B)$ принадлежат $\mathfrak{F}$. Покажем, что тогда $\xi^{-1}(B) \in \mathfrak{F}$ и для множеств B из $\mathfrak{B}^n$ (последнее будет обозначать, что $\xi = \xi(\omega)$ является случайной величиной).

Обозначим через $\mathfrak{S}$ класс подмножеств B пространства $\mathbb{R}^n$, для которых $\xi^{-1}(B) \in \mathfrak{F}$ (согласно лемме 8.1.1 класс $\mathfrak{S}$ является σ-алгеброй). Далее, по условию теоремы для каждого $B \in \mathfrak{M}$ его прообраз $\xi^{-1}(B) \in \mathfrak{F}$, поэтому

$$\mathfrak{M} \subset \mathfrak{S}.$$

Отсюда

$$\mathfrak{B}^n = \sigma(\mathfrak{M}) \subset \mathfrak{S}$$

($\sigma(\mathfrak{M}) \subset \mathfrak{S}$, так как $\sigma(\mathfrak{M})$ — наименьшая σ-алгебра, содержащая $\mathfrak{M}$, а $\mathfrak{B}^n = \sigma(\mathfrak{M})$ по условию теоремы). И поскольку класс $\mathfrak{S}$ состоит из таких B, для которых $\xi^{-1}(B) \in \mathfrak{F}$, а $\mathfrak{B}^n \subset \mathfrak{S}$, то, в частности, для каждого $B \in \mathfrak{B}^n$

$$\xi^{-1}(B) \in \mathfrak{F}.$$

Последнее, по определению, означает, что $\xi = \xi(\omega)$ является случайной величиной.

В приведенном далее следствии перечислены некоторые конкретные классы, часто используемые в качестве $\mathfrak{M}$.

Следствие (об эквивалентных определениях случайной величины). *Для того чтобы функция $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^1$, заданная на измеримом пространстве $\{\Omega, \mathfrak{F}\}$, была случайной величиной, достаточно, чтобы для множеств B из класса $\mathfrak{M}_i$ их прообразы $\xi^{-1}(B)$ принадлежали $\mathfrak{F}$, $i = 1, 2, \ldots, 8$, где*

$$\mathfrak{M}_1 = \left\{(-\infty, x),\ x \in \mathbb{R}^1\right\},\ \mathfrak{M}_2 = \left\{(-\infty, x],\ x \in \mathbb{R}^1\right\},$$

$$\mathfrak{M}_3 = \left\{(x, +\infty),\ x \in \mathbb{R}^1\right\},\ \mathfrak{M}_4 = \left\{[x, +\infty),\ x \in \mathbb{R}^1\right\},$$

$$\mathfrak{M}_5 = \left\{(a, b),\ a, b \in \mathbb{R}^1\right\},\ \mathfrak{M}_6 = \left\{[a, b),\ a, b \in \mathbb{R}^1\right\},$$

$$\mathfrak{M}_7 = \left\{(a, b],\ a, b \in \mathbb{R}^1\right\},\ \mathfrak{M}_8 = \left\{[a, b],\ a, b \in \mathbb{R}^1\right\}.$$

Д о к а з а т е л ь с т в о. Достаточно заметить, что каждый из перечисленных классов порождает σ-алгебру борелевских множеств.

8.2 Функция от случайной величины

Теорема. *Непрерывная функция $\varphi = \varphi(x)$ на $\mathbb{R}^l$ со значениями в $\mathbb{R}^1$ является борелевской.*

Доказательство. Функция $\varphi = \varphi(x)$ на $\mathbb{R}^l$ непрерывна, поэтому для каждого открытого множества, его прообраз — открытое, а следовательно, и борелевское множество, в частности, прообраз $\varphi^{-1}((a,b))$ открытого промежутка (a,b) борелевское множество. Класс $\mathfrak{M} = \left\{(a,b) : a,b \in \mathbb{R}^1\right\}$ открытых промежутков на $\mathbb{R}^1$ порождает σ-алгебру $\mathfrak{B}^1$. Отсюда в силу следствия из теоремы 8.1.1 получаем, что $\varphi = \varphi(x)$ является борелевской функцией на $\mathbb{R}^l$ со значениями в $\mathbb{R}^1$.

Утверждение имеет место и для непрерывной на $\mathbb{R}^l$ функции со значениями в $\mathbb{R}^n$: *непрерывная функция на $\mathbb{R}^l$ со значениями в $\mathbb{R}^n$ является борелевской.*

Так что борелевскими, к примеру, являются функции x^+, x^-, $|x|$, $\sin x$, $|x+y|$, $\cos(x+y)+e^z$, e^{xy}.

Теорема 8.2.1 (о функции от случайной величины). *Пусть $\xi = \xi(\omega)$ — случайная величина со значениями в $\mathbb{R}^1$ на $\{\Omega, \mathfrak{F}\}$, $\varphi = \varphi(x)$ — борелевская функция со значениями в $\mathbb{R}^1$ на $\mathbb{R}^1$. Тогда функция*

$$\eta = \eta(\omega) = \varphi(\xi(\omega))$$

на $\{\Omega, \mathfrak{F}\}$ также является случайной величиной (борелевская функция от случайной величины — случайная величина).

Доказательство. Пусть $B \in \mathfrak{B}^1$. Представим множество $\eta^{-1}(B)$ в виде

$$\eta^{-1}(B) = \{\omega : \eta(\omega) \in B\} = \{\omega : \varphi(\xi(\omega)) \in B\} =$$

$$= \left\{\omega : \xi(\omega) \in \varphi^{-1}(B)\right\} = \xi^{-1}(\varphi^{-1}(B)).$$

Так как φ — борелевская функция, то $\varphi^{-1}(B) \in \mathfrak{B}^1$, и поскольку ξ — случайная величина, то $\xi^{-1}(\varphi^{-1}(B)) \in \mathfrak{F}$. Так что $\eta^{-1}(B) \in \mathfrak{F}$. Последнее по определению обозначает, что η — случайная величина.

Из теоремы, в частности, следует, что если ξ — случайная величина, то ξ^n, e^ξ, $\sin\xi$, $\xi^+ = \max\{\xi, 0\}$, $\xi^- = -\min\{\xi, 0\}$, $|\xi|$ — также случайные величины.

Следующая теорема описывает многомерную случайную величину $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ в терминах ее компонент $\xi_1, \xi_2, \ldots, \xi_n$.

Теорема 8.2.2. *Для того чтобы функция* $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ *со значениями в* $\mathbb{R}^n$*, заданная на измеримом пространстве* $\{\Omega, \mathfrak{F}\}$*, была случайной величиной, необходимо и достаточно, чтобы случайными величинами были ее компоненты* $\xi_i = \xi_i(\omega), i = 1, 2, \ldots, n$.

Доказательство. Пусть $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots \ldots, \xi_n(\omega))$ — случайная величина со значениями в $\mathbb{R}^n$. Представим множество $\xi_i^{-1}(B_i),\ B_i \in \mathfrak{B}^1$, в виде

$$\xi_i^{-1}(B_i) = \{\omega : \xi_i(\omega) \in B_i\} =$$

$$= \{\omega : (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_{i-1}(\omega), \xi_i(\omega), \xi_{i+1}(\omega), \ldots, \xi_n(\omega)) \in C_i\} =$$

$$= \xi^{-1}(C_i),$$

где $C_i = \mathbb{R}^1 \times \mathbb{R}^1 \times \cdots \times \mathbb{R}^1 \times B_i \times \mathbb{R}^1 \times \cdots \times \mathbb{R}^1$. И так как C_i — борелевское множество в $\mathbb{R}^n$, а ξ — случайная величина со значениями в $\mathbb{R}^n$, то $\xi^{-1}(C_i) \in \mathfrak{F}$ а, значит, и $\xi_i^{-1}(B_i) \in \mathfrak{F}$. Поэтому по определению ξ_i — случайная величина со значениями в $\mathbb{R}^1$, $i = 1, 2, \ldots, n$.

Далее, пусть $\xi_i = \xi_i(\omega), i = 1, 2, \ldots, n$, — случайные величины на $\{\Omega, \mathfrak{F}\}$ со значениями в $\mathbb{R}^1$ и

$$\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$$

функция со значениями в $\mathbb{R}^n$. Обозначим через $\mathfrak{M}$ класс бесконечных интервалов

$$I_{a_1, a_2, \ldots, a_n} = \{(x_1, x_2, \ldots, x_n) : x_1 < a_1, x_2 < a_2, \ldots, x_n < a_n\}.$$

Класс $\mathfrak{M}$, во-первых, порождает σ-алгебру $\mathfrak{B}^n$ ($\mathfrak{B}^n = \sigma(\mathfrak{M})$), а во-вторых

$$\xi^{-1}(I_{a_1, a_2, \ldots, a_n}) = \{\omega : (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega)) \in I_{a_1, a_2, \ldots, a_n}\} =$$

$$= \{\omega : \xi_1(\omega) < a_1, \xi_2(\omega) < a_2, \ldots, \xi_n(\omega) < a_n\} =$$

$$= \bigcap_{k=1}^{n} \{\omega : \xi_k(\omega) < a_k\} \in \mathfrak{F}.$$

Поэтому в силу теоремы 8.1.1 функция $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ является случайной величиной со значениями в $\mathbb{R}^n$.

Теорема 8.2.3. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — конечные случайные величины со значениями в $\mathbb{R}^1$ на $\{\Omega, \mathfrak{F}\}$ и $\varphi = \varphi(t_1, t_2, \ldots, t_n)$ — борелевская функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$, тогда*

$$\eta = \eta(\omega) = \varphi(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$$

является случайной величиной (борелевская функция от случайных величин является случайной величиной).

Доказательство. Так как функции $\xi_i = \xi_i(\omega)$, $i = 1, 2, \ldots$ $\ldots, n$, являются случайными величинами со значениями в $\mathbb{R}^1$, то функция $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ со значениями в $\mathbb{R}^n$ является случайной величиной (со значениями в $\mathbb{R}^n$).

Для каждого $B \in \mathfrak{B}^1$ множество

$$\eta^{-1}(B) = \{\omega : \eta(\omega) \in B\} = \{\omega : \varphi(\xi(\omega)) \in B\} =$$

$$= \left\{\omega : \xi(\omega) \in \varphi^{-1}(B)\right\}$$

принадлежит $\mathfrak{F}$, поскольку $\varphi^{-1}(B)$ — борелевское множество, а $\xi = \xi(\omega)$ случайная величина. Поэтому, по определению, функция $\eta = \eta(\omega)$ является случайной величиной.

Следствие. *Если ξ и η — конечные случайные величины со значениями в $\mathbb{R}^1$ на $\{\Omega, \mathfrak{F}\}$, то*

$$\xi \pm \eta, \ \xi\eta, \ \xi/\eta$$

также случайные величины.

Доказательство следует из того, что функции $\varphi_1(x, y) = x \pm y$, $\varphi_2(x, y) = xy$, $\varphi_3(x, y) = x/y$ являются борелевскими. (при этом функцию $1/x$ необходимо однозначно определить в точке $x = 0$).

Резюме. *Алгебраические операции над случайными величинами не выводят из класса случайных величин.*

Пример. *Пусть ξ и η — случайные величины, тогда множества*

$$\{\omega : \xi(\omega) < \eta(\omega)\}, \{\omega : \xi(\omega) = \eta(\omega)\}, \{\omega : \xi(\omega) \geq \eta(\omega)\}$$

являются событиями.

Решение. Достаточно заметить, что $\xi - \eta$ является случайной величиной.

8.3 Предел последовательности случайных величин

Здесь будет целесообразно расширить класс случайных величин, допустив, что они могут принимать значения $\pm\infty$.

Напомним известные определения нижнего и верхнего пределов последовательности.

Определение. Пусть $\{a_n\}$ — числовая последовательность. Число $\overline{a}$ называется *верхним пределом* последовательности $\{a_n\}$ и обозначается символом $\overline{\lim} a_n$, если существует подпоследовательность $\{a_{n_k}\}$ последовательности $\{a_n\}$, сходящаяся к $\overline{a}$, и для любой другой подпоследовательности $\{a_{n_l}\}$ последовательности $\{a_n\}$, имеющей предел: $\lim a_{n_l} = a$, выполняется неравенство $a \le \overline{a}$.

Число $\underline{a}$ называется *нижним пределом* последовательности $\{a_n\}$ и обозначается $\underline{\lim} a_n$, если существует подпоследовательность $\{a_{n_k}\}$ последовательности $\{a_n\}$, сходящаяся к $\underline{a}$, и для любой другой подпоследовательности $\{a_{n_l}\}$ последовательности $\{a_n\}$, имеющей предел: $\lim a_{n_l} = a$, выполняется неравенство $\underline{a} \le a$.

Пусть $\xi_n = \xi_n(\omega)$, $n = 1, 2, \ldots,$ — последовательность случайных величин. При каждом фиксированном $\omega \in \Omega$ значения $\{\xi_n(\omega)\}$ образуют числовую последовательность. Оказывается, что функции $\sup \xi_n(\omega)$, $\inf \xi_n(\omega)$, $\overline{\lim} \xi_n(\omega)$, $\underline{\lim} \xi_n(\omega)$, $\omega \in \Omega$, на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ являются случайными величинами.

Теорема 8.3.1. *Пусть $\{\xi_n\}$ — последовательность случайных величин со значениями в $\mathbb{R}^1$, заданных на измеримом пространстве $\{\Omega, \mathfrak{F}\}$, тогда*

$$\sup_n \xi_n, \ \inf_n \xi_n, \ \overline{\lim} \xi_n = \inf_n \sup_{m \ge n} \xi_m, \ \underline{\lim} \xi_n = \sup_n \inf_{m \ge n} \xi_m$$

также являются случайными величинами.

Доказательство. Имеют место следующие равенства (в их справедливости убеждаемся непосредственной проверкой):

$$\left\{\omega : \sup_n \xi_n(\omega) > x\right\} = \bigcup_{n=1}^{\infty} \{\omega : \xi_n(\omega) > x\}, \tag{8.3.1}$$

$$\left\{\omega : \inf_n \xi_n(\omega) < x\right\} = \bigcup_{n=1}^{\infty} \{\omega : \xi_n(\omega) < x\}, \tag{8.3.2}$$

$$\left\{\omega : \overline{\lim}\xi_n(\omega) < x\right\} = \bigcup_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty}\left\{\omega : \xi_j(\omega) < x - \frac{1}{k}\right\}, \quad (8.3.3)$$

$$\{\omega : \underline{\lim}\xi_n(\omega) > x\} = \bigcup_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty}\left\{\omega : \xi_j(\omega) > x + \frac{1}{k}\right\}. \quad (8.3.4)$$

Далее, так как $\xi_1, \xi_2, \ldots$ — случайные величины, то каждое из множеств

$$\{\omega : \xi_n(\omega) > x\}, \ \{\omega : \xi_n(\omega) < x\},$$

$$\left\{\omega : \xi_j(\omega) > x + \frac{1}{k}\right\}, \ \left\{\omega : \xi_j(\omega) < x - \frac{1}{k}\right\}$$

принадлежит σ-алгебре $\mathfrak{F}$, поэтому $\mathfrak{F}$ принадлежат и множества, представляющие собой правые части равенств (8.3.1) — (8.3.4), а вместе с правыми частями этих равенств σ-алгебре $\mathfrak{F}$ принадлежат и левые части.

Осталось сослаться на теорему 8.1.1, заметив, что классы $\left\{(-\infty, t), t \in \mathbb{R}^1\right\}$ и $\left\{(t, +\infty), t \in \mathbb{R}^1\right\}$ порождают σ-алгебру борелевских множеств.

В справедливости равенств (8.3.1) — (8.3.4) убеждаемся непосредственной проверкой.

Следствие. *Если последовательность $\{\xi_n\}$ случайных величин имеет предел*

$$\lim_n \xi_n = \xi,$$

то он является случайной величиной.

Доказательство. Достаточно заметить, что если $\lim\limits_n \xi_n$ существует, то

$$\lim \xi_n = \overline{\lim}\xi_n,$$

а $\overline{\lim}\xi_n$ — случайная величина.

8.4 Распределение случайной величины

Определение. *Распределением случайной величины $\xi = \xi(\omega)$* со значениями в $\mathbb{R}^n$, заданной на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, будем называть вероятностное распределение P_ξ, определенное на σ-алгебре $\mathfrak{B}^n$ пространства $\mathbb{R}^n$ значений ξ равенством

$$\mathsf{P}_\xi(B) = \mathsf{P}\{\omega : \xi(\omega) \in B\} = \mathsf{P}(\xi^{-1}(B)), \ B \in \mathfrak{B}^n.$$

Определение корректно:

1° $\mathsf{P}_\xi(B) \geq 0$ для каждого $B \in \mathfrak{B}^n$;

2° если B_i из $\mathfrak{B}^n$, $i = 1, 2, \ldots$, попарно не пересекаются: $B_i \bigcap B_j = \emptyset, i \neq j$, то

$$\mathsf{P}_\xi\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathsf{P}\left(\xi^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = \mathsf{P}\left(\bigcup_{i=1}^{\infty} \xi^{-1}(B_i)\right) =$$

$$= \sum_{i=1}^{\infty} \mathsf{P}\left(\xi^{-1}\left(B_i\right)\right) = \sum_{i=1}^{\infty} \mathsf{P}_\xi\left(B_i\right);$$

3° $\mathsf{P}_\xi\left(\mathbb{R}^n\right) = \mathsf{P}\left\{\xi \in \mathbb{R}^n\right\} = 1.$

Распределение P_ξ случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ со значениями в $\mathbb{R}^n$ еще называют *совместным распределением случайных величин* $\xi_1, \xi_2, \ldots, \xi_n$.

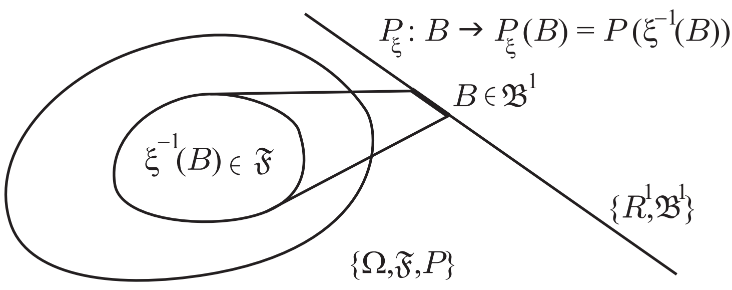

Рис. 8.4.1: К определению распределения случайной величины

Для любого вероятностного распределения Q можно указать вероятностное пространство и случайную величину $\xi = \xi(\omega)$ на нем, которая имеет Q своим распределением.

Теорема 8.4.1. *Пусть* Q — *вероятностное распределение на* $\left\{\mathbb{R}^1, \mathfrak{B}^1\right\}$. *Существуют вероятностное пространство* $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ *и случайная величина* $\xi = \xi(\omega)$ *со значениями в* $\mathbb{R}^1$ *на* $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ *такая, что ее распределение* P_ξ *совпадает с* Q.

Доказательство. В качестве вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ рассмотрим $\left\{\mathbb{R}^1, \mathfrak{B}^1, \mathsf{Q}\right\}$, т. е. $\Omega = \mathbb{R}^1$, $\mathfrak{F} = \mathfrak{B}^1$,

$\mathsf{P} = \mathsf{Q}$. А случайную величину $\xi = \xi(\omega)$ на $\left\{\mathbb{R}^1, \mathfrak{B}^1, \mathsf{Q}\right\}$ определим как тождественное отображение $\mathbb{R}^1$ на $\mathbb{R}^1$, т. е.

$$\xi : \omega \to \xi(\omega) = \omega.$$

По определению распределения P_ξ случайной величины ξ, с учетом того, что ξ — тождественное отображение, имеем:

$$\mathsf{P}_\xi(B) = \mathsf{P}(\xi^{-1}(B)) = \mathsf{Q}(\xi^{-1}(B)) = \mathsf{Q}(B)$$

для каждого $B \in \mathfrak{B}^1$. Так что распределение P_ξ случайной величины ξ совпадает с заданным распределением Q.

Тем самым теорема доказана.

Замечание 1. Теорема устанавливает соотношение между двумя моделями стохастического эксперимента с исходами в $\mathbb{R}^1$: 1° вероятностным распределением на $\mathbb{R}^1$ и 2° функцией со значениями в $\mathbb{R}^1$, заданной на пространстве элементарных событий некоторого вероятностного пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$.

Замечание 2. Пусть $\xi = \xi(\omega)$ — дискретная случайная величина со значениями в $\mathbb{R}^n$, $X = \{x_1, x_2, \ldots, x_n, \ldots\}$ — множество ее возможных значений:

$$\mathsf{P}_\xi(x_i) = \mathsf{P}\{\xi = x_i\} > 0, \ \sum_{x_i \in X} \mathsf{P}_\xi(x_i) = 1.$$

Для каждого $B \in \mathfrak{B}^n$ в силу счетной аддитивности вероятности

$$\mathsf{P}_\xi(B) = \mathsf{P}\{\omega : \xi(\omega) \in B\} =$$

$$= \mathsf{P}\{\omega : \xi(\omega) \in B \cap X\} = \mathsf{P}\Big(\bigcup_{x_i \in B} \{\omega : \xi(\omega) = x_i\}\Big) = \sum_{x_i : x_i \in B} \mathsf{P}_\xi(x_i),$$

т. е. по значениям $\mathsf{P}_\xi(x_i)$ можно вычислить $\mathsf{P}_\xi(B)$ для любого $B \in \mathfrak{B}^n$ (верно и обратное). Поэтому данное ранее определение распределения дискретной случайной величины как функции

$$\mathsf{P}_\xi : x_i \to \mathsf{P}_\xi(x_i) = \mathsf{P}\{\xi = x_i\}$$

на множестве $X \subset \mathbb{R}^n$ ее возможных значений эквивалентно приведенному общему определению.

Замечание 3. Для того чтобы можно было говорить о распределении случайной величины $\xi = \xi(\omega)$, она должна быть задана не на измеримом пространстве $\{\Omega, \mathfrak{F}\}$, а на вероятностном $\{\Omega, \mathfrak{F}, \mathsf{P}\}$.

Функция распределения случайной величины. Распределение P_ξ случайной величины ξ со значениями в $\mathbb{R}^1$, как и любое вероятностное распределение, имеет функцию распределения, которую называют функцией распределения случайной величины ξ.

Определение. *Функцией распределения случайной величины* ξ со значениями в $\mathbb{R}^1$ называется функция

$$F_\xi(x) = \mathsf{P}\{\omega : \xi(\omega) < x\} = \mathsf{P}_\xi((-\infty, x)),\ x \in \mathbb{R}^1.$$

Заметим, что распределение $\mathsf{P}_\xi(B)$, $B \in \mathfrak{B}^1$, случайной величины ξ своей функцией распределения имеет $\mathsf{P}_\xi((-\infty, x))$, $x \in \mathbb{R}^1$.

Функция распределения $F_\xi(x)$ случайной величины, как и каждая вероятностная функция распределения, обладает следующими свойствами (см. теорему 7.2.3):

1° $0 \le F_\xi(x) \le 1$;

2° $F_\xi(x)$ не убывает: если $x_1 \le x_2$, то $F_\xi(x_1) \le F_\xi(x_2)$;

3° $F_\xi(x)$ непрерывна слева: $F_\xi(x) = F_\xi(x - 0)$;

4° $F_\xi(-\infty) = 0$, $F_\xi(+\infty) = 1$.

Кроме того,

5° $\mathsf{P}\{\xi \in [a, b)\} = F_\xi(b) - F_\xi(a)$;

6° $\mathsf{P}\{\xi = x\} = F_\xi(x + 0) - F_\xi(x - 0)$.

Свойства 5° и 6° получаются из следующих соображений. Для распределения Q на $\mathbb{R}^1$ и его функции распределения $Q(x)$ всегда имеют место равенства (см. теорему 7.2.3)

$$\mathsf{Q}([a; b)) = Q(b) - Q(a),\ \mathsf{Q}(\{x\}) = Q(x + 0) - Q(x - 0).$$

В частности, для распределения P_ξ случайной величины ξ и ее функции распределения $F_\xi(x)$

$$\mathsf{P}_\xi([a, b)) = F_\xi(b) - F_\xi(a),\ \mathsf{P}_\xi(\{x\}) = F_\xi(x + 0) - F_\xi(x - 0).$$

А учитывая, что

$$\mathsf{P}_\xi([a, b)) = \mathsf{P}\{\xi \in [a, b)\}, \mathsf{P}_\xi(\{x\}) = \mathsf{P}\{\xi = x\},$$

получаем свойства 5° и 6°.

Из свойства 6° следует, что *случайная величина ξ принимает данное значение x с ненулевой вероятностью тогда и только тогда, когда функция распределения случайной величины ξ в точке x имеет скачок.*

Свойства $1^\circ - 6^\circ$ функции распределения $F_\xi(x)$ случайной величины ξ можно получить непосредственно из ее определения:

$$F_\xi(x) = \mathsf{P}\left\{\omega : \xi(\omega) < x\right\} = \mathsf{P}\left\{\xi < x\right\}.$$

1° Поскольку $F_\xi(x) = \mathsf{P}\{\xi < x\}$, то $0 \le F_\xi(x) \le 1$.

2° При $x_1 \le x_2$ имеет место включение

$$\{\omega : \xi(\omega) < x_1\} \subset \{\omega : \xi(\omega) < x_2\}.$$

Отсюда

$$F_\xi(x_1) = \mathsf{P}\{\omega : \xi(\omega) < x_1\} \le \mathsf{P}\{\omega : \xi(\omega) < x_2\} = F_\xi(x_2).$$

При доказательстве свойств $3^\circ,\ 4^\circ,\ 5^\circ$ используется свойство непрерывности вероятности.

3° Пусть $x_n \uparrow x$, тогда

$$\{\omega : \xi(\omega) < x\} = \bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\}.$$

Отсюда

$$F_\xi(x) = \mathsf{P}\{\omega : \xi(\omega) < x\} = \mathsf{P}\left(\bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\}\right) =$$

$$= \lim_{x_n \uparrow x} \mathsf{P}\{\omega : \xi(\omega) < x_n\} = \lim_{x_n \uparrow x} F_\xi(x_n) = F_\xi(x-0).$$

4° Пусть $x_n \uparrow +\infty$, тогда

$$\bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\} = \Omega.$$

Отсюда

$$1 = \mathsf{P}\left(\bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\}\right) =$$

$$= \lim_{x_n \uparrow +\infty} \mathsf{P}\{\omega : \xi(\omega) < x_n\} = \lim_{x_n \uparrow +\infty} F_\xi(x_n) = F_\xi(+\infty).$$

5° При $a < b$ из равенства

$$\{\omega : \xi(\omega) < b\} = \{\omega : \xi(\omega) \in [a, b)\} \cup \{\omega : \xi(\omega) < a\}$$

получаем:

$$F_\xi(b) = \mathsf{P}\{\omega : \xi(\omega) < b\} = \mathsf{P}\{\omega : \xi(\omega) \in [a, b)\} + \mathsf{P}\{\omega : \xi(\omega) < a\} =$$

$$= \mathsf{P}\{\omega : \xi(\omega) \in [a, b)\} + F_\xi(a).$$

6° Из равенства

$$\{\omega : \xi(\omega) = x\} = \bigcap_{n=1}^{\infty} \left\{\omega : \xi(\omega) \in \left[x - \frac{1}{n}, x + \frac{1}{n}\right)\right\},$$

аналогично доказательству свойства 3°, имеем:

$$\mathsf{P}\{\xi = x\} = F_\xi(x + 0) - F_\xi(x - 0).$$

Определение. Функцию

$$F(x_1, x_2, \ldots, x_n) = \mathsf{P}\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, \ldots, \xi_n(\omega) < x_n\},$$

определенную на $\mathbb{R}^n$, будем называть *функцией распределения* случайного вектора $\zeta = (\xi_1, \xi_2, \ldots, \xi_n)$.

Определение. Если функцию распределения случайного вектора $\zeta = (\xi_1, \xi_2, \ldots, \xi_n)$ можно представить в виде

$$F(x_1, x_2, \ldots, x_n) = \int\limits_{-\infty}^{x_1} \int\limits_{-\infty}^{x_2} \ldots \int\limits_{-\infty}^{x_n} p(t_1, t_2, \ldots, t_n) dt_n dt_{n-1} \ldots dt_1, \tag{8.4.1}$$

то будем говорить, что вектор $\zeta = (\xi_1, \xi_2, \ldots, \xi_n)$ имеет *абсолютно непрерывное распределение* (*вектор абсолютно непрерывный*), а функцию $p(x_1, x_2, \ldots, x_n)$ называют его *плотностью распределения, или совместной плотностью распределения случайных величин* $\xi_1, \xi_2, \ldots, \xi_n$.

Плотность распределения $p(x_1, x_2, \ldots, x_n)$ случайного вектора $\zeta = (\xi_1, \xi_2, \ldots, \xi_n)$ неотрицательна и

$$\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} \ldots \int\limits_{-\infty}^{+\infty} p(t_1, t_2, \ldots, t_n) dt_n dt_{n-1} \ldots dt_1 = 1.$$

Плотность распределения случайной величины со значениями в $\mathbb{R}^1$. Если функция распределения $F(x)$ случайной величины ξ представима в виде

$$F(x) = \int\limits_{-\infty}^{x} p(t)dt,$$

то говорят, что случайная величина имеет абсолютно непрерывное распределение, функцию $p(t)$ называют *плотностью распределения случайной величины*, а саму случайную величину еще называют *абсолютно непрерывной*. Пока интеграл понимаем как интеграл Римана.

Из определения плотности распределения случайной величины следует, что

1° плотность распределения $p(x)$ равна производной от функции распределения $F(x)$:

$$\frac{d}{dx}F(x) = p(x);$$

2° плотность распределения $p(x)$ неотрицательна;

3° $\int\limits_{-\infty}^{\infty} p(t)dt = 1$;

4° $\mathsf{P}\{\xi \in [a,b)\} = \int\limits_{a}^{b} p(t)dt.$

Примеры распределений случайных величин. Случайные величины называют по их распределениям. Случайную величину ξ со значениями в $\mathbb{R}^1$ будем называть *нормально распределенной* с параметрами $(a;\sigma^2)$ (*гауссовской случайной величиной*), если ее распределением является нормальное распределение с параметрами $(a;\sigma^2)$; *равномерно распределенной* на промежутке $[a,b]$, если ее распределением является равномерное на $[a,b]$ распределение; *гамма-распределенной* с параметрами $(\nu;\theta)$, если ее распределением является гамма-распределение с параметрами $(\nu;\theta)$; *показательно распределенной* с параметром θ, если ее распределением является показательное распределение с параметром θ, и т. д.

Приводимая далее теорема дает возможность по равномерно распределенной на $[0,1]$ случайной величине строить случайные величины с заданными распределениями.

Теорема 8.4.2. *Пусть ξ — равномерно распределенная на $[0,1]$ случайная величина, $F(x)$ — монотонно возрастающая непрерывная вероятностная функция распределения на $\mathbb{R}^1$, тогда случайная величина*

$$\eta = F^{-1}(\xi)$$

имеет своей функцией распределения $F(x)$.

Доказательство. Из определения функции распределения случайной величины и ее монотонности имеем

$$F_\eta(x) = P\{\eta < x\} = P\{F^{-1}(\xi) < x\} = P\{F(F^{-1}(\xi)) < F(x)\} =$$

$$= P\{\xi < F(x)\} = F_\xi(F(x)) = F(x).$$

Равенство $F_\xi(F(x)) = F(x)$ следует из того, что ξ равномерно распределена на $[0,1]$ — при $t \in [0,1]$ ее функция распределения $F_\xi(t) = t$.

Свойство отсутствия последействия показательного распределения. Будем говорить, что неотрицательная случайная величина ξ *обладает свойством отсутствия последействия,* если для любых $s > 0, t > 0$,

$$\mathsf{P}\{\xi < s + t \,/\, \xi \geq s\} = \mathsf{P}\{\xi < t\}. \qquad (8.4.1)$$

Теорема 8.4.3. *Для того чтобы неотрицательная случайная величина ξ с дифференцируемой на положительной полуоси функцией распределения, обращающейся в нуль в точке $x = 0$, обладала свойством отсутствия последействия, необходимо и достаточно, чтобы она была показательно распределенной.*

Доказательство. Пусть ξ — показательно распределенная случайная величина. Ее функция распределения

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0 \end{cases}$$

$(\lambda > 0)$. Для любых $s > 0, t > 0$

$$\mathsf{P}\{\xi < s + t \,/\, \xi \geq s\} = \frac{\mathsf{P}\{s \leq \xi < s + t\}}{\mathsf{P}\{\xi \geq s\}} = \frac{F(s+t) - F(s)}{1 - F(s)} =$$

$$= \frac{e^{-\lambda s} - e^{-\lambda(s+t)}}{e^{-\lambda s}} = 1 - e^{-\lambda t} = \mathsf{P}\{\xi < t\},$$

т. е.

$$\mathsf{P}\{\xi < s + t \,/\, \xi \geq s\} = \mathsf{P}\{\xi < t\}.$$

Последнее равенство означат, что ξ обладает свойством отсутствия последействия.

Далее. Пусть случайная величина ξ обладает свойством отсутствия последействия:

$$\mathsf{P}\{\xi < s + t \,/\, \xi \geq s\} = \mathsf{P}\{\xi < t\}, \; s > 0, t > 0,$$

или

$$\mathsf{P}\{\xi < t\} = \frac{\mathsf{P}\{s \leq \xi < s + t\}}{\mathsf{P}\{\xi \geq s\}},$$

в терминах функции $F(t)$:

$$\frac{F(s+t) - F(s)}{1 - F(s)} = F(t). \tag{8.4.2}$$

Последнее равенство в терминах функции $Q(t) = 1 - F(t)$ перепишется так:

$$\frac{Q(s+t) - Q(s)}{Q(s)} = Q(t) - Q(0) \tag{8.4.3}$$

(поскольку $F(0) = 0$, то $Q(0) = 1$.) Функция $Q(t) = 1 - F(t)$ дифференцируема, поскольку по условию функция $F(t)$ дифференцируема на положительной полуоси.

Пусть $s \in (0, +\infty)$. Разделив правую и левую части равенства (8.4.3) на t, получим:

$$\frac{(Q(s+t) - Q(s))/t}{Q(s)} = \frac{Q(t) - Q(0)}{t}. \tag{8.4.4}$$

Отсюда, переходя к пределу при $t \to 0 + 0$, получаем дифференциальное уравнение для $Q(s)$:

$$\frac{Q'(s)}{Q(s)} = Q'(0).$$

или

$$\frac{d}{ds} \ln Q(s) = Q'(0)$$

(функция $Q(s)$ дифференцируема справа в точке 0, что следует из существования предела при $t \to 0$ левой части равенства

(8.4.4)). Решая последнее дифференциальное уравнение при начальном условии $Q(0) = 1$, получаем:

$$Q(s) = e^{Q'(0)s}$$

(заметим еще, что $Q'(0) < 0$, поскольку $Q(t)$ — монотонно убывающая при $t > 0$ функция). Так что

$$F(s) = 1 - Q(s) = 1 - e^{Q'(0)s}.$$

Тем самым теорема доказана.

8.5 Независимые случайные величины

Определение. Случайные величины ξ_1 и ξ_2 со значениями в $\mathbb{R}^1$ называют *независимыми*, если для любых $B_1, B_2 \in \mathfrak{B}^1$

$$\mathsf{P}\{\xi_1 \in B_1, \xi_2 \in B_2\} = \mathsf{P}\{\xi_1 \in B_1\}\,\mathsf{P}\{\xi_2 \in B_2\}.$$

О независимых случайных величинах смотри также в параграфе 5.3 главы 5.

Определение. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ со значениями в $\mathbb{R}^1$ называют *независимыми* (*независимыми в совокупности*), если для любых $B_1, B_2, \ldots B_n \in \mathfrak{B}^1$

$$\mathsf{P}\{\xi_1 \in B_1, \xi_2 \in B_2, \ldots, \xi_n \in B_n\} =$$

$$= \mathsf{P}\{\xi_1 \in B_1\}\,\mathsf{P}\{\xi_2 \in B_2\} \ldots \mathsf{P}\{\xi_n \in B_n\}. \tag{8.5.1}$$

Будем говорить, что $\xi_1, \xi_2, \ldots, \xi_n$ *попарно независимы*, если для каждой пары случайных величин ξ_i, ξ_j $(i \neq j)$ и для любых множеств $B_i, B_j \in \mathfrak{B}^1$, $i, j = 1, 2, \ldots, n$,

$$\mathsf{P}\{\xi_i \in B_i, \xi_j \in B_j\} = \mathsf{P}\{\xi_i \in B_i\}\,\mathsf{P}\{\xi_j \in B_j\}.$$

Из независимости случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ следует их попарная независимость. Обратное утверждение не имеет места.

Далее нам понадобится понятие произведения распределений.

Пусть $\{\mathbb{R}^1, \mathfrak{B}^1, \mathsf{F}_i\}$ — *вероятностные пространства,* $i = 1, 2, \ldots, n$. *Существует, и притом единственное, распределение* F, *заданное на* σ*-алгебре* $\mathfrak{B}^n$ *борелевских множеств пространства* $\mathbb{R}^n$ *такое, что для любых борелевских множеств* $B_1, B_2, \ldots, B_n$ *из* $\mathbb{R}^1$

$$\mathsf{F}(B_1 \times B_2 \times \ldots \times B_n) = \mathsf{F}_1(B_1)\mathsf{F}_2(B_2)\ldots\mathsf{F}_n(B_n). \qquad (8.5.2)$$

Распределение F, значение которого на параллелепипедах $B_1 \times B_2 \times \ldots \times B_n$ определяется равенством (8.5.2), называют *произведением распределений* $\mathsf{F}_1, \mathsf{F}_2, \ldots, \mathsf{F}_n$ и обозначают через $\mathsf{F}_1 \times \mathsf{F}_2 \times \ldots \times \mathsf{F}_n$, т. е.

$$\mathsf{F} = \mathsf{F}_1 \times \mathsf{F}_2 \times \ldots \times \mathsf{F}_n.$$

Равенство (8.5.1), определяющее независимые случайные величины, можно переписать в виде

$$\mathsf{P}\{\xi_1 \in B_1, \xi_2 \in B_2, \ldots, \xi_n \in B_n\} = \mathsf{P}_{\xi_1}(B_1)\mathsf{P}_{\xi_2}(B_2)\ldots\mathsf{P}_{\xi_n}(B_n),$$

что приводит к следующему определению независимых случайных величин в терминах распределений.

Определение. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называют *независимыми*, если совместное распределение P_ξ случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ равно произведению их распределений:

$$\mathsf{P}_\xi = \mathsf{P}_{\xi_1} \times \mathsf{P}_{\xi_2} \times \ldots \times \mathsf{P}_{\xi_n},$$

т. е. для любых борелевских множеств $B_1, B_2, \ldots, B_n$

$$\mathsf{P}_\xi(B_1 \times B_2 \times \ldots \times B_n) = \mathsf{P}_{\xi_1}(B_1)\mathsf{P}_{\xi_2}(B_2)\ldots\mathsf{P}_{\xi_n}(B_n).$$

Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называют *попарно независимыми,* если совместное распределение случайных величин ξ_i, ξ_j равно произведению их распределений для всех $i \neq j$, $i, j = 1, 2, \ldots, n$:

$$\mathsf{P}_{\xi_i, \xi_j} = \mathsf{P}_{\xi_i} \times \mathsf{P}_{\xi_j}.$$

Теорема 8.5.1. *Пусть* ξ *и* η — *независимые случайные величины со значениями в* $\mathbb{R}^1$, f *и* g — *борелевские функции на* $\mathbb{R}^1$ *со значениями в* $\mathbb{R}^1$, *тогда* $f(\xi)$ *и* $g(\eta)$ — *независимые случайные величины.*

Доказательство. Пусть A и B — борелевские множества из $\mathfrak{B}^1$, тогда

$$\mathsf{P}\left\{f(\xi) \in A, g(\eta) \in B\right\} = \mathsf{P}\left\{\xi \in f^{-1}(A), \eta \in g^{-1}(B)\right\} =$$

$$= \mathsf{P}\left\{\xi \in f^{-1}(A)\right\} \mathsf{P}\left\{\eta \in g^{-1}(B)\right\} = \mathsf{P}\left\{f(\xi) \in A\right\} \mathsf{P}\left\{g(\eta) \in B\right\}.$$

Имеет место и более общее утверждение.

Теорема 8.5.2. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$, $\eta_1, \eta_2, \ldots, \eta_m$ — независимые случайные величины, f — борелевская функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$, g — борелевская функция на $\mathbb{R}^m$ со значениями в $\mathbb{R}^1$, тогда $f(\xi_1, \xi_2, \ldots, \xi_n)$ и $g(\eta_1, \eta_2, \ldots, \eta_m)$ — независимые случайные величины.*

Без доказательства.

8.6 Примеры и задачи

Примеры

Пример 8.6.1 . *Пусть ξ_i, $i = 1, 2, \ldots, n$, — независимые случайные величины, каждая с функцией распределения $F(x)$. Найти функцию распределения случайной величины*

$$\eta = \min\{\xi_1, \xi_2, \ldots, \xi_n\}.$$

Решение. По определению $F_\eta(x) = \mathsf{P}\{\eta < x\}$. Отсюда

$$1 - F_\eta(x) = 1 - \mathsf{P}\{\eta < x\} = \mathsf{P}\{\eta \geq x\} =$$

$$= \mathsf{P}\{\min\{\xi_1, \xi_2, \ldots, \xi_n\} \geq x\} =$$

$$= \mathsf{P}\{\xi_1 \geq x, \xi_2 \geq x, \ldots, \xi_n \geq x\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i \geq x\} =$$

$$= \prod_{i=1}^{n} (1 - F_{\xi_i}(x)) = \prod_{i=1}^{n} (1 - F(x)) = (1 - F(x))^n$$

(воспользовались независимостью случайных величин $\xi_1, \xi_2, \ldots$ $\ldots, \xi_n$). Так что

$$F_\eta(x) = 1 - (1 - F(x))^n.$$

Пример 8.6.2. *На отрезок* $[0;1]$ *наудачу бросают точку, делящую его на две части. Найти функцию распределения длины меньшей части отрезка* $[0;1]$.

Решение. Пусть ξ — координата брошенной точки, тогда длина η меньшей части отрезка равна $\min\{\xi, 1-\xi\}$. Заметим, что $F_\eta(x) = 0$ при $x \le 0$ и $F_\eta(x) = 1$ при $x \ge 1/2$.

При $0 < x \le 1/2$ имеем:

$$F_\eta(x) = \mathsf{P}\{\eta < x\} = \mathsf{P}\{\min\{\xi, 1-\xi\} < x\} =$$

$$= 1 - \mathsf{P}\{\min\{\xi, 1-\xi\} \ge x\} = 1 - \mathsf{P}\{\xi \ge x, 1-\xi \ge x\} =$$

$$= 1 - \mathsf{P}\{x \le \xi \le 1-x\}.$$

Вероятность $\mathsf{P}\{x \le \xi \le 1-x\}$ вычисляем как геометрическую вероятность (точку на отрезок $[0;1]$ бросают наудачу).

Следовательно,

$$F_\eta(x) = \begin{cases} 0, & \text{если } x \le 0; \\ 2x, & \text{если } 0 < x \le 1/2; \\ 1, & \text{если } x > 1/2. \end{cases}$$

Пример 8.6.3. *Случайная величина* η *имеет распределение* N_{a,σ^2}. *Доказать, что случайная величина* $\xi = (\eta - a)/\sigma$ *имеет распределение* $\mathsf{N}_{0;1}$.

Решение.

$$F_\xi(x) = \mathsf{P}\{\xi < x\} = \mathsf{P}\left\{\frac{\eta - a}{\sigma} < x\right\} = \mathsf{P}\{\eta < a + \sigma x\} =$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{a+\sigma x} \exp\left\{-\frac{(t-a)^2}{2\sigma^2}\right\} dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left\{-u^2/2\right\} du.$$

(воспользовались заменой $(t-a)/\sigma = u$).

Задачи

8.1°. Пусть ξ — случайная величина с плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \notin [-1;1]; \\ 1 - |x|, & \text{если } x \in [-1;1]. \end{cases}$$

Вычислить $\mathsf{P}\{\xi^2 > 1/4\}$.

Ответ: $1/4$.

8.2°. На отрезок $[0, l]$ наудачу бросают точку, ξ — ее координата. Найти функцию распределения и плотность распределения случайной величины ξ.

Ответ:
$$F_\xi(x) = \begin{cases} 0, & \text{если } x \le 0; \\ x/l, & \text{если } 0 < x \le l; \\ 1, & \text{если } x > l. \end{cases}$$

8.3. Случайная величина ξ равномерно распределена на отрезке $[-2; 2]$. Найти плотность распределения случайной величины $\eta = |\xi|$.

Ответ:
$$p_\eta(x) = \begin{cases} 0, & \text{если } x \notin [0; 2]; \\ 1/2, & \text{если } x \in [0; 2]. \end{cases}$$

8.4. Пусть $F(x)$ — функция распределения случайной величины ξ. Найти функцию распределения $\eta = \xi^2$.

Ответ:
$$F_\eta(x) = \begin{cases} F(\sqrt{x}) - F(-\sqrt{x} + 0), & \text{если } x > 0; \\ 0, & \text{если } x \le 0. \end{cases}$$

8.5. Функция распределения $F(x)$ случайной величины ξ строго монотонна и непрерывна. Найти функцию распределения случайной величины $\eta = F(\xi)$.

Ответ: случайная величина $\eta = F(\xi)$ распределена равномерно на отрезке $[0; 1]$.

8.6. На отрезок $[0; l]$ наудачу бросают точку, делящую его на две части. Найти функцию распределения длины большей части.

Указание. См. пример 8.6.2.

Ответ:
$$F_\eta(x) = \begin{cases} 0, & \text{если } x \le l/2; \\ (2x - l)/l, & \text{если } l/2 < x \le l; \\ 1, & \text{если } x > l. \end{cases}$$

8.7. Пусть $F(x)$ — функция распределения случайной величины ξ. Найти функцию распределения случайной величины $\eta = \min\{\xi, L\}$ (L — константа).

Решение.

$$F_\eta(x) = \mathsf{P}\{\eta < x\} = \mathsf{P}\{\min\{\xi, L\} < x\} =$$

$$= \int\limits_{t:\min\{t,L\}<x} \mathsf{P}_\xi(dt) = \begin{cases} \mathsf{P}_\xi(-\infty, x), & \text{если } x \le L; \\ 1, & \text{если } x > L. \end{cases}$$

Глава 9

Математическое ожидание

Прежде чем давать формальное определение математического ожидания (интеграла Лебега), рассмотрим несколько примеров, естественным образом приводящих к этому понятию.

Пример (доход телефонной станции). *На телефонной станции в каждый момент t промежутка времени I фиксируется число $f(t)$ ведущихся телефонных разговоров. Известна стоимость $\mathsf{F}(I_s)$ одного разговора за время I_s — функция, ставящая в соответствие каждой части I_s промежутка I стоимость одного разговора за время I_s ($\mathsf{F} : I_s \to \mathsf{F}(I_s)$). Найти накопленный доход телефонной станции за время I.*

Функция $f(t)$ принимает конечное число неотрицательных целочисленных значений. Если через $\{t : f(t) = k\}$ — обозначить часть промежутка I, на котором велось k разговоров, то очевидно, накопленный доход будет равен

$$\sum_k k\mathsf{F}\{t : f(t) = k\}.$$

Последнюю сумму принято обозначать $\int\limits_I f(t)\mathsf{F}(dt)$ и называть интегралом Лебега от $f(t)$ по мере F.

Пример (урожай с поля). *Пусть $f(x)$ — урожайность пшеницы на поле Q ($x \in Q$), F — площадь — функция множества, ставящая в соответствие каждой части Q_i поля Q число, равное ее площади $\mathsf{F}(Q_i)$ ($\mathsf{F} : Q_i \to \mathsf{F}(Q_i)$). Найти величину урожая с поля Q.*

В отличие от предыдущего примера значения функции $f(x)$ “сплошь заполняют” некоторый промежуток.

Измеряя урожайность, мы пользуемся той или иной единицей урожайности ε. Например, $\varepsilon = 1$ ц, 0,5 ц, 0,1 ц и т. д. (единица измерения ε, вообще говоря, может быть выбрана любой). Пусть $A_{n,\varepsilon} = \{x : (n-1)\varepsilon \le f(x) < n\varepsilon\}$ ($A_{n,\varepsilon}$ — часть Q, на которой урожайность f лежит в пределах от $(n-1)\varepsilon$ до $n\varepsilon$). Урожайность на $A_{n,\varepsilon}$ будем считать равной $(n-1)\varepsilon$. Рассмотрим на Q функцию $f_\varepsilon(x)$, принимающую конечное число значений, а именно $f_\varepsilon(x) = (n-1)\varepsilon$ при $x \in A_{n,\varepsilon}$ ($f_\varepsilon(x)$ — урожайность, измеренная с точностью до ε). В качестве величины урожая с Q естественно рассматривать

$$\sum_n (n-1)\varepsilon \mathsf{F}\{x : (n-1)\varepsilon \le f(x) < n\varepsilon\}.$$

Чем точнее мы хотим подсчитать величину урожая, тем меньше необходимо выбирать единицу измерения урожайности. В качестве истинной величины урожая с поля Q естественно рассматривать

$$\lim_{\varepsilon\to 0}\sum_n (n-1)\varepsilon \mathsf{F}\{x : (n-1)\varepsilon \le f(x) < n\varepsilon\}.$$

Этот предел принято называть интегралом Лебега от $f(x)$ по мере F и обозначать $\int\limits_Q f(x)\mathsf{F}(dx)$.

Пример (среднее случайной величины). *Пусть* $\xi = \xi(\omega)$ *— случайная величина со значениями в* $\mathbb{R}^1$*, заданная на* $\{\Omega, \mathfrak{F}, \mathsf{P}\}$*. Что понимать под средним значением* ξ*?*

Будем поступать аналогично тому, как мы поступали в предыдущем примере. Сначала выберем некоторую единицу измерения ε, которой будем пользоваться, измеряя значения, принимаемые случайной величиной ξ. Пусть

$$A_{n,\varepsilon} = \{\omega : (n-1)\varepsilon \le \xi(\omega) < n\varepsilon\}$$

($A_{n,\varepsilon}$ — часть множества точек Ω, в которых значения $\xi(\omega)$ лежат между $(n-1)\varepsilon$ и $n\varepsilon$, $n \in \mathbb{Z}$). Определим на Ω случайную величину $\xi_\varepsilon = \xi_\varepsilon(\omega)$, которая на множестве $A_{n,\varepsilon}$ принимает значение $\xi_\varepsilon(\omega) = (n-1)\varepsilon$ ($\xi_\varepsilon = \xi_\varepsilon(\omega)$ построена по $\xi = \xi(\omega)$ измерением последней с точностью до ε, т. е. измерением ξ "линейкой" с ценой деления ε). В качестве среднего значения случайной величины ξ естественно рассматривать предел

$$\lim_{\varepsilon\to 0}\sum_n (n-1)\varepsilon \mathsf{P}\{\omega : (n-1)\varepsilon \le \xi(\omega) < n\varepsilon\},$$

который принято называть интегралом Лебега от ξ по мере P или математическим ожиданием случайной величины ξ и обозначать $\int\limits_{\Omega} \xi(\omega)\mathsf{P}(d\omega)$.

Рассмотренные примеры иллюстрируют тот факт, что конструкция интеграла Лебега возникает в практической деятельности человека так же естественно как, например, понятие площади у древних египтян при измерении площадей полей различной формы после разливов Нила.

9.1 Математическое ожидание простой случайной величины

Напомним, что *индикатором* множества $A \subset \Omega$ называют функцию

$$I_A = I_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A; \\ 0, & \text{если } \omega \notin A. \end{cases}$$

Для индикаторов имеют место следующие очевидные соотношения:

1° Если $A \cap B = \emptyset$, то $I_{A \cup B} = I_A + I_B$.

2° $I_{A \cap B} = I_A I_B$.

3° $I_{\overline{A}} = 1 - I_A$.

4° Если $A \subset B$, то $I_A \leq I_B$.

5° Если $B_i \cap B_j = \emptyset$, $i \neq j$, $B = \bigcup\limits_{i=1}^{\infty} B_i$, то

$$I_B = \sum_{i=1}^{\infty} I_{B_i}.$$

Определение. Случайную величину ($\mathfrak{F}$-измеримую функцию) $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^1$, заданную на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, будем называть *простой*, если пространство Ω можно представить в виде объединения конечного числа непересекающихся множеств $A_i \in \mathfrak{F}$:

$$\Omega = \bigcup_{i=1}^{n} A_i, \ A_i \cap A_j = \emptyset, i \neq j,$$

на каждом из которых $\xi = \xi(\omega)$ принимает постоянное значение:

$$\xi(\omega) = a_i, \, \omega \in A_i, \, i = 1, 2, \ldots, n.$$

Каждую простую случайную величину можно представить в виде линейной комбинации индикаторов:

$$\xi = \xi(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega),\ \omega \in \Omega.$$

Это представление простой случайной величины не единственно.

Сначала определим математическое ожидание (интеграл Лебега) от неотрицательной простой случайной величины.

Определение. *Математическим ожиданием* (*интегралом Лебега по мере* P) от неотрицательной простой случайной величины

$$\xi = \xi(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega)$$

будем называть

$$\sum_{i=1}^{n} a_i \mathsf{P}(A_i).$$

Обозначать математическое ожидание от случайной величины ξ будем через $\mathsf{M}\xi$, а интеграл Лебега от ξ по мере P — через $\int\limits_{\Omega} \xi(\omega)\mathsf{P}(d\omega)$. Так что по определению

$$\mathsf{M}\xi = \sum_{i=1}^{n} a_i \mathsf{P}(A_i) = \int\limits_{\Omega} \xi(\omega)\mathsf{P}(d\omega).$$

На интеграл $\int\limits_{\Omega} \xi(\omega)\mathsf{P}(d\omega)$ мы будем смотреть как на интегральную форму записи математического ожидания $\mathsf{M}\xi$ случайной величины ξ.

Замечание. Случайная величина ξ, а вместе с ней и $\mathsf{M}\xi$, может принимать бесконечные значения.

Данное определение математического ожидания от простой случайной величины ξ корректно, а именно, не зависит от представления ξ в виде $\xi = \sum\limits_{i=1}^{n} a_i I_{A_i}$. Это следует из приведенной далее леммы.

Лемма 9.1.1. *Пусть ξ — неотрицательная простая случайная величина, для которой имеют место представления:*

$$\xi = \sum_{i=1}^{n} a_i I_{A_i},\ \bigcup_{i=1}^{n} A_i = \Omega, A_i \cap A_j = \emptyset, i \neq j, A_i \in \mathfrak{F},$$

$$\xi = \sum_{j=1}^{m} b_j I_{B_j},\ \bigcup_{j=1}^{m} B_j = \Omega, B_i \cap B_j = \emptyset, i \neq j, B_j \in \mathfrak{F},$$

тогда

$$\sum_{i=1}^{n} a_i \mathsf{P}(A_i) = \sum_{j=1}^{m} b_j \mathsf{P}(B_j).$$

Доказательство. Положим

$$D_{ij} = A_i \cap B_j, i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m.$$

Очевидно

$$\Omega = \bigcup_{i,j} D_{ij},\ A_i = \bigcup_{j=1}^{m} D_{ij},\ B_j = \bigcup_{i=1}^{n} D_{ij}$$

(некоторые из D_{ij} могут быть пустыми). Значение ξ на непустом D_{ij} обозначим через d_{ij}, ясно, что $d_{ij} = a_i = b_j$. Поэтому

$$\sum_{i,j} d_{ij} \mathsf{P}(D_{ij}) = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} \mathsf{P}(D_{ij}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} d_{ij} \mathsf{P}(D_{ij}) \right) =$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_i \mathsf{P}(D_{ij}) \right) = \sum_{i=1}^{n} \left(a_i \sum_{j=1}^{m} \mathsf{P}(D_{ij}) \right) = \sum_{i=1}^{n} a_i \mathsf{P}(A_i).$$

Аналогично устанавливаем, что

$$\sum_{i,j} d_{ij} \mathsf{P}(D_{ij}) = \sum_{j=1}^{m} b_j \mathsf{P}(B_j).$$

Тем самым лемма доказана.

Далее приводятся основные свойства математического ожидания от неотрицательной простой случайной величины.

Линейность математического ожидания. Эти свойства легко получаются из свойств конечных сумм.

Свойство 1. *Если k — постоянная, то*

$$\mathsf{M}k\xi = k\mathsf{M}\xi,$$

или, в интегральной форме,

$$\int_\Omega k\xi(\omega)\mathsf{P}(d\omega) = k\int_\Omega \xi(\omega)\mathsf{P}(d\omega)$$

(*константа выносится из-под знака математического ожидания*).

Свойство 2. *Пусть ξ_1 и ξ_2 — неотрицательные случайные величины, тогда*

$$\mathsf{M}(\xi_1 + \xi_2) = \mathsf{M}\xi_1 + \mathsf{M}\xi_2,$$

или, в интегральной форме,

$$\int_\Omega (\xi_1(\omega) + \xi_2(\omega))\mathsf{P}(d\omega) = \int_\Omega \xi_1(\omega)\mathsf{P}(d\omega) + \int_\Omega \xi_2(\omega)\mathsf{P}(d\omega)$$

(*математическое ожидание суммы случайных величин равно сумме математических ожиданий слагаемых*).

Доказательство. Так как ξ_1 и ξ_2 — простые случайные величины, то для них имеют место представления

$$\xi_1 = \sum_{i=1}^{n} a_i I_{A_i},\ \ \xi_2 = \sum_{j=1}^{m} b_j I_{B_j}.$$

Обозначим

$$D_{ij} = A_i \cap B_j,\ \ i = 1, 2, \ldots, n; j = 1, 2, \ldots, m.$$

На непустом D_{ij} сумма $\xi_1 + \xi_2$ постоянна и равна $a_i + b_j$. По определению $\mathsf{M}(\xi_1 + \xi_2)$ имеем:

$$\mathsf{M}(\xi_1 + \xi_2) = \sum_{i,j}(a_i + b_j)\mathsf{P}(D_{ij}) = \sum_{i,j} a_i\mathsf{P}(D_{ij}) + \sum_{i,j} b_j\mathsf{P}(D_{ij}) =$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_i \mathsf{P}(D_{ij}) \right) + \sum_{j=1}^{m} \left(\sum_{i=1}^{n} b_j \mathsf{P}(D_{ij}) \right) =$$

$$= \sum_{i=1}^{n} a_i \mathsf{P}(A_i) + \sum_{j=1}^{m} b_j \mathsf{P}(B_j) = \mathsf{M}\xi_1 + \mathsf{M}\xi_2.$$

Аддитивность математического ожидания как функции множества. Пусть ξ — неотрицательная простая случайная величина (произвольная, но фиксированная) со значениями в $\mathbb{R}^1$ на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$. Математическое ожидание $\mathsf{M}(\xi I_B)$, $B \in \mathfrak{F}$, является функцией множества на $\mathfrak{F}$. Функцией множества является и интеграл

$$\int_{\Omega} \xi(\omega) I_B(\omega) \mathsf{P}(d\omega),$$

который будем обозначать через $\int_B \xi(\omega)\mathsf{P}(d\omega)$ и и называть интегралом Лебега по множеству B. Заметим, что

$$\mathsf{M}(\xi I_B) = \int_B \xi(\omega) \mathsf{P}(d\omega).$$

Непосредственно из определения математического ожидания от простой случайной величины имеем:

$$\mathsf{M}(c I_B) = \int_{\Omega} c I_B(\omega) \mathsf{P}(d\omega) = c\mathsf{P}(B).$$

Свойство аддитивности. *Если* $B_1 \cap B_2 = \emptyset$ $(B_1, B_2 \in \mathfrak{F})$, *то*

$$\mathsf{M}(\xi I_{B_1 \cup B_2}) = \mathsf{M}(\xi I_{B_1}) + \mathsf{M}(\xi I_{B_2}),$$

или, в интегральной форме,

$$\int_{B_1 \cup B_2} \xi(\omega) \mathsf{P}(d\omega) = \int_{B_1} \xi(\omega) \mathsf{P}(d\omega) + \int_{B_2} \xi(\omega) \mathsf{P}(d\omega).$$

Другими словами, *математическое ожидание*

$$\mathsf{M}(\xi I_B) = \int\limits_B \xi(\omega)\mathsf{P}(d\omega), B \in \mathfrak{F},$$

как функция множества аддитивно.

Доказательство. Так как $B_1 \cap B_2 = \emptyset$, то

$$I_{B_1 \cup B_2} = I_{B_1} + I_{B_2}.$$

Поэтому

$$\mathsf{M}(\xi I_{B_1 \cup B_2}) = \mathsf{M}(\xi\,(I_{B_1} + I_{B_2})) = \mathsf{M}(\xi I_{B_1}) + \mathsf{M}(\xi I_{B_2}).$$

Свойство монотонности математического ожидания. *Если* $0 \le \xi \le \eta$, *то*

$$\mathsf{M}\xi \le \mathsf{M}\eta,$$

или, в интегральной форме,

$$\int\limits_\Omega \xi(\omega)\mathsf{P}(d\omega) \le \int\limits_\Omega \eta(\omega)\mathsf{P}(d\omega).$$

Доказательство. Представим случайную величину η в виде суммы неотрицательных случайных величин:

$$\eta = \xi + \zeta,$$

где

$$\zeta(\omega) = \begin{cases} \eta(\omega) - \xi(\omega), & \text{если } \xi(\omega) < +\infty; \\ 0, & \text{если } \xi(\omega) = +\infty. \end{cases}$$

Отсюда получаем:

$$\mathsf{M}\eta = \mathsf{M}(\xi + \zeta) = \mathsf{M}\xi + \mathsf{M}\zeta.$$

Далее, если $\mathsf{M}\xi = +\infty$, то и $\mathsf{M}\eta = +\infty$, поэтому

$$\mathsf{M}\xi \le \mathsf{M}\eta.$$

Если $\mathsf{M}\xi < \infty$, то $\mathsf{M}\eta - \mathsf{M}\xi = \mathsf{M}\zeta \ge 0$, т. е.

$$\mathsf{M}\xi \le \mathsf{M}\eta.$$

Тем самым свойство установлено.

Определение. *Положительной частью* случайной величины ξ будем называть

$$\xi^+ = \max\{\xi, 0\},$$

отрицательной частью —

$$\xi^- = -\min\{\xi, 0\}.$$

Случайную величину ξ всегда можно представить в виде разности двух неотрицательных случайных величин:

$$\xi = \xi^+ - \xi^-.$$

Определение. Для случайной величины ξ, у которой хотя бы одно из чисел $\mathsf{M}\xi^+$ или $\mathsf{M}\xi^-$ отлично от бесконечности, $\mathsf{M}\xi$ определяется равенством

$$\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^-.$$

Если $\mathsf{M}\xi^+ = +\infty$ и $\mathsf{M}\xi^- = +\infty$, то математическое ожидание случайной величины ξ не существует (не определено).

Приведенные выше свойства математического ожидания имеют место и для случайных величин, принимающих значения обоих знаков.

9.2 Математическое ожидание (общий случай)

Далее определим математическое ожидание (интеграл Лебега) от случайной величины со значениям в $\mathbb{R}^1$ в общем случае.

Поскольку любую случайную величину ξ всегда можно представить в виде разности двух неотрицательных:

$$\xi = \xi^+ - \xi^-,$$

то достаточно иметь "хорошее" определение математического ожидания от неотрицательной случайной величины. Математическое ожидание от произвольной случайной величины

$$\xi = \xi^+ - \xi^-,$$

определим равенством

$$\mathsf{M}\xi = \mathsf{M}\xi^{+} - \mathsf{M}\xi^{-},$$

при условии, что хотя бы одно из чисел $\mathsf{M}\xi^{+}$ или $\mathsf{M}\xi^{-}$ конечно.

Математическое же ожидание (интеграл Лебега) от неотрицательной случайной величины ξ естественно определить так: построить последовательность $\{\xi_n\}$ неотрицательных простых случайных величин, сходящуюся к ξ, и положить

$$\mathsf{M}\xi = \lim_{n} \mathsf{M}\xi_n.$$

При этом необходимо убедиться, что определение корректно, а именно:

1° для каждой неотрицательной случайной величины ξ можно указать последовательность $\{\xi_n\}$ неотрицательных простых случайных величин, сходящуюся к ξ;

2° $\lim \mathsf{M}\xi_n$ существует;

3° $\lim \mathsf{M}\xi_n$ зависит только от ξ и не зависит от последовательности $\{\xi_n\}$, аппроксимирующей ξ.

К тому же так определенное математическое ожидание случайной величины должно обладать "хорошими" свойствами: линейности, монотонности, аддитивности и т. д.

Далее реализуется намеченная программа построения математического ожидания.

Аппроксимирующая последовательность. Сначала установим, что любую случайную величину можно сколь угодно точно аппроксимировать простыми случайными величинами.

Лемма 9.2.1 (об аппроксимации). *Для неотрицательной случайной величины $\xi = \xi(\omega)$ существует монотонно неубывающая последовательность $\{\xi_n\}$ неотрицательных простых случайных величин, сходящаяся к $\xi(\omega)$ в каждой точке $\omega \in \Omega$.*

Доказательство. По $\xi = \xi(\omega)$ для каждого n $(n = 1, 2, \ldots)$ определим $\xi_n = \xi_n(\omega)$ следующим образом. Разобьем промежуток $[0, n)$ множества возможных значений $[0, +\infty]$ случайной величины ξ точками $j/2^n$, $j = 1, 2, \ldots, 2^n n$, на непересекающиеся части $[(j-1)/2^n, j/2^n)$, $j = 1, 2, \ldots, 2^n n$. Определим $\xi_n = \xi_n(\omega)$ в точке $\omega \in \Omega$ так: если значение $\xi(\omega) \geq n$, то $\xi_n(\omega)$ в этой точке положим равным n; если $\xi(\omega) < n$, а следовательно, $\xi(\omega)$ попадает в один из промежутков $[(j-1)/2^n, j/2^n)$, $j = 1, 2, \ldots, 2^n n$, то значение $\xi_n(\omega)$ полагаем равным $(j-1)/2^n$. Тем самым $\xi_n = \xi_n(\omega)$ определена (см. рис. 9.2.1). (Случайная величина $\xi_n = \xi_n(\omega)$ получается из $\xi = \xi(\omega)$ измерением значений ξ "линейкой" длиной n с ценой деления $1/2^n$.)

Убедимся, что построенная таким образом последовательность $\{\xi_n\}$ обладает перечисленными в лемме свойствами.

Очевидно, $\xi_n(\omega) \geq 0$ для каждого $\omega \in \Omega, n = 1, 2, \ldots$

Последовательность $\{\xi_n\}$ — монотонно неубывающая. Действительно, если в точке ω значение $\xi(\omega) < n$, то $\xi(\omega)$ принадлежит одному из промежутков $[(j-1)/2^n, j/2^n)$ $(j = 1, 2, \ldots, 2^n n)$, и по определению значение $\xi_n(\omega)$ случайной величины ξ_n в точке ω равно $(j-1)/2^n$. При построении ξ_{n+1} разбиение множества значений случайной величины ξ производится с шагом $1/2^{n+1}$, поэтому каждый из промежутков $[(j-1)/2^n, j/2^n)$ разбивается на два:

$$\left[\frac{j-1}{2^n}, \frac{j}{2^n}\right) = \left[\frac{j-1}{2^n}, \frac{j-1}{2^n} + \frac{1}{2^{n+1}}\right) \bigcup \left[\frac{j-1}{2^n} + \frac{1}{2^{n+1}}, \frac{j}{2^n}\right).$$

И коль скоро значение $\xi(\omega)$ принадлежит $[(j-1)/2^n, j/2^n)$, то оно будет принадлежать одному из промежутков:

$$\left[\frac{j-1}{2^n}, \frac{j-1}{2^n} + \frac{1}{2^{n+1}}\right) \text{ или } \left[\frac{j-1}{2^n} + \frac{1}{2^{n+1}}, \frac{j}{2^n}\right).$$

поэтому значение $\xi_{n+1}(\omega)$ равно $(j-1)/2^n$ или $(j-1)/2^n + 1/2^{n+1}$, а, следовательно,

$$\xi_n(\omega) \leq \xi_{n+1}(\omega).$$

Если в точке ω значение $\xi(\omega) \geq n$, то $\xi_n(\omega) = n$, а значение $\xi_{n+1}(\omega)$ в этой точке, очевидно, не меньше n.

Убедимся, что $\xi_n(\omega)$ сходится к $\xi(\omega)$ при $n \to \infty$ в каждой точке $\omega \in \Omega$.

Если в точке ω значение $\xi(\omega)$ конечно, то найдется такое натуральное n, что $\xi(\omega) < n$. При этом $\xi(\omega)$ принадлежит некоторому промежутку $[(j-1)/2^n, j/2^n)$, поэтому, по построению, $\xi_n(\omega)$ в точке ω равна $(j-1)/2^n$, и, следовательно,

$$|\xi(\omega) - \xi_n(\omega)| \leq \frac{1}{2^n}.$$

Последнее означает, что $\xi_n(\omega)$ сходится к $\xi(\omega)$ при $n \to \infty$.

Если в точке ω значение $\xi(\omega) = +\infty$, то для каждого n имеем $\xi(\omega) \geq n$, поэтому по определению $\xi_n(\omega) = n$. Последовательность $\xi_n(\omega) = n$ при $n \to +\infty$ сходится к $+\infty$, а значит, и к $\xi(\omega)$.

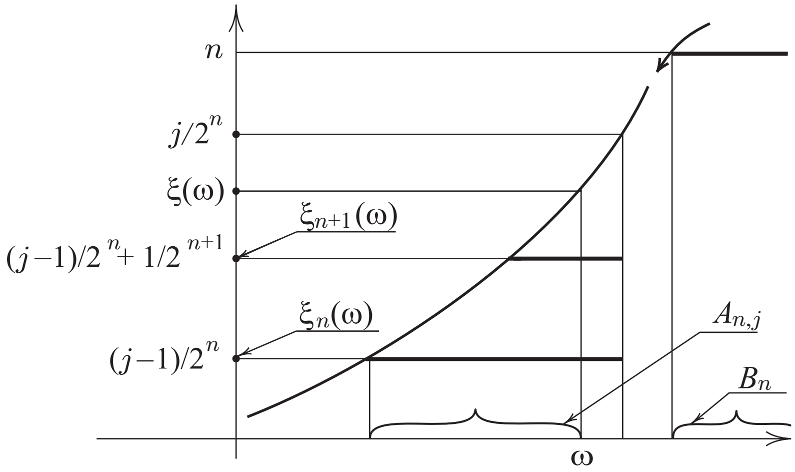

Рис. 9.2.1: Иллюстрация к построению ξ_n по ξ

Тем самым лемма доказана.

Замечание 1 к лемме. Если обозначить

$$B_n = \{\omega : \xi(\omega) \geq n\},$$

$$A_{n,j} = \left\{\omega : \frac{j-1}{2^n} \leq \xi(\omega) < \frac{j}{2^n}\right\},\ j = 1, 2, \ldots, 2^n n,$$

то случайную величину $\xi_n = \xi_n(\omega)$, аппроксимирующую ξ, можно представить в виде

$$\xi_n(\omega) = \sum_{j=1}^{2^n n} \frac{j-1}{2^n} I_{A_{n,j}}(\omega) + nI_{B_n}(\omega),\ \omega \in \Omega,$$

или короче

$$\xi_n = \sum_{j=0}^{2^n n} a_{n,j} I_{A_{n,j}}, \tag{9.2.1}$$

$$A_{n,0} = \{\omega : \xi(\omega) \geq n\},\ a_{n,0} = n,\ a_{n,j} = \frac{j-1}{2^n},\ j = 1, 2, \ldots, 2^n n.$$

Замечание 2 к лемме. Лемма утверждает, что значения случайной величины $\xi = \xi(\omega)$ можно измерить сколь угодно точно, выбирая достаточно малой цену деления “линейки”, которой производится измерение.

Следствие. *Для любой случайной величины $\xi = \xi(\omega)$ существует последовательность простых случайных величин, сходящаяся к ξ в каждой точке $\omega \in \Omega$.*

В самом деле, $\xi = \xi^+ - \xi^-$. Для ξ^+ существует монотонно неубывающая последовательность неотрицательных простых случайных величин $\eta_n = \eta_n(\omega)$, сходящаяся к ξ^+, причем $\eta_n(\omega) = 0$ на множестве $\{\omega : \xi(\omega) < 0\}$. Аналогично, существует монотонно неубывающая последовательность неотрицательных простых случайных величин $\zeta_n = \zeta_n(\omega)$, сходящаяся к ξ^-, причем $\zeta_n(\omega) = 0$ на множестве $\{\omega : \xi(\omega) > 0\}$. Тогда, очевидно, последовательность

$$\xi_n = \eta_n - \zeta_n$$

сходится к случайной величине $\xi = \xi^+ - \xi^-$ в каждой точке.

Существование $\lim\limits_n \mathsf{M}\xi_n$. Если $\{\xi_n\}$ — монотонно неубывающая последовательность неотрицательных случайных величин, то последовательность $\{\mathsf{M}\xi_n\}$ также монотонно неубывающая и, следовательно, существует предел $\lim\limits_n \mathsf{M}\xi_n$ (возможно равный $+\infty$).

Независимость $\lim\limits_n \mathsf{M}\xi_n$ от последовательности $\{\xi_n\}$. Пусть $\{\xi_n\}$ — монотонно неубывающая последовательность неотрицательных случайных величин, сходящаяся к ξ. Определение $\mathsf{M}\xi$ как $\lim\limits_n \mathsf{M}\xi_n$ будет корректным, если для любой другой монотонно неубывающей последовательности $\{\eta_n\}$ неотрицательных простых случайных величин, сходящейся к ξ, имеет место равенство

$$\lim_n \mathsf{M}\xi_n = \lim_n \mathsf{M}\eta_n.$$

Это равенство есть следствие приводимых далее лемм 9.2.2 и 9.2.3.

Лемма 9.2.2. *Пусть η — неотрицательная простая случайная величина, $\{\xi_n\}$ — монотонно неубывающая последовательность неотрицательных простых случайных величин, и*

$$B = \{\omega : \lim_n \xi_n(\omega) \geq \eta(\omega)\},$$

тогда

$$\lim_n \mathsf{M}(\xi_n I_B) \geq \mathsf{M}(\eta I_B),$$

или, в интегральной форме,

$$\lim_n \int\limits_B \xi_n(\omega)\mathsf{P}(d\omega) \geq \int\limits_B \eta(\omega)\mathsf{P}(d\omega).$$

Доказательство. Пусть $0 \leq h_1 < h_2 < \dots < h_s$ — значения, принимаемые случайной величиной η (не ограничивая общности, будем считать, что $h_1 > 0$, в противном случае неравенство достаточно доказать для множества $B \setminus \{\omega : \eta(\omega) = 0\}$). Далее, пусть $0 < \varepsilon < h_1/2$. Для каждого $m \geq 1$ определим множество

$$Q_m = \{\omega : \omega \in B, \xi_m(\omega) \geq \eta(\omega) - \varepsilon\}.$$

Поскольку $\{\xi_n\}$ — монотонно неубывающая последовательность, то: 1) на Q_m все ξ_n, начиная с номера m, будут больше или равны $\eta - \varepsilon$; 2) последовательность $\{Q_m\}$ монотонно возрастающая и

$$B = \bigcup_{m=1}^{\infty} Q_m.$$

При $n \geq m$, на множестве Q_m

$$\mathsf{M}(\xi_n I_B) \geq \mathsf{M}(\xi_n I_{Qm}) \geq \mathsf{M}((\eta - \varepsilon)I_{Qm}) = \mathsf{M}(\eta I_{Qm}) - \mathsf{M}(\varepsilon I_{Qm}) =$$

$$= \mathsf{M}(\eta I_{Qm}) - \varepsilon\mathsf{P}(Q_m) \geq \mathsf{M}(\eta I_{Qm}) - \varepsilon\mathsf{P}(B).$$

Так что

$$\mathsf{M}(\xi_n I_B) \geq \mathsf{M}(\eta I_{Qm}) - \varepsilon\mathsf{P}(B).$$

Отсюда, заменив $\mathsf{M}(\eta I_{Qm})$ его выражением из равенства

$$\mathsf{M}(\eta I_B) = \mathsf{M}(\eta I_{Qm}) + \mathsf{M}(\eta I_{B\setminus Q_m}),$$

получим:

$$\mathsf{M}(\xi_n I_B) \geq \mathsf{M}(\eta I_B) - \mathsf{M}(\eta I_{B\setminus Q_m}) - \varepsilon\mathsf{P}(B) \geq$$

$$\geq \mathsf{M}(\eta I_B) - h_s\mathsf{P}(B \setminus Q_m) - \varepsilon\mathsf{P}(B).$$

Перейдем к пределу при $m \to \infty$ (так как $n \geq m$, то и $n \to \infty$). При этом в силу непрерывности вероятности $\lim\limits_m \mathsf{P}(B \setminus Q_m) = 0$. Поэтому

$$\lim_n \mathsf{M}(\xi_{nB}) \geq \mathsf{M}(\eta I_B) - \varepsilon\mathsf{P}(B),$$

и, так как ε произвольно, то

$$\lim_n \mathsf{M}(\xi_n I_B) \geq \mathsf{M}(\eta I_B).$$

Лемма доказана.

Лемма 9.2.3. *Если $\{\xi_n\}$ и $\{\eta_n\}$ — монотонно неубывающие последовательности неотрицательных простых случайных величин и*

$$B = \left\{\omega : \lim_n \xi_n(\omega) = \lim_n \eta_n(\omega)\right\},$$

то

$$\lim_n \mathsf{M}(\xi_n I_B) = \lim_n \mathsf{M}(\eta_n I_B),$$

или, в интегральной форме,

$$\lim_n \int\limits_B \xi_n(\omega)\mathsf{P}(d\omega) = \lim_n \int\limits_B \eta_n(\omega)\mathsf{P}(d\omega).$$

Доказательство. Из условия леммы следует, что на множестве B для каждого натурального m

$$\lim_n \xi_n(\omega) \geq \eta_m(\omega).$$

Поэтому в силу леммы 9.2.2

$$\lim_n \mathsf{M}(\xi_n I_B) \geq \mathsf{M}(\eta_m I_B),$$

и так как m произвольно, то

$$\lim_n \mathsf{M}(\xi_n I_B) \geq \lim_m \mathsf{M}(\eta_m I_B).$$

Но последовательность $\{\eta_n\}$ ничем не хуже и не лучше последовательности $\{\xi_n\}$, поэтому

$$\lim_m \mathsf{M}(\eta_m I_B) \geq \lim_n \mathsf{M}(\xi_n I_B).$$

Так что

$$\lim_n \mathsf{M}(\xi_n I_B) = \lim_m \mathsf{M}(\eta_m I_B).$$

Теперь можно определить математическое ожидание от произвольной случайной величины.

Определение. Математическим ожиданием $\mathsf{M}\xi$ от неотрицательной случайной величины $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^1$, заданной на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, будем называть предел

$$\lim_n \mathsf{M}\xi_n = \mathsf{M}\xi,$$

где $\{\xi_n\}$ — монотонно неубывающая последовательность неотрицательных простых случайных величин, сходящаяся к ξ.

Если случайная величина ξ принимает значения обоих знаков, и хотя бы одно из чисел $\mathsf{M}\xi^+$ или $\mathsf{M}\xi^-$ конечно, то $\mathsf{M}\xi$ определяется равенством

$$\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^-.$$

Если $\mathsf{M}\xi^+$ и $\mathsf{M}\xi^-$ равны бесконечности, то $\mathsf{M}\xi$ не определено (не существует).

Согласно определению, математическое ожидание может принимать бесконечные значения.

Интеграл Лебега $\int\limits_\Omega \xi(\omega)\mathsf{P}(d\omega)$ по мере P от неотрицательной случайной величины $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^1$, заданной на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, определяется равенством

$$\int\limits_\Omega \xi(\omega)\mathsf{P}(d\omega) = \lim_n \int\limits_\Omega \xi_n(\omega)\mathsf{P}(d\omega),$$

где $\{\xi_n\}$ — монотонно неубывающая последовательность неотрицательных простых случайных величин, сходящаяся к ξ.

Если ξ принимает значения обоих знаков и хотя бы одно из чисел $\int\limits_\Omega \xi^+(\omega)\mathsf{P}(d\omega)$ или $\int\limits_\Omega \xi^-(\omega)\mathsf{P}(d\omega)$ конечно, то $\int\limits_\Omega \xi(\omega)\mathsf{P}(d\omega)$ определяется равенством:

$$\int\limits_\Omega \xi(\omega)\mathsf{P}(d\omega) = \int\limits_\Omega \xi^+(\omega)\mathsf{P}(d\omega) - \int\limits_\Omega \xi^-(\omega)\mathsf{P}(d\omega).$$

Если $\int\limits_\Omega \xi^+(\omega)\mathsf{P}(d\omega)$ и $\int\limits_\Omega \xi^-(\omega)\mathsf{P}(d\omega)$ равны бесконечности, то интеграл Лебега не определен (не существует).

Интеграл Лебега $\int\limits_B \xi(\omega)\mathsf{P}(d\omega)$ по множеству $B \in \mathfrak{F}$ определяем равенством

$$\int\limits_B \xi(\omega)\mathsf{P}(d\omega) = \int\limits_\Omega \xi(\omega) I_B(\omega)\mathsf{P}(d\omega).$$

З а м е ч а н и я к определению интеграла Лебега.

1° Рассмотренное выше построение интеграла Лебега по мере P от случайной величины $\xi = \xi(\omega)$, заданной на вероятностном пространстве $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, без изменений переносится на $\mathfrak{F}$-измеримые функции $f = f(\omega)$, заданные на пространстве с мерой $\{\Omega, \mathfrak{F}, \mu\}$. Причем мера μ может быть как конечной, так и бесконечной ($\mu(\Omega) = +\infty$), а функция f может принимать бесконечные значения. Обозначать интеграл Лебега по мере μ от функции $f = f(\omega)$ будем так: $\int\limits_{\Omega} f(\omega)\mu(d\omega)$.

В частности, если $\Omega = \mathbb{R}^n, \mathfrak{F} = \mathfrak{B}^n$, μ — мера Лебега на $\mathfrak{B}^n$, то интеграл Лебега от $f(x)$ обозначают через $\int\limits_{\mathbb{R}^n} f(x)\mathsf{L}(dx)$ или $\int\limits_{\mathbb{R}^n} f(x)dx$, подробнее так: $\int\limits_{\mathbb{R}^n} f(x_1, x_2, \ldots, x_n)d(x_1, x_2, \ldots, x_n)$, при $n = 1$ так $\int\limits_{\mathbb{R}^1} f(x)d(x)$.

2° Пусть $\{\Omega, \mathfrak{F}, \mu\}$ — пространство с мерой, $B \in \mathfrak{F}, \mu(B) > 0$. Наряду с пространством $\{\Omega, \mathfrak{F}, \mu\}$ можно рассматривать и пространство с мерой $\{B, \mathfrak{F}_B, \mu\}$, $\mathfrak{F}_B$ образовано множествами вида $B \cap A$, где $A \in \mathfrak{F}$. Причем интеграл $\int\limits_{B} f(\omega)\mu(d\omega)$ от $f(\omega)$, заданной на $\{B, \mathfrak{F}_B, \mu\}$, совпадает с интегралом $\int\limits_{\Omega} f(\omega)I_B(\omega)\mu(d\omega)$ от функции $f(\omega)I_B(\omega)$, заданной на $\{\Omega, \mathfrak{F}, \mu\}$.

Действительно, не нарушая общности, будем считать, что $f \geq 0$. Пусть $\{f_n\}$ — монотонно неубывающая последовательность неотрицательных $\mathfrak{F}$-измеримых функций, сходящихся к f на Ω. По определению

$$\int\limits_{\Omega} f(\omega)I_B(\omega)\mu(d\omega) = \lim_n \int\limits_{\Omega} f_n(\omega)I_B(\omega)\mu(d\omega) =$$

$$= \lim_n \left(\sum_{j=1}^{2^n n} \frac{j-1}{2^n}\mu(A_j \cap B) + n\mu(B_n \cap B) \right).$$

Далее fI_B — это $\mathfrak{F}_B$-измеримая функция на $\{B, \mathfrak{F}_B, \mu\}$; $\{f_n I_B\}$ — монотонно неубывающая на B последовательность неотрицательных простых измеримых функций, сходящихся к fI_B, поэтому по определению интеграла Лебега от fI_B на $\{B, \mathfrak{F}_B, \mu\}$

$$\int\limits_B f(\omega)\mu(d\omega) = \lim_n \int\limits_B f_n(\omega)\mu(d\omega) =$$

$$= \lim_n \left(\sum_{j=1}^{2^n n} \frac{j-1}{2^n} \mu(A_j \cap B) + n\mu(B_n \cap B) \right).$$

И, следовательно, интеграл $\int\limits_B f(\omega)\mu(d\omega)$ от функции $f(\omega)$, заданной на $\{B, \mathfrak{F}_B, \mu\}$, совпадает с интегралом $\int\limits_\Omega f(\omega)I_B(\omega)\mu(d\omega)$ от функции $f(\omega)I_B(\omega)$, заданной на $\{\Omega, \mathfrak{F}, \mu\}$.

Приводимые далее свойства аддитивности и линейности математического ожидания получаются предельным переходом из соответствующих свойств математического ожидания для простых случайных величин.

Теорема 9.2.1 (свойство аддитивности). *Если* $B_1 \cap B_2 = \emptyset$, $B_1, B_2 \in \mathfrak{F}$, *то*

$$\mathsf{M}(\xi I_{B_1 \cup B_2}) = \mathsf{M}(\xi I_{B_1}) + \mathsf{M}(\xi I_{B_2}),$$

или, в интегральной форме,

$$\int\limits_{B_1 \cup B_2} \xi(\omega)\mathsf{P}(d\omega) = \int\limits_{B_1} \xi(\omega)\mathsf{P}(d\omega) + \int\limits_{B_2} \xi(\omega)\mathsf{P}(d\omega)$$

при условии, что в правой части не стоят бесконечности разных знаков.

Теорема 9.2.2. *Если* α *— константа, то*

$$\mathsf{M}(\alpha\xi I_B) = \alpha\mathsf{M}(\xi I_B),$$

или, в интегральной форме,

$$\int\limits_B \alpha\xi(\omega)\mathsf{P}(d\omega) = \alpha \int\limits_B \xi(\omega)\mathsf{P}(d\omega),\ B \in \mathfrak{F}$$

(*константа выносится из-под знака математического ожидания*).

Теорема 9.2.3.

$$\mathsf{M}((\xi+\eta)I_B) = \mathsf{M}(\xi I_B) + \mathsf{M}(\eta I_B),$$

или, в интегральной форме,

$$\int\limits_B (\xi(\omega) + \eta(\omega)\mathsf{P}(d\omega) = \int\limits_B \xi(\omega)\mathsf{P}(d\omega) + \int\limits_B \eta(\omega)\mathsf{P}(d\omega),$$

если в правой части не стоят бесконечности разных знаков (математическое ожидание суммы случайных величин равно сумме их математических ожиданий).

Свойства монотонности математического ожидания. Эти свойства доказываются аналогично доказательству свойств монотонности для простых случайных величин.

Теорема 9.2.4. *Если $\xi \geq 0$ на множестве $B \in \mathfrak{F}$, то*

$$\mathsf{M}(\xi I_B) \geq 0$$

(математическое ожидание неотрицательной случайной величины неотрицательно), или, в интегральной форме,

$$\int\limits_B \xi(\omega)\mathsf{P}(d\omega) \geq 0.$$

Доказательство очевидно.

Теорема 9.2.5. *Если $\xi \leq \eta$ на B, то*

$$\mathsf{M}(\xi I_B) \leq \mathsf{M}(\eta I_B),$$

или, в интегральной форме,

$$\int\limits_B \xi(\omega)\mathsf{P}(d\omega) \leq \int\limits_B \eta(\omega)\mathsf{P}(d\omega)$$

Следствие. *Если $\mathsf{M}\xi$ существует, то*

$$|\mathsf{M}\xi| \leq \mathsf{M}|\xi|,$$

или, в интегральной форме,

$$\left|\int\limits_\Omega \xi(\omega)\mathsf{P}(d\omega)\right| \leq \int\limits_\Omega |\xi(\omega)|\,\mathsf{P}(d\omega).$$

Из неравенств
$$-|\xi| \le \xi \le |\xi|,$$
имеем:
$$-\mathsf{M}|\xi| \le \mathsf{M}\xi \le \mathsf{M}|\xi|$$
или, что то же,
$$|\mathsf{M}\xi| \le \mathsf{M}|\xi|.$$

Теорема 9.2.6 (интеграл по множеству P-меры нуль). *Если $\mathsf{P}(B) = 0$, то для любой случайной величины ξ ($\mathfrak{F}$-измеримой функции)*
$$\mathsf{M}(\xi I_B) = 0$$
или в интегральной форме:
$$\int\limits_B \xi(\omega)\mathsf{P}(d\omega) = 0$$
(*интеграл по множеству* P-*меры нуль равен нулю*).

Доказательство. Так как
$$\mathsf{M}(\xi I_B) = \mathsf{M}(\xi^+ I_B) - \mathsf{M}(\xi^- I_B),$$
то утверждение достаточно установить для неотрицательной случайной величины ξ.

Пусть
$$\xi_n = \sum_{j=0}^{2^n n} a_{n,j} I_{A_{n,j}}$$
— монотонно неубывающая последовательность неотрицательных случайных величин, сходящаяся к ξ (см. лемму 9.2.1 и представление (9.2.1)). Тогда
$$\xi_n I_B = I_B \sum_j a_{n,j} I_{A_{n,j}} = \sum_j a_{n,j} I_{A_{n,j}} I_B = \sum_j a_{n,j} I_{A_{n,j} \bigcap B},$$
а так как $\mathsf{P}(B) = 0$, то
$$\mathsf{M}(\xi_n I_B) = \sum_j a_{n,j} \mathsf{P}(A_{n,j} \cap B) = 0,$$

и, следовательно,

$$\mathsf{M}(\xi I_B) = \lim_n \mathsf{M}(\xi_n I_B) = 0.$$

Эквивалентные случайные величины. Случайные величины ($\mathfrak{F}$-измеримые функции) ξ и η будем называть *эквивалентными* на множестве B, если

$$\mathsf{P}\{\omega : \omega \in B, \xi(\omega) \neq \eta(\omega)\} = 0.$$

Теорема 9.2.7. *Если случайные величины ($\mathfrak{F}$-измеримые функции) ξ и η эквивалентны на B, то*

$$\mathsf{M}(\xi I_B) = \mathsf{M}(\eta I_B)$$

(*математические ожидания эквивалентных случайных величин равны*), *или, в интегральной форме,*

$$\int\limits_B \xi(\omega)\mathsf{P}(d\omega) = \int\limits_B \eta(\omega)\mathsf{P}(d\omega).$$

Доказательство.

$$\mathsf{M}(\xi I_B) = \mathsf{M}((\eta + (\xi - \eta))I_B) = \mathsf{M}(\eta I_B) + \mathsf{M}((\xi - \eta)I_B),$$

$$\mathsf{M}((\xi - \eta)I_B) = \mathsf{M}((\xi - \eta)I_{\{\omega\in B,\xi(\omega)=\eta(\omega)\}}) +$$

$$+\mathsf{M}((\xi - \eta)I_{\{\omega\in B,\xi(\omega)\neq\eta(\omega)\}}) =$$

$$= \mathsf{M}(0 \cdot I_{\{\omega\in B,\xi(\omega)=\eta(\omega)\}}) + \mathsf{M}((\xi - \eta)I_{\{\omega\in B,\xi(\omega)\neq\eta(\omega)\}}) = 0,$$

$\mathsf{M}((\xi-\eta)I_{\{\omega\in B,\xi(\omega)\neq\eta(\omega)\}}) = 0$ поскольку $\mathsf{P}\{\omega \in B, \xi(\omega) \neq \eta(\omega)\} = 0$.

Определение. Будем говорить, что некоторое свойство выполняется с вероятностью 1, если оно имеет место для всех точек $\omega \in \Omega$ пространства $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, исключая, быть может, множество точек P-меры нуль.

Интегрируемые случайные величины. Будем говорить, что случайная величина ($\mathfrak{F}$-измеримая функция) ξ *интегрируема* на множестве B по мере P (не обязательно вероятностной), если

$$\left|\int\limits_B \xi(\omega)\mathsf{P}(d\omega)\right| < \infty.$$

Теорема 9.2.8. *Случайная величина ξ ($\mathfrak{F}$-измеримая функция) интегрируема на множестве B тогда и только тогда, когда интегрируема случайная величина $|\xi|$.*

Доказательство. Интеграл

$$\int\limits_B \xi(\omega)\mathsf{P}(d\omega) = \int\limits_B \xi^+(\omega)\mathsf{P}(d\omega) - \int\limits_B \xi^-(\omega)\mathsf{P}(d\omega)$$

конечен (случайная величина ξ — интегрируема), тогда и только тогда, когда конечны интегралы $\int\limits_B \xi^+(\omega)\mathsf{P}(d\omega)$ и $\int\limits_B \xi^-(\omega)\mathsf{P}(d\omega)$.

Интеграл

$$\int\limits_B |\xi|(\omega)\mathsf{P}(d\omega) = \int\limits_B \xi^+\mathsf{P}(d\omega) + \int\limits_B \xi^-(\omega)\mathsf{P}(d\omega)$$

конечен (случайная величина $|\xi|$ — интегрируема) тогда и только тогда, когда конечны интегралы $\int\limits_B \xi^+(\omega)\mathsf{P}(d\omega)$ и $\int\limits_B \xi^-(\omega)\mathsf{P}(d\omega)$.

Определение. Будем говорить, что случайная величина ξ *мажорируема* на множестве B случайной величиной η, если на B

$$|\xi| \leq \eta.$$

Теорема 9.2.9. *Случайная величина ξ, мажорируемая на множестве B интегрируемой случайной величиной η, интегрируема на B.*

Доказательство. Пусть $|\xi| \leq \eta$ и $\int\limits_B \eta(\omega)\mathsf{P}(d\omega) < \infty$. Из неравенства $|\xi| \leq \eta$ следует неравенство

$$\int\limits_B |\xi(\omega)|\mathsf{P}(d\omega) \leq \int\limits_B \eta(\omega)\mathsf{P}(d\omega) < \infty,$$

поэтому $|\xi|$ — интегрируемая случайная величина, а вместе с ней интегрируема и ξ.

Среднее случайной величины и ее значения. Следующие две теоремы показывают, что некоторые выводы о значениях, принимаемых случайной величиной, можно сделать по ее математическому ожиданию (среднему).

Теорема 9.2.10. *Если математическое ожидание неотрицательной случайной величины равно нулю, то и сама случайная величина равна нулю (с вероятностью 1).*

Доказательство. Пусть $\xi(\omega) \geq 0$, $\omega \in \Omega$, и

$$B = \{\omega : \xi(\omega) > 0\},\ B_n = \{\omega : \xi(\omega) > 1/n\},\ n = 1, 2, \ldots$$

Очевидно

$$B = \bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) > 1/n\} = \bigcup_{n=1}^{\infty} B_n, \tag{9.2.2}$$

причем

$$B_n \subset B_{n+1}, n = 1, 2, \ldots$$

Далее

$$0 = \mathsf{M}\xi = \int_{\Omega} \xi(\omega)\mathsf{P}(d\omega) \geq$$

$$\geq \int_{B_n} \xi(\omega)\mathsf{P}(d\omega) \geq \int_{B_n} \frac{1}{n}\mathsf{P}(d\omega) = \frac{1}{n}\mathsf{P}(B_n),$$

поэтому $\mathsf{P}(B_n) = 0,\ n = 1, 2, \ldots$ Отсюда, учитывая представление (9.2.2) и свойство непрерывности вероятности, получаем:

$$\mathsf{P}\{\omega : \xi(\omega) > 0\} = \mathsf{P}(B) = \mathsf{P}\Big(\bigcup_{n=1}^{\infty} B_n\Big) = \lim_n \mathsf{P}(B_n) = 0.$$

Замечание. Аналогичное утверждение имеет место и для неположительных случайных величин.

Теорема 9.2.11. *Если математическое ожидание случайной величины отлично от $+\infty$, то и сама случайная величина отлична от $+\infty$ (с вероятностью 1)).*

Доказательство. Пусть $\mathsf{M}\xi \neq +\infty$. Поскольку $\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^-$, то соотношение $\mathsf{M}\xi \neq +\infty$ равносильно $\mathsf{M}\xi^+ \neq \neq +\infty$. Заметим еще, что $\{\omega : \xi(\omega) = +\infty\} = \{\omega : \xi^+(\omega) = +\infty\}$. Поэтому достаточно доказать, что если $\mathsf{M}\xi^+ \neq +\infty$, то и

$$\mathsf{P}\{\omega : \xi^+(\omega) \neq +\infty\} = 1.$$

или

$$\mathsf{P}\{\omega : \xi^+(\omega) = +\infty\} = 0.$$

По случайной величине ξ^+ определим последовательность $\{h_n\}$ случайных величин:

$$h_n(\omega) = \begin{cases} n, \text{ если } \xi^+(\omega) = +\infty, \\ 0, \text{ если } \xi^+(\omega) < +\infty. \end{cases}$$

Очевидно, $\xi^+ \geq h_n$ для каждого n, поэтому

$$\mathsf{M}\xi^+ \geq \mathsf{M}h_n = 0 \cdot \mathsf{P}\{\omega : \xi^+(\omega) < \infty\} + n \cdot \mathsf{P}\{\omega : \xi^+(\omega) = +\infty\}.$$

В предположении, что $\mathsf{P}\{\omega : \xi^+(\omega) = +\infty\} > 0$, правая часть последнего неравенства за счет выбора n может быть сделана сколь угодно большой, левая же часть $\mathsf{M}\xi^+$ конечна, поэтому $\mathsf{P}\{\omega : \xi^+(\omega) = +\infty\} = 0$.

Тем самым теорема полностью доказана.

Аналогично имеем.

Если $\mathsf{M}\xi \neq -\infty$, *то и*

$$\mathsf{P}\{\omega : \xi(\omega) \neq -\infty\} = 1$$

(*если математическое ожидание случайной величины отлично от* $-\infty$, *то и сама случайная величина отлична от* $-\infty$ (*с вероятностью* 1)).

Следствие. *Если* $\mathsf{M}|\xi| \neq \infty$, *то*

$$\mathsf{P}\{\omega : |\xi(\omega)| \neq \infty\} = 1.$$

Теорема 9.2.12 (мультипликативное свойство математического ожидания). *Если* ξ *и* η *— независимые случайные величины с конечными* $\mathsf{M}\xi$ *и* $\mathsf{M}\eta$, *то* $\mathsf{M}\xi\eta$ *также конечно и*

$$\mathsf{M}\xi\eta = \mathsf{M}\xi \cdot \mathsf{M}\eta$$

(*математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий*).

Доказательство. Учитывая, что $\xi = \xi^+ - \xi^-$ и $\eta = \eta^+ - \eta^-$, теорему достаточно доказать для неотрицательных случайных величин.

Пусть ξ и η — неотрицательные случайные величины и

$$\xi_n = \sum_{i=0}^{2^n n} a_{n,i} I_{A_{n,i}}, \ \ \eta_n = \sum_{j=0}^{2^n n} b_{n,j} I_{B_{n,j}}$$

— монотонно неубывающие последовательности неотрицательных простых случайных величин, сходящиеся соответственно к ξ, η (см. лемму 9.2.1 и представление (9.2.1)). Монотонно неубывающая последовательность $\{\xi_n\eta_n\}$ сходится к $\xi\eta$.

Покажем, что
$$\mathsf{M}\xi_n\eta_n = \mathsf{M}\xi_n\mathsf{M}\eta_n.$$

Поскольку случайные величины ξ и η — независимы, то независимы и события $A_{n,i}$ и $B_{n,j}$, так как $A_{n,0} = \{\omega : \xi(\omega) \geq n\}$, $A_{n,i} = \{\omega : (i-1)/2^n \leq \xi(\omega) < i/2^n\}$, $i = 1, 2, \ldots, 2^n n$, $B_{n,0} = \{\omega : \eta(\omega) \geq n\}$, $B_{n,j} = \{\omega : (j-1)/2^n \leq \eta(\omega) < j/2^n\}$, $j = 1, 2, \ldots, 2^n n$ (см. лемму 9.2.1). Отсюда получаем, что

$$\mathsf{M}I_{A_{n,i}}I_{B_{n,j}} = \mathsf{M}I_{A_{n,i}\bigcap B_{n,j}} = \mathsf{P}(A_{n,i} \cap B_{n,j}) =$$

$$= \mathsf{P}(A_{n,i})\mathsf{P}(B_{n,j}) = \mathsf{M}I_{A_{n,i}}\mathsf{M}I_{B_{n,j}}.$$

Далее

$$\mathsf{M}\xi_n\eta_n = \mathsf{M}\left(\left(\sum_{i=0}^{2^n n} a_{n,i}I_{A_{n,i}}\right)\cdot\left(\sum_{j=0}^{2^n n} b_{n,j}I_{B_{n,j}}\right)\right) =$$

$$= \mathsf{M}\left(\sum_{i=0}^{2^n n}\sum_{j=0}^{2^n n} a_{n,i}b_{n,j}I_{A_{n,i}}I_{B_{n,j}}\right) = \sum_{i=0}^{2^n n}\sum_{j=0}^{2^n n} a_{n,i}b_{n,j}\mathsf{M}I_{A_{n,i}}I_{B_{n,j}} =$$

$$= \sum_{i=0}^{2^n n}\sum_{j=0}^{2^n n} a_{n,i}b_{n,j}\mathsf{M}I_{A_{n,i}}\mathsf{M}I_{B_{n,j}} =$$

$$= \sum_{i=0}^{2^n n}\sum_{j=0}^{2^n n} a_{n,i}b_{n,j}\mathsf{P}\left(A_{n,i}\right)\mathsf{P}\left(B_{n,j}\right) =$$

$$= \left(\sum_{i=0}^{2^n n} a_{n,i}\mathsf{P}\left(A_{n,i}\right)\right)\left(\sum_{j=0}^{2^n n} b_{n,j}\mathsf{P}\left(B_{n,j}\right)\right) = \mathsf{M}\xi_n\cdot\mathsf{M}\eta_n.$$

Так что
$$\mathsf{M}\xi_n\eta_n = \mathsf{M}\xi_n\cdot\mathsf{M}\eta_n.$$

Переходя в правой и левой частях последнего равенства к пределу при $n \to \infty$, по определению математического ожидания от неотрицательной случайной величины, получаем:

$$\mathsf{M}\xi\eta = \mathsf{M}\xi\cdot\mathsf{M}\eta. \tag{9.2.3}$$

Из последнего равенства и неравенств $\mathsf{M}|\xi| < \infty, \mathsf{M}|\eta| < \infty$ имеем $\mathsf{M}|\xi\eta| < \infty$.

Произведение пространств с мерой. Пусть имеется два пространства с мерой: $\{\Omega^{(1)}, \mathfrak{F}^{(1)}, \mu^{(1)}\}$ и $\{\Omega^{(2)}, \mathfrak{F}^{(2)}, \mu^{(2)}\}$. Через $\Omega^{(1)} \times \Omega^{(2)}$ обозначим декартово произведение пространств $\Omega^{(1)}$ и $\Omega^{(2)}$ — множество упорядоченных пар $(\omega^{(1)}, \omega^{(2)})$, где $\omega^{(1)} \in \Omega^{(1)}$, $\omega^{(2)} \in \Omega^{(2)}$. Через $\mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}$ обозначим σ-алгебру подмножеств множества $\Omega^{(1)} \times \Omega^{(2)}$, порожденную "прямоугольниками" $A^{(1)} \times A^{(2)}$ ($A^{(1)} \in \mathfrak{F}^{(1)}$, $A^{(2)} \in \mathfrak{F}^{(2)}$), будем ее называть *произведением σ-алгебр* $\mathfrak{F}^{(1)}$ и $\mathfrak{F}^{(2)}$.

На σ-алгебре $\mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}$ существует, и притом единственная, мера ν такая, что

$$\nu(A^{(1)} \times A^{(2)}) = \mu^{(1)}(A^{(1)})\mu^{(2)}(A^{(2)}).$$

Меру ν называют *произведением мер* $\mu^{(1)}$ и $\mu^{(2)}$ и обозначают $\mu^{(1)} \times \mu^{(2)}$.

Пространство с мерой $\{\Omega, \mathfrak{F}, \nu\}$, где $\Omega = \Omega^{(1)} \times \Omega^{(2)}, \mathfrak{F} = \mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}, \nu = \mu^{(1)} \times \mu^{(2)}$ называется произведением пространств с мерой $\{\Omega^{(1)}, \mathfrak{F}^{(1)}, \mu^{(1)}\}$ и $\{\Omega^{(2)}, \mathfrak{F}^{(2)}, \mu^{(2)}\}$.

Пример. *Пространство с мерой $\{\mathbb{R}^2, \mathfrak{B}^2, \mathsf{L}^2\}$, где L^2 — мера Лебега на $\mathfrak{B}^2$, можно рассматривать как произведение пространств с мерой $\{\mathbb{R}^1, \mathfrak{B}^1, \mathsf{L}\}$ и $\{\mathbb{R}^1, \mathfrak{B}^1, \mathsf{L}\}$, где L — мера Лебега на прямой.*

Теорема 9.2.13 (Фубини). *Пусть $\{\Omega, \mathfrak{F}, \nu\}$ — произведение пространств с мерой $\{\Omega^{(1)}, \mathfrak{F}^{(1)}, \mu^{(1)}\}$ и $\{\Omega^{(2)}, \mathfrak{F}^{(2)}, \mu^{(2)}\}$: $\Omega = \Omega^{(1)} \times \Omega^{(2)}$, $\mathfrak{F} = \mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}$, $\nu = \mu^{(1)} \times \mu^{(2)}$; $f(\omega^{(1)}, \omega^{(2)})$ — $\mathfrak{F}^{(1)} \times \mathfrak{F}^{(2)}$-измеримая функция, заданной на $\Omega^{(1)} \times \Omega^{(2)}$.*

Если интеграл

$$\int\limits_{\Omega^{(1)} \times \Omega^{(2)}} f(\omega^{(1)}, \omega^{(2)})\nu(d(\omega^{(1)}, \omega^{(2)}))$$

конечен, то

$$\int\limits_{\Omega^{(1)} \times \Omega^{(2)}} f(\omega^{(1)}, \omega^{(2)})\nu(d(\omega^{(1)}, \omega^{(2)})) =$$

$$= \int\limits_{\Omega^{(2)}} \left(\int\limits_{\Omega^{(1)}} f(\omega^{(1)}, \omega^{(2)}) \mu^{(1)}(d\omega^{(1)}) \right) \mu^{(2)}(d\omega^{(2)}) =$$

$$= \int\limits_{\Omega^{(1)}} \left(\int\limits_{\Omega^{(2)}} f(\omega^{(1)}, \omega^{(2)}) \mu^{(2)}(d\omega^{(2)}) \right) \mu^{(1)}(d\omega^{(1)}).$$

Замечание. Если $f(x, y)$ — борелевская функция на $\mathbb{R}^2$, $\mu^{(1)} = \mathsf{L}$, $\mu^{(2)} = \mathsf{L}$ — меры Лебега на $\mathbb{R}^1$, $\mathsf{L}^2 = \mathsf{L} \times \mathsf{L}$ — произведение мер L и L, то

$$\int\limits_{\mathbb{R}^2} f(x, y) \mathsf{L}^2(d(x, y)) =$$

$$= \int\limits_{\mathbb{R}^1} \left(\int\limits_{\mathbb{R}^1} f(x, y) \mathsf{L}(dx) \right) \mathsf{L}(dy) = \int\limits_{\mathbb{R}^1} \left(\int\limits_{\mathbb{R}^1} f(x, y) \mathsf{L}(dy) \right) \mathsf{L}(dx)$$

или в привычной записи

$$\int\limits_{\mathbb{R}^2} f(x, y) d(x, y) = \int\limits_{\mathbb{R}^1} \left(\int\limits_{\mathbb{R}^1} f(x, y) dx \right) dy = \int\limits_{\mathbb{R}^1} \left(\int\limits_{\mathbb{R}^1} f(x, y) dy \right) dx.$$

9.3 Неравенства для математических ожиданий

Определение. Функцию $g(x)$ со значениями в $\mathbb{R}^1$, заданную на промежутке $I \subset \mathbb{R}^1$, будем называть выпуклой на I, если для любых $x_1, x_2 \in I$ и $t \in [0, 1]$

$$g((1-t)x_1 + tx_2) \leq (1-t)g(x_1) + tg(x_2)$$

и вогнутой на I, если

$$g((1-t)x_1 + tx_2) \geq (1-t)g(x_1) + tg(x_2).$$

Заметим, что

$$x = (1-t)x_1 + tx_2,\ y = (1-t)g(x_1) + tg(x_2),\ 0 \le t \le 1$$

— параметрическое уравнение отрезка, соединяющего точки $(x_1, g(x_1))$ и $(x_2, g(x_2))$.

Теорема 9.3.1 (неравенство Гёльдера). *Пусть действительные числа* $p > 0$, $q > 0$ *удовлетворяют соотношению* $1/p + 1/q = 1$, *тогда*

$$\mathsf{M}|\xi\eta| \le (\mathsf{M}|\xi|^p)^{1/p} (\mathsf{M}|\eta|^q)^{1/q}, \tag{9.3.1}$$

или, в интегральной форме,

$$\int\limits_\Omega |\xi(\omega)\eta(\omega)|\mathsf{P}(d\omega) \le \left(\int\limits_\Omega |\xi(\omega)|^p\mathsf{P}(d\omega)\right)^{1/p} \left(\int\limits_\Omega |\eta(\omega)|^q\mathsf{P}(d\omega)\right)^{1/q}.$$

Доказательство. Если хотя бы одно из чисел $\mathsf{M}|\xi|^p$ или $\mathsf{M}|\eta|^q$ равно нулю, то

$$\mathsf{P}\{\omega : |\xi(\omega)|^p = 0\} = 1$$

или

$$\mathsf{P}\{\omega : |\eta(\omega)|^q = 0\} = 1,$$

см. теорему 9.2.10. Отсюда $\mathsf{P}\{\omega : |\xi(\omega)\eta(\omega)| = 0\} = 1$, а значит, $\mathsf{M}|\xi(\omega)\eta(\omega)| = 0$ и неравенство Гёльдера имеет место.

Если хотя бы одно из математических ожиданий $\mathsf{M}|\xi|^p$ или $\mathsf{M}|\eta|^q$ равно бесконечности, то неравенство очевидно.

Пусть $0 < \mathsf{M}|\xi|^p < \infty$, $0 < \mathsf{M}|\eta|^q < \infty$. Для доказательства неравенства (9.3.1) достаточно доказать, что

$$\mathsf{M}\frac{|\xi|}{(\mathsf{M}|\xi|^p)^{1/p}}\frac{|\eta|}{(\mathsf{M}|\eta|^q)^{1/q}} \le 1,$$

или, обозначив,

$$\tilde{\xi} = \frac{|\xi|}{(\mathsf{M}|\xi|^p)^{1/p}} \quad \text{и} \quad \tilde{\eta} = \frac{|\eta|}{(\mathsf{M}|\eta|^q)^{1/q}},$$

что

$$\mathsf{M}\tilde{\xi}\tilde{\eta} \le 1;$$

заметим, что

$$\tilde{\xi} > 0, \tilde{\eta} > 0,\ \mathsf{M}\tilde{\xi}^p = 1, \mathsf{M}\tilde{\eta}^q = 1.$$

Функция $y = \ln x$ вогнута на $(0, +\infty)$, т. е. для произвольных $x_1 > 0$ и $x_2 > 0$ и произвольного $t \in [0; 1]$ имеет место неравенство

$$\ln((1-t)x_1 + tx_2) \geq (1-t)\ln x_1 + t\ln x_2 = \ln(x_1^{1-t}x_2^t),$$

а вместе с ним и неравенство

$$(1-t)x_1 + tx_2 \geq x_1^{1-t}x_2^t.$$

Положив в последнем неравенстве $t = 1/q$, $x_1 = \tilde{\xi}^p$, $x_2 = \tilde{\eta}^q$, получим:

$$\tilde{\xi}\tilde{\eta} \leq \frac{1}{p}\tilde{\xi}^p + \frac{1}{q}\tilde{\eta}^q.$$

Отсюда, учитывая, что $1/p + 1/q = 1$, имеем

$$\mathsf{M}\tilde{\xi}\tilde{\eta} \leq \frac{1}{p}\mathsf{M}\tilde{\xi}^p + \frac{1}{q}\mathsf{M}\tilde{\eta}^q = \frac{1}{p} + \frac{1}{q} = 1.$$

Тем самым неравенство Гёльдера доказано.

Следствие (неравенство Коши—Буняковского).

$$\mathsf{M}|\xi\eta| \leq \left(\mathsf{M}\xi^2\right)^{1/2}\left(\mathsf{M}\eta^2\right)^{1/2}.$$

Неравенство Коши—Буняковского получается из неравенства Гёльдера при $p = 2$, $q = 2$.

Из неравенства Коши—Буняковского, в частности, следует, что если $\mathsf{M}\xi^2 < \infty$, $\mathsf{M}\eta^2 < \infty$, то и

$$\mathsf{M}|\xi\eta| < \infty.$$

Следствие (неравенство Ляпунова). *Если* $0 < s < t$, *то*

$$\left(\mathsf{M}|\zeta|^s\right)^{1/s} \leq \left(\mathsf{M}|\zeta|^t\right)^{1/t}. \qquad (9.3.2)$$

Доказательство. Из неравенства Гёльдера

$$\mathsf{M}|\xi\eta| \leq \left(\mathsf{M}|\xi|^p\right)^{1/p}\left(\mathsf{M}|\eta|^q\right)^{1/q},$$

$1/p + 1/q = 1\,(p > 0, q > 0)$ при $\eta = 1$ получаем неравенство

$$\mathsf{M}|\xi| \le (\mathsf{M}|\xi|^p)^{1/p}\,,$$

которое, положив $\xi = |\zeta|^s$, перепишем следующим образом

$$\mathsf{M}|\zeta|^s \le (\mathsf{M}|\zeta|^{sp})^{1/p}\,. \tag{9.3.3}$$

Далее выберем p так, чтобы $sp = t$ (поскольку $s < t$ это можно сделать). Тогда неравенство (9.3.3) запишется в виде

$$\mathsf{M}|\zeta|^s \le \left(\mathsf{M}|\zeta|^t\right)^{s/t}$$

или

$$(\mathsf{M}|\zeta|^s)^{1/s} \le \left(\mathsf{M}|\zeta|^t\right)^{1/t}.$$

Следствие. *Если* $\mathsf{M}|\zeta|^t < \infty$, *то для* $0 < s < t$ *и* $\mathsf{M}|\zeta|^s < \infty$.

Далее нам понадобится следующее определение выпуклой функции, эквивалентное приведенному выше.

Определение. Будем говорить, что функция $g = g(x)$ со значениями в $\mathbb{R}^1$, заданная на промежутке $I \subset \mathbb{R}^1$, *выпукла на* I, если для каждой точки x_0 из I существует такое число $\lambda(x_0)$, что для всех $x \in I$

$$g(x_0) + \lambda(x_0)(x - x_0) \le g(x).$$

Заметим, что

$$y = g(x_0) + \lambda(x_0)(x - x_0)$$

— уравнение прямой, проходящей через точку $(x_0, g(x_0))$ с угловым коэффициентом $\lambda(x_0)$, см. также рис. 9.3.1.

Если $g(x)$ *дважды дифференцируема на промежутке* I *и* $g''(x) \ge 0$, *то* $g(x)$ *выпукла на промежутке* I.

Действительно, если $g''(x) \ge 0$, то для любых $x, x_0 \in I$, $x > x_0$,

$$\int_{x_0}^{x} g''(t)dt = g'(x) - g'(x_0) \ge 0,$$

$$\int_{x_0}^{x} (g'(t) - g'(x_0))dt = g(x) - g(x_0) - g'(x_0)(x - x_0) \ge 0,$$

т. е.

$$g(x) \geq g(x_0) + g'(x_0)(x - x_0).$$

Доказательство этого неравенства для $x < x_0$ аналогично.

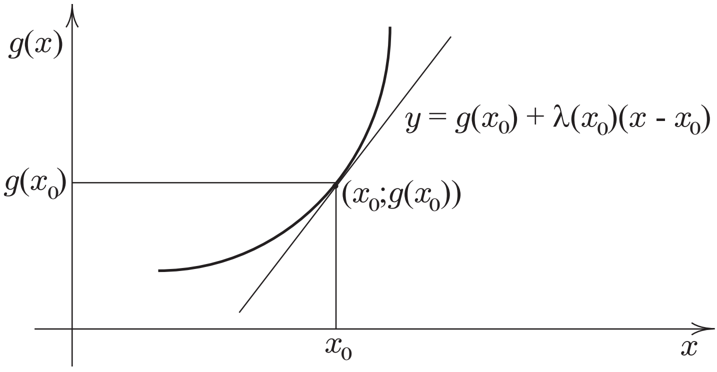

Рис. 9.3.1: К определению выпуклой функции

Теорема 9.3.2 (неравенство Иенсена). *Для выпуклой на $\mathbb{R}^1$ борелевской функции $g(x)$ и случайной величины ξ с конечным* $\mathsf{M}\xi$

$$g(\mathsf{M}\xi) \leq \mathsf{M}g(\xi).$$

В частности,

$$(\mathsf{M}\xi)^2 \leq \mathsf{M}\xi^2, \ |\mathsf{M}\xi|^r \leq \mathsf{M}|\xi|^r, \ r \geq 1.$$

Доказательство. Функция $g(x)$ выпукла, т. е. для каждого $x_0 \in \mathbb{R}^1$ существует число $\lambda(x_0)$ такое, что

$$g(x_0) + \lambda(x_0)(x - x_0) \leq g(x), \, x \in \mathbb{R}^1.$$

Поскольку последнее неравенство выполняется для любых $x_0, x \in \mathbb{R}^1$, то, положив $x_0 = \mathsf{M}\xi$ и $x = \xi$, получим:

$$g(\mathsf{M}\xi) + \lambda(\mathsf{M}\xi)(\xi - \mathsf{M}\xi) \leq g(\xi).$$

Вычислив математическое ожидание от правой и левой частей, имеем

$$g(\mathsf{M}\xi) \leq \mathsf{M}g(\xi).$$

В частности, если в качестве $g(x)$ рассмотреть выпуклые функции $g(x) = x^2$, $g(x) = |x|^r$, $r \geq 1$, получим неравенства

$$(\mathsf{M}\xi)^2 \leq \mathsf{M}\xi^2, \ |\mathsf{M}\xi|^r \leq \mathsf{M}|\xi|^r, \ r \geq 1.$$

9.4 Предельный переход под знаком математического ожидания

Определение. Будем говорить, что последовательность случайных величин $\{\xi_n\}$, заданная на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, *сходится с вероятностью* 1 (*сходится* P-*почти всюду, сходится почти наверное*) к случайной величине ξ (конечной), если $\{\xi_n(\omega)\}$ сходится к $\xi(\omega)$ для всех $\omega \in \Omega$, исключая, быть может, множество точек P-меры нуль, т. е.

$$\mathsf{P}\{\omega : \lim_n \xi_n(\omega) = \xi(\omega)\} = 1.$$

Обозначать эту сходимость будем так:

$$\lim_n \xi_n = \xi \quad (\text{п.в.}).$$

Пример. *Пусть* $\Omega = [0,1]$, $\mathfrak{B}^1_{[0;1]}$ — σ-*алгебра борелевских множеств на* $[0;1]$, P — *мера Лебега на* $\mathfrak{B}^1_{[0;1]}$. *На вероятностном пространстве* $\{\Omega, \mathfrak{B}^1_{[0;1]}, \mathsf{P}\}$ *рассмотрим последовательность случайных величин*

$$\xi_n = \xi_n(\omega) = \omega^n, \ n = 1, 2, \ldots$$

Исследовать последовательность $\{\xi_n\}$ *на сходимость с вероятностью* 1 *к* $\xi = \xi(\omega) = 0$.

Решение. Последовательность $\{\xi_n(\omega)\}$ сходится к случайной величине

$$\xi = \xi(\omega) = 0, \ \omega \in [0,1],$$

в каждой точке $\omega \in [0;1]$, исключая точку $\omega = 1$. И поскольку $\mathsf{P}(\{1\}) = 0$, то по определению $\{\xi_n\}$ на $[0;1]$ сходится к $\xi(\omega) = 0$ с вероятностью 1.

Если вероятность Q на $[0;1]$ задана функцией распределения

$$\mathsf{Q}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ x/2, & \text{если } 0 < x \leq 1; \\ 1, & \text{если } x > 1; \end{cases}$$

(при этом значение $\mathsf{Q}(\{1\}) = 1/2$), то на вероятностном пространстве $\{[0,1], \mathfrak{B}^1_{[0;1]}, \mathsf{Q}\}$ последовательность $\xi_n(\omega) = \omega^n$ не сходится с вероятностью 1 к $\xi = \xi(\omega) = 0$ ($\xi_n(\omega)$ не сходится к $\xi(\omega)$ только в одной точке 1, но $\mathsf{Q}(\{1\}) = 1/2 \neq 0$).

О предельном переходе под знаком математического ожидания. Пусть последовательность случайных величин $\{\xi_n\}$ сходится п.в. к случайной величине ξ. Можно ли утверждать, что последовательность математических ожиданий $\{\mathsf{M}\xi_n\}$ сходится к математическому ожиданию $\mathsf{M}\xi$ предельной случайной величины ξ? Оказывается, что из сходимости последовательности $\{\xi_n\}$, вообще говоря, не следует даже сходимость последовательности математических ожиданий $\{\mathsf{M}\xi_n\}$ (см. пример 9.4.1), но если последовательность $\{\xi_n\}$ монотонно неубывающая, то последовательность $\{\mathsf{M}\xi_n\}$ сходится и, более того,

$$\lim_n \mathsf{M}\xi_n = \mathsf{M}\lim_n \xi_n.$$

Пример 9.4.1. *Построить сходящуюся с вероятностью* 1 *последовательность неотрицательных случайных величин, для которой* $\lim \mathsf{M}\xi_n$ *не существует.*

Пусть $\Omega = [0,1]$, $\mathfrak{F} = \mathfrak{B}^1_{[0;1]}$ — σ-алгебра борелевских множеств на $[0;1]$, P — мера Лебега на $\mathfrak{B}^1_{[0;1]}$.

Последовательность случайных величин $\xi_n = \xi_n(\omega)$, заданная на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ равенствами

$$\xi_n = \xi_n(\omega) = \begin{cases} 2n, & \text{если } \omega \in [0;1/n]; \\ 0, & \text{если } \omega \notin [0;1/n], \end{cases}$$

для четных n и равенствами

$$\xi_n = \xi_n(\omega) = \begin{cases} n, & \text{если } \omega \in [0;1/n]; \\ 0, & \text{если } \omega \notin [0;1/n], \end{cases}$$

для нечетных n, сходится с вероятностью 1, но $\lim \mathsf{M}\xi_n$ не существует.

Теорема 9.4.1 (Лебега о монотонной сходимости). *Если сходящаяся п. в. последовательность* $\{\xi_k\}$ *неотрицательных случайных величин монотонно неубывающая, то сходится и последовательность* $\{\mathsf{M}\xi_k\}$ *и*

$$\lim_k \mathsf{M}\xi_k = \mathsf{M}\lim_k \xi_k,$$

или, в интегральной форме,

$$\lim_k \int\limits_\Omega \xi_k(\omega)\mathsf{P}(d\omega) = \int\limits_\Omega \lim_k \xi_k(\omega)\mathsf{P}(d\omega).$$

Доказательство. Поскольку последовательность $\{\xi_k\}$ монотонно неубывающая, то монотонно неубывающая и последовательность $\{\mathsf{M}\xi_k\}$, а следовательно существует $\lim\limits_k \mathsf{M}\xi_k$. Убедимся, что $\lim\limits_k \mathsf{M}\xi_k = \mathsf{M}\lim\limits_k \xi_k$.

Для каждой случайной величины ξ_k найдется монотонно неубывающая последовательность $\xi_k^{(n)}, n = 1, 2, \ldots$, неотрицательных простых случайных величин, сходящаяся к ξ_k при $n \to \infty$:

$$\xi_k^{(n)} \uparrow \xi_k.$$

(см. лемму об аппроксимации 9.2.1). Для всех n, k $(n = 1, 2, \ldots;$ $k = 1, 2, \ldots)$ имеют место неравенства

$$\xi_k^{(n)} \leq \xi_k,$$

а вместе с ними и неравенства

$$\xi_k^{(n)} \leq \max_{k:k\leq n} \xi_k^{(n)} \leq \max_{k:k\leq n} \xi_k = \xi_n \tag{9.4.1}$$

(равенство $\max\limits_{k:k\leq n} \xi_k = \xi_n$ имеет место в силу монотонности $\{\xi_k\}$) или

$$\xi_k^{(n)} \leq \eta_n \leq \xi_n,\ k \leq n, \tag{9.4.2}$$

где

$$\eta_n = \max_{k:k\leq n} \xi_k^{(n)}.$$

Последовательность $\{\eta_n\}$ простых (по построению) случайных величин монотонно неубывающая — поскольку

$$\xi_k^{(n)} \leq \xi_k^{(n+1)},$$

то

$$\eta_n = \max_{k:k\leq n} \xi_k^{(n)} \leq \max_{k:k\leq n} \xi_k^{(n+1)} \leq \max_{k:k\leq n+1} \xi_k^{(n+1)} = \eta_{n+1}.$$

Убедимся, что последовательность $\{\eta_n\}$ сходится к предельной случайной величине

$$\xi = \lim_k \xi_k.$$

Переходя в неравенствах (9.4.2) к пределу сначала по n:

$$\xi_k = \lim_n \xi_k^{(n)} \le \lim_n \eta_n \le \lim_n \xi_n = \xi,$$

$$\xi_k \le \lim_n \eta_n \le \xi;$$

затем по k:

$$\xi = \lim_k \xi_k \le \lim_n \eta_n \le \xi,$$

получим

$$\lim_n \eta_n = \xi.$$

Из последнего равенства, учитывая, что $\{\eta_n\}$ — монотонно неубывающая последовательность неотрицательных простых случайных величин, сходящаяся к ξ, по определению $\mathsf{M}\xi$ имеем

$$\lim_n \mathsf{M}\eta_n = \mathsf{M}\xi. \tag{9.4.3}$$

Далее, из неравенств (9.4.2) имеем

$$\mathsf{M}\xi_k^{(n)} \le \mathsf{M}\eta_n \le \mathsf{M}\xi_n,\ k \le n. \tag{9.4.4}$$

Дважды переходя к пределу в неравенствах (9.4.4), сначала по n:

$$\mathsf{M}\xi_k = \lim_n \mathsf{M}\xi_k^{(n)} \le \lim_n \mathsf{M}\eta_n \le \lim_n \mathsf{M}\xi_n,$$

затем по k:

$$\lim_k \mathsf{M}\xi_k \le \lim_n \mathsf{M}\eta_n \le \lim_n \mathsf{M}\xi_n,$$

получим

$$\lim_k \mathsf{M}\xi_k = \lim_n \mathsf{M}\eta_n = \lim_n \mathsf{M}\xi_n,$$

или

$$\lim_k \mathsf{M}\xi_k = \mathsf{M}\xi = \lim_n \mathsf{M}\xi_n$$

— равенство $\lim_n \mathsf{M}\xi_k^{(n)} = \mathsf{M}\xi_k$ следует из определения математического ожидания случайной величины ξ_k, а равенство

$$\lim_n \mathsf{M}\eta_n = \mathsf{M}\xi$$

из определения математического ожидания случайной величины ξ (см. 9.4.3). Тем самым теорема доказана.

Пример. *Показать, что условие монотонности последовательности в теореме* 9.4.1 *является существенным.*

Решение. Пусть $\Omega = [0,1]$, $\mathfrak{B}^1_{[0;1]}$ — σ-алгебра борелевских множеств на $[0;1]$, P — мера Лебега на $\mathfrak{B}^1_{[0;1]}$. На вероятностном пространстве $\{\Omega, \mathfrak{B}^1_{[0;1]}, \mathsf{P}\}$ рассмотрим последовательность случайных величин

$$\xi_n = \xi_n(\omega) = \begin{cases} n, & \text{если } \omega \in [0, 1/n), \\ 0, & \text{если } \omega \in [1/n, 1], \end{cases}$$

$n = 1, 2, \ldots$ Для этой последовательности $\lim\limits_n \xi_n = 0$ (п.в.), но $\mathsf{M} \lim\limits_n \xi_n \neq \lim\limits_n \mathsf{M}\xi_n$.

Следствие 1 (математическое ожидание суммы ряда). *Если* $\xi_k \geq 0,\ k = 1, 2, \ldots,$ *и ряд* $\sum\limits_{k=1}^{\infty} \xi_k$ *сходится с вероятностью* 1, *то*

$$\mathsf{M} \sum_{k=1}^{\infty} \xi_k = \sum_{k=1}^{\infty} \mathsf{M}\xi_k$$

и для каждого $A \in \mathfrak{F}$

$$\mathsf{M}\left(\left(\sum_{k=1}^{\infty} \xi_k\right) I_A\right) = \sum_{k=1}^{\infty} \mathsf{M}(\xi_k I_A)$$

или, в интегральной форме,

$$\int\limits_A \left(\sum_{k=1}^{\infty} \xi_k(\omega)\right) \mathsf{P}(d\omega) = \sum_{k=1}^{\infty} \int\limits_A \xi_k(\omega) \mathsf{P}(d\omega)$$

(*кратко — почти всюду сходящиеся ряды из неотрицательных случайных величин можно почленно интегрировать*).

Доказательство. Пусть $\eta_n = \sum\limits_{k=1}^{n} \xi_k$. Так как $\xi_k \geq 0$, то $\{\eta_n\}$ — монотонно неубывающая последовательность неотрицательных случайных величин, сходящаяся к конечному пределу $\sum\limits_{k=1}^{\infty} \xi_k$. Тогда в силу теоремы Лебега о монотонной сходимости

$$\mathsf{M} \lim_n \eta_n = \lim_n \mathsf{M}\eta_n$$

или

$$\mathsf{M}\lim_n \sum_{k=1}^{n} \xi_k = \lim_n \mathsf{M} \sum_{k=1}^{n} \xi_k = \lim_n \sum_{k=1}^{n} \mathsf{M}\xi_k.$$

Так что

$$\mathsf{M} \sum_{k=1}^{\infty} \xi_k = \sum_{k=1}^{\infty} \mathsf{M}\xi_k.$$

Для доказательства второго равенства достаточно заметить, что

$$I_A \sum_{k=1}^{\infty} \xi_k = \sum_{k=1}^{\infty} I_A \xi_k$$

почти всюду.

Следствие 2 (счетная аддитивность интеграла Лебега). *Если* $A_i \cap A_j = \emptyset,\ i \neq j,$ *то*

$$\int\limits_{\bigcup\limits_{i=1}^{\infty} A_i} \xi(\omega)\mathsf{P}(d\omega) = \sum_{i=1}^{\infty} \int\limits_{A_i} \xi(\omega)\mathsf{P}(d\omega),$$

в предположении существования интеграла в левой части, другими словами, интеграл Лебега

$$\rho(A) = \int\limits_A \xi(\omega)\mathsf{P}(d\omega),\ \ A \in \mathfrak{F},$$

является счетно-аддитивной функцией множества:

$$\rho\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \rho(A_i).$$

Доказательство. Пусть сначала $\xi \geq 0$.

Если $A = \bigcup\limits_{i=1}^{\infty} A_i$ и $A_i \cap A_j = \emptyset, i \neq j$, то

$$I_A = \sum_{i=1}^{\infty} I_{A_i},\ \ \xi I_A = \sum_{i=1}^{\infty} \xi I_{A_i}.$$

Отсюда в силу следствия 1

$$\int\limits_{\Omega} \xi(\omega) I_A(\omega) \mathsf{P}(d\omega) = \sum_{i=1}^{\infty} \int\limits_{\Omega} \xi(\omega) I_{A_i}(\omega) \mathsf{P}(d\omega)$$

или

$$\int\limits_{A} \xi(\omega) \mathsf{P}(d\omega) = \sum_{i=1}^{\infty} \int\limits_{A_i} \xi(\omega) \mathsf{P}(d\omega)$$

или

$$\rho(A) = \sum_{i=1}^{\infty} \rho(A_i).$$

Если случайная величина принимает значения обоих знаков и $\int\limits_A \xi(\omega)\mathsf{P}(d\omega)$ существует, утверждение следует из его справедливости для неотрицательных случайных величин и равенства

$$\int\limits_{A} \xi(\omega) \mathsf{P}(d\omega) = \int\limits_{A} \xi^{+}(\omega) \mathsf{P}(d\omega) - \int\limits_{A} \xi^{-}(\omega) \mathsf{P}(d\omega).$$

Следствие 3 (счетная аддитивность математического ожидания). *Если* $\Omega = \bigcup\limits_{i=1}^{\infty} A_i$ *и* $A_i \cap A_j = \emptyset, i \neq j$, *то*

$$\mathsf{M}\xi = \sum_{i=1}^{\infty} \mathsf{M}(\xi I_{A_i}),$$

в предположении существования $\mathsf{M}\xi$, *другими словами, математическое ожидание*

$$\rho(A) = \mathsf{M}(\xi I_A), \ A \in \mathfrak{F},$$

как функция множества, счетно-аддитивно.

Следствие 3 получается из следствия 2, если заметить, что

$$\mathsf{M}(\xi I_A) = \int\limits_{A} \xi(\omega) \mathsf{P}(d\omega).$$

Из счетной аддитивности интеграла Лебега как функции множества, в частности, имеем:

Следствие 4. *Если ξ — неотрицательная случайная величина, заданная на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, то функция множества*

$$\rho(A) = \int\limits_A \xi(\omega)\mathsf{P}(d\omega),\ A \in \mathfrak{F},$$

является мерой на $\mathfrak{F}$ и, как следствие, для монотонно возрастающей последовательности $\{A_n\}$

$$\int\limits_{\bigcup A_n} \xi(\omega)\mathsf{P}(d\omega) = \lim_n \int\limits_{A_n} \xi(\omega)\mathsf{P}(d\omega),$$

для монотонно убывающей последовательности $\{A_n\}$

$$\int\limits_{\bigcap A_n} \xi(\omega)\mathsf{P}(d\omega) = \lim_n \int\limits_{A_n} \xi(\omega)\mathsf{P}(d\omega).$$

Теорема 9.4.2 (лемма Фату). *Для любой последовательности $\{\xi_n\}$ неотрицательных случайных величин*

$$\mathsf{M}\underline{\lim}\,\xi_n \leq \underline{\lim}\,\mathsf{M}\xi_n,$$

или, в интегральной форме,

$$\int\limits_\Omega \underline{\lim}\,\xi_n(\omega)\mathsf{P}(d\omega) \leq \underline{\lim} \int\limits_\Omega \xi_n(\omega)\mathsf{P}(d\omega).$$

Доказательство. Заметим, что поскольку последовательность $\{\xi_k\}$ ограничена снизу ($\xi_k \geq 0$), то $\underline{\lim}\,\xi_n$ конечен. Представим его в виде предела монотонно неубывающей последовательности случайных величин $\zeta_n = \inf\{\xi_n, \xi_{n+1}, \ldots\}$:

$$\underline{\lim}\,\xi_n = \lim \zeta_n.$$

Воспользовавшись теоремой Лебега о монотонной сходимости, применительно к монотонно неубывающей последовательности $\{\zeta_n\}$, сходящейся к конечному пределу $\underline{\lim}\,\xi_n$, получаем:

$$\mathsf{M}\underline{\lim}\,\xi_n = \mathsf{M}\lim_n \zeta_n = \lim_n \mathsf{M}\zeta_n = \underline{\lim}\,\mathsf{M}\zeta_n \leq \underline{\lim}\,\mathsf{M}\xi_n$$

(неравенство $\underline{\lim}\,\mathsf{M}\zeta_n \le \underline{\lim}\,\mathsf{M}\xi_n$ следует из $\zeta_n \le \xi_n$). Так что

$$\mathsf{M}\underline{\lim}\,\xi_n \le \underline{\lim}\,\mathsf{M}\xi_n.$$

Замечание. Лемма Фату имеет место и для любой ограниченной снизу последовательности случайных величин.

Следствие 1. *Если последовательность $\{\xi_n\}$ неотрицательных случайных величин имеет предел п.в., то*

$$\mathsf{M}\lim_n \xi_n \le \underline{\lim}\,\mathsf{M}\xi_n.$$

Следствие 2. *Если последовательность $\{\xi_n\}$ случайных величин ограничена сверху, то*

$$\overline{\lim}\,\mathsf{M}\xi_n \le \mathsf{M}\overline{\lim}\,\xi_n,$$

Теорема 9.4.3 (Лебега о мажорируемой сходимости). *Если последовательность случайных величин $\{\xi_n\}$ мажорируема интегрируемой случайной величиной:*

$$|\xi_n| \le \eta,\ \ \mathsf{M}\eta < \infty,\ \ n = 1, 2, \ldots,$$

то

$$\mathsf{M}\underline{\lim}\,\xi_n \le \underline{\lim}\,\mathsf{M}\xi_n \le \overline{\lim}\,\mathsf{M}\xi_n \le \mathsf{M}\overline{\lim}\,\xi_n.$$

Если последовательность случайных величин $\{\xi_n\}$ мажорируема интегрируемой случайной величиной и сходится п.в., то сходится и последовательность $\{\mathsf{M}\xi_n\}$ и

$$\lim \mathsf{M}\xi_n = \mathsf{M}\lim \xi_n,$$

в интегральной форме

$$\lim_n \int\limits_\Omega \xi_n(\omega)\mathsf{P}(d\omega) = \int\limits_\Omega \lim_n \xi_n(\omega)\mathsf{P}(d\omega).$$

Доказательство. По условию теоремы $|\xi_n| \le \eta$ или, что то же, $-\eta \le \xi_n \le \eta$. Отсюда $\eta + \xi_n \ge 0, n = 1, 2, \ldots$ Применяя лемму Фату к последовательности $\{\eta + \xi_n\}$, получаем:

$$\mathsf{M}\underline{\lim}\,(\eta + \xi_n) \le \underline{\lim}\,\mathsf{M}(\eta + \xi_n) \tag{9.4.5}$$

или

$$\mathsf{M}\eta + \mathsf{M}\underline{\lim}\,\xi_n \le \mathsf{M}\eta + \underline{\lim}\,\mathsf{M}\xi_n.$$

А поскольку $\mathsf{M}\eta$ конечно, то вычитая $\mathsf{M}\eta$ из обеих частей неравенства, получим

$$\mathsf{M}\varliminf \xi_n \le \varliminf \mathsf{M}\xi_n. \tag{9.4.6}$$

Последовательность $\{-\xi_n\}$ имеет ту же интегрируемую мажоранту η, поэтому

$$\mathsf{M}\varliminf (-\xi_n) \le \varliminf \mathsf{M}(-\xi_n),$$

а так как $\varlimsup a_n = -\varliminf (-a_n)$, то $\mathsf{M}(-\varlimsup \xi_n) \le -\varlimsup \mathsf{M}\xi_n$, или

$$\varlimsup \mathsf{M}\xi_n \le \mathsf{M}\varlimsup \xi_n,$$

Из последнего неравенства и (9.4.6) имеем:

$$\mathsf{M}\varliminf \xi_n \le \varliminf \mathsf{M}\xi_n \le \varlimsup \mathsf{M}\xi_n \le \mathsf{M}\varlimsup \xi_n.$$

Если последовательность $\{\xi_n\}$ сходится, то

$$\varliminf \xi_n = \varlimsup_n \xi_n = \lim_n \xi_n.$$

Поэтому

$$\mathsf{M}\lim_n \xi_n = \mathsf{M}\varliminf \xi_n \le \varliminf \mathsf{M}\xi_n \le \varlimsup \mathsf{M}\xi_n \le \mathsf{M}\varlimsup \xi_n = \mathsf{M}\lim_n \xi_n.$$

Так что

$$\lim_n \mathsf{M}\xi_n = \mathsf{M}\lim_n \xi_n.$$

Тем самым теорема доказана.

Соотношение между интегралами Лебега и Римана. Прежде всего заметим, что интеграл Лебега определяется для измеримых функций, заданных на любом пространстве с мерой $\{\Omega, \mathfrak{F}, \mu\}$. Интеграл Римана на абстрактных пространствах $\{\Omega, \mathfrak{F}, \mu\}$ вообще не определяется, а его построение на $\mathbb{R}^n$ зависит от n.

В основу построения интегралов Римана и Лебега положены различные идеи. Первый шаг в конструкции интеграла (как Римана, так и Лебега) — построение интегральной суммы. Но при построении интеграла Лебега точки ω из области интегрирования Ω группируются по признаку близости значений интегрируемой функции в этих точках, а при построении интеграла Римана точки группируются по признаку их близости в $\Omega = \mathbb{R}^n$. В результате оказывается, что интеграл Римана определен для “не слишком разрывных” функций, в то время как интеграл Лебега определен для измеримых функций — широкого класса, включающего, в частности, непрерывные функции. Если же интегрируемая функция $g(x)$ достаточно гладкая, то интегралы Лебега и Римана от $g(x)$ совпадают:

$$(L)\int_a^b g(x)dx = (R)\int_{[a,b]} g(x)dx,$$

для таких функций интеграл Лебега можно вычислять как интеграл Римана.

Теорема. *Если $g = g(x)$ — непрерывная функция со значениями в $\mathbb{R}^1$, заданная на конечном промежутке $[a,b] \subset \mathbb{R}^1$, то значение интеграла Лебега $(L)\int_{[a,b]} g(x)dx$ совпадает со значением интеграла Римана $(R)\int_a^b g(x)dx$.*

Доказательство. Для каждого натурального n разобьем промежуток $[a,b]$ точками $x_0, x_1, \ldots, x_n$ $(a = x_0 < x_1 < \ldots \ldots < x_n = b)$ причем так, чтобы $\max(x_{i+1} - x_i) \to 0$ при $n \to \infty$. На $[a,b]$ рассмотрим последовательность простых функций

$$g_n(x) = \sum_{i=1}^{n-1} g(\widetilde{x}_i) I_{[x_i, x_{i+1})}(x),\ \widetilde{x}_i \in [x_i, x_{i+1}), g(b) = g(\widetilde{x}_{n-1}).$$

Поскольку функция $g(x)$ непрерывна на $[a,b]$, а следовательно, и равномерно непрерывна, то

$$\sup_{a \le x \le b} |g_n(x) - g(x)| \to 0 \text{ при } n \to \infty,$$

т. е. для данного $\varepsilon > 0$, начиная с некоторого N $(n \ge N)$,

$$\sup_{a \le x \le b} |g_n(x) - g(x)| \le \varepsilon.$$

Отсюда, во-первых,

$$\lim_n g_n(x) = g(x),$$

во-вторых, при $n \geq N$

$$\sup_{x\in[a,b]} |g_n(x)| \leq \sup_{x\in[a,b]} |g_n(x) - g(x)| + \sup_{x\in[a,b]} |g(x)| \leq \varepsilon + C,$$

т. е. последовательность функций $\{g_n(x)\}$ при $n \geq N$ мажорируема интегрируемой функцией $\varepsilon + C$ (константа на конечном промежутке интегрируема по мере Лебега). Поэтому в силу теоремы Лебега о мажорируемой сходимости

$$\lim_n (L)\int\limits_{[a,b]} g_n(x)dx = (L)\int\limits_{[a,b]} g(x)dx.$$

С другой стороны,

$$\lim_n (L)\int\limits_{[a,b]} g_n(x)dx = \lim_n \sum_{i=0}^{n-1} g(\widetilde{x}_i)(x_{i+1} - x_i) = (R)\int\limits_a^b g(x)dx.$$

Так что

$$(L)\int\limits_{[a,b]} g(x)dx = (R)\int\limits_a^b g(x)dx.$$

9.5 Атомическое распределение

Определение. Пусть F — распределение на $\mathbb{R}^n$. Точку $x_0 \in \mathbb{R}^n$ называют *атомом* распределения F, если $\mathsf{F}(\{x_0\}) > 0$, а само число $\mathsf{F}(\{x_0\})$ называют *массой атома* x_0.

Будем говорить, что распределение F сосредоточено на множестве $A \in \mathfrak{B}^n$, если $\mathsf{F}(\overline{A}) = 0$.

Распределение F называется *атомическим* (*дискретным*), если оно сосредоточено на множестве своих атомов.

Распределение, не имеющее атомов, будем называть *непрерывным.*

З а м е ч а н и е. Разумеется можно говорить не только об атомических распределениях на $\{\mathbb{R}^n, \mathfrak{B}^n\}$, но и об атомических мерах на $\{\Omega, \mathfrak{F}\}$.

Теорема 9.5.1. *Число атомов вероятностного распределения не более чем счетно.*

Доказательство. Множество A атомов вероятностного распределения F представим в виде объединения

$$A = \bigcup_{k=2}^{\infty} A_k,$$

где A_k — множество атомов с массой, большей $1/k$, $k = 2, 3, \ldots$ Поскольку F — вероятностное распределение ($\mathsf{F}(\mathbb{R}^1) = 1$), то число атомов с массой, большей $1/2$, не превышает одного; число атомов с массой, большей $1/3$, не превышает двух; число атомов с массой, большей $1/k$, не превышает $(k-1)$ и т. д. Поэтому число атомов (т. е. число элементов в A) не более чем счетно, как объединение не более чем счетного числа конечных множеств.

Замечание. Атомическое вероятностное распределение F задается своими значениями на множестве A атомов — для каждого $B \in \mathfrak{B}^n$

$$\mathsf{F}(B) = \mathsf{F}((B \cap A) \cup (B \cap \overline{A})) = \mathsf{F}(B \cap A) + \mathsf{F}(B \cap \overline{A}) =$$

$$= \mathsf{F}(B \cap A) = \sum_{x_i \in B \cap A} \mathsf{F}(\{x_i\}).$$

Поэтому вероятностное атомическое распределение часто определяют как неотрицательную функцию точки, заданную на не более чем счетном множестве $A \subset \mathbb{R}^n$, такую, что

$$\sum_{x_i \in A} \mathsf{F}(\{x_i\}) = 1.$$

Атомическое распределение на $\mathbb{R}^1$. Пусть F — вероятностное распределение на $\mathbb{R}^1$ и $F(x)$ — его функция распределения.

Точку x_0 будем называть *точкой разрыва функции распределения* $F(x)$, если

$$F(x_0 + 0) - F(x_0 - 0) > 0.$$

Величину $F(x_0+0) - F(x_0-0)$ будем называть *скачком функции распределения* $F(x)$ в точке x_0.

Для любой точки x_0

$$\mathsf{F}(\{x_0\}) = F(x_0 + 0) - F(x_0 - 0). \tag{9.5.1}$$

(см. теорему 7.2.3). Левая часть равенства — значение распределения в точке x_0, правая — величина скачка функции распределения в точке x_0.

Равенство (9.5.1) устанавливает взаимно однозначное соответствие между множеством атомов распределения и множеством точек разрыва его функции распределения $F(x)$. Отсюда и из теоремы 9.5.1 следует, что *число точек разрыва вероятностной функции распределения не более чем счетно.*

Непосредственно из равенства (9.5.1) получаем следующее утверждение.

Теорема 9.5.2 . *Точка $x_0 \in \mathbb{R}^1$ является атомом с массой p_0 вероятностного распределения F, тогда и только тогда, когда x_0 является точкой разрыва функции распределения $F(x)$, при этом величина скачка $F(x_0 + 0) - F(x_0 - 0)$ равна массе p_0 атома x_0 :*

$$F(x_0 + 0) - F(x_0 - 0) = p_0.$$

Точка x_0 не является атомом распределения F тогда и только тогда, когда функция распределения $F(x)$ в точке x_0 непрерывна.

Теорема 9.5.3 (о функции распределения атомического распределения). *Функция распределения $F(x)$ вероятностного атомического распределения F с конечным числом атомов $x_1, x_2, \ldots, x_i, \ldots, x_n$ кусочно постоянна, со скачками в атомах и величиной скачка $F(x_i + 0) - F(x_i - 0)$ в атоме x_i равной его массе $\mathsf{F}(\{x_i\})$.*

Кусочно-постоянная вероятностная функция распределения $F(x)$ с конечным числом точек разрыва $x_1, x_2, \ldots, x_i, \ldots, x_n$ является функцией распределения атомического распределения F с атомами в точках разрыва функции распределения и массой атома x_i равной величине скачка $F(x_i + 0) - F(x_i - 0)$ функции распределения $F(x)$ в точке x_i.

Доказательство. Пусть атомы вероятностного атомического распределения F перенумерованы в порядке возрастания: $x_1 < \ldots < x_i < x_{i+1} < \ldots < x_n$, так что между соседними атомами x_i и x_{i+1} нет других атомов. Функция распределения $F(x)$ постоянна на каждом промежутке $(x_i, x_{i+1}]$, $i = 1, 2, \ldots, n$. Если бы нашлись точки y', y'' $(y' < y'')$ из промежутка $(x_i, x_{i+1}]$, такие что $F(y'') - F(y') > 0$, то $\mathsf{F}([y', y'')) = F(y'') - F(y') > 0$ (см. теорему 7.2.3), и, следовательно, на промежутке $[y', y'')$ была бы сосредоточена положительная масса, что противоречит предположению об отсутствии атомов атомического распределения F между точками x_i и x_{i+1}. Так что на промежутке $(x_i, x_{i+1}]$ между соседними атомами x_i и x_{i+1} функция распределения $F(x)$ постоянна и

$$F(x) = F(x_{i+1}).$$

Тот факт, что $F(x)$ в атоме x_i имеет скачек p_i следует из равенства (9.5.1).

Для доказательства второй части утверждения достаточно заметить, что функции распределения $F(x)$ соответствует распределение F (см. теорему 7.2.2), причем

$$\mathsf{F}([a, b)) = F(b) - F(a), \ a < b,$$

$$\mathsf{F}(\{x_i\}) = F(x_i + 0) - F(x_i - 0).$$

И так как $F(x)$ кусочно постоянна со скачками в точках x_i, $i = 1, 2, \ldots, n$, то точки x_i являются атомами и вся масса распределения сосредоточена в них — распределение F атомическое.

График функции атомического распределения. График функции распределения $F(x)$ вероятностного атомического с конечным числом атомов

$$x_1 < x_2 < \ldots < x_i < x_{i+1} < \ldots < x_n$$

можно построить так.

Функция распределения атомического распределения постоянна на промежутках $(x_i, x_{i+1}]$, поэтому

$$F(x_i + 0) = F(x) = F(x_{i+1} - 0) = F(x_{i+1}). \tag{9.5.2}$$

По определению функция распределения $F(x) = F((-\infty, x))$, поэтому $F(x_1) = F((-\infty, x_1)) = 0$ и, следовательно, при $x \in (-\infty, x_1]$

$$F(x) = 0.$$

По известному значению $F(x) = 0$ на $(-\infty, x_1]$ и известной величине скачка функции $F(x)$ в атоме x_1:

$$F(x_1 + 0) - F(x_1 - 0) = p_1$$

определим значение $F(x)$ на промежутке $(x_1, x_2]$. Поскольку

$$F(x_1 + 0) = F(x_1 - 0) + p_1 = F(x_1) + p_1 = 0 + p_1 = p_1,$$

и функция распределения $F(x)$ постоянна на промежутке $(x_1, x_2]$ (см. 9.5.2), то при всех $x \in (x_1, x_2]$

$$F(x) = F(x_1 + 0) = p_1.$$

По известному значению $F(x) = p_1$ на $(x_1, x_2]$ и известной величине скачка функции $F(x)$ в атоме x_2 :

$$F(x_2 + 0) - F(x_2 - 0) = p_2$$

определим значение $F(x)$ на промежутке $(x_2, x_3]$. Поскольку

$$F(x_2 + 0) = F(x_2 - 0) + p_2 = F(x_2) + p_2 = p_1 + p_2,$$

и функция распределения $F(x)$ постоянна на $(x_2, x_3]$ (см. (9.5.2)), то при $x \in (x_2, x_3]$

$$F(x) = F(x_2 + 0) = p_1 + p_2.$$

И так далее — при переходе через атом x_i функция распределения $F(x)$ возрастает скачком на величину $p_i = F(\{x_i\})$ массы атома x_i (см. рис. 9.5.1).

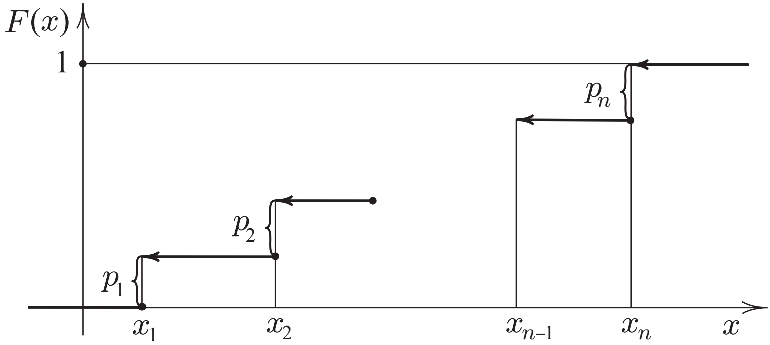

Рис. 9.5.1: График функции вероятностного атомического распределения F, сосредоточенного в атомах x_i ($\mathsf{F}(\{x_i\}) = p_i$, $i = 1, 2, \ldots, n$)

Теорема 9.5.4 (о вычислении интеграла по атомическому распределению). *Если* F — *атомическое распределение на* $\mathbb{R}^n$, *сосредоточенное на не более чем счетном множестве* $A = \{x_1, x_2, \dots\}$ *своих атомов, то для любой борелевской функции* g *на* $\mathbb{R}^n$ *со значениями в* $\mathbb{R}^1$

$$\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx) = \sum_{x_j} g(x_j)\mathsf{F}(\{x_j\}), \tag{9.5.3}$$

в предположении, что интеграл $\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx)$ *конечен или ряд* $\sum\limits_{x_j} g(x_j)\mathsf{F}(\{x_j\})$ *сходится абсолютно.*

Доказательство. Заметим, что функция $g(x)$ интегрируема вместе со своим модулем $|g(x)|$.

Поскольку F — атомическое распределение и $A = \bigcup\limits_{j=1}^{\infty}\{x_j\}$ — множество его атомов, то $\mathsf{F}(\mathbb{R}^n \setminus A) = 0$ и, следовательно,

$$\int\limits_{\mathbb{R}^n} |g(x)|\mathsf{F}(dx) = \int\limits_{A} |g(x)|\mathsf{F}(dx) + \int\limits_{\mathbb{R}^n\setminus A} |g(x)|\mathsf{F}(dx) = \int\limits_{A} |g(x)|\mathsf{F}(dx).$$

В силу счетной аддитивности интеграл Лебега

$$\int\limits_{A} |g(x)|\mathsf{F}(dx) = \int\limits_{\bigcup\{x_j\}} |g(x)|\mathsf{F}(dx) =$$

$$= \sum_{j=1}^{\infty} \int\limits_{\{x_j\}} |g(x)|\mathsf{F}(dx) = \sum_{j=1}^{\infty} |g(x_j)|\mathsf{F}(\{x_j\}),$$

т. е.

$$\int\limits_{A} |g(x)|\mathsf{F}(dx) = \sum_{j=1}^{\infty} |g(x_j)|\mathsf{F}(\{x_j\}), \tag{9.5.4}$$

здесь

$$\int\limits_{\{x_j\}} |g(x)|\mathsf{F}(dx) = |g(x_j)|\mathsf{F}(\{x_j\})$$

как интеграл Лебега от простой функции (на конечном множестве каждая борелевская функция простая). Так что $\int\limits_{\mathbb{R}^n} |g(x)|\mathsf{F}(dx)$ конечен тогда и только тогда, когда ряд $\sum\limits_{x_j} |g(x_j)|\mathsf{F}(\{x_j\})$ сходится.

Равенство (9.5.3) получаем аналогично равенству (9.5.4), пользуясь свойством счетной аддитивности интеграла Лебега.

9.6 Абсолютно непрерывное распределение

Определение. Распределение F на $\mathbb{R}^n$ называется *абсолютно непрерывным относительно меры* L на $\mathbb{R}^n$, если оно представимо в виде:

$$\mathsf{F}(B) = \int\limits_B f(x)\mathsf{L}(dx),\ \ B \in \mathfrak{B}^n, \tag{9.6.1}$$

где $f(x)$ — борелевская функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$, называемая *плотностью* распределения F относительно меры L.

В частности, если L — мера Лебега, то F называют *абсолютно непрерывным распределением*, а f — плотностью.

Равенство (9.6.1) коротко будем записывать так:

$$\mathsf{F}(dx) = f(x)\mathsf{L}(dx),$$

а если L — мера Лебега, то так

$$\mathsf{F}(dx) = f(x)dx,$$

Обычно распределение и его плотность будем обозначать одной буквой — распределение прописной, плотность строчной.

Разумеется, можно говорить и об абсолютной непрерывности не только распределений, но и об абсолютной непрерывности мер.

Пусть μ и ν — меры на измеримом пространстве $\{\Omega, \mathfrak{F}\}$. Меру μ будем называть *абсолютно непрерывной* относительно меры ν, если она представима в виде

$$\mu(B) = \int\limits_B f(\omega)\nu(d\omega),\ \ B \in \mathfrak{F},$$

$f(\omega)$ — $\mathfrak{F}$-измеримая функция на $\{\Omega, \mathfrak{F}\}$ со значениями в $\mathbb{R}^1$, называемая *плотностью меры* μ относительно меры ν.

Свойства плотности распределения. *Плотность данного абсолютно непрерывного распределения* F *на* $\mathbb{R}^n$ *относительно данной меры* L

1) *определяется однозначно, исключая, быть может, множество точек* L*-меры нуль;*

2) *неотрицательна, исключая, быть может, множество точек* L*-меры нуль.*

Действительно, пусть f и φ — плотности распределения F относительно меры L. Обозначим

$$B_1 = \{x : \varphi(x) > f(x)\}$$

(заметим что $B_1 \in \mathfrak{B}^n$). Из определения плотности распределения имеем:

$$\mathsf{F}(B_1) = \int\limits_{B_1} f(x)\mathsf{L}(dx),\ \mathsf{F}(B_1) = \int\limits_{B_1} \varphi(x)\mathsf{L}(dx).$$

Отсюда:

$$\int\limits_{B_1} (\varphi(x) - f(x))\mathsf{L}(dx) = 0,$$

а так как на B_1 значения функции $\varphi(x) - f(x)$ больше нуля, то $\mathsf{L}(B_1) = 0$.

Аналогично для множества $B_2 = \{x : \varphi(x) < f(x)\}$ имеем $\mathsf{L}(B_2) = 0$. Поэтому

$$\mathsf{L}\{x : f(x) \neq \varphi(x)\} = \mathsf{L}(B_1 \cup B_2) = 0.$$

Так что

$$\varphi(x) = f(x),$$

исключая, быть может, множество точек L-меры нуль.

Далее, пусть f — плотность распределения F и $B = \{x : f(x) < 0\}$, тогда

$$0 \leq \mathsf{F}(B) = \int\limits_{B} f(x)\mathsf{L}(dx) \leq 0.$$

Поэтому

$$\int\limits_{B} f(x)\mathsf{L}(dx) = 0,$$

а так как на B значения функции $f(x)$ меньше нуля, то

$$\mathsf{L}(B) = 0.$$

Свойства плотности распределения вероятностей. *Плотность f данного абсолютно непрерывного вероятностного распределения F на $\mathbb{R}^n$ относительно данной меры L*

1) *определяется однозначно, исключая, быть может, множество точек L-меры нуль;*

2) *неотрицательна, исключая, быть может, множество точек L-меры нуль;*

3) $\int\limits_{\mathbb{R}^n} f(x)\mathsf{L}(dx) = 1.$

Справедливо и обратное, а именно: *неотрицательная борелевская функция f на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$ такая, что*

$$\int\limits_{\mathbb{R}^n} f(x)\mathsf{L}(dx) = 1,$$

является плотностью некоторого вероятностного распределения на $\mathbb{R}^n$ относительно меры L.

В самом деле, функция множества F, определенная равенством

$$\mathsf{F}(B) = \int\limits_{B} f(x)\mathsf{L}(dx),\ B \in \mathfrak{B}^n, \tag{9.6.2}$$

является счетно-аддитивной (как интеграл Лебега), неотрицательной и к тому же $\mathsf{F}(\mathbb{R}^n) = 1$, т. е. F — вероятностное распределение на $\mathbb{R}^n$. А равенство (9.6.2) означает, что $f(x)$ — плотность распределения F относительно меры L.

Поэтому часто неотрицательную борелевскую функцию на $\mathbb{R}^n$ такую, что

$$\int\limits_{\mathbb{R}^n} f(x)\mathsf{L}(dx) = 1,$$

называют плотностью.

Теорема 9.6.1 (достаточное условие абсолютной непрерывности распределения на $\mathbb{R}^n$). *Для того чтобы распределение F на $\mathbb{R}^n$ было абсолютно непрерывным относительно меры L и борелевская функция $f(x) \geq 0$, $x \in \mathbb{R}^n$, была его* *плотностью, достаточно, чтобы на классе $\mathfrak{K}$ параллелепипедов $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n) = B$, $a_i, b_i \in \mathbb{R}^1$, $i = 1, 2, \ldots, n$,*

$$\mathsf{F}(B) = \int_B f(x)\mathsf{L}(dx). \tag{9.6.3}$$

В частности, для того чтобы распределение F на $\mathbb{R}^1$ было абсолютно непрерывным относительно меры L и борелевская функция $f(x) \geq 0$ была его плотностью, достаточно, чтобы на классе $\mathfrak{K}$ промежутков $[a, b), a, b \in \mathbb{R}^1$,

$$\mathsf{F}([a, b)) = \int_{[a,b)} f(t)\mathsf{L}(dt). \tag{9.6.4}$$

Доказательство. Если распределения $\mathsf{F}(B)$ и $\int_B f(x)\mathsf{L}(dx)$ совпадают на параллелепипедах $[a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n) = B$ (см. (9.6.3)), то они совпадают и на алгебре $\mathfrak{A}$ конечных объединений параллелепипедов, а в силу теоремы Каратеодори (см. теорему 7.1.3) и на σ-алгебре $\sigma(\mathfrak{A}) = \mathfrak{B}^n$:

$$\mathsf{F}(B) = \int_B f(x)\mathsf{L}(dx), \ B \in \mathfrak{B}^n.$$

Последнее и обозначает, что F абсолютно непрерывно относительно меры L.

Следствие 1 (функция и плотность распределения). *Вероятностное распределение F на $\mathbb{R}^1$ абсолютно непрерывно, тогда и только тогда, когда его функция распределения $F(x)$ представима в виде:*

$$F(x) = \int_{-\infty}^{x} f(t)dt, \ x \in \mathbb{R}^1, \tag{9.6.5}$$

где $f(t)$ — борелевская функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$, при этом $f(t)$ является плотностью распределения F.

Доказательство. Пусть имеет место равенство (9.6.5).

Заметим, что $f(x)$ неотрицательна как производная от монотонно неубывающей функции $F(x)$.

Из равенства (9.6.5) следует, что для промежутков $[a, b)$, $a, b \in \mathbb{R}^1$,

$$\mathsf{F}([a, b)) = \int\limits_{[a,b)} f(t)dt,$$

поэтому в силу теоремы распределение F абсолютно непрерывно и $f(t)$ его плотность.

Если F абсолютно непрерывно с плотностью f, то

$$F(x) = \mathsf{F}((-\infty, x)) = \int\limits_{-\infty}^{x} f(t)dt, \ x \in \mathbb{R}^1,$$

З а м е ч а н и е. Следствие 1 дает эквивалентное определение абсолютно непрерывного распределения на $\mathbb{R}^1$.

Следствие 2. *Если F — абсолютное непрерывное распределение на $\mathbb{R}^1$, то его плотность $f(t)$ равна (п.в.) производной от функции распределения $F(x)$:*

$$f(x) = F'(x).$$

Д о к а з а т е л ь с т в о. Если F — абсолютное непрерывное распределение с плотностью $f(x)$, то функция распределения

$$F(x) = \mathsf{F}((-\infty, x)) = \int\limits_{-\infty}^{x} f(t)dt. \tag{9.6.6}$$

Поэтому, во-первых, $F(x)$ дифференцируема почти всюду (относительно меры Лебега), а, во-вторых, восстанавливается по своей производной посредством операции интегрирования, а именно:

$$\int\limits_{a}^{x} F'(t)dt = F(x) - F(a).$$

Отсюда

$$\mathsf{F}([a, x)) = \int\limits_{[a,x)} \mathsf{F}'(t)dt.$$

Поэтому в силу теоремы 1 функция $F'(x)$ является плотностью распределения F, а в силу единственности плотности

$$f(x) = F'(x) \text{ (п.в.).}$$

Теорема 9.6.2 (о вычислении интеграла по абсолютно непрерывному распределению). *Если* F *— абсолютно непрерывное относительно меры* L *распределение на* $\mathbb{R}^n$, f *— его плотность, то для любой борелевской функции* g *на* $\mathbb{R}^n$ *со значениями в* $\mathbb{R}^1$

$$\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^n} g(x)f(x)\mathsf{L}(dx), \tag{9.6.7}$$

в предположении, что определен интеграл $\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx)$ *или интеграл* $\int\limits_{\mathbb{R}^n} g(x)f(x)\mathsf{L}(dx)$.

Доказательство. В силу свойства линейности интеграла Лебега и представления $g = g^+ - g^-$ функции g в виде разности неотрицательных функций g^+ и g^-, теорему достаточно доказать для неотрицательных g.

Теорема имеет место, если g — индикатор, т. е. $g(x) = I_B(x)$, $B \in \mathfrak{B}^n$. В самом деле, с одной стороны,

$$\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^n} I_B(x)\mathsf{F}(dx) = \mathsf{F}(B),$$

с другой —

$$\int\limits_{\mathbb{R}^n} g(x)f(x)\mathsf{L}(dx) = \int\limits_{\mathbb{R}^n} I_B(x)f(x)\mathsf{L}(dx) = \int\limits_{B} f(x)\mathsf{L}(dx) = \mathsf{F}(B),$$

поэтому

$$\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^n} g(x)f(x)\mathsf{L}(dx).$$

В силу свойства линейности интеграла Лебега теорема справедлива и для линейных комбинаций индикаторов, т. е. для простых борелевских функций.

Далее, пусть g — неотрицательная борелевская функция. Монотонно неубывающая последовательность неотрицательных простых борелевских функций

$$g_k(x) = \sum_{j=1}^{2^k k} \frac{j-1}{2^k} I_{\left\{x:(j-1)/2^k \le g(x) < j/2^k\right\}}(x) + kI_{\{x:g(x)\ge k\}}(x)$$

сходится в каждой точке $x \in \mathbb{R}^1$ к $g(x)$ (см. лемму 9.2.1). Для простых $g_k(x)$ теорема имеет место:

$$\int\limits_{\mathbb{R}^n} g_k(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^n} g_k(x)f(x)\mathsf{L}(dx).$$

Воспользовавшись теоремой Лебега о монотонной сходимости, перейдем к пределу в правой и левой частях последнего неравенства:

$$\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^n} g(x)f(x)\mathsf{L}(dx).$$

Так что в предположениях теоремы $\int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx)(dx)$ определен тогда и только тогда, когда определен интеграл $\int\limits_{\mathbb{R}^n} g(x)f(x)\mathsf{L}(dx)$ и имеет место равенство (9.6.7).

Тем самым теорема доказана.

Замечание. Теорема имеет место и в случае интегрирования относительно абсолютно непрерывной меры μ $\mathfrak{F}$-измеримой функции, заданной на пространстве с мерой $\{\Omega, \mathfrak{F}\}$.

Интеграл по сумме распределений. Пусть F и G — распределения на $\mathbb{R}^n$. Функция множества Q, заданная на σ-алгебре $\mathfrak{B}^n$ равенством

$$\mathsf{Q}(B) = \mathsf{F}(B) + \mathsf{G}(B),\ B \in \mathfrak{B}^n,$$

очевидно является распределением. Распределение Q будем называть *суммой распределений* F и G и будем обозначать $\mathsf{F} + \mathsf{G}$.

Непосредственно из определения интеграла Лебега получаем, что

$$\int\limits_{\mathbb{R}^n} g(x)\mathsf{Q}(dx) = \int\limits_{\mathbb{R}^n} g(x)\mathsf{F}(dx) + \int\limits_{\mathbb{R}^n} g(x)\mathsf{G}(dx)$$

(в предположении существования интегралов).

В частности, если распределение Q является суммой абсолютно непрерывного распределения F (с плотностью f) и атомического G, то

$$\int\limits_{\mathbb{R}^n} g(x)\mathsf{Q}(dx) = \int\limits_{\mathbb{R}^n} g(x)f(x)dx + \sum_{x_i} g(x_i)\mathsf{G}(\{x_i\}),$$

в последней сумме суммирование ведется по множеству атомов распределения G.

9.7 Абсолютно непрерывные и дискретные случайные величины

Каждая случайная величина имеет распределение. В зависимости от того, каким является распределение случайной величины — абсолютно непрерывным или дискретным, — мы будем различать абсолютно непрерывные и дискретные случайные величины.

Дискретные случайные величины. Случайная величина ξ со значениями в $\mathbb{R}^n$ называется *дискретной*, если ее распределение P_ξ дискретное (атомическое).

Другими словами, случайную величину ξ со значениями в $\mathbb{R}^n$ будем называть дискретной, если существует не более чем счетное множество точек $x_i \in \mathbb{R}^n$ таких, что

$$\mathsf{P}_\xi(\{x_i\}) = \mathsf{P}\{\xi = x_i\} > 0,\ i = 1, 2, \ldots,\ \sum_{x_i} \mathsf{P}_\xi(\{x_i\}) = 1. \quad (9.7.1)$$

Точки $x_i \in \mathbb{R}^n$, для которых выполняется неравенство

$$\mathsf{P}\{\xi = x_i\} > 0,$$

называют возможными значениями дискретной случайной величины.

Замечание относительно обозначений. Значение

$$\mathsf{P}_\xi(\{x_i\}) = \mathsf{P}\{\xi \in \{x_i\}\} = \mathsf{P}\{\xi = x_i\}$$

дискретного распределения P_ξ на атоме $\{x_i\}$ еще будем обозначать $\mathsf{P}_\xi(x_i)$ (если запись $\mathsf{P}_\xi(x_i)$ понимается однозначно), т. е.

$$\mathsf{P}_\xi(\{x_i\}) = \mathsf{P}_\xi(x_i).$$

Теорема 9.7.1. *Пусть $\xi = \xi(\omega)$ — функция на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ со значениями в $\mathbb{R}^1$, принимающая конечное или счетное число значений $x_1, x_2, \ldots$ Если для каждого x_i*

$$\{\omega : \xi(\omega) = x_i\} \in \mathfrak{F}, \quad (9.7.2)$$

то $\xi = \xi(\omega)$ является $\mathfrak{F}$-измеримой (является случайной величиной).

Для каждого вещественного x

$$\{\omega : \xi(\omega) < x\} = \bigcup_{x_i : x_i < x} \{\omega : \xi(\omega) = x_i\} \in \mathfrak{F},$$

поэтому согласно следствию из теоремы 8.1.1 функция $\xi = \xi(\omega)$ является $\mathfrak{F}$ измеримой.

Абсолютно непрерывные случайные величины. Случайная величина ξ со значениями в $\mathbb{R}^n$ называется *абсолютно непрерывной* относительно меры L, если ее распределение P_ξ абсолютно непрерывно относительно L.

Другими словами, случайная величина ξ со значениями в $\mathbb{R}^n$ абсолютно непрерывна относительно меры L, если ее распределение P_ξ представимо в виде:

$$\mathsf{P}_\xi(B) = \int_B p(x)\mathsf{L}(dx), \; B \in \mathfrak{B}^n, \; x \in \mathbb{R}^n, \tag{9.7.3}$$

где $p(x)$ — борелевская функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$, называемая *плотностью распределения* случайной величины ξ относительно меры L.

В частности, если распределение P_ξ абсолютно непрерывно относительно относительно меры Лебега, то $p(x)$ называют плотностью распределения случайной величины ξ.

Плотность распределения $p(x)$ случайной величины относительно меры L, будучи плотностью вероятностного распределения,

1) *определяется однозначно, исключая, быть может, множество точек* L*-меры нуль;*

2) *неотрицательна, исключая, быть может, множество точек* L*-меры нуль;*

3) $\int\limits_{\mathbb{R}^n} p(x)\mathsf{L}(dx) = 1.$

Имеет место и обратное утверждение, а именно, неотрицательная борелевская функция $f(x)$ на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$ такая, что

$$\int_{\mathbb{R}^n} f(x)\mathsf{L}(dx) = 1,$$

является плотностью распределения некоторой случайной величины со значениями в $\mathbb{R}^n$.

Теорема 9.7.2 (о функции и плотности распределения случайной величины). *Для того чтобы случайная величина ξ со значениями в $\mathbb{R}^1$ была абсолютно непрерывной, необходимо и достаточно, чтобы ее функция распределения $F(x)$ была представима в виде*

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt, \ x \in \mathbb{R}^1,$$

при этом борелевская функция f является плотностью распределения случайной величины ξ.

Следствие. *Если распределение F случайной величины со значениями в $\mathbb{R}^1$ абсолютно непрерывно, то его плотность $f(t)$ равна (п.в.) производной от функции распределения $F(x)$:*

$$f(x) = F'(x).$$

Теорема и следствие из нее являются частным случаем следствий 1 и 2 из теоремы 9.6.1, поскольку распределение случайной величины является частным случаем вероятностного распределения.

Распределение функции от случайной величины. По распределению P_ξ случайной величины ξ всегда можно найти распределение любой функции $g(\xi)$ от ξ.

Теорема 9.7.3 (о распределении функции от $\boldsymbol{\xi}$). *Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — случайная величина со значениями в $\mathbb{R}^n$ и P_ξ — ее распределение, $g(x)$ — борелевская функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^l$. Для любого борелевского множества B из $\mathbb{R}^l$*

$$\mathsf{P}\{g(\xi) \in B\} = \int\limits_{x: g(x) \in B} \mathsf{P}_\xi(dx), \ x \in \mathbb{R}^n;$$

если $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — абсолютно непрерывная случайная величина и $p_\xi(x)$ — ее плотность, то

$$\mathsf{P}\{g(\xi) \in B\} = \int\limits_{x: g(x) \in B} p_\xi(x)dx, \ x \in \mathbb{R}^n;$$

если $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ *— дискретная случайная величина и* $\mathsf{P}_\xi : x \to \mathsf{P}_\xi(x)$ *— ее распределение, то*

$$\mathsf{P}\{g(\xi) \in B\} = \sum_{x: g(x) \in B} \mathsf{P}_\xi(x)$$

(*суммирование ведется по атомам распределения* P_ξ, *принадлежащим множеству* $\{x : g(x) \in B\}$).

Доказательство.

$$\mathsf{P}\{g(\xi) \in B\} = \mathsf{P}\{\xi \in g^{-1}(B)\} = \mathsf{P}_\xi(g^{-1}(B)) =$$

$$= \int\limits_{g^{-1}(B)} \mathsf{P}_\xi(dx) = \int\limits_{x: x \in g^{-1}(B)} \mathsf{P}_\xi(dx) = \int\limits_{x: g(x) \in B} \mathsf{P}_\xi(dx).$$

В частности, если случайная величина $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ имеет плотность распределения $p_\xi(x) = p_\xi(x_1, x_2, \ldots, x_n)$, то

$$\mathsf{P}\{g(\xi) \in B\} = \int\limits_{x: g(x) \in B} \mathsf{P}_\xi(dx) = \int\limits_{x: g(x) \in B} p_\xi(x) dx, \ x \in \mathbb{R}^n$$

(в силу теоремы 9.6.2 о вычислении интеграла Лебега по абсолютно непрерывному распределению).

Если $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — дискретная случайная величина и $\mathsf{P}_\xi : x \to \mathsf{P}_\xi(x)$ — ее распределение, то

$$\mathsf{P}\{g(\xi) \in B\} = \int\limits_{x: g(x) \in B} \mathsf{P}_\xi(dx) = \sum_{x: g(x) \in B} \mathsf{P}_\xi(x),$$

суммирование ведется по атомам распределения P_ξ из множества $\{x : g(x) \in B\}$ (в силу теоремы 9.5.4 о вычислении интеграла Лебега по дискретному распределению).

Следствие. *Пусть* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ *— случайная величина со значениями в* $\mathbb{R}^n$ *и* P_ξ *— ее распределение. Для любого борелевского множества* B *из* $\mathbb{R}^n$

$$\mathsf{P}\{\xi \in B\} = \int\limits_B \mathsf{P}_\xi(dx), \ x \in \mathbb{R}^n;$$

если ξ абсолютно непрерывна и $p_\xi(x)$ — ее плотность распределения, то

$$\mathsf{P}\{\xi \in B\} = \int\limits_B p_\xi(x)dx, \ x \in \mathbb{R}^n;$$

если ξ — дискретна и $\mathsf{P}_\xi : x \to \mathsf{P}_\xi(x)$ — ее распределение, то

$$\mathsf{P}\{\xi \in B\} = \sum_{x \in B} \mathsf{P}_\xi(x)$$

(*суммирование ведется по атомам распределения P_ξ, принадлежащим множеству B*).

Достаточно в качестве g рассмотреть тождественное отображение $\mathbb{R}^n$ на $\mathbb{R}^n$.

Пример 9.7.1. *Случайная величина ξ распределена показательно с параметром λ. Найти плотность распределения случайной величины $\eta = 1/(1-\xi)$.*

Решение. Сначала найдем функцию распределения η, воспользовавшись теоремой 9.7.3

$$F_\eta(x) = \begin{cases} \int\limits_1^{1-1/x} \lambda e^{-\lambda t}dt, & \text{если } x < 0, \\ \int\limits_1^{+\infty} \lambda e^{-\lambda t}dt, & \text{если } 0 \le x \le 1, \\ \int\limits_0^{1-1/x} \lambda e^{-\lambda t}dt + \int\limits_1^{+\infty} \lambda e^{-\lambda t}dt, & \text{если } x > 1, \end{cases}$$

$$p_\eta(x) = \begin{cases} 0, & \text{если } x \in [0;1]; \\ \lambda \exp\{-\lambda(1-1/x)\}/x^2, & \text{если } x \notin [0;1]. \end{cases}$$

Совместное распределение независимых случайных величин. Распределение вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ еще называют *совместным распределением* случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, а плотность распределения вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, если она существует, — *плотностью совместного распределения*[1] $\xi_1, \xi_2, \ldots, \xi_n$.

По совместному распределению $\xi_1, \xi_2, \ldots, \xi_n$ всегда можно найти распределение каждой из случайных величин ξ_i, поскольку компонента ξ_i вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ является функцией от ξ. Обратное, вообще говоря, неверно, исключая частный, но

[1]Плотность совместного распределения случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ еще называют *совместной плотностью распределения* $\xi_1, \xi_2, \ldots, \xi_n$.

важный и часто встречающийся случай независимых случайных величин, когда совместное распределение случайных величин $\xi_1, \xi_2, \ldots \xi_n$ равно произведению их распределений:

$$\mathsf{P}_\xi(B_1 \times B_2 \times \ldots \times B_n) = \mathsf{P}_{\xi_1}(B_1)\mathsf{P}_{\xi_2}(B_2)\ldots\mathsf{P}_{\xi_n}(B_n)$$

для любых борелевских множеств $B_1, B_2, \ldots, B_n$ из $\mathbb{R}^1$.

В частности, для дискретных случайных величин имеем.

Теорема 9.7.4. *Совместное распределение*

$$\mathsf{P}_\xi(x_i, y_j, \ldots, z_k) = \mathsf{P}\{(\xi_1, \xi_2, \ldots, \xi_n) = (x_i, y_j, \ldots, z_k)\}$$

независимых дискретных случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ равно произведению их распределений $\mathsf{P}_{\xi_1}(x_i), \mathsf{P}_{\xi_2}(y_j), \ldots, \mathsf{P}_{\xi_n}(z_k)$:

$$\mathsf{P}_\xi(x_i, y_j, \ldots, z_k) = \mathsf{P}_{\xi_1}(x_i)\mathsf{P}_{\xi_2}(y_j)\ldots\mathsf{P}_{\xi_n}(z_k) \qquad (9.7.4)$$

для всех возможных значений $x_i, y_j, \ldots, z_k$ случайных величин $\xi_1, \xi_2, \ldots, \xi_n$.

Действительно, если случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы, то

$$\mathsf{P}_\xi(x_i, y_j, \ldots, z_k) = \mathsf{P}\{\xi_1 = x_i, \xi_2 = y_j, \ldots, \xi_n = z_k\} =$$

$$= \mathsf{P}\{\xi_1 = x_i\}\mathsf{P}\{\xi_2 = y_j\}\ldots\mathsf{P}\{\xi_n = z_k\} =$$

$$= \mathsf{P}_{\xi_1}(x_i)\mathsf{P}_{\xi_2}(y_j)\ldots\mathsf{P}_{\xi_n}(z_k).$$

Пример 9.7.2. *Найти совместное распределение независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, каждая из которых имеет геометрическое распределение с параметром p.*

Р е ш е н и е. Для случайных величин ξ_i, $i = 1, 2, \ldots, n$,

$$\mathsf{P}_{\xi_i}(k_i) = p(1-p)^{k_i},\ k_i = 0, 1, \ldots$$

Совместное распределение ξ_i, $i = 1, 2, \ldots, n$,

$$\mathsf{P}(k_1, k_2, \ldots, k_n) = \prod_{i=1}^{n} \mathsf{P}_{\xi_i}(k_i) = \prod_{i=1}^{n} p(1-p)^{k_i} = p^n(1-p)^{k_1+k_2+\ldots+k_n},$$

$k_1 = 0, 1, 2, \ldots;\ k_2 = 0, 1, 2, \ldots;\ \ldots k_n = 0, 1, 2, \ldots$

Теорема 9.7.5. *Если* $\xi_1, \xi_2, \ldots, \xi_n$ *— независимые абсолютно непрерывные случайные величины соответственно с плотностями* $p_1(x_1), p_2(x_2), \ldots, p_n(x_n)$*, то существует плотность* $p(x_1, x_2, \ldots, x_n)$ *совместного распределения случайных величин* $\xi_1, \xi_2, \ldots, \xi_n$*, и она равна произведению их плотностей:*

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)\, p_2(x_2) \ldots p_n(x_n).$$

Доказательство. Поскольку случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и абсолютно непрерывны соответственно с плотностями $p_1(x_1), p_2(x_2), \ldots, p_n(x_n)$, то для любых $[a_i, b_i)$, $i = 1, 2, \ldots, n$, на $B = [a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$ значение

$$\mathsf{P}_\xi(B) = \mathsf{P}_\xi([a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)) =$$

$$= \mathsf{P}\{(\xi_1, \xi_2, \ldots, \xi_n) \in [a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)\} =$$

$$= \mathsf{P}\{\xi_1 \in [a_1, b_1), \xi_2 \in [a_2, b_2), \ldots, \xi_n \in [a_n, b_n)\} =$$

$$= \mathsf{P}\{\xi_1 \in [a_1, b_1)\}\mathsf{P}\{\xi_2 \in [a_2, b_2)\} \ldots \mathsf{P}\{\xi_n \in [a_n, b_n)\} =$$

$$= \int\limits_{[a_1, b_1)} p_1(x_1) dx_1 \int\limits_{[a_2, b_2)} p_2(x_2) dx_2 \ldots \int\limits_{[a_n, b_n)} p_n(x_n) dx_n =$$

$$= \int\limits_B p_1(x_1)\, p_2(x_2) \ldots p_n(x_n) d(x_1, x_2, \ldots, x_n),$$

последнее равенство имеет место в силу теоремы Фубини. Таким образом, два распределения, а именно, распределение

$$\mathsf{P}_\xi(B), \ B \in \mathfrak{B}^n,$$

и распределение

$$\mathsf{G}(B) = \int\limits_B p_1(x_1) p_2(x_2) \ldots p_n(x_n) d(x_1, x_2, \ldots, x_n), \ B \in \mathfrak{B}^n$$

совпадают на классе параллелепипедов $B = [a_1, b_1) \times [a_2, b_2) \times \ldots \times [a_n, b_n)$, а значит, и на алгебре $\mathfrak{A}$ конечных объединений непересекающихся параллелепипедов, а в силу теоремы Каратеодори и на σ-алгебре $\sigma(\mathfrak{A}) = \mathfrak{B}^n$ борелевских множеств:

$$\mathsf{P}_\xi(B) = \int\limits_B p_1(x_1) p_2(x_2) \ldots p_n(x_n)\, d(x_1, x_2, \ldots, x_n), \ B \in \mathfrak{B}^n.$$

Последнее равенство означает, что распределение P_ξ векторной случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ абсолютно непрерывно, и функция $p_1(x_1)p_2(x_2)\ldots p_n(x_n)$ является плотностью распределения P_ξ — плотностью совместного распределения случайных величин $\xi_1, \xi_2, \ldots, \xi_n$.

Легко убедиться в справедливости обратного утверждения: *если совместная плотность случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ равна произведению их плотностей, то случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ — независимы.*

Пример 9.7.3. *Найти совместную плотность распределения независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, каждая из которых имеет показательное распределение с параметром θ :*

$$p(x;\theta) = \begin{cases} \theta \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0, \end{cases}$$

где $\theta > 0$.

Р е ш е н и е. Совместная плотность распределения $\xi_1, \xi_2, \ldots, \xi_n$

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i) =$$

$$= \prod_{i=1}^{n} \theta \exp\{-\theta x_i\} = \theta^n \exp\left\{-\theta \sum_{i=1}^{n} x_i\right\},$$

если все x_i положительны, и равна нулю, если хотя бы одно из x_i неположительно.

9.8 Примеры и задачи

Пример 9.8.1. *Пусть P — мера Лебега на $\mathbb{R}^1$. Если функция $f(x)$ интегрируема по Лебегу на действительной прямой, то для данного $\varepsilon > 0$ найдется такое $\delta > 0$, что для любого борелевского множества A такого, что $\mathsf{P}(A) < \delta$,*

$$\int_A |f(x)| \mathsf{P}(dx) < \varepsilon.$$

Решение. Интеграл

$$\mathsf{Q}(B) = \int\limits_B |f(x)| \mathsf{P}(dx), \ B \in \mathfrak{B}^1,$$

как функция множества является мерой.

Поскольку функция $f(x)$ интегрируема по мере P на $\mathbb{R}^1$, то

$$\mathsf{P}\{x : |f(x)| = \infty\} = 0$$

(см. теорему 9.2.11 и следствия из нее).

Пусть $A_n = \{x : \ |f(x)| > n\}$. Очевидно $A_{n+1} \subset A_n$ и

$$\{x : |f(x)| = \infty\} = \bigcap_{n=1}^{\infty} A_n.$$

В силу непрерывности меры Q

$$0 = \mathsf{Q}(\{x : |f(x)| = \infty\}) = \lim_n \mathsf{Q}(A_n).$$

Поэтому для данного $\varepsilon > 0$ найдется такое N, что

$$\mathsf{Q}(A_N) = \int\limits_{A_N} |f(x)| \mathsf{P}(dx) < \varepsilon/2.$$

Для любого A

$$\mathsf{Q}(A) = \int\limits_{A \cap \overline{A}_N} |f(x)| \mathsf{P}(dx) + \int\limits_{A \cap A_N} |f(x)| \mathsf{P}(dx) \le$$

$$\le N\mathsf{P}(A \cap \overline{A}_N) + \int\limits_{A_N} |f(x)| \mathsf{P}(dx) \le N\mathsf{P}(A) + \varepsilon/2.$$

Поэтому, если в качестве δ рассмотреть $\delta = \varepsilon/2N$, то для любого множества A, такого, что $\mathsf{P}(A) \le \varepsilon/2N = \delta$, значение

$$\mathsf{Q}(A) = \int\limits_A |f(x)| \mathsf{P}(dx) < \varepsilon.$$

Глава 10

Вычисление математического ожидания

10.1 Вычисление математического ожидания по распределению

По распределению случайной величины ξ всегда можно найти математическое ожидание $\mathsf{M}g(\xi)$ любой функции $g(\xi)$ от ξ, лишь бы только $\mathsf{M}g(\xi)$ существовало.

Теорема 10.1.1 (о вычислении $\mathsf{M}g(\xi)$ по распределению ξ). *Пусть ξ — случайная величина со значениями в $\mathbb{R}^1$ и P_ξ — ее распределение, для любой борелевской функции $g(x)$ на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$*

$$\mathsf{M}g(\xi) = \int\limits_{\mathbb{R}^1} g(x)\mathsf{P}_\xi(dx), \tag{10.1.1}$$

в предположении, что определено математическое ожидание $\mathsf{M}g(\xi)$ или определен интеграл $\int\limits_{\mathbb{R}^1} g(x)\mathsf{P}_\xi(dx)$.

Доказательство. Из представления $g = g^+ - g^-$, свойства линейности математического ожидания и интеграла Лебега следует, что теорему достаточно доказать для неотрицательных функций.

Теорема имеет место, если $g(x)$ — индикатор, т. е.

$$g(x) = I_B(x),\ B \in \mathfrak{B}^1.$$

Действительно, с одной стороны,

$$\int\limits_{\mathbb{R}^1} g(x)\mathsf{P}_\xi(dx) = \int\limits_{\mathbb{R}^1} I_B(x)\mathsf{P}_\xi(dx) =$$

$$= 1 \cdot \mathsf{P}_\xi(B) + 0 \cdot \mathsf{P}_\xi(\overline{B}) = \mathsf{P}\{\omega : \xi(\omega) \in B\},$$

с другой —

$$\mathsf{M}g(\xi) = \mathsf{M}I_B(\xi) = 1 \cdot \mathsf{P}\{\omega : \xi(\omega) \in B\} + 0 \cdot \mathsf{P}\{\omega : \xi(\omega) \in \overline{B}\} =$$

$$= \mathsf{P}\{\omega : \xi(\omega) \in B\}.$$

В силу свойства линейности математического ожидания и интеграла Лебега равенство (10.1.1) имеет место и для линейных комбинаций индикаторов, т. е. для простых борелевских функций.

Пусть $g(x)$ — неотрицательная борелевская функция. Монотонно неубывающая последовательность неотрицательных простых борелевских функций,

$$g_n(x) = \sum_{j=1}^{2^n n} \frac{j-1}{2^n} I_{\{x:(j-1)/2^n \le g(x) < j/2^n\}}(x) + nI_{\{x:g(x)\ge n\}}(x),$$

$n = 1, 2, \ldots$, сходится к $g(x)$ в каждой точке $x \in \mathbb{R}^1$ (см. лемму 9.2.1). И так как $g_n(x)$ — простая, то

$$\mathsf{M}g_n(\xi) = \int\limits_{\mathbb{R}^1} g_n(x)\mathsf{P}_\xi(dx).$$

Воспользовавшись теоремой Лебега о монотонной сходимости, перейдем к пределу в правой и левой частях последнего равенства, получим

$$\mathsf{M}g(\xi) = \int\limits_{\mathbb{R}^1} g(x)\mathsf{P}_\xi(dx).$$

Так что в предположениях теоремы математическое ожидание $\mathsf{M}g(\xi)$ определено тогда и только тогда, когда определен интеграл $\int\limits_{\mathbb{R}^1} g(x)\mathsf{P}_\xi(dx)$, и имеет место равенство (10.1.1).

Следствие. *Пусть ξ — случайная величина со значениями в $\mathbb{R}^1$ и P_ξ — ее распределение, тогда*

$$\mathsf{M}\xi = \int\limits_{\mathbb{R}^1} x\mathsf{P}_\xi(dx),$$

в предположении, что определено математическое ожидание $\mathsf{M}\xi$ или определен интеграл $\int\limits_{\mathbb{R}^1} x\mathsf{P}_\xi(dx)$.

Достаточно в теореме в качестве функции $g(x)$ рассмотреть $g(x) = x$.

Аналогичная теорема о вычислении $\mathsf{M}g(\xi)$ по распределению ξ имеет место и для случайной величины со значениями в $\mathbb{R}^n$.

Теорема 10.1.2. *Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — случайная величина со значениями в $\mathbb{R}^n$ и P_ξ — ее распределение, для любой борелевской функции $g(x) = g(x_1, x_2, \ldots, x_n)$ на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$*

$$\mathsf{M}g(\xi) = \int\limits_{\mathbb{R}^n} g(x)\mathsf{P}_\xi(dx),\ x \in \mathbb{R}^n, \tag{10.1.2}$$

в предположении, что определено математическое ожидание $\mathsf{M}g(\xi)$ или определен интеграл $\int\limits_{\mathbb{R}^n} g(x)\mathsf{P}_\xi(dx)$.

Доказательство аналогично доказательству теоремы 10.1.1.

В координатной записи равенство (10.1.2), записывается так

$$\mathsf{M}g(\xi_1, \xi_2, \ldots, \xi_n) = \int\limits_{\mathbb{R}^n} g(x_1, x_2, \ldots, x_n)\mathsf{P}_\xi(d(x_1, x_2, \ldots, x_n)).$$

Далее приведены два наиболее часто встречающихся случая вычисления $\mathsf{M}g(\xi)$ по распределению P_ξ: когда распределение P_ξ абсолютно непрерывно и когда P_ξ дискретно.

Теорема 10.1.3 (о вычислении $\mathsf{M}g(\xi)$ по плотности распределения ξ). *Если ξ — абсолютно непрерывная случайная величина со значениями в $\mathbb{R}^1$ и $p_\xi(x)$ — ее плотность распределения, то для любой борелевской функции $g(x)$ на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$*

$$\mathsf{M}g(\xi) = \int\limits_{\mathbb{R}^1} g(x)p_\xi(x)dx, \tag{10.1.3}$$

в предположении, что определено математическое ожидание $\mathsf{M}g(\xi)$ *или определен интеграл* $\int\limits_{\mathbb{R}^1} g(x)p_\xi(x)dx$.

Доказательство. В предположениях теоремы о вычислении $\mathsf{M}g(\xi)$ по распределению ξ (теорема 10.1.1) и теоремы о вычислении интеграла Лебега по абсолютно непрерывному распределению (теорема 9.6.2)

$$\mathsf{M}g(\xi) = \int\limits_{\mathbb{R}^1} g(x)\mathsf{P}_\xi(dx) = \int\limits_{\mathbb{R}^1} g(x)p_\xi(x)dx.$$

Так что $\mathsf{M}g(\xi)$ определено тогда и только тогда, когда определен интеграл $\int\limits_{\mathbb{R}^1} g(x)p_\xi(x)dx$ и имеет место равенство (10.1.3).

Следствие. *Если* ξ *— абсолютно непрерывная случайная величина со значениями в* $\mathbb{R}^1$ *и* $p_\xi(x)$ *— ее плотность распределения, то*

$$\mathsf{M}\xi = \int\limits_{\mathbb{R}^1} xp_\xi(x)dx,$$

в предположении, что определено математическое ожидание $\mathsf{M}\xi$ *или определен интеграл* $\int\limits_{\mathbb{R}^1} xp_\xi(x)dx$.

Достаточно в теореме в качестве функции $g(x)$ рассмотреть $g(x) = x$.

Аналогичная теорема о вычислении $\mathsf{M}g(\xi)$ по плотности распределения ξ имеет место и для случайной величины со значениями в $\mathbb{R}^n$.

Теорема 10.1.4. *Если* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ *— абсолютно непрерывная случайная величина со значениями в* $\mathbb{R}^n$ *и* $p_\xi(x) = p_\xi(x_1, x_2, \ldots, x_n)$ *— ее плотность распределения, то для любой борелевской функции* $g(x) = g(x_1, x_2, \ldots, x_n)$ *на* $\mathbb{R}^n$ *со значениями в* $\mathbb{R}^1$.

$$\mathsf{M}g(\xi) = \int\limits_{\mathbb{R}^n} g(x)p_\xi(x)dx,\ x \in \mathbb{R}^n, \tag{10.1.4}$$

в предположении, что определено математическое ожидание $\mathsf{M}g(\xi)$ *или определен интеграл* $\int\limits_{\mathbb{R}^1} g(x)p_\xi(x)dx$.

В координатной записи равенство (10.1.4) записывается так

$$\mathsf{M}g(\xi_1, \xi_2, \ldots, \xi_n) =$$

$$= \int\limits_{\mathbb{R}^n} g(x_1, x_2, \ldots, x_n) p_\xi(x_1, x_2, \ldots, x_n) d(x_1, x_2, \ldots, x_n).$$

Доказательство теоремы аналогично доказательству теоремы 10.1.3.

При вычислении математических ожиданий случайных величин часто бывает полезным следующее утверждение.

Теорема 10.1.5. *Если плотность распределения $f(x)$ абсолютно непрерывной случайной величины ξ со значениями в $\mathbb{R}^1$ симметрична относительно прямой $x = a$:*

$$f(a+t) = f(a-t),\ t \in \mathbb{R}^1,$$

то

$$\mathsf{M}\xi = a$$

в предположении, что $\mathsf{M}\xi$ конечно.

Если $f(x)$ — плотность распределения абсолютно непрерывной случайной величины ξ со значениями в $\mathbb{R}^1$, то $f(x-c)$ — плотность распределения случайной величины $\eta = \xi + c$, при этом

$$\mathsf{M}\eta = \mathsf{M}\xi + c.$$

Доказательство. Симметричность плотности $f(x)$ относительно прямой $x = a$:

$$f(a+t) = f(a-t),\ t \in \mathbb{R}^1,$$

означает, что функция $\varphi(t) = f(a+t)$ четная.

Далее,

$$\mathsf{M}\xi = \int\limits_{-\infty}^{+\infty} x f(x) dx.$$

В последнем интеграле сделаем замену $x = t + a$. Имеем

$$\int\limits_{-\infty}^{+\infty} x f(x) dx = \int\limits_{-\infty}^{+\infty} (t+a) f(t+a) dt =$$

$$= a \int\limits_{-\infty}^{+\infty} f(t+a)dt + \int\limits_{-\infty}^{+\infty} tf(t+a)dt = a \cdot 1 + 0 = a,$$

интеграл $\int\limits_{-\infty}^{+\infty} tf(t+a)dt$ равен нулю как интеграл от нечетной функции $tf(t+a)$ по промежутку, симметричному относительно нуля.

Чтобы убедится в справедливости второго утверждения, достаточно заметить, что функция распределения случайной величины η

$$F_\eta(x) = \mathsf{P}\{\xi + c < x\} = F_\xi(x - c),$$

а плотность распределения

$$f_\eta(x) = \frac{d}{dx} F_\eta(x) = \frac{d}{dx} F_\xi(x-c) = f(x-c).$$

Справедливость равенства

$$\mathsf{M}\eta = \mathsf{M}\xi + c$$

очевидна.

Например, как следствие из теоремы получаем: математическое ожидание нормально распределенной с параметрами (a, σ^2) случайной величины равно a (ее плотность симметрична относительно прямой $x = a$); математическое ожидание равномерно распределенной на промежутке $[a, b]$ случайной величины равно $(a + b)/2$ (ее плотность симметрична относительно прямой $x = (a + b)/2$).

Пример 10.1.1. *На отрезок* $[0; 1]$ *наудачу бросают точку, она делит его на две части. Найти распределение и вычислить математическое ожидание длины окружности с радиусом, равным длине большей части отрезка.*

Решение. Пусть ξ — координата наудачу брошенной на отрезок $[0; 1]$ точки, тогда $\eta = \max\{\xi, 1-\xi\}$ — длина большей части отрезка, $\zeta = 2\pi\eta$ — длина окружности с радиусом η.

Сначала найдем функцию распределения случайной величины $\eta = \max\{\xi, 1 - \xi\}$. При $x < 1/2$, очевидно, $\mathsf{P}\{\eta < x\} = 0$, а при $x > 1$ значение $\mathsf{P}\{\eta < x\} = 1$. Если $1/2 < x \le 1$, имеем:

$$\mathsf{P}\{\eta < x\} = \mathsf{P}\{\max\{\xi, 1-\xi\} < x\} = \mathsf{P}\{\xi < x, 1 - \xi < x\} =$$

$$= \mathsf{P}\{1 - x < \xi < x\} = (x - (1 - x))/(1 - 0) = 2x - 1.$$

Вероятность $\mathsf{P}\{1-x < \xi < x\}$ вычислена как геометрическая вероятность события “точка, брошенная наудачу на отрезок $[0;1]$, попадет на отрезок $[1-x;x]$”. Следовательно, функцией распределения случайной величины η является

$$F_\eta(x) = \begin{cases} 0, & \text{если } x \le 1/2; \\ 2x-1, & \text{если } 1/2 < x \le 1; \\ 1, & \text{если } x > 1 \end{cases}$$

(η распределена равномерно на отрезке $[1/2;1]$). Отсюда получаем функцию распределения ζ:

$$F_\zeta(x) = \mathsf{P}\{\zeta < x\} = \mathsf{P}\{2\pi\eta < x\} = \mathsf{P}\left\{\eta < \frac{x}{2\pi}\right\} = F_\eta\left(\frac{x}{2\pi}\right).$$

Поскольку η равномерно распределена на $[1/2;1]$, то

$$\mathsf{M}\zeta = \mathsf{M}2\pi\eta = 2\pi\mathsf{M}\eta = 3\pi/2.$$

Отметим, что $\mathsf{M}\zeta$ можно вычислить и как математическое ожидание функции $\zeta = 2\pi\max\{\xi, 1-\xi\}$ случайной величины ξ по ее распределению (равномерное на отрезке $[0;1]$), см. теорему 10.1.3:

$$\mathsf{M}\zeta = \mathsf{M}2\pi\max\{\xi,1-\xi\} = 2\pi\int\limits_{-\infty}^{+\infty}\max\{x,1-x\}p_\xi(x)dx =$$

$$= 2\pi\int\limits_0^1 \max\{x,1-x\}dx = 3\pi/2.$$

Теорема 10.1.6 (о вычислении $\mathsf{M}g(\xi)$, когда ξ дискретна). *Если ξ — дискретная случайная величина со значениями в $\mathbb{R}^1$ и*

$$\mathsf{P}_\xi : x_i \to \mathsf{P}_\xi(x_i),\ x_i \in X,$$

— ее распределение, то для любой борелевской функции $g(x)$ на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$

$$\mathsf{M}g(\xi) = \sum_{x_i} g(x_i)\mathsf{P}_\xi(x_i), \tag{10.1.5}$$

в предположении, что математическое ожидание $\mathsf{M}g(\xi)$ конечно или ряд $\sum\limits_{x_i} g(x_i)\mathsf{P}_\xi(x_i)$ сходится абсолютно.

Доказательство. В предположениях теоремы о вычислении $\mathsf{M}g(\xi)$ по распределению ξ (теорема 10.1.1) и теоремы о вычислении интеграла Лебега по дискретному распределению (теорема 9.5.4)

$$\mathsf{M}g(\xi) = \int\limits_{\mathbb{R}^1} g(x)\mathsf{P}_\xi(dx) = \sum_{x_i} g(x_i)\mathsf{P}_\xi(x_i).$$

Так что математическое ожидание $\mathsf{M}g(\xi)$ конечно тогда и только тогда, когда ряд $\sum\limits_{x_i} g(x_i)\mathsf{P}_\xi(x_i)$ абсолютно сходится имеет место равенство (10.1.5).

Следствие. *Если ξ — дискретная случайная величина со значениями в $\mathbb{R}^1$ и*

$$\mathsf{P}_\xi : x_i \to \mathsf{P}_\xi(x_i),\ x_i \in X,$$

— ее распределение, то

$$\mathsf{M}\xi = \sum_{x_i} x_i \mathsf{P}_\xi(x_i), \tag{10.1.6}$$

в предположении, что математическое ожидание $\mathsf{M}\xi$ конечно или ряд $\sum\limits_{x_i} x_i\mathsf{P}_\xi(x_i)$ сходится абсолютно.

Достаточно в теореме в качестве функции $g(x)$ рассмотреть $g(x) = x$.

Аналогичная теорема о вычислении $\mathsf{M}g(\xi)$ по распределению ξ, когда ξ дискретна, имеет место и для случайной величины со значениями в $\mathbb{R}^n$.

Теорема 10.1.7. *Если $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — дискретная случайная величина со значениями в $\mathbb{R}^n$ и*

$$\mathsf{P}_\xi : (x_1, x_2, \ldots, x_n) \to \mathsf{P}_\xi(x_1, x_2, \ldots, x_n),\ x = (x_1, x_2, \ldots, x_n) \in X,$$

— ее распределение, то для любой борелевской функции $g(x) = g(x_1, x_2, \ldots, x_n)$ на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$

$$\mathsf{M}g(\xi) = \sum_{x} g(x)\mathsf{P}_\xi(x), \tag{10.1.7}$$

в предположении, что математическое ожидание $\mathsf{M}g(\xi)$ конечно или ряд $\sum\limits_{x} g(x)\mathsf{P}_\xi(x)$ сходится абсолютно.

В координатной записи равенство (10.1.7) записывается так

$$\mathsf{M}g(\xi_1, \xi_2, \ldots, \xi_n) = \sum g(x_1, x_2, \ldots, x_n)\mathsf{P}_\xi(x_1, x_2, \ldots, x_n),$$

суммирование ведется по множеству всех атомов дискретного распределения P_ξ.

Пример 10.1.2. *Случайная величина ξ распределена показательно с параметром θ.*

Найти распределение случайной величины $\eta = [\xi]$, вычислить $\mathsf{M}\eta$ ($[x]$ — целая часть x).

Решение. Случайная величина $\eta = [\xi]$ принимает значения $0, 1, 2, \ldots$ (является дискретной). Ее распределение:

$$\mathsf{P}_\eta(k) = \mathsf{P}\{\eta = k\} = \mathsf{P}\{[\xi] = k\} = \mathsf{P}\{k \le \xi < k+1\} =$$

$$= \int\limits_k^{k+1} \theta e^{-\theta x} dx = e^{-\theta k}(1 - e^{-\theta}) = p(1-p)^k, \; k = 0, 1, 2, \ldots,$$

где $p = 1 - e^{-\theta}$. Так что $\eta = [\xi]$ имеет геометрическое распределение с параметром $p = 1 - e^{-\theta}$.

По известному распределению дискретной случайной величины η ее математическое ожидание вычисляется так (см. теорему 10.1.6):

$$\mathsf{M}\eta = \sum_{k=0}^{\infty} k\mathsf{P}_\eta(k) = \sum_{k=0}^{\infty} k(1-p)^k p = \frac{1-p}{p} = \frac{e^{-\theta}}{1 - e^{-\theta}}.$$

10.2 Моменты. Дисперсия. Неравенство Чебышёва

Определение. Пусть ξ — случайная величина со значениями в $\mathbb{R}^1$ на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$, F — ее распределение, $r = 1, 2, \ldots$ Будем называть

$$\mathsf{M}\xi^r = \int\limits_{\mathbb{R}^1} x^r \mathsf{F}(dx) = m_r$$

моментом порядка r случайной величины ξ (моментом порядка r распределения F);

$$\mathsf{M}|\xi|^r = \int\limits_{\mathbb{R}^1} |x|^r \mathsf{F}(dx) = \mathsf{M}_r$$

— *абсолютным моментом порядка* r случайной величины ξ (абсолютным моментом порядка r распределения F);

$$\mathsf{M}(\xi - m_1)^r = \int\limits_{\mathbb{R}^1} (x - m_1)^r \mathsf{F}(dx)$$

— *центральным моментом порядка* r случайной величины ξ (центральным моментом порядка r распределения F);

$$\mathsf{M}|\xi - m_1|^r = \int\limits_{\mathbb{R}^1} |x - m_1|^r \mathsf{F}(dx)$$

— *абсолютным центральным моментом порядка* r случайной величины ξ (абсолютным центральным моментом порядка r распределения F).

Часто используемой характеристикой случайной величины (распределения) является центральный момент второго порядка, называемый дисперсией случайной величины (дисперсией распределения).

Определение. Пусть ξ — случайная величина с распределением F. *Дисперсией* случайной величины ξ (дисперсией распределения F) будем называть

$$\mathsf{M}(\xi - \mathsf{M}\xi)^2 = \int\limits_{\mathbb{R}^1} (x - m_1)^2 \mathsf{F}(dx).$$

Дисперсия обозначается через $\mathsf{D}\xi$ или σ^2, т. е. по определению

$$\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2 = \int\limits_{\mathbb{R}^1} (x - m_1)^2 \mathsf{F}(dx) = \sigma^2.$$

Величина

$$\sigma = \sqrt{\mathsf{D}\xi}$$

называется *средним квадратическим уклонением* или *стандартным уклонением* случайной величины ξ (стандартным уклонением распределения F).

Определение. Пусть $\mathsf{M}|\xi|^r < \infty$, $\mathsf{M}|\xi_n|^r < \infty$, $n = 1, 2, \ldots$, r — положительное. Будем говорить, что последовательность случайных величин $\{\xi_n\}$ *сходится в среднем порядка* r к случайной величине ξ, если

$$\mathsf{M}|\xi_n - \xi|^r \to 0, \text{ при } n \to \infty.$$

В случае $r = 2$ эта сходимость называется сходимостью *в среднем квадратическом.*

Из неравенства Ляпунова

$$(\mathsf{M}|\zeta|^s)^{1/s} \leq \left(\mathsf{M}|\zeta|^t\right)^{1/t}, \ 0 < s < t,$$

следует, что при $s < t$ сходимость в среднем порядка t влечет сходимость в среднем порядка s.

Теорема 10.2.1 (неравенство Чебышёва). *Пусть $f(x)$ — заданная на положительной полуоси неотрицательная монотонно неубывающая функция и ξ — случайная величина со значениями в $\mathbb{R}^1$. Для любого $a > 0$*

$$\mathsf{P}\{|\xi| \geq a\} \leq \frac{\mathsf{M}f(|\xi|)}{f(a)},$$

в частности,

$$\mathsf{P}\{|\xi| \geq a\} \leq \frac{\mathsf{M}|\xi|}{a},$$

$$\mathsf{P}\{|\xi - \mathsf{M}\xi| \geq a\} \leq \frac{\mathsf{D}\xi}{a^2}.$$

Доказательство. С учетом неотрицательности и монотонности $f(x)$ имеем:

$$\mathsf{M}f(|\xi|) = \int\limits_{\Omega} f(|\xi(\omega)|)\mathsf{P}(d\omega) =$$

$$= \int\limits_{\{\omega:|\xi(\omega)|\geq a\}} f(|\xi(\omega)|)\mathsf{P}(d\omega) + \int\limits_{\{\omega:|\xi(\omega)|<a\}} f(|\xi(\omega)|)\mathsf{P}(d\omega) \geq$$

$$\geq \int\limits_{\{\omega:|\xi(\omega)|\geq a\}} f(|\xi(\omega)|)\mathsf{P}(d\omega) \geq \int\limits_{\{\omega:|\xi(\omega)|\geq a\}} f(a)\mathsf{P}(d\omega) =$$

$$= f(a)\mathsf{P}\{|\xi| \geq a\}.$$

Отсюда

$$\mathsf{P}\{|\xi| \geq a\} \leq \frac{\mathsf{M}f(|\xi|)}{f(a)}.$$

В частности, если $f(x) = x$, то

$$\mathsf{P}\{|\xi| \geq a\} \leq \frac{\mathsf{M}|\xi|}{a}.$$

Если $f(x) = x^2$, а в качестве случайной величины рассмотреть $(\xi - \mathsf{M}\xi)$, то

$$\mathsf{P}\{|\xi - \mathsf{M}\xi| \geq a\} \leq \frac{\mathsf{M}(\xi - \mathsf{M}\xi)^2}{a^2} = \frac{\mathsf{D}\xi}{a^2}$$

(часто используемая форма неравенства Чебышёва).

Из неравенства Чебышёва:

$$\mathsf{P}\{|\xi - \mathsf{M}\xi| \geq a\} \leq \frac{\mathsf{M}|\xi - \mathsf{M}\xi|^r}{a^r}$$

получаем, что из сходимости в среднем следует сходимость по вероятности, обратное, вообще говоря, неверно.

Неравенство Чебышёва для распределений. *Пусть* F *— вероятностное распределение на* $\mathbb{R}^1$ *с конечной дисперсией* σ^2*. Для любого* $a > 0$

$$\mathsf{F}\{x : |x - m_1| \geq a\} \leq \frac{\sigma^2}{a^2} \tag{10.2.1}$$

(масса распределения F*, сосредоточенная на его “хвостах”* $(-\infty, m_1 - a] \cup [m_1 + a, +\infty)$*, не превосходит* σ^2/a^2*).*

Доказательство. Очевидно,

$$\sigma^2 = \int\limits_{\mathbb{R}^1} (x - m_1)^2\mathsf{F}(dx) =$$

$$= \int\limits_{\{x:|x-m_1|\geq a\}} (x-m_1)^2 \mathsf{F}(dx) + \int\limits_{\{x:|x-m_1|<a\}} (x-m_1)^2 \mathsf{F}(dx) \geq$$

$$\geq \int\limits_{\{x:|x-m_1|\geq a\}} (x-m_1)^2 \mathsf{F}(dx) \geq \int\limits_{\{x:|x-m_1|\geq a\}} a^2 \mathsf{F}(dx) =$$

$$= a^2 \mathsf{F}\{x : |x-m_1| \geq a\}.$$

Так что

$$\mathsf{F}\{x : |x-m_1| \geq a\} \leq \frac{\sigma^2}{a^2}.$$

Свойства дисперсии. Далее предполагается, что дисперсии случайных величин конечны.

1. *Для любых a и b (a, b — константы)*

$$\mathsf{D}(a\xi + b) = a^2 \mathsf{D}\xi.$$

В частности,

$$\mathsf{D}(a\xi) = a^2 \mathsf{D}\xi$$

(*константа выносится из-под знака дисперсии с квадратом*) *и*

$$\mathsf{D}b = 0$$

(*дисперсия константы равна нулю*).

Действительно,

$$\mathsf{D}(a\xi + b) = \mathsf{M}(a\xi + b - \mathsf{M}(a\xi + b))^2 = \mathsf{M}(a\xi + b - a\mathsf{M}\xi - b)^2 =$$

$$= a^2 \mathsf{M}(\xi - \mathsf{M}\xi)^2 = a^2 \mathsf{D}\xi.$$

В частности, если $b = 0$, то

$$\mathsf{D}(a\xi) = a^2 \mathsf{D}\xi,$$

если $a = 0$, то

$$\mathsf{D}b = 0.$$

2. *Дисперсия суммы попарно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ равна сумме их дисперсий:*

$$\mathsf{D}\left(\sum_{i=1}^n \xi_i\right) = \sum_{i=1}^n \mathsf{D}\xi_i,$$

Действительно,

$$\mathsf{D}\left(\sum_{i=1}^{n}\xi_i\right) = \mathsf{M}\left(\sum_{i=1}^{n}\xi_i - \mathsf{M}\sum_{i=1}^{n}\xi_i\right)^2 =$$

$$= \mathsf{M}\left(\sum_{i=1}^{n}(\xi_i - \mathsf{M}\xi_i)\right)^2 = \mathsf{M}\sum_{i,j=1}^{n}(\xi_i - \mathsf{M}\xi_i)(\xi_j - \mathsf{M}\xi_j) =$$

$$= \sum_{i=1}^{n}\mathsf{M}(\xi_i - \mathsf{M}\xi_i)(\xi_i - \mathsf{M}\xi_i) + \sum_{i\neq j}\mathsf{M}((\xi_i - \mathsf{M}\xi_i)(\xi_j - \mathsf{M}\xi_j)) =$$

$$= \sum_{i=1}^{n}\mathsf{D}\xi_i + \sum_{i\neq j}\mathsf{M}((\xi_i - \mathsf{M}\xi_i)(\xi_j - \mathsf{M}\xi_j)) = \sum_{i=1}^{n}\mathsf{D}\xi_i.$$

При $i \neq j$ в силу свойства мультипликативности математического ожидания

$$\mathsf{M}(\xi_i - \mathsf{M}\xi_i)(\xi_j - \mathsf{M}\xi_j) = \mathsf{M}(\xi_i - \mathsf{M}\xi_i)\mathsf{M}(\xi_j - \mathsf{M}\xi_j) = 0.$$

3. *Если* $\mathsf{D}\xi = 0$, *то*

$$\mathsf{P}\{\xi = \mathsf{M}\xi\} = 1.$$

Очевидно,

$$\{\omega : |\xi(\omega) - \mathsf{M}\xi| > 0\} = \bigcup_{n=1}^{\infty}\left\{\omega : |\xi(\omega) - \mathsf{M}\xi| > \frac{1}{n}\right\},$$

причем

$$\left\{\omega : |\xi(\omega) - \mathsf{M}\xi| > \frac{1}{n}\right\} \subset \left\{\omega : |\xi(\omega) - \mathsf{M}\xi| > \frac{1}{n+1}\right\},$$

$n = 1, 2, \ldots$ Отсюда, пользуясь свойством непрерывности вероятности, неравенством Чебышёва и учитывая, что $\mathsf{D}\xi = 0$, получаем:

$$\mathsf{P}\{|\xi - \mathsf{M}\xi| > 0\} = \lim_{n}\mathsf{P}\left\{|\xi - \mathsf{M}\xi| > \frac{1}{n}\right\} \leq \lim_{n}\frac{\mathsf{D}\xi}{1/n^2} = 0.$$

Поэтому
$$\mathsf{P}\{|\xi - \mathsf{M}\xi| = 0\} = 1$$
или, что то же,
$$\mathsf{P}\{\xi = \mathsf{M}\xi\} = 1.$$

4. *Для любой константы* c
$$\mathsf{M}(\xi - c)^2 \geq \mathsf{M}(\xi - \mathsf{M}\xi)^2 = \mathsf{D}\xi$$
$(\mathsf{M}(\xi - c)^2$ *принимает наименьшее значение, если* $c = \mathsf{M}\xi)$.

Действительно,
$$\mathsf{M}(\xi - c)^2 = \mathsf{M}((\xi - \mathsf{M}\xi) + (\mathsf{M}\xi - c))^2 = \mathsf{M}(\xi - \mathsf{M}\xi)^2 +$$
$$+2(\mathsf{M}\xi - c)\mathsf{M}(\xi - \mathsf{M}\xi) + (\mathsf{M}\xi - c)^2 = \mathsf{D}\xi + (\mathsf{M}\xi - c)^2 \geq \mathsf{D}\xi.$$

Теорема 10.2.2 (о моментах нормального распределения). *Пусть* ξ *— нормально распределенная с параметрами* $(a; \sigma^2)$ *случайная величина. Тогда*
$$\mathsf{M}(\xi - a)^{2k-1} = 0,\ \mathsf{M}(\xi - a)^{2k} = (2k-1)!!\,\sigma^{2k},$$
$k = 1, 2, \ldots,$ *где* $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2k-1)$. *В частности,*
$$\mathsf{M}\xi = a,\ \mathsf{D}\xi = \sigma^2.$$

Доказательство.
$$\mathsf{M}(\xi - a)^n = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{\mathbb{R}^1} (x-a)^n \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{\mathbb{R}^1} \sigma^{n+1} s^n \exp\left\{-\frac{1}{2}s^2\right\} ds = \frac{\sigma^n}{\sqrt{2\pi}} \int\limits_{\mathbb{R}^1} s^n \exp\left\{-\frac{1}{2}s^2\right\} ds$$
(мы воспользовались заменой переменной $(x-a)/\sigma = s$), т. е.
$$\mathsf{M}(\xi - a)^n = \frac{\sigma^n}{\sqrt{2\pi}} \int\limits_{\mathbb{R}^1} s^n e^{-s^2/2} ds.$$

Если $n = 2k - 1$, то

$$\mathsf{M}(\xi - a)^n = \frac{\sigma^n}{\sqrt{2\pi}} \int\limits_{\mathbb{R}^1} s^n e^{-s^2/2} ds =$$

$$= \frac{\sigma^n}{\sqrt{2\pi}} \lim_m \int\limits_{[-m,m]} s^n e^{-s^2/2} ds = 0$$

(последний интеграл равен нулю как интеграл от нечетной функции по симметричному промежутку).

В частности, при $n = 1$ имеем $\mathsf{M}(\xi - a) = 0$, или

$$\mathsf{M}\xi = a.$$

Следовательно, у нормально распределенной с параметрами $(a; \sigma^2)$ случайной величины математическое ожидание равно a (первый момент $\mathsf{N}_{a;\sigma^2}$-распределения равен a).

Если n четное ($n = 2k$), то

$$\mathsf{M}\,(\xi - a)^n = \mathsf{M}\,(\xi - a)^{2k} = \frac{\sigma^{2k}}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} s^{2k} \exp\left\{-\frac{1}{2}s^2\right\} ds =$$

$$= \frac{2\sigma^{2k}}{\sqrt{2\pi}} \int\limits_{0}^{+\infty} s^{2k} \exp\left\{-\frac{1}{2}s^2\right\} ds.$$

Выполнив в последнем интеграле замену переменной $s^2/2 = t$, получим:

$$\frac{2\sigma^{2k}}{\sqrt{2\pi}} \int\limits_{0}^{+\infty} (2t)^k e^{-t} (2t)^{-1/2} dt = \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \int\limits_{0}^{+\infty} t^{(k+1/2)-1} e^{-t} dt =$$

$$= \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) = \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \left(k + \frac{1}{2} - 1\right) \cdot \left(k + \frac{1}{2} - 2\right) \cdots$$

$$\ldots \left(k + \frac{1}{2} - k\right) \Gamma\left(\frac{1}{2}\right) = \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \cdot \frac{2k-1}{2} \cdot \frac{2k-3}{2} \ldots \frac{1}{2} \Gamma\left(\frac{1}{2}\right) =$$

$$= \frac{\sigma^{2k}}{\sqrt{\pi}}\,(2k-1)!!\,\Gamma\left(\frac{1}{2}\right).$$

Заметим, что

$$\sqrt{2\pi} = \int\limits_{-\infty}^{+\infty} e^{-t^2/2}dt = \sqrt{2}\int\limits_{0}^{+\infty} u^{-1/2}e^{-u}du = \sqrt{2}\,\Gamma\left(\frac{1}{2}\right)$$

(мы воспользовались заменой переменной $t^2/2 = u$). Отсюда

$$\Gamma(1/2) = \sqrt{\pi}.$$

И следовательно,

$$\mathsf{M}\,(\xi - a)^{2k} = (2k-1)!!\,\sigma^{2k}.$$

В частности, при $k = 1$ имеем $\mathsf{M}\,(\xi - a)^2 = 1!!\sigma^2 = \sigma^2$, а учитывая, что $\mathsf{M}\xi = a$, получаем:

$$\mathsf{D}\xi = \mathsf{M}\,(\xi - a)^2 = \sigma^2.$$

Следовательно, у нормально распределенной с параметрами $(a; \sigma^2)$ случайной величины дисперсия равна σ^2 (второй центральный момент $\mathsf{N}_{a;\sigma^2}$-распределения равен σ^2).

Замечание. Значение $\mathsf{M}(\xi - a)^2$ можно вычислить и так:

$$\mathsf{M}(\xi - a)^2 = \frac{\sigma^2}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} s^2 e^{-s^2/2}ds = -\frac{\sigma^2}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} s\,de^{-s^2/2} =$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}}\left(se^{-s^2/2}\Big|_{-\infty}^{+\infty} - \int\limits_{-\infty}^{+\infty} e^{-s^2/2}ds\right) = \sigma^2.$$

Пример 10.2.1 . *Вычислить моменты k-го порядка ($k = 1, 2, 3, \ldots$) и дисперсию случайной величины ξ, имеющей гамма-распределение с параметрами $(\nu; \theta)$:*

$$p(x) = \begin{cases} \dfrac{\theta^\nu}{\Gamma(\nu)}x^{\nu-1}\exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0. \end{cases}$$

Р е ш е н и е. В силу теоремы 10.1.3 о вычислении $\mathsf{M}g(\xi)$ по распределению ξ имеем:

$$\mathsf{M}\xi = \int\limits_{-\infty}^{+\infty} xp(x)dx = \int\limits_{0}^{+\infty} x\frac{\theta^{\nu}}{\Gamma(\nu)}x^{\nu-1}\exp\{-\theta x\}dx =$$

$$= \frac{\Gamma(\nu+1)}{\theta\Gamma(\nu)}\int\limits_{0}^{+\infty}\frac{\theta^{\nu+1}}{\Gamma(\nu+1)}x^{(\nu+1)-1}\exp\{-\theta x\}dx =$$

$$= \frac{\Gamma(\nu+1)}{\theta\Gamma(\nu)}\cdot 1 = \frac{\nu\Gamma(\nu)}{\theta\Gamma(\nu)} = \frac{\nu}{\theta}.$$

$$\int\limits_{0}^{+\infty}\frac{\theta^{\nu+1}}{\Gamma(\nu+1)}x^{\nu}\exp\{-\theta x\}\,dx = 1$$

как интеграл от плотности гамма-распределения с параметрами $(\nu+1;\theta)$.

Аналогично имеем:

$$\mathsf{M}\xi^{k} = \frac{\nu(\nu+1)\dots(\nu+(k-1))}{\theta^{k}},\ k = 2,3,\dots$$

Дисперсия

$$\mathsf{D}\xi = \mathsf{M}\xi^{2} - (\mathsf{M}\xi)^{2} = \frac{\nu(\nu+1)}{\theta^{2}} - \left(\frac{\nu}{\theta}\right)^{2} = \frac{\nu}{\theta^{2}}.$$

Пример 10.2.2 . *Примером распределения, у которого не определен момент первого порядка, является распределение Коши.*

По определению момент первого порядка распределения F

$$m_1 = \int\limits_{-\infty}^{+\infty} x^{+}\mathsf{F}(dx) - \int\limits_{-\infty}^{+\infty} x^{-}\mathsf{F}(dx),$$

если только интегралы от положительной части x^{+} и отрицательной части x^{-} не равны ∞.

Для распределения Коши имеем

$$\int\limits_{-\infty}^{+\infty}(x^+)\frac{1}{\pi}\frac{1}{1+x^2}dx = \int\limits_{0}^{+\infty}x\frac{1}{\pi}\frac{1}{1+x^2}dx =$$

$$= \frac{1}{2\pi}\ln(1+x^2)\Big|_0^{+\infty} = +\infty,$$

и аналогично

$$\int\limits_{-\infty}^{+\infty}(x^-)\frac{1}{\pi}\frac{1}{1+x^2}dx = \int\limits_{-\infty}^{0}(-x)\frac{1}{\pi}\frac{1}{1+x^2}dx = +\infty.$$

Поэтому момент первого порядка распределения Коши не определен (не существует).

Пример 10.2.3. *Пусть ξ_i, $i = 1, 2, \ldots,$ — независимые одинаково распределенные случайные величины с конечными моментами второго порядка, ν — пуассоновская случайная величина, не зависящая от случайных величин ξ_i, $i = 1, 2, \ldots$.*

Случайная величина S_ν определена равенством

$$S_\nu = \sum_{j=0}^{\nu} \xi_j,\ \xi_0 = 0.$$

Вычислить $\mathsf{M}S_\nu$, $\mathsf{D}S_\nu$.

Решение. Равенство

$$S_\nu = \sum_{j=0}^{\nu} \xi_j,$$

следует понимать так

$$S_\nu = \sum_{k=0}^{\infty} S_k I_{\{\nu=k\}},\ S_k = \sum_{j=0}^{k} \xi_j,\ \xi_0 = 0.$$

Обозначим первый и второй моменты случайной величины ξ_i соответственно через m_1 и m_2, параметр пуассоновского распределения обозначим через θ.

В силу свойства счётной аддитивности и мультипликативности математического ожидания имеем

$$\mathsf{M}S_\nu = \sum_{k=1}^{\infty} \mathsf{M}(S_\nu I_{\{\nu=k\}}) =$$

$$= \sum_{k=1}^{\infty} \mathsf{M}(S_k I_{\{\nu=k\}}) = \sum_{k=1}^{\infty} \mathsf{M}S_k \mathsf{M}I_{\{\nu=k\}} =$$

$$\sum_{k=1}^{\infty} \mathsf{M}S_k \cdot \frac{\theta^k}{k!} e^{-\theta} = \sum_{k=1}^{\infty} m_1 k \frac{\theta^k}{k!} e^{-\theta} = m_1\theta.$$

Так что

$$\mathsf{M}S_\nu = m_1\theta.$$

Далее,

$$\mathsf{D}S_\nu = \mathsf{M}S_\nu^2 - (\mathsf{M}S_\nu)^2.$$

Вычислим $\mathsf{M}S_\nu^2$.

$$\mathsf{M}S_\nu^2 = \sum_{k=1}^{\infty} \mathsf{M}(S_\nu^2 I_{\{\nu=k\}}) =$$

$$= \sum_{k=1}^{\infty} \mathsf{M}(S_k^2 I_{\{\nu=k\}}) = \sum_{k=1}^{\infty} \mathsf{M}S_k^2 \mathsf{M}I_{\{\nu=k\}} = \sum_{k=1}^{\infty} \mathsf{M}S_k^2 \frac{\theta^k}{k!} e^{-\theta};$$

$$\mathsf{M}S_k^2 = \mathsf{M}\left(\sum_{i=1}^{k} \xi_i\right)^2 = \mathsf{M}\left(\sum_{i=1}^{k} \xi_i^2 + \sum_{i,j,i\neq j}^{k} \xi_i\xi_j\right) =$$

$$= km_2 + k(k-1)m_1^2,$$

поэтому

$$\mathsf{M}S_\nu^2 = \sum_{k=1}^{\infty} (km_2 + k(k-1)m_1^2) \frac{\theta^k}{k!} e^{-\theta} =$$

$$= m_2 \sum_{k=1}^{\infty} k \frac{\theta^k}{k!} e^{-\theta} + m_1^2 \left(\sum_{k=1}^{\infty} k^2 \frac{\theta^k}{k!} e^{-\theta} - \sum_{k=1}^{\infty} k \frac{\theta^k}{k!} e^{-\theta}\right) =$$

$$= m_2\theta + m_1^2(\theta^2 + \theta - \theta) = m_2\theta + m_1^2\theta^2.$$

Так что
$$\mathsf{M}S_\nu^2 = m_2\theta + m_1^2\theta^2,$$
$$\mathsf{D}S_\nu = m_2\theta + m_1^2\theta^2 - (m_1\theta)^2 = m_2\theta.$$

Моменты некоторых абсолютно непрерывных распределений и случайных величин. Далее приведены вычисленные ранее моменты некоторых абсолютно непрерывных распределений (см. теорему 10.2.2, пример 10.2.1).

Моменты нормального распределения с параметрами $(a;\sigma^2)$ (*моменты нормально распределенной с параметрами* $(a;\sigma^2)$ *случайной величины* ξ)*:*

$$\mathsf{M}\xi = a,\ \mathsf{D}\xi = \sigma^2,$$

$$\mathsf{M}(\xi - a)^{2k-1} = 0,\ \mathsf{M}(\xi - a)^{2k} = (2k-1)!!\sigma^{2k},\ k = 1, 2, \ldots$$

Моменты гамма-распределения с параметрами (ν, θ) (*моменты гамма-распределенной с параметрами* (ν, θ) *случайной величины* ξ)*:*
$$m_1 = \frac{\nu}{\theta},\ \sigma^2 = \frac{\nu}{\theta^2}.$$
$$m_k = \frac{\nu(\nu+1)\ldots(\nu+(k-1))}{\theta^k},\ k = 1, 2, \ldots$$

Моменты распределения Эрланга с параметрами (m, θ) *— гамма-распределения с параметрами* (m, θ)*:*

$$m_1 = \frac{m}{\theta},\ \sigma^2 = \frac{m}{\theta^2},$$

$$m_k = \frac{m(m+1)\ldots(m+(k-1))}{\theta^k},\ k = 1, 2, \ldots$$

Моменты показательного распределения с параметром θ *— гамма-распределения с параметрами* $(1, \theta)$*:*

$$m_1 = \frac{1}{\theta},\ \sigma^2 = \frac{1}{\theta^2},$$

$$m_k = \frac{1\cdot 2\ldots k}{\theta^k},\ k = 1, 2, 3, \ldots$$

Моменты χ^2-распределения с параметром n — гамма-распределения с параметрами $(n/2, 1/2)$

$$m_1 = n,\ \sigma^2 = 2n.$$

Моменты равномерного на промежутке $[a, b]$ распределения, моменты равномерно распределенной на промежутке $[a, b]$ случайной величины ξ:

$$m_1 = \frac{a+b}{2},\ \sigma^2 = \frac{(b-a)^2}{12}.$$

10.3 Примеры и задачи

Примеры

Пример 10.3.1. *Случайная величина ξ с распределением F имеет конечное математическое ожидание $\mathsf{M}\xi = a$.*

Пусть

$$\xi_n = \xi I_{\{y:|y|\le n\}}(\xi),\ \eta_n = \xi I_{\{y:|y|>n\}}(\xi).$$

Доказать, что

$$\lim_n \mathsf{M}\eta_n = 0,\ \lim_n \mathsf{M}\xi_n = \mathsf{M}\xi = a.$$

Решение. Если

$$a = \int\limits_{\mathbb{R}^1} x\mathsf{F}(dx) \neq \infty,$$

то функция множества

$$\mathsf{Q}(A) = \int\limits_A x\mathsf{F}(dx),$$

является непрерывной по монотонно убывающей последовательности множеств

$$A_n = (-\infty, -n) \cup (n, +\infty),\ n = 1, 2, \ldots, \bigcap_{n=1}^{\infty} A_n = \emptyset$$

Поэтому

$$\lim_n \mathsf{Q}(A_n) = \lim_n \int\limits_{A_n} x\mathsf{F}(dx) = \int\limits_{\emptyset} x\mathsf{F}(dx) = 0,$$

И следовательно,

$$\lim_n \mathsf{M}\eta_n = \lim_n \int\limits_{\mathbb{R}^1} xI_{A_n}(x)\mathsf{F}(dx) = \lim_n \int\limits_{A_n} x\mathsf{F}(dx) = 0.$$

Далее, поскольку

$$\xi = \xi_n + \eta_n,$$

то

$$\mathsf{M}\xi = \mathsf{M}\xi_n + \mathsf{M}\eta_n.$$

Отсюда

$$\lim_n \mathsf{M}\xi_n = \mathsf{M}\xi = a.$$

Задачи

10.1. Случайная величина η имеет распределение $\mathsf{N}_{0;\sigma^2}$. Вычислить математическое ожидание случайной величины $\eta^+ = \max\{0, \eta\}$.

Ответ: $\mathsf{M}\eta^+ = \sigma/\sqrt{2\pi}$.

10.2. Пусть $\zeta = (\xi, \eta)$ — абсолютно непрерывный вектор с плотностью $f_\zeta(x, y)$. Вычислить $\mathsf{M}\xi$, $\mathsf{M}\eta$.

Указание. Случайная величина ξ является функциею вектора $\zeta = (\xi, \eta)$:

$$\xi = g(\zeta) = g(\xi, \eta),$$

поэтому

$$\mathsf{M}\xi = \mathsf{M}g(\zeta) = \int\limits_{\mathbb{R}^2} g(x, y) f_\zeta(x, y) d(x, y) = \int\limits_{\mathbb{R}^2} x f_\zeta(x, y) d(x, y).$$

10.3. На окружности с радиусом R наудачу берут две точки. Найти функцию распределения расстояния η между ними и вычислить $\mathsf{M}\eta$.

Указание. Поскольку при любом повороте окружности вероятность события, зависящего от от расстояния между двумя точками окружности, остается неизменной, то одну из точек можно считать фиксированной.

Ответ:

$$F_\eta(x) = \begin{cases} 0, & \text{если } x \le 0; \\ \dfrac{2}{\pi}\arcsin\dfrac{x}{2R}, & \text{если } 0 < x \le 2R; \\ 1, & \text{если } x > 2R. \end{cases}$$

10.4. Пусть ξ — случайная величина с плотностью распределения

$$p(x) = \frac{1}{\pi(1+x^2)}.$$

Вычислить $\mathsf{M}\min\{|\xi|, 1\}$.

Ответ: $(\ln 2)/\pi + 1/2$.

10.5. На отрезок $[0; 1]$ наудачу бросают точку. Она делит его на две части. Найти распределение и вычислить математическое ожидание случайной величины η:

1. Длины окружности с радиусом, равным длине: а) меньшей части отрезка; б) большей части отрезка.

2. Площади круга с радиусом, равным длине: а) меньшей части отрезка; б) большей части отрезка.

10.6. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, каждая из которых имеет своей плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \le \theta; \\ \dfrac{1}{\alpha}\exp\left\{-\dfrac{1}{\alpha}(x-\theta)\right\}, & \text{если } x > \theta,\ \alpha > 0. \end{cases}$$

Вычислить математическое ожидание случайных величин:

$$1)\, \overline{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i;\ \ 2)\ \min\{\xi_i\};$$

$$3)\, \hat{\theta}_1 = \min\{\xi_i\} - \frac{\overline{\xi} - \min\{\xi_i\}}{n};\ \ 4)\, \hat{\theta}_2 = \overline{\xi} - \hat{\theta}_1.$$

Ответы: 1) $\mathsf{M}\overline{\xi} = \theta + \alpha$; 2) $\mathsf{M}\min\{\xi_i\} = \theta + \alpha/n$; 3) $\mathsf{M}\hat{\theta}_1 = \theta + \alpha/n^2$; 4) $\mathsf{M}\hat{\theta}_2 = \alpha(1 - 1/n^2)$.

10.7. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, каждая из которых имеет своей плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \notin [\theta - h; \theta + h]; \\ 1/2h, & \text{если } x \in [\theta - h; \theta + h]. \end{cases}$$

Вычислить математическое ожидание случайных величин:

$$1)\ \min\{\xi_1, \xi_2, \ldots, \xi_n\};\ \ 2)\ \max\{\xi_1, \xi_2, \ldots, \xi_n\};$$

$$3)\ (\max\{\xi_1, \xi_2, \ldots, \xi_n\} - \min\{\xi_1, \xi_2, \ldots, \xi_n\})/2.$$

Ответы: 1) $\theta - \dfrac{n-1}{n+1}h$; 2) $\theta + \dfrac{n-1}{n+1}h$; 3) $\dfrac{n-1}{n+1}h$.

10.8*. Пусть $A_x = \{(u, v) : u + v < x\}$ — множество в $\mathbb{R}^2$, x — произвольное, но фиксированное число, $\zeta = (\xi, \eta)$ — случайная величина со значениями в $\mathbb{R}^2$.

Вычислить $\mathsf{M}I_{A_x}(\xi, \eta)$, если известно, что:

1) случайные величины ξ, η независимы и имеют своими распределениями F и G соответственно;

2) случайные величины ξ, η независимы и абсолютно непрерывны с плотностями f и g соответственно.

Указание. Поскольку ξ и η независимые случайные величины, распределением случайной величины (ξ, η) является $\mathsf{F} \times \mathsf{G}$. Согласно теореме о вычислении математического ожидания функции от случайной величины по ее распределению, имеем

$$\mathsf{M}I_{A_x}(\xi, \eta) = \int\limits_{\mathbb{R}^2} I_{A_x}(u, v)\mathsf{F} \times \mathsf{G}(d(u, v)) =$$

$$= \int\limits_{\mathbb{R}^1} \left(\int\limits_{\mathbb{R}^1} I_{A_x}(u, v)\mathsf{F}(du) \right) \mathsf{G}(dv) = \int\limits_{\mathbb{R}^1} F(x - v)\mathsf{G}(dv).$$

А если ξ и η абсолютно непрерывные независимые случайные величины, то для последнего интеграла имеем:

$$\int\limits_{\mathbb{R}^1} F(x - v)\mathsf{G}(dv) = \int\limits_{-\infty}^{+\infty} \left(\int\limits_{-\infty}^{x-v} f(s)ds \right) g(v)dv.$$

Заметим, что $\mathsf{M}I_{A_x}(\xi, \eta)$, как математическое ожидание от простой случайной величины $I_{A_x}(\xi, \eta)$, равно $\mathsf{P}\{\xi + \eta < x\}$ — функции распределения случайной величины $\xi + \eta$ в точке x.

10.9. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые одинаково распределенные случайные величины (каждая с распределением F) со значениями в $\mathbb{R}^1$;

$$\mathbb{R}^1 = \bigcup_{i=1}^{r} X_i,\ \ X_i \cap X_j = \emptyset,\ \ i \neq j;$$

ν_i — число случайных величин $\xi_1, \xi_2, \dots, \xi_n$, попавших в X_i, $p_i = \mathsf{F}(X_i) = P\{\xi_k \in X_i\}$ — вероятность, того, что ξ_k попадет в X_i, $i = 1, 2, \dots, r$.

Вычислить

$$\mathsf{M}\nu_i,\ \mathsf{D}\nu_i,\ i = 1, 2, \dots, r.$$

Выписать распределение вектора $\nu = (\nu_1, \nu_2, \dots, \nu_r)$.

У к а з а н и е. Очевидно

$$\nu_i = \sum_{k=1}^{n} I_{X_i}(\xi_k),\ i = 1, 2, \dots, r.$$

10.10. Пусть $\xi_1, \xi_2, \dots, \xi_n$, — независимые одинаково распределенные случайные величины, каждая с функцией распределения $F(x)$.

Функция $\hat{F}_n(x)$, определенная на $\mathbb{R}^1$ равенством

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{(-\infty,x)}(\xi_k),$$

называется эмпирической функцией распределения.

При каждом фиксированном x эмпирическая функция распределения $\hat{F}_n(x)$, как функция от случайных величин $\xi_1, \xi_2, \dots$ $\dots, \xi_n$, является случайной величиной.

1° Найти распределение $\hat{F}_n(x)$.

2° Показать, что

$$\mathsf{M}\hat{F}_n(x) = F(x),$$

$$\mathsf{D}\hat{F}_n(x) = \frac{1}{n}F(x)(1 - F(x)).$$

Глава 11

Свертка

11.1 Свертка и распределение суммы независимых случайных величин

Определение. *Сверткой* борелевской функции φ на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$ с вероятностным распределением F на $\mathbb{R}^1$ называется функция $u = u(x)$, определяемая для каждого $x \in \mathbb{R}^1$ равенством

$$u(x) = \int_{\mathbb{R}^1} \varphi(x-y)\mathsf{F}(dy),$$

обозначается свертка φ с F так:

$$u = \mathsf{F} * \varphi.$$

Свертка ограниченной борелевской функции φ с вероятностным распределением F всегда определена — ограниченная функция интегрируема по вероятностному распределению.

Если вероятностное распределение F абсолютно непрерывно и f его плотность, то

$$u(x) = \mathsf{F} * \varphi(x) = \int_{\mathbb{R}^1} \varphi(x-y)\mathsf{F}(dy) = \int_{\mathbb{R}^1} \varphi(x-y)f(y)dy,$$

а если F — атомическое распределение, сосредоточенное в атомах $y_1, y_2, \ldots$, то

$$u(x) = \mathsf{F} * \varphi(x) = \int_{\mathbb{R}^1} \varphi(x-y)\mathsf{F}(dy) = \sum_{y_i} \varphi(x-y_i)\mathsf{F}\left(\{y_i\}\right).$$

Заметим, что свертка функции φ с распределением F определяется и для распределений F, отличных от вероятностных.

З а м е ч а н и е. Порядок компонент в свертке $\mathsf{F} * \varphi$ существенен. Символ $\varphi * \mathsf{F}$, вообще говоря, смысла не имеет.

Пример 11.1.1. *Пусть* $\varphi(x) = I_{[0;1)}(x)$, F — *атомическое распределение:*

$$\mathsf{F}(\{i\}) = 1/3,\ i = 0, 1, 2.$$

Найти $\mathsf{F} * \varphi$.

Р е ш е н и е. По определению свертки $u(x) = \mathsf{F} * \varphi(x)$

$$u(x) = \int\limits_{-\infty}^{+\infty} \varphi(x-y)\mathsf{F}(dy) = \int\limits_{-\infty}^{+\infty} I_{[0;1)}(x-y)\mathsf{F}(dy) =$$

$$= I_{[0;1)}(x)\cdot\mathsf{F}(\{0\}) + I_{[0;1)}(x-1)\cdot\mathsf{F}(\{1\}) + I_{[0;1)}(x-2)\cdot\mathsf{F}(\{2\}) =$$

$$= I_{[0;1)}(x)\cdot\frac{1}{3} + I_{[0;1)]}(x-1)\cdot\frac{1}{3} + I_{[0;1)}(x-2)\cdot\frac{1}{3} = \frac{1}{3}\cdot I_{[0;3)}(x).$$

Теорема 11.1.1 (свойства свертки). *Свертка* $\mathsf{F} * \varphi$ *функции* φ *с вероятностным распределением* F

1) *ограничена, если* φ *ограничена;*

2) *ограничена и непрерывна, если* φ *ограничена и непрерывна;*

3) *является вероятностной функцией распределения, если* φ — *вероятностная функция распределения.*

Д о к а з а т е л ь с т в о. Пусть

$$u = \mathsf{F} * \varphi.$$

Если φ ограничена ($|\varphi(x)| \le C < +\infty$, $x \in \mathbb{R}^1$), то

$$|u(x)| \le \int\limits_{\mathbb{R}^1} |\varphi(x-y)|\mathsf{F}(dy) \le C < +\infty.$$

Далее, пусть φ ограничена и непрерывна. Для каждого $x \in \mathbb{R}^1$ (произвольного, но фиксированного)

$$u(x+h) - u(x) = \int\limits_{\mathbb{R}^1} \varphi(x+h-y)\mathsf{F}(dy) - \int\limits_{\mathbb{R}^1} \varphi(x-y)\mathsf{F}(dy) =$$

$$= \int\limits_{\mathbb{R}^1} (\varphi(x+h-y) - \varphi(x-y))\mathsf{F}(dy).$$

Подынтегральная функция

$$\psi_h(y) = \varphi(x+h-y) - \varphi(x-y),\ y \in \mathbb{R}^1,$$

во-первых, мажорируема интегрируемой по вероятностному распределению функцией $2C$ и, во-вторых, сходится при $h \to 0$ к функции $\psi(y) = 0$ поскольку φ непрерывна. Поэтому согласно теореме Лебега о мажорируемой сходимости при $h \to 0$

$$u(x+h) - u(x) = \int\limits_{\mathbb{R}^1} \psi_h(y)\mathsf{F}(dy) \to 0.$$

Последнее означает, что функция u непрерывна в точке x.

Пусть φ — вероятностная функция распределения — неотрицательная монотонно неубывающая непрерывная слева функция, имеющая пределы на "бесконечностях": $\varphi(-\infty) = 0$, $\varphi(+\infty) = 1$. Тогда $u = \mathsf{F} * \varphi$ непрерывна слева (следует из ограниченности и непрерывности слева φ), и монотонна (следует из монотонности φ). Далее, в интеграле

$$u(x) = \int\limits_{\mathbb{R}^1} \varphi(x-y)\mathsf{F}(dy)$$

подынтегральная функция $\psi_x(y) = \varphi(x-y), y \in \mathbb{R}^1$, мажорируема на $\mathbb{R}^1$ интегрируемой по вероятностному распределению F функцией $\psi(y) \equiv 1$ и при $x \to +\infty$ сходится к функции $\psi(y) \equiv 1$. Поэтому в силу теоремы Лебега о мажорируемой сходимости

$$\lim_{x\to+\infty} u(x) = \lim_{x\to+\infty} \int\limits_{\mathbb{R}^1} \varphi(x-y)\mathsf{F}(dy) =$$

$$= \int\limits_{\mathbb{R}^1} \lim_{x\to+\infty} \varphi(x-y)\mathsf{F}(dy) = \int\limits_{\mathbb{R}^1} 1\mathsf{F}(dy) = 1.$$

Аналогично проверяется, что

$$\lim_{x\to-\infty} u(x) = 0.$$

Таким образом, $u(x) = \mathsf{F} * \varphi(x)$ — вероятностная функция распределения.

Свертка вероятностных распределений. *Сверткой вероятностного распределения* G *с вероятностным распределением* F будем называть вероятностное распределение Q, функция распределения $Q(x)$ которого является сверткой функции распределения $G(x)$ с распределением F:

$$Q(x) = \int\limits_{\mathbb{R}^1} G(x-y)\mathsf{F}(dy).$$

Обозначать свертку вероятностного распределения G с вероятностным распределением F будем так $\mathsf{F} * \mathsf{G}$:

$$\mathsf{Q} = \mathsf{F} * \mathsf{G}$$

Корректность приведенного определения следует из теоремы 11.1.1.

Напомним, что распределение, не имеющего атомов, называется непрерывным.

Из теоремы 11.1.1 как следствие, в частности, получаем следующее утверждение:

Свертка $\mathsf{Q} = \mathsf{F} * \mathsf{G}$ *непрерывного вероятностного распределения* G *с вероятностным распределением* F *— является непрерывным вероятностным распределением.*

Далее мы установим, что для вероятностных распределений F и G

$$\mathsf{F} * \mathsf{G} = \mathsf{G} * \mathsf{F},$$

поэтому можно говорить о свертке вероятностных распределений G и F.

Напомним, что распределение F на $\mathbb{R}^n$ сосредоточено на множестве $A \subset \mathbb{R}^n$, если $\mathsf{F}(\overline{A}) = 0$.

Теорема 11.1.2 (о свертке атомических распределений). *Свертка*

$$\mathsf{Q} = \mathsf{F} * \mathsf{G}$$

атомических вероятностных распределений G *и* F*, сосредоточенных на множестве* $\{0, 1, \dots\}$ *является атомическим распределением, сосредоточенным на том же множестве, причем*

$$\mathsf{Q}(\{k\}) = \sum_{j=0}^{k} \mathsf{G}(\{k-j\})\mathsf{F}(\{j\}),\ k = 0, 1, \dots \tag{11.1.1}$$

Доказательство. По определению функция распределения свертки $\mathsf{Q} = \mathsf{F} * \mathsf{G}$ вероятностных распределений F и G

$$Q(x) = \int\limits_{\mathbb{R}^1} G(x-y)\mathsf{F}(dy).$$

Значение распределения Q в точке $x \in \mathbb{R}^1$ равно

$$\mathsf{Q}(\{x\}) = Q(x+0) - Q(x-0).$$

Воспользовавшись теоремой Лебега о мажорируемой сходимости получаем

$$\mathsf{Q}(\{x\}) = \int\limits_{\mathbb{R}^1} (G(x-y+0)-G(x-y-0))\mathsf{F}(dy) = \int\limits_{\mathbb{R}^1} \mathsf{G}(\{x-y\})\,\mathsf{F}(dy)$$

Так как F — атомическое распределение, сосредоточенное на множестве $\{0, 1, \dots\}$, то в силу теоремы о вычислении интеграла Лебега по атомическому распределению (теорема 9.5.4)

$$\mathsf{Q}(\{x\}) = \sum_{j=0}^{\infty} \mathsf{G}(\{x-j\})\,\mathsf{F}(\{j\}).$$

Но G — также атомическое распределение, сосредоточенное на множестве $\{0, 1, \dots\}$, поэтому значение $\sum\limits_{j=0}^{\infty} \mathsf{G}(\{x-j\})\,\mathsf{F}(\{j\}) = \mathsf{Q}(\{x\})$ может быть отлично от нуля только для целых неотрицательных x, а следовательно, распределение Q сосредоточено на множестве $\{0, 1, \dots\}$ при этом

$$\mathsf{Q}(\{k\}) = \sum_{j=0}^{k} \mathsf{G}(\{k-j\})\mathsf{F}(\{j\}),\ k = 0, 1, \dots$$

Замечание. Если обозначить

$$\mathsf{G}(\{i\}) = g_i,\ i = 0, 1, \dots;\ \mathsf{F}(\{j\}) = f_j,\ j = 0, 1, \dots;$$

$$\mathsf{Q}(\{k\}) = q_k,\ k = 0, 1, \dots$$

то равенство (11.1.1) запишется так

$$q_k = \sum_{j=0}^{k} g_{k-j} f_j,\ k = 0, 1, \dots$$

Теорема 11.1.3 (о плотности свертки). *Свертка*

$$\mathsf{V} = \mathsf{F} * \mathsf{G}$$

абсолютно непрерывного вероятностного распределения G *с вероятностным распределением* F *является абсолютно непрерывным вероятностным распределением и его плотность* v *равна свертке плотности* g *абсолютно непрерывного распределения* G *с распределением* F*:*

$$v = \mathsf{F} * g.$$

Доказательство. Функция распределения $V(x)$ свертки $V = \mathsf{F} * \mathsf{G}$ определяется равенством

$$V(x) = \int\limits_{-\infty}^{+\infty} G(x-y)\mathsf{F}(dy).$$

Отсюда, учитывая, что G — абсолютно непрерывное распределение с плотностью g, имеем:

$$V(x) = \int\limits_{-\infty}^{+\infty} G(x-y)\mathsf{F}(dy) = \int\limits_{-\infty}^{+\infty} \left(\int\limits_{-\infty}^{x-y} g(t)dt \right) \mathsf{F}(dy).$$

Выполнив замену переменной $s = t + y$, получим:

$$\int\limits_{-\infty}^{+\infty} \left(\int\limits_{-\infty}^{x-y} g(t)dt \right) \mathsf{F}(dy) = \int\limits_{-\infty}^{+\infty} \left(\int\limits_{-\infty}^{x} g(s-y)ds \right) \mathsf{F}(dy) =$$

$$= \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{+\infty} g(s-y)\mathsf{F}(dy) \right) ds$$

(воспользовались теоремой Фубини).

Итак, для каждого $x \in \mathbb{R}^1$ функция распределения $V(x)$ представима в виде:

$$V(x) = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{+\infty} g(s-y)\mathsf{F}(dy) \right) ds = \int\limits_{-\infty}^{x} v(s)\, ds,$$

поэтому в силу следствия 1 из теоремы о достаточном условии абсолютной непрерывности распределения (см. теорему 9.6.1) распределение V абсолютно непрерывно, и его плотность $v(s)$ равна $\int\limits_{-\infty}^{+\infty} g(s-y)\mathsf{F}(dy)$, т. е.

$$v = \mathsf{F} * g.$$

Следствие. *Свертка* $\mathsf{V} = \mathsf{F} * \mathsf{G}$ *абсолютно непрерывных вероятностных распределений* G *и* F *соответственно с плотностями* g *и* f *абсолютно непрерывна, и ее плотность* v *равна свертке плотностей* g *и* f*:*

$$v(x) = \int\limits_{\mathbb{R}^1} g(x-t)f(t)dt.$$

Действительно,

$$v(x) = \int\limits_{-\infty}^{+\infty} g(x-t)\mathsf{F}(dt) = \int\limits_{-\infty}^{+\infty} g(x-t)f(t)dt.$$

Свертку плотностей g *и* f *будем обозначать через* $f * g$*, т. е.*

$$v = f * g.$$

Следующая теорема дает вероятностную интерпретацию свертки $\mathsf{F} * \mathsf{G}$ вероятностных распределений.

Теорема 11.1.4 . *Распределение суммы независимых случайных величин равно свертке распределений слагаемых.*

Другими словами, если ξ *и* η *— независимые случайные величины соответственно с распределениями* F *и* G*, то распределение* Q *суммы* $\xi + \eta$ *равно* $\mathsf{F} * \mathsf{G}$*, т. е.*

$$Q(x) = \int\limits_{\mathbb{R}^1} G(x-u)\mathsf{F}(du) = \int\limits_{\mathbb{R}^1} F(x-u)\mathsf{G}(du). \tag{11.1.2}$$

Доказательство. Пусть $Q(x)$ — функция распределения случайной величины $\xi + \eta$:

$$Q(x) = \mathsf{P}\{\xi + \eta < x\}.$$

Если через $I_{A_x} = I_{A_x}(u, v)$ обозначить индикатор множества $A_x = \{(u, v) : u + v < x\} \subset \mathbb{R}^2$, то для $Q(x)$ получим представление

$$Q(x) = \mathsf{P}\{\xi + \eta < x\} = \mathsf{M} I_{A_x}(\xi, \eta)$$

— достаточно вычислить $\mathsf{M} I_{A_x}(\xi, \eta)$ как математическое ожидание простой случайной величины $I_{A_x}(\xi, \eta)$, принимающей значение 1, если $\xi + \eta < x$, и значение 0, если $\xi + \eta \geq x$.

Далее вычислим $\mathsf{M} I_{A_x}(\xi, \eta)$ как математическое ожидание функции $I_{A_x}(\xi, \eta)$ от случайной величины (ξ, η) со значениями в $\mathbb{R}^2$ по ее распределению, см. теорему 10.1.2. (Распределение (ξ, η), как совместное распределение независимых случайных величин ξ и η соответственно с распределениями F и G, равно произведению $\mathsf{F} \times \mathsf{G}$ распределений F и G.) Имеем

$$\mathsf{M} I_{A_x}(\xi, \eta) = \int\limits_{\mathbb{R}^2} I_{A_x}(u, v) \mathsf{F} \times \mathsf{G}(d(u, v)).$$

Последний интеграл вычислим, воспользовавшись теоремой Фубини:

$$\int\limits_{\mathbb{R}^2} I_{A_x}(u, v) \mathsf{F} \times \mathsf{G}(d(u, v)) = \int\limits_{\mathbb{R}^1} \left(\int\limits_{\mathbb{R}^1} I_{A_x}(u, v) \mathsf{G}(dv) \right) \mathsf{F}(du).$$

При фиксированном u функция $I_{A_x}(u, v)$ является индикатором множества $(-\infty, x - u)$, поэтому

$$\int\limits_{\mathbb{R}^1} \left(\int\limits_{\mathbb{R}^1} I_{A_x}(u, v) \mathsf{G}(dv) \right) \mathsf{F}(du) =$$

$$= \int\limits_{\mathbb{R}^1} \big(1 \cdot \mathsf{G}((-\infty, x - u)) + 0 \cdot \mathsf{G}([x - u, +\infty))\big) \mathsf{F}(du) =$$

$$= \int\limits_{\mathbb{R}^1} G(x - u) \mathsf{F}(du)$$

(см. также рис. 11.1.1). В итоге получаем

$$Q(x) = \int\limits_{\mathbb{R}^1} G(x-u)\mathsf{F}(du).$$

Аналогично получаем

$$Q(x) = \int\limits_{\mathbb{R}^1} F(x-u)\mathsf{G}(du).$$

Тем самым теорема доказана.

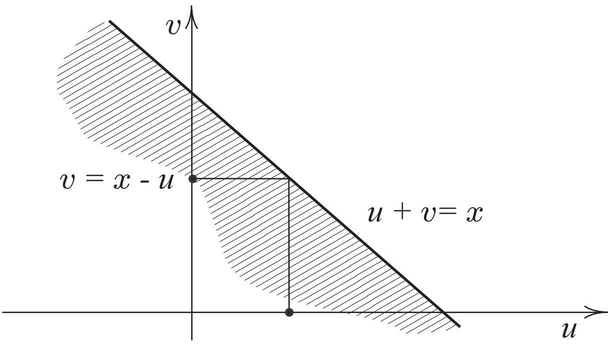

Рис. 11.1.1: К вычислению интеграла $\int\limits_{\mathbb{R}^1}\Big(\int\limits_{\mathbb{R}^1} I_{A_x}(u,v)\mathsf{G}(dv)\Big)\mathsf{F}(du)$

Следствие 1. *В классе вероятностных распределений операция свертки коммутативна:*

$$\mathsf{F} * \mathsf{G} = \mathsf{G} * \mathsf{F}$$

и ассоциативна:

$$\mathsf{F} * (\mathsf{G} * \mathsf{Q}) = (\mathsf{F} * \mathsf{G}) * \mathsf{Q}.$$

Доказательство. Коммутативность операции свертки доказана в теореме 11.1.4 (см. (11.1.2)).

Ассоциативность операции свертки: следует из ассоциативности сложения случайных величин: для независимых случайных величин ξ, η, ζ соответственно с распределениями $\mathsf{F}, \mathsf{G}, \mathsf{Q}$,

$$\mathsf{F} * (\mathsf{G} * \mathsf{Q}) = (\mathsf{F} * \mathsf{G}) * \mathsf{Q},$$

поскольку

$$\xi + (\eta + \zeta) = (\xi + \eta) + \zeta.$$

Из теоремы 11.1.4 и теоремы о плотности свертки (теорема 11.1.3) получим ряд утверждений о распределении суммы независимых абсолютно непрерывных случайных величин.

Следствие 2. *Сумма независимых случайных величин, хотя бы одна из которых абсолютно непрерывна, является абсолютно непрерывной случайной величиной, и ее плотность равна свертке плотности абсолютно непрерывной случайной величины с распределением другой.*

Другими словами, если ξ и η — независимые случайные величины с распределениями F и G, причем ξ абсолютно непрерывна и f — ее плотность, то сумма $\xi+\eta$ абсолютно непрерывна и ее плотность p равна свертке плотности f с распределением G:

$$p(x) = \int\limits_{\mathbb{R}^1} f(x-y)\mathsf{G}(dy)$$

или

$$p = \mathsf{G} * f.$$

Распределение P суммы $\xi+\eta$ равно $\mathsf{G}*\mathsf{F}$ (см. теорему 11.1.4). А так как F абсолютно непрерывно и f — плотность распределения F, то в силу теоремы о плотности свертки (теорема 11.1.3)

$$p(x) = \int\limits_{\mathbb{R}^1} f(x-y)\mathsf{G}(dy).$$

Следствие 3. *Плотность суммы независимых абсолютно непрерывных случайных величин равна свертке плотностей слагаемых.*

Другими словами, если ξ и η — независимые абсолютно непрерывные случайные величины соответственно с плотностями f

и g, то плотность p распределения суммы $\xi+\eta$ равна свертке плотностей f и g :

$$p(x)=\int\limits_{\mathbb{R}^1} f(x-y)g(y)dy=\int\limits_{\mathbb{R}^1} g(x-y)f(y)dy$$

или

$$p=g*f=f*g.$$

Для доказательства достаточно воспользоваться следствием 2 и теоремой о вычислении интеграла Лебега по абсолютно непрерывному распределению.

Следствие 4. *Пусть ξ и η — независимые абсолютно непрерывные случайные величины соответственно с плотностями f и g, тогда плотность распределения разности $\zeta=\xi-\eta$*

$$p(x)=\int\limits_{\mathbb{R}^1} f(x+y)g(y)dy.$$

Действительно, сумма $\zeta=\xi-\eta=\xi+(-\eta)$ имеет своей плотностью функцию

$$p(x)=\int\limits_{\mathbb{R}^1} f(x-t)s(t)dt, \tag{11.1.3}$$

где $s(t)$ — плотность случайной величины $-\eta$. Выразим $s(t)$ через плотность $g(t)$ случайной величины η:

$$F_{(-\eta)}(t)=\mathsf{P}\{-\eta<t\}=\mathsf{P}\{\eta>-t\}=\int\limits_{-t}^{+\infty} g(u)du=1-\int\limits_{-\infty}^{-t} g(u)du,$$

$$s(t)=\frac{d}{dt}F_{(-\eta)}(t)=\frac{d}{dt}\left(1-\int\limits_{-\infty}^{-t} g(u)du\right)=g(-t).$$

Поэтому равенство (11.1.3) можно переписать так:

$$p(x)=\int\limits_{-\infty}^{+\infty} f(x-t)g(-t)dt,$$

или, после замены переменных $y = -t$,

$$p(x) = \int\limits_{-\infty}^{+\infty} f(x+y)g(y)dy.$$

Пример 11.1.2. *Пусть* F *и* G *— вероятностные распределения на* $\mathbb{R}^1$, $u(s)$ *— борелевская функция на* $\mathbb{R}^1$ *со значениями в* $\mathbb{R}^1$. *Доказать, что*

$$\int\limits_{\mathbb{R}^1} u(s)\mathsf{F} * \mathsf{G}(ds) = \int\limits_{\mathbb{R}^2} u(x+y)\mathsf{F} \times \mathsf{G}(d(x,y)),$$

в предположении, что интегралы определены.

Решение. Для независимых случайных величин ξ и η соответственно с распределениями F и G, вычислим $\mathsf{M}u(\xi+\eta)$ сначала как математическое ожидание функции от случайной величины $\zeta = \xi + \eta$ (с распределением $\mathsf{F} * \mathsf{G}$), а затем как математическое ожидание $\mathsf{M}u(\xi+\eta)$ функции от случайной величины (ξ, η) (с распределением $\mathsf{F} \times \mathsf{G}$).

Теорема 11.1.5. *Семейство гамма-плотностей замкнуто относительно операции свертки:*

$$f_{\mu;\theta} * f_{\nu;\theta} = f_{\mu+\nu;\theta}$$

(*сумма независимых гамма-распределенных с параметрами* $(\nu;\theta)$ *и* $(\mu;\theta)$, *случайных величин есть гамма-распределенная с параметрами* $(\mu+\nu;\theta)$ *случайная величина*).

Доказательство. Гамма-распределение с параметрами $(\nu;\theta)$ абсолютно непрерывно и его плотность

$$f_{\nu;\theta}(x) = \begin{cases} \dfrac{\theta^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\theta x} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

При $x \leq 0$ равенство

$$f_{\mu;\theta} * f_{\nu;\theta} = f_{\mu+\nu;\theta}$$

очевидно.

При $x > 0$ имеем

$$f_{\mu;\theta} * f_{\nu;\theta}(x) = \int_0^\infty f_{\nu;\theta}(x-y) f_{\mu;\theta}(y) dy = \int_0^x f_{\nu;\theta}(x-y) f_{\mu;\theta}(y) dy =$$

$$= \int_0^x \frac{\theta^\nu}{\Gamma(\nu)} (x-y)^{\nu-1} e^{-\theta(x-y)} \frac{\theta^\mu}{\Gamma(\mu)} y^{\mu-1} e^{-\theta y} dy =$$

$$= \frac{\theta^{\nu+\mu}}{\Gamma(\nu)\Gamma(\mu)} e^{-\theta x} \int_0^x (x-y)^{\nu-1} y^{\mu-1} dy.$$

Выполнив в последнем интеграле замену переменной $y = xt$, получим:

$$\frac{\theta^{\nu+\mu}}{\Gamma(\nu)\Gamma(\mu)} e^{-\theta x} \int_0^1 (x-xt)^{\nu-1} (xt)^{\mu-1} x dt =$$

$$= \frac{\theta^{\nu+\mu}}{\Gamma(\nu)\Gamma(\mu)} e^{-\theta x} x^{\nu+\mu-1} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt =$$

$$= \frac{\theta^{\nu+\mu}}{\Gamma(\nu+\mu)} x^{\nu+\mu-1} e^{-\theta x} \frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt =$$

$$= f_{\mu+\nu;\theta}(x) \cdot \frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt.$$

Следовательно,

$$f_{\mu;\theta} * f_{\nu;\theta}(x) = f_{\mu+\nu;\theta}(x) \cdot \frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt. \quad (11.1.4)$$

И поскольку $f_{\mu;\theta} * f_{\nu;\theta}(x)$ и $f_{\mu+\nu;\theta}(x)$ — плотности, то, проинтегрировав (11.1.4) по $\mathbb{R}^1$, получаем, что

$$\frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int\limits_0^1 t^{\mu-1}(1-t)^{\nu-1}dt = 1. \qquad (11.1.5)$$

Так что

$$f_{\mu;\theta} * f_{\nu;\theta} = f_{\mu+\nu;\theta}.$$

Тем самым теорема доказана.

11.2 Примеры и задачи

Примеры

Пример 11.2.1. *Независимым образом бросают две монеты, одна из которых симметричная, вероятность выпадения герба на другой монете равна 3/4. Пусть ξ — число выпавших гербов на первой монете, η — на второй. Найти распределение случайной величины $\zeta = \xi + \eta$.*

Решение. Распределения случайных величин ξ и η:

$$\mathsf{P}_\xi(i) = f_i, i = 0, 1;\ f_0 = 1/2,\ f_1 = 1/2;$$

$$\mathsf{P}_\eta(j) = g_j, j = 0, 1;\ g_0 = 1/4,\ g_1 = 3/4.$$

Согласно теореме 11.1.2 распределение случайной величины $\zeta = \xi + \eta$ равно свертке распределений P_ξ и P_η:

$$\mathsf{P}_\zeta(k) = q_k = \sum_{i=0}^{k} f_{k-i}g_i = \sum_{i=0}^{k} g_{k-i}f_i,\ k = 0, 1, 2.$$

Подробнее:

$$\mathsf{P}_\zeta(0) = q_0 = \sum_{i=0}^{0} f_{0-i}g_i = f_0g_0 = \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8},$$

$$\mathsf{P}_\zeta(1) = q_1 = \sum_{i=0}^{1} f_{1-i}g_i = f_1g_0 + f_0g_1 = \frac{1}{2}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{3}{4} = \frac{1}{2},$$

$$\mathsf{P}_\zeta(2) = q_2 = \sum_{i=0}^{2} f_{2-i} g_i = f_2 g_0 + f_1 g_1 + f_0 g_2 = f_1 g_1 = \frac{1}{2}\cdot\frac{3}{4} = \frac{3}{8}$$

($\mathsf{P}_\xi(2) = f_2 = 0$ так как $f_0 + f_1 = 1$; $\mathsf{P}_\eta(2) = g_2 = 0$, поскольку $g_0 + g_1 = 1$).

Пример 11.2.2. *Пусть ξ и η — независимые случайные величины, каждая из которых равномерно распределена на $[0;1]$.*

Найти:

1) *плотность распределения суммы $\zeta = \xi + \eta$;*

2) *функцию распределения суммы $\zeta = \xi + \eta$;*

3) $\mathsf{P}\{|\xi + \eta - 1/2| < 1\}$.

Решение.

1) По известным плотностям распределений

$$f_\xi(y) = \begin{cases} 1, & \text{если } y \in [0;1]; \\ 0, & \text{если } y \notin [0;1], \end{cases}$$

$$f_\eta(y) = \begin{cases} 1, & \text{если } y \in [0;1]; \\ 0, & \text{если } y \notin [0;1] \end{cases}$$

случайных величин ξ и η плотность распределения суммы $\zeta = \xi + \eta$ находим как свертку плотностей слагаемых (см. следствие из теоремы 11.1.4):

$$f_\zeta(x) = \int\limits_{\mathbb{R}^1} f_\xi(x-y) f_\eta(y)dy = \int\limits_0^1 f_\xi(x-y)dy = \int\limits_{x-1}^{x} f_\xi(t)dt.$$

Вычислим последний интеграл для каждого значения $x \in (-\infty, +\infty)$. Имеем:

если $x < 0$, то $\int\limits_{x-1}^{x} f_\xi(t)dt = \int\limits_{x-1}^{x} 0dt = 0$;

если $0 < x < 1$, то $\int\limits_{x-1}^{x} f_\xi(t)dt = \int\limits_{x-1}^{0} 0dt + \int\limits_{0}^{x} 1dt = x$;

если $0 < x-1 < 1$, то $\int\limits_{x-1}^{x} f_\xi(t)dt = \int\limits_{x-1}^{1} 1dt + \int\limits_{1}^{x} 0dt = 2 - x$;

если $x - 1 > 1$, то $\int\limits_{x-1}^{x} f_\xi(t)dt = \int\limits_{x-1}^{x} 0dt = 0$.

Так что плотность распределения случайной величины $\zeta = \xi + \eta$

равна

$$f_\zeta(x) = \begin{cases} 0, & \text{если } x < 0; \\ x, & \text{если } 0 \le x < 1; \\ 2 - x, & \text{если } 1 \le x < 2; \\ 0, & \text{если } x \ge 2. \end{cases}$$

2) Функция распределения $F_\zeta(x)$ случайной величины ζ по ее плотности распределения $f_\zeta(t)$ находится так:

$$F_\zeta(x) = \int\limits_{-\infty}^{x} f_\zeta(t)dt =$$

$$= \begin{cases} \int\limits_{-\infty}^{x} 0dt = 0, & \text{если } x < 0; \\ \int\limits_{-\infty}^{0} 0dt + \int\limits_{0}^{x} tdt = x^2/2, & \text{если } 0 \le x < 1; \\ \int\limits_{0}^{1} tdt + \int\limits_{1}^{x} (2-t)dt = -(x-2)^2/2 + 1, & \text{если } 1 \le x < 2; \\ \int\limits_{-\infty}^{0} 0dt + \int\limits_{0}^{1} tdt + \int\limits_{1}^{2} (2-t)dt = 1, & \text{если } x \ge 2. \end{cases}$$

3) Вероятность того, что значение ζ попадает в множество B, по известной плотности распределения $f_\zeta(t)$ случайной величины ζ вычисляется следующим образом:

$$\mathsf{P}\{\zeta \in B\} = \int\limits_{B} f_\zeta(t)dt$$

(см. следствие из теоремы 9.7.3). В частности, в данной задаче

$$\mathsf{P}\left\{|\xi + \eta - 1/2| < 1\right\} = \mathsf{P}\left\{|\zeta - 1/2| < 1\right\} =$$

$$= \mathsf{P}\left\{-1/2 \le \zeta \le 3/2\right\} =$$

$$= \int\limits_{-1/2}^{3/2} f_\zeta(t)dt = \int\limits_{-1/2}^{0} 0dt + \int\limits_{0}^{1} tdt + \int\limits_{1}^{3/2} (2-t)dt = \frac{7}{8}.$$

Пример 11.2.3. *Пусть* $\xi_1, \xi_2, \ldots, \xi_n$ — *независимые случайные величины, каждая с распределением* $N_{0;1}$. *Найти распределение случайной величины*

$$\eta = \sum_{i=1}^{n} \xi_i^2.$$

Решение. Сначала покажем, что если ξ распределена $N_{0;1}$, то $\eta = \xi^2$ имеет гамма-распределение с параметрами $(1/2; 1/2)$.

В самом деле,

$$F_\eta(x) = \mathsf{P}\{\eta < x\} = \mathsf{P}\{\xi^2 < x\}.$$

При $x \le 0$ значение $F_\eta(x) = \mathsf{P}\{\xi^2 < x\}$ равно нулю, а при $x > 0$

$$F_\eta(x) = \mathsf{P}\{|\xi| < \sqrt{x}\} = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} \exp\{-t^2/2\} dt.$$

Или после замены $t^2 = s$ под знаком интеграла:

$$F_\eta(x) = \frac{1}{\sqrt{2\pi}} \int_0^x s^{-1/2} e^{-s/2} ds = \frac{(1/2)^{1/2}}{\Gamma(1/2)} \int_0^x s^{1/2-1} e^{-s/2} ds,$$

а это функция гамма-распределения с параметрами $(1/2; 1/2)$ (воспользовались тем, что $\Gamma(1/2) = \sqrt{\pi}$).

Далее, поскольку гамма-распределение замкнуто относительно операции свертки (см. теорему 11.1.5), то сумма n независимых гамма-распределенных с параметрами $(1/2; 1/2)$ случайных величин является гамма-распределенной случайной величиной с параметрами $(n/2; 1/2)$.

Гамма-распределение с параметрами $(n/2; 1/2)$ еще называют χ^2-распределением с n степенями свободы.

Пример 11.2.4 . *Пусть* ξ *и* η — *независимые случайные величины,* ξ *распределена равномерно на* $[0; 1)$, η *имеет своим распределением* $\mathsf{P}_\eta(k) = \mathsf{P}\{\eta = k\} = 1/2$, $k = 0, 1$. *Найти распределение случайной величины* $\zeta = \xi + \eta$.

Р е ш е н и е. Случайная величина ξ абсолютно непрерывна (ее плотность $f_\xi(x) = I_{[0,1)}(x)$), поэтому $\zeta = \xi + \eta$ также абсолютно непрерывна и ее плотность

$$f_\zeta(x) = \int\limits_{\mathbb{R}^1} I_{[0,1)}(x-y)\mathsf{P}_\eta(dy) = \frac{1}{2}I_{[0,1)}(x) + \frac{1}{2}I_{[0,1)}(x-1).$$

или

$$f_\zeta(x) = \begin{cases} 0, & \text{если } x \notin [0,2]; \\ 1/2, & \text{если } x \in [0,2]. \end{cases}$$

Задачи

11.1. Найти свертку функции $\varphi = \varphi(x)$, заданной на $\mathbb{R}^1$, с атомическим распределением, сосредоточенным в точке a.

Ответ: $\varphi(x-a)$.

11.2. Пусть Q — собственное атомическое распределение, сосредоточенное в точке 1, а распределение Q_n задано равенством

$$\mathsf{Q}_n = \mathsf{Q}^{n*} = \mathsf{Q}^{(n-1)*} * \mathsf{Q},\ n = 2, 3, \ldots,\ \mathsf{Q}^{1*} = \mathsf{Q}.$$

Найти Q_n.

Ответ: собственное атомическое распределение, сосредоточенное в точке n.

11.3. Пусть $\mathsf{U}_{[0,a]}$ — равномерное на промежутке $[0,a]$ распределение, P_θ — пуассоновское распределение с параметром θ, G_p — геометрическое распределение с параметром p.

Найти свертки:

1) $\mathsf{P}_\theta * \mathsf{U}_{[0,1]}$,

2) $\mathsf{P}_\theta * \mathsf{U}_{[0,1/2]} * \mathsf{U}_{[0,1/2]}$,

3) $\mathsf{G}_p * \mathsf{U}_{[0,1]}$,

4) $\mathsf{G}_p * \mathsf{U}_{[0,1/2]} * \mathsf{U}_{[0,1/2]}$,

Построить графики плотностей полученных сверток.

Ответы:

1) Плотность свертки $\mathsf{P}_\theta * \mathsf{U}_{[0,1]}$:

$$f_1(x) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} e^{-\theta} I_{[0,1)}(x-k).$$

2) Сначала находим плотность свертки $\mathsf{U}_{[0,1/2]} * \mathsf{U}_{[0,1/2]}$:

$$\varphi(x) = I_{[0,1)}(x)(2 - 4|x - 1/2|).$$

Плотность свертки $\mathsf{P}_\theta * \mathsf{U}_{[0,1/2]} * \mathsf{U}_{[0,1/2]} = \mathsf{P}_\theta * \varphi(x)$ равна

$$f_2(x) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} e^{-\theta} \varphi(x-k).$$

11.4. Пусть μ — считающая мера — атомическая мера, сосредоточенная на множестве $\mathbb{N}$, такая, что

$$\mu(\{k\}) = 1,\ k \in \mathbb{N}.$$

и пусть $\varphi = \varphi(x)$ — борелевская функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$.

Найти свертку $\mu * \varphi$, если

1) $\varphi(x) = 4I_{[0,1]}(x)(x-1/2)^2$,

2) $\varphi(x) = I_{[0,1]}(x)(1-4(x-1/2)^2)$,

3) $\varphi(x) = I_{[0,1]}(x)|x-1/2|$,

4) $\varphi(x) = I_{[0,1]}(x)(1/2-|x-1/2|)$.

Построить графики функций $\mu * \varphi(x)$.

11.5. Подбрасывают симметричную игральную кость и наудачу бросают точку на отрезок $[0;1]$. Пусть η — число выпавших очков, а ξ — координата точки на отрезке $[0;1]$. Найти распределение случайной величины $\zeta = \xi + \eta$.

Ответ: равномерное на промежутке $[1;7]$ распределение.

11.6. Найти плотность распределения суммы n независимых случайных величин, каждая из которых имеет показательное распределение с параметром λ. Сравните ее с плотностью распределения Эрланга.

11.7. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые равномерно распределенные на $[0;1]$ случайные величины. Найти распределение случайной величины

$$\eta = \xi_1 \xi_2 \ldots \xi_n.$$

Вычислить $\mathsf{M}\eta$.

Ответ:

$$\mathsf{M}\eta = \mathsf{M}\xi_1\xi_2\ldots\xi_n = \mathsf{M}\xi_1\mathsf{M}\xi_2\ldots\mathsf{M}\xi_n = (1/2)^n.$$

Чтобы найти распределение η, воспользуйтесь тем, что случайная величина $-\ln \xi_i$ имеет показательное распределение.

11.8. Пусть F и G — вероятностные распределения на $\mathbb{R}^1$. Доказать, что

$$\int\limits_{\mathbb{R}^1} s\,\mathsf{F} * \mathsf{G}(ds) = \int\limits_{\mathbb{R}^1} x\,\mathsf{F}d(x) + \int\limits_{\mathbb{R}^1} y\,\mathsf{G}d(y),$$

при условии, что интегралы определены.

Указание. Утверждение можно получить и из равенства

$$\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta,$$

записав его в терминах распределений соответствующих случайных величин.

11.9. Пусть ξ и η — независимые случайные величины, случайная величина ξ распределена равномерно на $[0; 2]$, случайная величина η — на $[-2; 0]$. Вычислить

$$\mathsf{P}\left\{|\xi - 1| + |\eta + 1| > 1\right\}.$$

11.10. Пусть ξ и η — независимые случайные величины; ξ распределена равномерно на $[-1; 1]$, η — на $[0; 1]$.

Вычислить:

1) $\mathsf{P}\{\xi^2 + \eta > 1/2\}$;
2) $\mathsf{P}\{\xi + \eta > 1\}$;
3) $\mathsf{P}\{|\xi + \eta| > 1/2\}$;
4) $\mathsf{P}\{|\eta - \xi| < 1/2\}$;
5) $\mathsf{P}\{\eta^2 - \xi > 0\}$;
6) $\mathsf{P}\{|\xi| + \eta > 1\}$.

Указание. Сначала найти распределение суммы (разности) соответствующих случайных величин.

Другой подход — найти совместную плотность распределения ξ и η, а затем воспользоваться теоремой 9.7.3.

Ответы: 1) $1 - 1/(3\sqrt{2})$; 2) $1/4$; 3) $9/16$; 4) $7/16$; 5) $2/3$; 6) $1/2$.

Глава 12

Сходимость распределений

12.1 Сходимость распределений: определения, примеры

Излагаемые далее результаты не зависят от размерности n пространства $\mathbb{R}^n$. Но мы для наглядности будем рассматривать одномерный случай.

Определение. Интервал I вида $[a, b)$ будем называть *интервалом непрерывности распределения* F, если его концы a и b не являются атомами F.

Напомним, что точка x_0 является атомом распределения F, если $\mathsf{F}(\{x_0\}) > 0$.

Определение. Будем говорить, что последовательность распределений $\{\mathsf{F}_n\}$ *сходится к распределению* F, если для каждого ограниченного интервала непрерывности I распределения F

$$\mathsf{F}_n(I) \to \mathsf{F}(I)$$

при $n \to \infty$.

Эту сходимость распределений будем обозначать так:

$$\mathsf{F}_n \to \mathsf{F}$$

или

$$\lim_n \mathsf{F}_n = \mathsf{F}.$$

Предел сходящейся последовательности распределений единственен.

Введенное определение сходимости распределений, разумеется, имеет место и для вероятностных распределений.

Сходимость вероятностных распределений. Далее мы расширим понятия вероятностного распределения и вероятностной функции распределения.

Распределение F на $\mathbb{R}^1$ будем называть *вероятностным*, если $\mathsf{F}(\mathbb{R}^1) \leq 1$. Вероятностное распределение F будем называть *собственным*, если $\mathsf{F}(\mathbb{R}^1) = 1$ и *несобственным*, если $\mathsf{F}(\mathbb{R}^1) < 1$.

Функцию распределения $F(x)$ на $\mathbb{R}^1$ будем называть *вероятностной*, если $0 \leq F(x) \leq 1$. Вероятностную функцию распределения будем называть *собственной*, если $F(+\infty) = 1$ и $F(-\infty) = 0$, и *несобственной*, если $F(+\infty) < 1$ или $F(-\infty) > 0$.

Несобственные вероятностные распределения естественным образом возникают как пределы последовательностей вероятностных распределений.

Лемма 12.1.1. *Предел последовательности вероятностных распределений является вероятностным распределением* (*собственным или несобственным*).

Действительно, пусть $\{\mathsf{F}_n\}$ — последовательность вероятностных распределений, сходящаяся к некоторому распределению F. Для каждого интервала непрерывности $[a_s, b_s)$ распределения F

$$\mathsf{F}([a_s, b_s)) = \lim_n \mathsf{F}_n([a_s, b_s)) \leq 1.$$

Отсюда при $a_s \downarrow -\infty, b_s \uparrow +\infty$ в силу свойства непрерывности распределения имеем

$$\mathsf{F}(\mathbb{R}^1) = \lim_s \mathsf{F}([a_s, b_s)) \leq 1.$$

Так что F — вероятностное распределение.

Предел последовательности $\{\mathsf{F}_n\}$ вероятностных распределений может быть как собственным, так и несобственным (это иллюстрируют приводимые далее примеры). Мы будем различать сходимость последовательности $\{\mathsf{F}_n\}$ вероятностных распределений к собственному вероятностному распределению и к несобственному.

Определение. Если последовательность вероятностных распределений $\{\mathsf{F}_n\}$ сходится и предельное распределение F собственное, то будем говорить, что $\{\mathsf{F}_n\}$ сходится к F *в собственном смысле* (*сходимость* $\mathsf{F}_n \to \mathsf{F}$ *собственная*).

Если $\{\mathsf{F}_n\}$ сходится и предельное распределение F несобственное, то будем говорить, что $\{\mathsf{F}_n\}$ сходится к F *в несобственном смысле* (*сходимость* $\mathsf{F}_n \to \mathsf{F}$ *несобственная*).

Далее мы будем рассматривать сходимость вероятностных распределений.

Пример 12.1.1. *Пусть* F — *собственное непрерывное (не имеющее атомов) вероятностное распределение. Рассмотрим последовательность* $\{\mathsf{F}_n\}$ *вероятностных распределений соответственно с функциями распределения*

$$F_n(x) = F(x + 1/n).$$

Покажем, что $\{\mathsf{F}_n\}$ *сходится к распределению* F.

Так как $F(x)$ непрерывна, то для каждого $x \in \mathbb{R}^1$

$$\lim_n F_n(x) = \lim_n F(x + 1/n) = F(x),$$

поэтому для любого интервала $[a, b)$

$$\lim_n \mathsf{F}_n([a, b)) = \lim_n (F_n(b) - F_n(a)) =$$

$$= \lim_n F_n(b) - \lim_n F_n(a) = F(b) - F(a) = \mathsf{F}([a, b)).$$

Последнее по определению означает, что $\mathsf{F}_n \to \mathsf{F}$ при $n \to \infty$. И поскольку предельное распределение F — собственное, то сходимость $\mathsf{F}_n \to \mathsf{F}$ собственная.

Пример 12.1.2. *Пусть* F — *собственное непрерывное вероятностное распределение. Рассмотрим последовательность вероятностных распределений* $\{\mathsf{F}_n\}$ *соответственно с функциями распределения*

$$F_n(x) = F(x/n).$$

Покажем, что при $n \to \infty$ *последовательность распределений* $\{\mathsf{F}_n\}$ *сходится к несобственному вероятностному распределению* Q, *тождественно равному нулю (распределение* $\mathsf{Q} \equiv 0$, *если для каждого* $B \in \mathfrak{B}^1$ *выполняется равенство* $\mathsf{Q}(B) = 0$).

Действительно, учитывая непрерывность $F(x)$, имеем:

$$\lim_n \mathsf{F}_n([a, b)) = \lim_n (F_n(b) - F_n(a)) =$$

$$= \lim_n (F(b/n) - F(a/n)) = F(0) - F(0) = 0 = \mathsf{Q}([a, b))$$

для любого промежутка $[a, b)$, т. е. $\mathsf{F}_n \to \mathsf{Q}$ при $n \to \infty$. А так как Q — несобственное вероятностное распределение, то сходимость $\mathsf{F}_n \to \mathsf{Q}$ несобственная.

Сходимость в основном. Последовательность $\{F_n(x)\}$ функций распределения *сходится в основном* к функции распределения $F(x)$, если $\{F_n(x)\}$ сходится к $F(x)$ в каждой точке непрерывности $F(x)$. (Сходимость $F_n(x)$ в основном к $F(x)$ это сходимость $\mathsf{F}_n((-\infty, x))$ к $\mathsf{F}((-\infty, x)$ в каждой точке непрерывности $F(x)$.)

Достаточное условие сходимости распределений. *Сходимость в основном последовательности $\{F_n(x)\}$ вероятностных функций распределения к функции распределения $F(x)$ влечет сходимость соответствующей последовательности распределений $\{\mathsf{F}_n\}$ к распределению F (из сходимости в основном функций распределения следует сходимость распределений).*

Действительно, если $\{F_n(x)\}$ сходится к $F(x)$ в основном, то для каждого ограниченного интервала непрерывности $[a, b)$ распределения F

$$\mathsf{F}_n([a, b)) = F_n(b) - F_n(a) \to F(b) - F(a) = \mathsf{F}([a, b)),$$

последнее обозначает, что $\mathsf{F}_n \to \mathsf{F}$.

Обратное неверно. *Из сходимости распределений $\mathsf{F}_n \to \mathsf{F}$, вообще говоря, не следует сходимость в основном функций распределений $F_n(x)$ к $F(x)$.* Приведем пример.

Пусть $F(x)$ — собственная вероятностная функция распределения. Рассмотрим последовательность распределений $\{\mathsf{F}_n\}$ с функциями распределения

$$F_n(x) = F\left(x + (-1)^n n\right).$$

Для каждого $x \in \mathbb{R}^1$ имеем:

$$\lim_{k\to\infty} F_{2k}(x) = 1, \ \lim_{k\to\infty} F_{2k+1}(x) = 0.$$

Поэтому последовательность функций распределений $F_n(x)$ не сходится ни в одной точке $x \in \mathbb{R}^1$, хотя легко убедиться, что

$$\mathsf{F}_n \to \mathsf{Q} \equiv 0.$$

Собственная же сходимость последовательности вероятностных распределений эквивалентна сходимости в основном их функций распределения.

Теорема 12.1.1. *Собственная сходимость последовательности вероятностных распределений $\{\mathsf{F}_n\}$ к распределению F равносильна сходимости в основном $\{F_n(x)\}$ к собственной функции распределения $F(x)$.*

Доказательство. Из сходимости в основном $\{F_n(x)\}$ к собственной вероятностной функции распределения $F(x)$ в силу достаточного условия сходимости следует собственная сходимость $\{\mathsf{F}_n\}$ к F.

Покажем обратное: из собственной сходимости $\{\mathsf{F}_n\}$ к F следует сходимость $\{F_n(x)\}$ к $F(x)$ в основном.

Пусть $\varepsilon > 0$. Так как F — собственное вероятностное распределение ($\mathsf{F}((-\infty, +\infty)) = 1$), то в силу непрерывности вероятности найдется такой конечный интервал $[a, b)$ (дополнительно выберем его интервалом непрерывности распределения F), что

$$\mathsf{F}[a, b)) \geq 1 - \varepsilon. \tag{12.1.1}$$

А поскольку $\mathsf{F}_n \to \mathsf{F}$, то, начиная с некоторого N ($n \geq N$), и

$$\mathsf{F}_n([a, b)) \geq 1 - 2\varepsilon. \tag{12.1.2}$$

Из неравенства (12.1.1) следует, что

$$\mathsf{F}\left((-\infty, a) \cup [b, +\infty)\right) \leq \varepsilon,$$

а из неравенства (12.1.2) следует, что при $n \geq N$

$$\mathsf{F}_n((-\infty, a) \cup [b, +\infty)) \leq 2\varepsilon.$$

Пусть x — точка непрерывности функции распределения $F(x)$ и $n \geq N$.

1. Если $x < a$, то

$$|F_n(x) - F(x)| = |\mathsf{F}_n((-\infty, x)) - \mathsf{F}((-\infty, x))| \leq$$

$$\leq \mathsf{F}_n((-\infty, a)) + \mathsf{F}((-\infty, a)) \leq 3\varepsilon.$$

2. Если $x \geq b$, то

$$|F_n(x) - F(x)| = |\mathsf{F}_n((-\infty, x)) - \mathsf{F}((-\infty, x)) + 1 - 1| \leq$$

$$\leq |1 - \mathsf{F}_n((-\infty, x))| + |1 - \mathsf{F}((-\infty, x))| =$$

$$= \mathsf{F}_n([x, +\infty)) + \mathsf{F}([x, +\infty)) \leq \mathsf{F}_n([b, +\infty)) + \mathsf{F}([b, +\infty)) \leq 3\varepsilon.$$

3. Если $x \in [a, b)$, то

$$|F_n(x) - F(x)| = |F_n((-\infty, x)) - F((-\infty, x))| =$$

$$= |\mathsf{F}_n((-\infty, a)) + \mathsf{F}_n([a, x)) - \mathsf{F}((-\infty, a)) - \mathsf{F}([a, x))| \leq$$
$$\leq \mathsf{F}_n((-\infty, a)) + \mathsf{F}((-\infty, a)) + |\mathsf{F}_n([a, x)) - \mathsf{F}([a, x))| \leq$$
$$\leq 2\varepsilon + \varepsilon + \varepsilon = 4\varepsilon,$$

($|\mathsf{F}_n([a, x)) - \mathsf{F}([a, x))| < \varepsilon$ при достаточно больших n, поскольку $\mathsf{F}_n \to \mathsf{F}$, а $[a, x)$ — промежуток непрерывности распределения F).

Так что для каждой точки непрерывности $x \in \mathbb{R}^1$ функции распределения $F(x)$ при $n \to \infty$ имеем $F_n(x) \to F(x)$.

Пример 12.1.3. *Пусть $\{\mathsf{F}_n\}$ — последовательность распределений с функциями распределения*

$$F_n(x) = \begin{cases} 0, & \text{\textit{если} } x < 0; \\ nx, & \text{\textit{если} } 0 \leq x \leq 1/n; \\ 1, & \text{\textit{если} } x > 1/n. \end{cases}$$

Исследовать последовательность $\{\mathsf{F}_n\}$ на сходимость.

Решение. При $n \to \infty$ последовательность функций $F_n(x)$ сходится в каждой точке (исключая, быть может, точку $x = 0$) к функции распределения

$$F(x) = \begin{cases} 1, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

атомического распределения, сосредоточенного в точке 0 ($x = 0$ является точкой разрыва $F(x)$).

В самом деле, пусть x — произвольная, но фиксированная точка, лежащая слева от нуля. По условию в этой точке $F_n(x) = 0$ при каждом n, поэтому $\lim\limits_n F_n(x) = 0 = F(x)$. Далее, пусть x — произвольная, но фиксированная точка, лежащая справа от нуля. Поскольку $n \to \infty$, то, начиная с некоторого N ($n \geq N$), выполняется неравенство $1/n < x$ и, следовательно, $F_n(x) = 1$. Поэтому $\lim\limits_n F_n(x) = 1 = F(x)$ и при $x > 0$. То есть при $n \to \infty$

$$F_n(x) \to F(x)$$

в каждой точке $x \in \mathbb{R}^1$, исключая, быть может, точку 0. Отсюда в силу достаточного условия сходимости распределений имеем:

$$\mathsf{F}_n \to \mathsf{F}.$$

Так что последовательность распределений $\{\mathsf{F}_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 0.

Стохастическая ограниченность. Последовательность вероятностных распределений $\{\mathsf{F}_n\}$ будем называть *стохастически ограниченной*, если для каждого $\varepsilon > 0$ найдется такое a, что для всех n начиная с некоторого N

$$\mathsf{F}_n([-a, a)) \geq 1 - \varepsilon.$$

Собственная сходимость и стохастическая ограниченность вероятностных распределений тесно связаны между собой. А именно, имеет место теорема.

Теорема. *Если последовательность распределений $\{\mathsf{F}_n\}$ сходится в собственном смысле, то она стохастически ограничена, и если сходящаяся последовательность $\{\mathsf{F}_n\}$ стохастически ограничена, то она сходится в собственном смысле.*

Доказательство. Пусть последовательность вероятностных распределений $\{\mathsf{F}_n\}$ сходится к собственному распределению F. Докажем, что $\{\mathsf{F}_n\}$ стохастически ограничена.

Так как F — собственное вероятностное распределение, то для $\varepsilon > 0$ найдется интервал $[-a, a)$ такой, что

$$\mathsf{F}([-a, a)) \geq 1 - \varepsilon,$$

дополнительно потребуем, чтобы $[-a, a)$ был интервалом непрерывности F. А поскольку $\mathsf{F}_n([-a, a)) \to \mathsf{F}([-a, a))$, то, начиная с некоторого N,

$$\mathsf{F}_n([-a, a)) > 1 - 2\varepsilon,$$

т. е. последовательность $\{\mathsf{F}_n\}$ стохастически ограничена.

Пусть теперь $\mathsf{F}_n \to \mathsf{F}$ и последовательность $\{\mathsf{F}_n\}$ стохастически ограничена. Докажем, что сходимость $\mathsf{F}_n \to \mathsf{F}$ собственная, т. е. F — собственное распределение.

По определению стохастически ограниченной последовательности, для данного $\varepsilon > 0$ найдется такой промежуток $[-a, a)$ (дополнительно выберем его промежутком непрерывности F), что начиная с некоторого N $(n \geq N)$

$$\mathsf{F}_n([-a, a)) \geq 1 - \varepsilon.$$

Переходя к пределу в последнем неравенстве, сначала при $n \to \infty$, а затем при $a \to \infty$ получаем:

$$\mathsf{F}\left(\mathbb{R}^1\right) \geq 1 - \varepsilon.$$

Из последнего следует, что F — собственное распределение.

Теорема 12.1.2 (первое достаточное условие сходимости к атомическому распределению). *Последовательность собственных вероятностных распределений* F_n *с одним и тем же средним* a *и дисперсиями* σ_n^2*, сходящимися к нулю, сходится к собственному вероятностному атомическому распределению, сосредоточенному в точке* a.

Доказательство. Воспользовавшись неравенством Чебышёва, убедимся, что функция распределения $F_n(x)$ при $\sigma_n^2 \to 0$ сходится в основном к функции распределения

$$F_0(x) = \begin{cases} 0, & \text{если } x \le a, \\ 1, & \text{если } x > a \end{cases}$$

собственного атомического распределения F_0, сосредоточенного в точке a (выполняется достаточное условие сходимости распределения F_n к F_0).

Пусть x — произвольная, но фиксированная точка, лежащая левее a. Представив x в виде $x = a - t$ $(t > 0)$ получим

$$F_n(x) - F_0(x) = F_n(a-t) - 0 = F_n(a-t) = \mathsf{F}_n\{y:\ y < a-t\} =$$

$$= \mathsf{F}_n\{y:\ y - a < -t\} \le \mathsf{F}_n\{y:\ |y-a| > t\} \le \frac{\sigma_n^2}{t^2}.$$

Отсюда при $n \to \infty$ следует, что $F_n(x) \to F_0(x)$.

Если x — произвольная, но фиксированная точка, лежащая правее a, то представив x в виде $x = a + t$ $(t > 0)$, получим

$$F_0(x) - F_n(x) = 1 - F_n(x) = 1 - F_n(a+t) = 1 - \mathsf{F}_n\{y:\ y < t+a\} =$$

$$= \mathsf{F}_n\{y:\ y \ge a+t\} = \mathsf{F}_n\{y:\ y-a \ge t\} \le \mathsf{F}_n\{y:\ |y-a| \ge t\} \le \frac{\sigma_n^2}{t^2}.$$

Так что $F_n(x) \to F_0(x)$ при $\sigma_n^2 \to 0$ в каждой точке непрерывности $F_0(x)$.

Тем самым теорема доказана.

Теорема 12.1.3 (второе достаточное условие сходимости к атомическому распределению). *Последовательность* $\{\mathsf{F}_n\}$ *собственных вероятностных распределений соответственно со средними* a_n*, сходящимися к* a*, и дисперсиями* σ_n^2*, сходящимися к* 0*, сходится к собственному вероятностному атомическому распределению, сосредоточенному в точке* a.

Доказательство. Достаточно доказать, что последовательность $\{F_n(x)\}$ функций распределения сходится к функции распределения

$$F_0(x) = \begin{cases} 0, & \text{если } x \le a; \\ 1, & \text{если } x > a \end{cases}$$

атомического распределения, сосредоточенного в точке a, во всех точках $x \neq a$.

Пусть $t > 0$, $x = a - 2t$. Поскольку $a_n \to a$, при $n \to \infty$, то a_n при достаточно больших n попадает в окрестность $(a - t, +\infty)$ точки a, т. е.

$$a - t < a_n,$$

и, следовательно,

$$a - 2t < a_n - t.$$

Учитывая последнее неравенство, имеем:

$$F_n(x) = F_n(a - 2t) = \mathsf{F}_n((-\infty, a - 2t)) \le \mathsf{F}_n((-\infty, a_n - t)) =$$

$$= \mathsf{F}_n\{y : y < a_n - t\} \le \mathsf{F}_n\{y : |y - a_n| \ge t\} \le \frac{\sigma_n^2}{t^2}$$

(воспользовались неравенством Чебышева). Поэтому

$$F_n(a - 2t) \to 0$$

при $\sigma_n^2 \to 0$.

Аналогично убеждаемся, что $F_n(a + 2t) \to 1$ при $n \to \infty$.

Пример 12.1.4. *Исследовать на сходимость последовательность распределений $\{\mathsf{Q}_n\}$ с плотностями*

$$q_n(x) = \frac{n}{\sqrt{2\pi}} \exp\left\{-\frac{(x - (-1)^n)^2 n^2}{2}\right\}, \ n = 1, 2, \ldots,$$

соответственно.

Решение. В силу достаточного условия сходимости к атомическому распределению (см. теорему 12.1.2) для четных n последовательность распределений $\{\mathsf{Q}_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 1, а при нечетных n — к собственному атомическому распределению, сосредоточенному в точке -1. Поэтому последовательность $\{\mathsf{Q}_n\}$ не является сходящейся.

12.2 Теорема Хелли о выборе

Стандартный метод доказательства сходимости числовой последовательности состоит в следующем: сначала устанавливается существование у последовательности хотя бы одной предельной точки, затем ее единственность. Сходный прием применим и к последовательностям распределений. Аналогом теоремы о существовании предельной точки последовательности является теорема Хелли о выборе. Теорема Хелли верна в пространстве $\mathbb{R}^s$ любого конечного числа измерений. Мы докажем ее для одномерного случая.

Теорема 12.2.1 (Хелли). *Из любой последовательности вероятностных распределений на* $\mathbb{R}^1$ *можно выбрать сходящуюся в собственном или несобственном смысле подпоследовательность.*

Пусть $\{\mathsf{F}_n\}$ — последовательность вероятностных распределений. Для того, чтобы установить, что из $\{\mathsf{F}_n\}$ можно выбрать сходящуюся подпоследовательность $\{\mathsf{F}_{n_k}\}$, достаточно убедиться, что из последовательности функций распределения $\{F_n(x)\}$ можно выбрать подпоследовательность $\{F_{n_k}(x)\}$, сходящуюся в основном к некоторой вероятностной функции распределения $F(x)$ (собственной или несобственной).

Доказательство использует следующую лемму.

Лемма. *Из любой последовательности функций* $\{u_n\}$ *(со значениями в* $\mathbb{R}^1$*), заданной на счетном подмножестве* $\{a_i\}$ *прямой* $\mathbb{R}^1$*, можно выбрать подпоследовательность, сходящуюся в каждой точке* $a_j \in \{a_i\}$ *(быть может, к* $\pm\infty$*).*

Для доказательства леммы воспользуемся “диагональным методом” Кантора.

Из последовательности функций $\{u_n\}$ можно выбрать подпоследовательность $\{u_{n_k}\}$, сходящуюся в точке a_1 (быть может, к $\pm\infty$). Чтобы избежать сложных индексов, обозначим $u_{n_k} = u_k^{(1)}$, так что $\left\{u_k^{(1)}\right\}$ — подпоследовательность последовательности $\{u_n\}$, сходящаяся в точке a_1. Из последовательности функций $\left\{u_k^{(1)}\right\}$ выберем подпоследовательность $\left\{u_k^{(2)}\right\}$, сходящуюся в точке a_2, как подпоследовательность последовательности $\left\{u_k^{(1)}\right\}$ она сходится и в точке a_1. Продолжая по индукции, построим для каждого s $(s = 1, 2, \ldots)$ последовательность функций $\left\{u_k^{(s)}\right\}$, сходящуюся в точке a_s и являющуюся подпоследовательностью последовательности $\left\{u_k^{(s-1)}\right\}$, а следовательно, сходящуюся и в точках $a_1, a_2, \ldots, a_{s-1}$.

Рассмотрим теперь "диагональную" последовательность функций: $u_1^{(1)}, u_2^{(2)}, \ldots$ Эта последовательность сходится в каждой точке a_j множества $\{a_i\}$. Действительно, пусть a_s — произвольная, но фиксированная точка из $\{a_i\}$. По построению последовательность $u_k^{(s)}$, $k = 1, 2, \ldots$, сходится в точке a_s. Последовательность же $u_1^{(1)}, u_2^{(2)}, \ldots, u_s^{(s)}, \ldots, u_l^{(l)}, \ldots$, начиная с номера s, является подпоследовательностью последовательности $u_k^{(s)}$, $k = 1, 2, \ldots$ (сходящейся в точке a_s), и поэтому она сходится в точке a_s. А поскольку точка a_s выбрана произвольно, то лемма доказана.

Доказательство теоремы о выборе.

Пусть $\{a_i\}$ — счетная всюду плотная в $\mathbb{R}^1$ последовательность точек. В силу леммы из последовательности функций $\{F_n(y)\}$ можно выбрать подпоследовательность $\{F_{n_k}(y)\}$, сходящуюся в каждой точке $a_j \in \{a_i\}$. Определим на множестве $\{a_i\}$ функцию $F(y)$ равенством:

$$F(a_j) = \lim_{n_k} F_{n_k}(a_j), \ a_j \in \{a_i\}.$$

Функция $F(y)$, заданная на $\{a_i\}$, очевидно, является монотонно неубывающей и $0 \le F(a_i) \le 1$. Доопределим $F(y)$ на $\mathbb{R}^1 \setminus \{a_i\}$, полагая в каждой точке $y \in \mathbb{R}^1 \setminus \{a_i\}$

$$F(y) = \sup_{a_k\colon\ a_k < y} F(a_k). \tag{12.2.1}$$

Так определенная на $\mathbb{R}^1$ функция $F(y)$ принимает значения из $[0; 1]$ и является монотонно неубывающей.

Далее докажем, что подпоследовательность $\{F_{n_k}(y)\}$, $y \in \mathbb{R}^1$, последовательности $\{F_n(y)\}$ сходится к $F(y)$ не только в точках множества $\{a_i\}$ (на $\{a_i\}$ она сходится по построению), но и в каждой точке непрерывности x функции $F(y)$.

Пусть $\varepsilon > 0$, x — точка непрерывности функции $F(y)$, не принадлежащая множеству $\{a_i\}$. В силу плотности множества $\{a_i\}$ в $\mathbb{R}^1$ и непрерывности $F(y)$ в точке x найдется пара точек a_i, a_j из $\{a_k\}$ таких, что $a_i < x < a_j$ и

$$|F(a_j) - F(a_i)| \le |F(a_j) - F(x)| + |F(x) - F(a_i)| \le \varepsilon. \tag{12.2.2}$$

В силу монотонности функций $F(y)$ и $F_{n_k}(y)$ имеют место неравенства
$$F(a_i) \le F(x) \le F(a_j), \tag{12.2.3}$$
$$F_{n_k}(a_i) \le F_{n_k}(x) \le F_{n_k}(a_j),$$
а вместе с последним неравенством, и неравенство,
$$F(a_i) \le \underline{\lim}\, F_{n_k}(x) \le \overline{\lim}\, F_{n_k}(x) \le F(a_j), \tag{12.2.4}$$
полученное из него предельным переходом. Из неравенств (12.2.3) и (12.2.4) следует, что точки $F(x)$, $\underline{\lim} F_{n_k}(x)$, $\overline{\lim} F_{n_k}(x)$ принадлежат отрезку $[F(a_i), F(a_j)]$, длина которого не превосходит ε (см. неравенство (12.2.2))). И поскольку $\varepsilon > 0$ произвольно, то все три точки $F(x)$, $\underline{\lim}\, F_{n_k}(x)$, $\overline{\lim}\, F_{n_k}(x)$ совпадают:
$$\underline{\lim}\, F_{n_k}(x) = \overline{\lim}\, F_{n_k}(x) = F(x).$$
Последнее обозначает, что $F_{n_k}(x)$ сходится к $F(x)$ (в каждой точке непрерывности x функции $F(y)$).

На этом доказательство теоремы Хелли заканчивалось бы, если бы $F(y)$ была функцией распределения, чего мы утверждать, вообще говоря, не можем — функция распределения должна быть непрерывной слева, а $F(y)$ таковой, вообще говоря, не является. “Подправим” функцию $F(y)$ в точках разрыва так, чтобы она стала непрерывной слева. Точнее, определим на $\mathbb{R}^1$ функцию $\widetilde{F}(y)$, которая совпадает с $F(y)$ в точках непрерывности $F(y)$, а в точках разрыва функции $F(y)$ значения $\widetilde{F}(y)$ определим так, чтобы $\widetilde{F}(y)$ была непрерывной слева. А именно, если в точке z функция $F(y)$ терпит разрыв, то положим
$$\widetilde{F}(z) = \lim_{y\uparrow z} F(y).$$

Функция $\widetilde{F}(y)$ монотонно неубывающая, непрерывная слева и $0 \le \widetilde{F}(y) \le 1$, т. е. является функцией распределения. Причем поскольку в точках непрерывности функции $F(y)$ выполняется равенство $\widetilde{F}(y) = F(y)$, то в этих точках
$$\lim_{n_k\to\infty} F_{n_k}(y) = F(y) = \widetilde{F}(y).$$
Из последнего равенства в силу достаточного условия сходимости распределений получаем, что $F_{n_k} \to \widetilde{F}$ при $n_k \to \infty$.

12.3 Слабая сходимость распределений

Лемма 12.3.1 (о задании распределения значениями интегралов). *Распределение на $\mathbb{R}^1$ задается значениями интегралов (по этому распределению) на классе $\{\varphi\}$ непрерывных ограниченных функций φ, обращающихся в нуль вне ограниченных промежутков.*

Если значения интегралов по распределениям F и G на классе $\{\varphi\}$ непрерывных ограниченных функций φ, обращающихся в нуль вне ограниченных промежутков, совпадают:

$$\int\limits_{\mathbb{R}^1} \varphi(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} \varphi(x)\mathsf{G}(dx), \tag{12.3.1}$$

то совпадают и сами распределения:

$$\mathsf{F} = \mathsf{G}.$$

Доказательство. Пусть для распределения F, заданного на $\mathbb{R}^1$ ($\mathsf{F}(\mathbb{R}^1) < \infty$), известны значения интегралов

$$I(\varphi) = \int\limits_{\mathbb{R}^1} \varphi(x)\mathsf{F}(dx)$$

на классе $\{\varphi\}$ непрерывных ограниченных функций, обращающихся в нуль вне ограниченных промежутков.

Распределение на $\mathbb{R}^1$ однозначно задается своими значениями на промежутках вида $[a, b)$, $a < b$, $a, b \in \mathbb{R}^1$ (см. следствие 2 из теоремы 7.2.1). Поэтому достаточно установить, что значения $\mathsf{F}([a, b))$ распределения F на промежутках $[a, b)$ задаются значениями $I(\varphi)$.

Значение

$$\mathsf{F}([a, b)) = \int\limits_{\mathbb{R}^1} I_{[a,b)}(x)\mathsf{F}(dx).$$

Представим индикатор $I_{[a,b)}(x)$ промежутка $[a, b)$ в виде предела последовательности $\{\varphi_n(x)\}$ непрерывных ограниченных функций, обращающихся в нуль вне конечных промежутков:

$$I_{[a,b)}(x) = \lim_{n} \varphi_n(x).$$

В качестве $\varphi_n(x)$ можно рассмотреть, например, функции вида:

$$\varphi_n(x) = \begin{cases} 0, & \text{при } x \in (-\infty, a-1/n), \\ n(x-(a-1/n)), & \text{при } x \in [a-1/n, a), \\ 1, & \text{при } x \in [a, b-1/n), \\ -n(x-b), & \text{при } x \in [b-1/n, b), \\ 0, & \text{при } x \in [b, +\infty). \end{cases}$$

Тогда

$$\mathsf{F}([a,b)) = \int\limits_{\mathbb{R}^1} I_{[a,b)}(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} \big(\lim_n \varphi_n(x)\big)\mathsf{F}(dx)$$

или, воспользовавшись теоремой Лебега о мажорируемой сходимости,

$$\mathsf{F}([a,b)) = \lim_n \int\limits_{\mathbb{R}^1} \varphi_n(x)\mathsf{F}(dx) = \lim_n I(\varphi_n).$$

Последнее равенство и обозначает, что распределение F однозначно определяется значениями интегралов по этому распределению на классе $\{\varphi\}$ непрерывных ограниченных функций φ, обращающихся в нуль вне ограниченных промежутков.

Далее, если для распределений F и G имеет место (12.3.1), то оно имеет место, в частности, если $\varphi(x) = \varphi_n(x)$:

$$\int\limits_{\mathbb{R}^1} \varphi_n(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} \varphi_n(x)\mathsf{G}(dx).$$

Переходя в этом равенстве к пределу в правой и левой частях и пользуясь теоремой Лебега о мажорируемой сходимости, получаем:

$$\int\limits_{\mathbb{R}^1} I_{[a,b)}(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} I_{[a,b)}(x)\mathsf{G}(dx)$$

или

$$\mathsf{F}([a,b)) = \mathsf{G}([a,b)).$$

Из совпадения F и G на промежутках $[a,b)$ в силу следствия 2 из теоремы Каратеодори (теорема 7.2.1) следует, что

$$\mathsf{F} = \mathsf{G}.$$

Обозначения. Мы будем различать на $\mathbb{R}^1$ три класса функций: класс непрерывных ограниченных функций (обозначение $C(-\infty,+\infty)$); класс непрерывных ограниченных функций u, имеющих конечные пределы $u(-\infty)$ и $u(+\infty)$ соответственно на $-\infty$ и на $+\infty$ (обозначение $C[-\infty,+\infty]$); класс непрерывных ограниченных функций u, имеющих пределы $u(-\infty)=0$ и $u(+\infty)=0$ на $-\infty$ и на $+\infty$ (обращающихся в нуль на бесконечности), обозначение $C_0[-\infty,+\infty]$. Очевидно,

$$C_0[-\infty,+\infty] \subset C[-\infty,+\infty] \subset C(-\infty,+\infty).$$

Например, функция $\sin x \in C(-\infty,+\infty)$, но не принадлежит классу $C[-\infty,+\infty]$; функция $1+e^{-|x|}\sin x \in C[-\infty,+\infty]$, но $1+e^{-|x|}\sin x \notin C_0[-\infty,+\infty]$; $e^{-|x|}\sin x \in C_0[-\infty,+\infty]$.

Определение. Пусть U — класс борелевских функций на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$. Будем говорить, что последовательность распределений $\{\mathsf{F}_n\}$ *слабо сходится* к распределению F относительно класса U, если для каждой $u \in U$ при $n \to \infty$

$$\int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_n(dx) \to \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}(dx).$$

Мы докажем, что введенная нами ранее сходимость распределений эквивалентна слабой сходимости относительно класса $C_0[-\infty,+\infty]$. Понятие слабой сходимости имеет то преимущество, что применимо к произвольным вероятностным пространствам, но наше определение сходимости больше соответствует интуиции.

Далее будем использовать обозначения

$$\mathsf{M}_n u = \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_n(dx),\ \ \mathsf{M}u = \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}(dx).$$

Теорема 12.3.1 (о сходимости и слабой сходимости распределений). *Сходимость последовательности $\{\mathsf{F}_n\}$ вероятностных распределений к распределению F (собственная или несобственная) эквивалентна слабой сходимости $\{\mathsf{F}_n\}$ к F относительно класса $C_0[-\infty,+\infty]$.*

Собственная сходимость последовательности $\{\mathsf{F}_n\}$ вероятностных распределений к распределению F влечет слабую сходимость $\{\mathsf{F}_n\}$ к F относительно класса $C(-\infty,+\infty)$.

Доказательство. Сначала докажем, что из сходимости $\mathsf{F}_n \to \mathsf{F}$ (собственной или несобственной) следует слабая сходимость F_n к F относительно класса $C_0[-\infty, +\infty]$.

Пусть $\mathsf{F}_n \to \mathsf{F}$ (собственно или несобственно) и пусть $u \in C_0[-\infty, +\infty]$. Покажем, что при $n \to \infty$

$$\mathsf{M}_n u = \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_n(dx) \to \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}(dx) = \mathsf{M}u.$$

Поскольку $u \in C_0[-\infty, +\infty]$, то

$$\lim_{x\to+\infty} u(x) = 0, \ \lim_{x\to-\infty} u(x) = 0,$$

и для данного $\varepsilon > 0$ найдется конечный промежуток J такой, что при всех $x \notin J$ значение $|u(x)| \le \varepsilon$ (дополнительно выберем J интервалом непрерывности F).

Для функции $u(x)$ построим на $\mathbb{R}^1$ кусочно-постоянную функцию $\varphi(x)$, принимающую конечное число значений и отличающуюся от $u(x)$ не более чем на ε. Это можно сделать, например, следующим образом. Из непрерывности $u(x)$ на $\mathbb{R}^1$ следует ее равномерная непрерывность на каждом конечном промежутке, в частности и на J. Поэтому J можно разбить на конечное число r непересекающихся интервалов J_k:

$$J = \bigcup_{k=1}^{r} J_k, \ J_i \cap J_j = \emptyset, \ i \ne j$$

(дополнительно выберем их интервалами непрерывности распределения F) так, чтобы на каждом из них

$$\sup_{(x', x''):x', x'' \in J_k} |u(x') - u(x'')| \le \varepsilon.$$

Из последнего неравенства следует, что для каждого J_k можно указать такое число φ_k, что

$$|u(x) - \varphi_k| \le \varepsilon, \ x \in J_k, \ k = 1, 2, \ldots, r.$$

Определим теперь на $\mathbb{R}^1$ функцию

$$\varphi(x) = \sum_{k=1}^{r} \varphi_k I_{J_k}(x). \tag{12.3.2}$$

Функция $\varphi(x)$ кусочно-постоянна — равна φ_k на $J_k \subset J$, $k = 1, 2, \ldots, r$, и нулю вне J — и по построению на промежутке J уклоняется от $u(x)$ не больше, чем на ε:

$$|u(x) - \varphi(x)| \leq \varepsilon.$$

А поскольку при $x \notin J$ значение $|u(x)| \leq \varepsilon$, то и при всех $x \in \mathbb{R}^1$

$$|u(x) - \varphi(x)| \leq \varepsilon,$$

кратко

$$|u - \varphi| \leq \varepsilon.$$

Используя функцию $\varphi(x)$, оценим

$$|\mathsf{M}_n u - \mathsf{M} u|.$$

Для этого оценим каждое слагаемое в правой части неравенства

$$|\mathsf{M}_n u - \mathsf{M} u| \leq |\mathsf{M}_n u - \mathsf{M}_n \varphi| + |\mathsf{M} u - \mathsf{M} \varphi| + |\mathsf{M}_n \varphi - \mathsf{M} \varphi|.$$

Поскольку $|u - \varphi| \leq \varepsilon$, то

$$|\mathsf{M}_n u - \mathsf{M}_n \varphi| \leq \mathsf{M}_n |u - \varphi| \leq \varepsilon,$$

$$|\mathsf{M} u - \mathsf{M} \varphi| \leq \mathsf{M} |u - \varphi| \leq \varepsilon.$$

Слагаемое $|\mathsf{M}_n \varphi - \mathsf{M} \varphi|$ оценим так:

$$|\mathsf{M}_n \varphi - \mathsf{M} \varphi| =$$

$$= \left| \int\limits_{\mathbb{R}^1} \left(\sum_{k=1}^{r} \varphi_k I_{J_k}(x) \right) \mathsf{F}_n(dx) - \int\limits_{\mathbb{R}^1} \left(\sum_{k=1}^{r} \varphi_k I_{J_k}(x) \right) \mathsf{F}(dx) \right| =$$

$$= \left| \sum_{k=1}^{r} \varphi_k \mathsf{F}_n(J_k) - \sum_{k=1}^{r} \varphi_k \mathsf{F}(J_k) \right| \leq \sum_{k=1}^{r} |\varphi_k| |\mathsf{F}_n(J_k) - \mathsf{F}(J_k)|.$$

Поскольку $\mathsf{F}_n \to \mathsf{F}$, то $\mathsf{F}_n(J_k) \to \mathsf{F}(J_k)$ для каждого $k = 1, 2, \ldots$ $\ldots, r$, и поэтому при достаточно больших n

$$\sum_{k=1}^{r} |\varphi_k| |\mathsf{F}_n(J_k) - \mathsf{F}(J_k)| \leq \varepsilon,$$

а следовательно, и
$$|\mathsf{M}_n\varphi - \mathsf{M}\varphi| \le \varepsilon. \tag{12.3.3}$$

Так что для функции $u \in C_0[-\infty, +\infty]$, начиная с некоторого N $(n \ge N)$,
$$|\mathsf{M}_n u - \mathsf{M}u| \le 3\varepsilon,$$
т. е. при $n \to \infty$
$$\mathsf{M}_n u \to \mathsf{M}u.$$

Пусть теперь $\mathsf{F}_n \to \mathsf{F}$ в собственном смысле (предельное распределение F — собственное) и пусть $u \in C(-\infty, +\infty)$ (не ограничивая общности будем считать, что $|u| \le 1$). Докажем, что при $n \to \infty$
$$\mathsf{M}_n u \to \mathsf{M}u.$$

Так как F — собственное вероятностное распределение ($\mathsf{F}(\mathbb{R}^1) = 1$), то для данного $\varepsilon > 0$ в силу непрерывности вероятности можно выбрать ограниченный интервал J такой, что $\mathsf{F}(J) \ge 1 - \varepsilon$, а следовательно,
$$\mathsf{F}(\overline{J}) \le \varepsilon \tag{12.3.4}$$
(дополнительно выберем J интервалом непрерывности распределения F). При этом поскольку $\{\mathsf{F}_n\}$ сходится к F, то, начиная с некоторого N $(n \ge N)$, значения $\mathsf{F}_n(J) \ge 1-2\varepsilon$, а следовательно,
$$\mathsf{F}_n(\overline{J}) \le 2\varepsilon. \tag{12.3.5}$$

Далее, для функции $u(x)$ строим на $\mathbb{R}^1$ кусочно постоянную функцию $\varphi(x)$, отличающуюся от $u(x)$ на промежутке J не более чем на ε (как это делали ранее, см. (12.3.2)).

Используя $\varphi(x)$, оценим
$$|\mathsf{M}_n u - \mathsf{M}u|.$$
Для этого оценим каждое слагаемое в правой части неравенства
$$|\mathsf{M}_n u - \mathsf{M}u| \le |\mathsf{M}_n u - \mathsf{M}_n\varphi| + |\mathsf{M}u - \mathsf{M}\varphi| + |\mathsf{M}_n\varphi - \mathsf{M}\varphi|.$$
Для первого слагаемого имеем:
$$|\mathsf{M}_n u - \mathsf{M}_n\varphi| = \left| \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_n(dx) - \int\limits_{\mathbb{R}^1} \varphi(x)\mathsf{F}_n(dx) \right| \le$$

$$\leq \int_{\mathbb{R}^1} |u(x) - \varphi(x)|\, \mathsf{F}_n(dx) =$$

$$= \int_J |u(x) - \varphi(x)|\, \mathsf{F}_n(dx) + \int_{\overline{J}} |u(x) - \varphi(x)|\, \mathsf{F}_n(dx) \leq$$

$$\leq \int_J \varepsilon \mathsf{F}_n(dx) + \int_{\overline{J}} 1 \mathsf{F}_n(dx) \leq \varepsilon + \mathsf{F}_n(\overline{J}) \leq \varepsilon + 2\varepsilon = 3\varepsilon,$$

$\mathsf{F}_n(\overline{J}) \leq 2\varepsilon$, см. (12.3.5). Аналогично

$$|\mathsf{M}u - \mathsf{M}\varphi| \leq \varepsilon + \mathsf{F}(\overline{J}) \leq 2\varepsilon,$$

$\mathsf{F}(\overline{J}) \leq \varepsilon$, см. (12.3.4). Слагаемое $|\mathsf{M}_n\varphi - \mathsf{M}\varphi|$ при достаточно больших n не превосходит ε, см. (12.3.3). Поэтому, начиная с некоторого N $(n \geq N)$,

$$|\mathsf{M}_n u - \mathsf{M}u| \leq 6\varepsilon,$$

т. е. при $n \to \infty$

$$\mathsf{M}_n u \to \mathsf{M}u.$$

Так что из собственной сходимости $\mathsf{F}_n \to \mathsf{F}$ следует слабая сходимость последовательности $\{\mathsf{F}_n\}$ к F относительно класса непрерывных ограниченных функций.

Осталось установить, что из слабой сходимости $\mathsf{F}_n \to \mathsf{F}$ относительно класса $C_0[-\infty, +\infty]$ следует сходимость $\mathsf{F}_n \to \mathsf{F}$ (собственная или несобственная).

Пусть для каждой функции u из класса $C_0[-\infty, +\infty]$

$$\mathsf{M}_n u = \int_{\mathbb{R}^1} u(x) \mathsf{F}_n(dx) \to \int_{\mathbb{R}^1} u(x) \mathsf{F}(dx) = \mathsf{M}u. \qquad (12.3.6)$$

Докажем, что тогда последовательность $\{\mathsf{F}_n\}$ сходится к F (собственно или несобственно).

В силу теоремы Хелли о выборе из последовательности $\{\mathsf{F}_n\}$ можно выбрать подпоследовательность $\{\mathsf{F}_{n_k}\}$, сходящуюся к некоторому распределению G (в собственном или несобственном смысле). Покажем, что, во-первых, $\mathsf{G} = \mathsf{F}$, а, во-вторых, и сама последовательность $\{\mathsf{F}_n\}$ сходится к распределению F.

Из сходимости $\mathsf{F}_{n_k} \to \mathsf{G}$ следует, что для каждой функции $u \in C_0[-\infty, +\infty]$

$$\lim_{n_k} \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_{n_k}(dx) = \int\limits_{\mathbb{R}^1} u(x)\mathsf{G}(dx) \qquad (12.3.7)$$

(см. доказательство первой части теоремы). С другой стороны, последовательность $\int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_{n_k}(dx)$, будучи подпоследовательностью последовательности $\int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_n(dx)$, сходящейся к $\int\limits_{\mathbb{R}^1} u(x)\mathsf{F}(dx)$ (см. (12.3.6)), сходится к $\int\limits_{\mathbb{R}^1} u(x)\mathsf{F}(dx)$:

$$\lim_{n_k} \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}_{n_k}(dx) = \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}(dx).$$

Отсюда и из равенства (12.3.7) имеем

$$\int\limits_{\mathbb{R}^1} u(x)\mathsf{G}(dx) = \int\limits_{\mathbb{R}^1} u(x)\mathsf{F}(dx)$$

для каждой функции $u \in C_0[-\infty, +\infty]$. Поэтому в силу леммы 12.3.1 о задании распределения значениями интегралов

$$\mathsf{G} = \mathsf{F}.$$

Так что если последовательность $\{\mathsf{F}_n\}$ слабо сходится к F относительно класса $C_0[-\infty, \infty]$, то каждая ее сходящаяся подпоследовательность $\{\mathsf{F}_{n_k}\}$ сходится к F. Сама последовательность F_n также сходится к F — если бы $\{\mathsf{F}_n\}$ не сходилась к F, то в силу теоремы о выборе Хелли нашлась бы подпоследовательность $\{\mathsf{F}_{n'_s}\}$ последовательности $\{\mathsf{F}_n\}$ сходящаяся к некоторому распределению $\mathsf{G}' \neq \mathsf{F}$, что невозможно, т. к. каждая сходящаяся подпоследовательность последовательности $\{\mathsf{F}_n\}$ сходится к F.

Тем самым теорема полностью доказана.

Пусть последовательность $\{\mathsf{F}_n\}$ вероятностных распределений сходится в собственном смысле к распределению F и $u_t = u_t(x)$ — семейство функций из класса $C(-\infty, +\infty)$, зависящих от параметра $t \in T$. Для каждого фиксированного $t \in T$ при $n \to \infty$

$$\mathsf{M}_n u_t \to \mathsf{M} u_t.$$

Далее приводятся условия, при которых сходимость последовательности функций $\mathsf{M}_n u_t$ к функции $\mathsf{M} u_t$ равномерная относительно $t \in T$.

Напомним, что семейство функций $u_t = u_t(x)$, зависящее от параметра $t \in T$, равномерно ограничено, если существует такое число M, что

$$|u_t| \leq M$$

для всех $t \in T$.

Семейство функций $u_t = u_t(x)$, $t \in T$, называется равностепенно непрерывным, если для каждого $\varepsilon > 0$ найдется $\delta > 0$ (общее для всех t) такое, что из $|x_2 - x_1| \leq \delta$ следует

$$|u_t(x_2) - u_t(x_1)| \leq \varepsilon$$

для всех $t \in T$.

Следствие (критерий равномерной сходимости $\mathsf{M}_n u_t$). *Пусть $\mathsf{F}_n \to \mathsf{F}$ в собственном смысле. Если семейство функций $u_t = u_t(x)$, $t \in T$, из класса $C(-\infty, +\infty)$, зависящее от параметра t, является равностепенно непрерывным и равномерно ограниченным, то сходимость*

$$\mathsf{M}_n u_t \to \mathsf{M} u_t$$

при $n \to \infty$ равномерная по $t \in T$.

Доказательство. В доказательстве теоремы было использовано разбиение интервала J на интервалы $J_1, J_2, \ldots, J_r$, на каждом из которых колебание

$$\sup_{(x',\, x''): x',\, x'' \in J_k} |u(x') - u(x'')|$$

функции u меньше ε. В силу равностепенной непрерывности семейства u_t это разбиение может быть выбрано одним и тем же для для всех функций семейства u_t, $t \in T$, поэтому, повторяя рассуждения теоремы, получаем:

$$|\mathsf{M}_n u_t - \mathsf{M} u_t| \leq 6\varepsilon$$

при достаточно больших n для всех t.

Тем самым следствие доказано.

12.4 Примеры и задачи

Примеры

Пример 12.4.1 . *Пусть* F_h *— семейство вероятностных распределений, заданных плотностями*

$$f_h(x) = \frac{1}{\sqrt{2\pi}h} \exp\left\{-\frac{(x-a)^2}{2h^2}\right\}, \ h > 0.$$

Исследовать F_h *на сходимость* 1° *при* $h \to \infty$*;* 2° *при* $h \to 0$*.*

Решение. 1° Заметим, что функция $F(x) = c$, $x \in \mathbb{R}^1$, где c — константа ($0 \le c \le 1$) является функцией распределения несобственного вероятностного распределения, тождественно равного нулю. В самом деле, значение $\mathsf{F}([a,b))$ распределения F на промежутках $[a,b)$ равно $F(b) - F(a) = c - c = 0$, а следовательно, $\mathsf{F} = 0$ и на множествах из алгебры $\mathfrak{A}$ конечных объединений непересекающихся промежутков вида $[a,b)$, а в силу теоремы Каратеодори и на борелевских множествах.

Далее,

$$F_h(x) = \int\limits_{-\infty}^{x} f_h(t)dt = \frac{1}{\sqrt{2\pi}h} \int\limits_{-\infty}^{x} \exp\left\{-\frac{(t-a)^2}{2h^2}\right\} dt =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{(x-a)/h} \exp\left\{-\frac{u^2}{2}\right\} du$$

(воспользовались заменой $(t-a)/h = u$). При каждом фиксированном x

$$\lim_{h\to\infty} F_h(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{0} \exp\left\{-\frac{u^2}{2}\right\} du = \frac{1}{2}.$$

Поэтому F_h при $h \to \infty$ сходится к несобственному распределению, тождественно равному нулю.

2° При $h \to 0$ семейство распределений F_h сходится к собственному атомическому распределению, сосредоточенному в точке a (в силу теоремы 12.1.2 о сходимости к атомическому распределению).

Задачи

12.1. Пусть $\{\mathsf{F}_n\}$ — последовательность распределений заданных своими функциями распределений

$$F_n(x) = \begin{cases} 0, & \text{если } x \le -1/n; \\ n(x+1/n)/2, & \text{если } -1/n < x \le 1/n; \\ 1, & \text{если } x > 1/n. \end{cases}$$

Исследовать последовательность распределений $\{\mathsf{F}_n\}$ на сходимость.

12.2. Исследовать на сходимость при $n \to \infty$ последовательность распределений F_n с плотностями

$$f_n(x) = \begin{cases} n/2, & \text{если } x \in [-1/n; 1/n]; \\ 0, & \text{если } x \notin [-1/n; 1/n]. \end{cases}$$

Ответ: F_n сходится к собственному атомическому распределению, сосредоточенному в точке 0.

12.3. Пусть $F(x)$ — функция распределения непрерывного собственного вероятностного распределения.

Исследовать на сходимость каждую из перечисленных далее последовательностей распределений, заданных своими функциями распределения:

1) $F_n(x) = F(x + 1/n),\ n = 1, 2, \ldots;$
2) $G_n(x) = F(x + n),\ n = 1, 2, \ldots;$
3) $S_n(x) = F(x - n),\ n = 1, 2, \ldots;$
4) $P_n(x) = F(x/n),\ n = 1, 2, \ldots;$
5) $Q_n(x) = F(x + (-1)^n n),\ n = 1, 2, \ldots$

Ответ: F_n сходится в собственном смысле к F; распределения G_n, S_n, P_n, Q_n сходятся к несобственному распределению, тождественно равному нулю.

12.4. Обозначим через Q_n собственное атомическое распределение, сосредоточенное в точке a_n. Исследовать на сходимость последовательность распределений $\{\mathsf{Q}_n\}$ 1° при $a_n \to +\infty$; 2° при $a_n \to -\infty$; 3° при $a_n \to a$.

12.5. Пусть $\{\mathsf{F}_n\}$ — последовательность распределений с плотностями

$$f_n(x) = \begin{cases} n/2, & \text{если } x \in [(-1)^n - 1/n; (-1)^n + 1/n]; \\ 0, & \text{если } x \notin [(-1)^n - 1/n; (-1)^n + 1/n], \end{cases}$$

соответственно.

Исследовать на сходимость $\{\mathsf{F}_n\}$ при $n \to \infty$.

Ответ: последовательность распределений F_n не является сходящейся.

12.6. Пусть $\{\mathsf{F}_n\}$ — последовательность распределений заданных своими функциями распределения:

$$F_n(x) = \begin{cases} 0, & \text{если } x \le -1/n; \\ G(x), & \text{если } -1/n < x \le 1/n; \\ 1, & \text{если } x > 1/n, \end{cases}$$

$G(x)$ — монотонно неубывающая непрерывная слева функция со значениями в $[0,1]$, заданная на $[-1,1]$.

Исследовать последовательность распределений $\{\mathsf{F}_n\}$ на сходимость.

12.7. Пусть D_n — атомическое распределение:

$$\mathsf{D}_n(k) = 1/n,\ k = 0, 1, 2, \ldots, n-1.$$

Исследовать последовательность распределений D_n на сходимость при $n \to \infty$.

12.8. Исследовать на сходимость последовательность распределений:

1) $\mathsf{F}_n : \begin{pmatrix} -n & n \\ 1/2 & 1/2 \end{pmatrix},\ n = 1, 2, \ldots;$

2) $\mathsf{G}_n : \begin{pmatrix} -1/n & 1/n \\ 1/2 & 1/2 \end{pmatrix},\ n = 1, 2, \ldots;$

3) $\mathsf{F}_n : \begin{pmatrix} n & n^2 \\ 1/2 & 1/2 \end{pmatrix},\ n = 1, 2, \ldots;$

4) $\mathsf{S}_n : \begin{pmatrix} -n^{(-1)^n} & n^{(-1)^n} \\ 1/2 & 1/2 \end{pmatrix},\ n = 1, 2, \ldots$

Ответы: 1) последовательность распределений $\{\mathsf{F}_n\}$ сходится к несобственному распределению, тождественно равному нулю; 2) последовательность распределений $\{\mathsf{G}_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 0.

12.9. Пусть

$$N_{x;\sigma^2}(y) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{y} \exp\left\{-\frac{(t-x)^2}{2\sigma^2}\right\} dt;\ x, y \in \mathbb{R}^1, \sigma > 0.$$

Вычислить $\lim\limits_{\sigma \to 0} N_{x;\sigma^2}(y)$.

Ответ: $I_{(-\infty;y)}(x)$ для каждой пары $(x, y), x \ne y$.

12.10*. Пусть F_λ — распределение, заданное функцией распределения

$$F_\lambda(y) = \sum_{k:0\le k<y} \frac{\lambda^k}{k!} e^{-\lambda},\ y \in \mathbb{R}^1,\ \lambda > 0,$$

k — целые неотрицательные.

Исследовать F_λ на сходимость при $\lambda \to 0$.

Р е ш е н и е. F_λ — пуассоновское распределение с параметром λ. Математическое ожидание и дисперсия пуассоновского распределения с параметром λ равны λ. При $\lambda \to 0$, пуассоновское распределение сходится к собственному атомическому распределению, сосредоточенному в точке 0.

12.11. Пусть $\mathsf{N}_{0,ct}$ $(t > 0)$ — семейство нормальных распределений со средним 0 и дисперсией ct, $t > 0$, $c > 0$, c — константа.

Семейство распределений Q_t $(t > 0)$ определяется равенством

$$\mathsf{Q}_t(dy) = \frac{1}{ct} y^2 \mathsf{N}_{0,ct}(dy)$$

или, что то же,

$$\mathsf{Q}_t(B) = \int\limits_B \frac{1}{ct} y^2 \mathsf{N}_{0,ct}(dy),\ B \in \mathfrak{B}^1$$

(распределение Q_t абсолютно непрерывно относительно распределения $\mathsf{N}_{0,ct}$ с плотностью $\frac{1}{ct}y^2$).

Исследовать Q_t на сходимость при $t \to 0$.

12.12. Пусть F — собственное вероятностное распределение на $\mathbb{R}^1$, $\mathsf{U}_{[-h,h]}$ — равномерное на $[-h, h]$ распределение. Доказать, что при $h \to 0$ распределение

$$\mathsf{F}_h = \mathsf{F} * \mathsf{U}_{[-h,h]}$$

сходится к распределению F.

12.13. Пусть Q_n — атомическое распределение, заданное равенством

$$\mathsf{Q}_n(\{k/n\}) = 1/n,\ k = 0, 1, \ldots, n-1.$$

Исследовать Q_n на сходимость при $n \to \infty$.

У к а з а н и е. Распределение Q_n сходится к равномерному на промежутке $[0, 1]$ распределению. Достаточно заметить, что при

$0 \le x \le 1$ для функции распределения $Q_n(x)$ имеют место неравенства

$$x - 1/n \le Q_n(x) \le x + 1/n.$$

12.14*. Пусть F_λ —распределение, заданное функцией распределения

$$F_\lambda(y) = \sum_{k:0\le k<y} \frac{\lambda^k}{k!} e^{-\lambda}, \ y \in \mathbb{R}^1, \ \lambda > 0,$$

k — целые неотрицательные.

Вычислить

$$\lim_{\lambda\to 0} \int\limits_{-\infty}^{+\infty} e^{-y^2/2}\, \mathsf{F}_\lambda(dy).$$

Глава 13

Характеристическая функция

13.1 Комплекснозначные случайные величины

Комплексные числа. *Комплексным числом* $z = (a, b)$ называется упорядоченная пара действительных чисел (a, b); число a будем называть действительной частью комплексного числа $z = (a, b)$, b — мнимой.

Суммой $z_1 + z_2$ чисел $z_1 = (a_1, b_1)$ и $z_2 = (a_2, b_2)$ называется комплексное число $(a_1 + a_2, b_1 + b_2)$.

Произведением $z_1 z_2$ чисел $z_1 = (a_1, b_1)$ и $z_2 = (a_2, b_2)$ называется комплексное число $(a_1 a_2 - b_1 b_2, a_1 b_2 + b_1 a_2)$.

Комплексное число вида $(a, 0)$ будем отождествлять с действительным числом a.

Комплексное число вида $(0, 1)$ называется *мнимой единицей* и обозначается i, т. е. $i = (0, 1)$. Очевидно,

$$i^2 = (0, 1)(0, 1) = (-1, 0) = -1.$$

Комплексное число $z = (a, b)$ можно представить в виде

$$z = (a, b) = a + ib$$

(*алгебраическая форма* записи комплексного числа). Операции над комплексными числами $z_1 = a_1 + ib_1$ и $z_2 = a_2 + ib_2$ удобно выполнять как операции над многочленами, учитывая, что $i^2 = -1$.

Модулем комплексного числа $z = a + ib$ называется число

$$|z| = \sqrt{a^2 + b^2}.$$

В декартовой системе координат число $z = a+ib$ изображают точкой плоскости с координатами (a, b).

Из соотношений между декартовыми (a, b) и полярными (ρ, φ) координатами точки на плоскости ($a = \rho\cos\varphi, b = \rho\sin\varphi$) получаем:

$$z = (a, b) = a + ib = \rho(\cos\varphi + i\sin\varphi)$$

(*тригонометрическая форма* записи комплексного числа).

Элементарные функции комплексного переменного определяются как суммы абсолютно сходящихся рядов, а именно:

$$e^z = \sum_{k=0}^{\infty} \frac{1}{k!} z^k,$$

$$\sin z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1},$$

$$\cos z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} z^{2k},$$

$$\ln(z+1) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} z^{k+1}, \ |z| < 1.$$

Мы часто будем пользоваться формулой Эйлера:

$$e^{i\varphi} = \cos\varphi + i\sin\varphi,$$

которая очевидным образом следует из приведенных выше определений $e^z, \sin z, \cos z$. Из нее, в частности, получаем, что комплексное число $z = (a, b)$ можно представить в виде:

$$z = \rho e^{i\varphi}$$

(*показательная форма* записи комплексного числа).

Комплекснозначные случайные величины. Пусть $\xi_1 = \xi_1(\omega)$ и $\xi_2 = \xi_2(\omega)$ — случайные величины со значениями в $\mathbb{R}^1$

на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$. Функцию $\xi = \xi_1 + i\xi_2$ со значениями в $\mathbb{C}$ на $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ будем называть *комплекснозначной случайной величиной.*

Математическое ожидание комплекснозначной случайной величины $\xi = \xi_1 + i\xi_2$ определяется равенством

$$\mathsf{M}\xi = \mathsf{M}(\xi_1 + i\xi_2) = \mathsf{M}\xi_1 + i\mathsf{M}\xi_2$$

и обладает теми же свойствами, что и математическое ожидание действительной случайной величины (за исключением свойства монотонности):

$$\mathsf{M}(\xi I_B) = 0, \text{ если } \mathsf{P}(B) = 0,$$

$$\mathsf{M}(\xi I_{B_1 \cup B_2}) = \mathsf{M}(\xi I_{B_1}) + \mathsf{M}(\xi I_{B_2}), \text{ если } B_1 \cap B_2 = \emptyset,$$

$$\mathsf{M}(\alpha\xi I_B) = \alpha\mathsf{M}(\xi I_B),\ \alpha \text{ — константа},$$

$$\mathsf{M}((\xi + \eta)I_B) = \mathsf{M}(\xi I_B) + \mathsf{M}(\eta I_B),$$

$$|\mathsf{M}\xi| \leq \mathsf{M}|\xi|.$$

Перечисленные свойства математического ожидания комплекснозначной случайной величины, за исключением последнего, очевидным образом следуют из соответствующих свойств математического ожидания действительной случайной величины.

Докажем, что

$$|\mathsf{M}\xi| \leq \mathsf{M}|\xi|.$$

Представим случайную величину ξ и ее математическое ожидание $\mathsf{M}\xi$ в показательном виде:

$$\xi = \rho e^{i\varphi}, |\xi| = \rho;\ \mathsf{M}\xi = re^{i\beta}, |\mathsf{M}\xi| = r$$

($\rho = \rho(\omega),\ \varphi = \varphi(\omega)$ — случайные величины; r, β — постоянные).

Из последних представлений имеем

$$|\mathsf{M}\xi| = r = e^{-i\beta}\mathsf{M}\xi = e^{-i\beta}\mathsf{M}\rho e^{i\varphi} = \mathsf{M}\rho e^{i(\varphi-\beta)} =$$

$$= \mathsf{M}\rho\cos(\varphi - \beta) + i\mathsf{M}\rho\sin(\varphi - \beta).$$

И так как r — действительное число, то $r = \mathsf{M}\rho\cos(\varphi - \beta)$, поэтому

$$|\mathsf{M}\xi| = r = \mathsf{M}\rho\cos(\varphi - \beta) \leq \mathsf{M}\rho = \mathsf{M}|\xi|.$$

Так что

$$|\mathsf{M}\xi| \leq \mathsf{M}|\xi|.$$

Определение. Комплекснозначные случайные величины $\xi = \xi_1 + i\xi_2$ и $\eta = \eta_1 + i\eta_2$ будем называть *независимыми*, если для любых борелевских множеств B_1, B_2, D_1, D_2

$$\mathsf{P}\{\xi_1 \in B_1, \xi_2 \in B_2, \eta_1 \in D_1, \eta_2 \in D_2\} =$$

$$= \mathsf{P}\{\xi_1 \in B_1, \xi_2 \in B_2\}\mathsf{P}\{\eta_1 \in D_1, \eta_2 \in D_2\}.$$

Для независимых комплекснозначных случайных величин имеет место мультипликативное свойство математического ожидания:

$$\mathsf{M}\xi\eta = \mathsf{M}\xi\mathsf{M}\eta.$$

13.2 Характеристическая функция — определение, свойства

Приводимое далее определение характеристической функции вводится для конечных с вероятностью 1 случайных величин и для собственных вероятностных распределений.

Характеристические функции являются мощным аппаратом исследования распределений. Связано это с тем, что между распределениями и характеристическими функциями существует взаимно однозначное и взаимно непрерывное соответствие. Поэтому свойства распределений можно изучать, изучая свойства их характеристических функций.

Определение. Пусть ξ — случайная величина со значениями в $\mathbb{R}^1$ и F — ее распределение. *Характеристической функцией случайной величины* ξ (*характеристической функцией распределения* F) будем называть комплекснозначную функцию $\varphi(t)$, определенную равенством

$$\varphi(t) = \mathsf{M}e^{it\xi} = \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}(dx).$$

Характеристическая функция любой случайной величины (любого вероятностного распределения F) определена, поскольку $|e^{it\xi}| = 1$, а математическое ожидание ограниченной случайной величины существует и конечно.

Если распределение F (распределение случайной величины ξ) дискретно — существует не более чем счетное множество точек x_k таких, что $\mathsf{P}_\xi(x_k) = \mathsf{F}(\{x_k\}) > 0$ и $\sum\limits_k \mathsf{F}(\{x_k\}) = 1$, то

$$\varphi(t) = \mathsf{M}e^{it\xi} = \sum_k e^{itx_k}\mathsf{F}(\{x_k\}).$$

Если распределение F (распределение случайной величины ξ) абсолютно непрерывно с плотностью f, то

$$\varphi(t) = \mathsf{M}e^{it\xi} = \int\limits_{\mathbb{R}^1} e^{itx} f(x)dx. \tag{13.2.1}$$

Пример 13.2.1. *Вычислить характеристическую функцию пуассоновского распределения с параметром λ.*

Р е ш е н и е. Пуассоновское распределение P_λ с параметром λ задается равенством

$$\mathsf{P}_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, \ldots$$

Поэтому

$$\varphi(t) = \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{P}_\lambda(dx) = \sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!}e^{-\lambda} =$$

$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^{it}\right)^k}{k!} = e^{-\lambda}e^{\lambda e^{it}} = \exp\left\{\lambda(e^{it}-1)\right\}.$$

Пример 13.2.2. *Вычислить характеристическую функцию атомического распределения F_a, сосредоточенного в точке a.*

Р е ш е н и е.

$$\varphi(t) = \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}_a(dx) = e^{ita}\mathsf{F}_a(a) = e^{ita}.$$

Основные свойства характеристической функции получаются непосредственно из ее определения.

Лемма 13.2.1. 1° *Характеристическая функция $\varphi(t)$ случайной величины ξ равномерно непрерывна и ограничена:*

$$|\varphi(t)| \leq 1,\ \varphi(0) = 1,\ t \in \mathbb{R}^1.$$

2° *Характеристическая функция $\psi(t)$ суммы*

$$\eta = \sigma\xi + a$$

равна произведению $\varphi(\sigma t)e^{ita}$.

Доказательство.

1° $|\varphi(t)| = |\mathsf{M}e^{it\xi}| \leq \mathsf{M}|e^{it\xi}| = 1$; $\varphi(0) = \mathsf{M}e^{i0\xi} = 1, t \in \mathbb{R}^1$.

Далее,

$$|\varphi(t+h) - \varphi(t)| = |\mathsf{M}e^{i(t+h)\xi} - \mathsf{M}e^{it\xi}| =$$

$$= |\mathsf{M}e^{it\xi}(e^{ih\xi} - 1)| \leq \mathsf{M}|e^{ih\xi} - 1|.$$

Случайная величина $|e^{ih\xi} - 1|$ мажорируема интегрируемой по вероятностной мере функцией $\gamma(\omega) \equiv 2$ и $|e^{ih\xi} - 1| \to 0$ при $h \to 0$ в каждой точке ω. Поэтому в силу теоремы Лебега о мажорируемой сходимости $\mathsf{M}|e^{ih\xi} - 1| \to 0$ при $h \to 0$. Так что $\varphi(t)$ равномерно непрерывна.

2° $\psi(t) = \mathsf{M}e^{it\eta} = \mathsf{M}e^{it(\sigma\xi + a)} = e^{ita}\mathsf{M}e^{i(t\sigma)\xi} = \varphi(\sigma t)e^{ita}$.

Лемма 13.2.2 (мультипликативное свойство характеристических функций). *Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых, в терминах сверток: характеристическая функция свертки вероятностных распределений равна произведению характеристических функций этих распределений.*

Доказательство. Пусть ξ и η — независимые случайные величины соответственно с распределениями F и G и характеристическими функциями $\varphi(t)$ и $\psi(t)$. Характеристическая функция суммы $\xi + \eta$:

$$\mathsf{M}e^{it(\xi+\eta)} = \mathsf{M}e^{it\xi}e^{it\eta} = \mathsf{M}e^{it\xi}\mathsf{M}e^{it\eta} = \varphi(t)\psi(t)$$

— мы воспользовались свойством мультипликативности математического ожидания (случайные величины $e^{it\xi}$ и $e^{it\eta}$ независимы как функции независимых случайных величин ξ и η).

Далее, поскольку $\mathsf{F} * \mathsf{G}$ — распределение суммы $\xi + \eta$ независимых случайных величин ξ и η (см. теорему 11.1.4), то характеристическая функция распределения $\mathsf{F} * \mathsf{G}$ совпадает с характеристической функцией суммы $\xi + \eta$, а последняя равна произведению $\varphi(t)\psi(t)$ характеристических функций распределений F и G:

$$\varphi(t)\psi(t) = \mathsf{M}e^{it(\xi+\eta)} = \mathsf{M}e^{it\zeta} = \int\limits_{\mathbb{R}^1} e^{itx} \mathsf{F} * \mathsf{G}(dx).$$

13.3 Дифференцируемость характеристической функции

Далее нам понадобится оценка погрешности, с которой функция e^{it} аппроксимируется конечным отрезком своего ряда Тейлора.

Лемма 13.3.1 (об оценке погрешности). *Погрешность при замене функции e^{it}, $t \in \mathbb{R}^1$, отрезком ее ряда Тейлора не превосходит модуля первого отброшенного члена:*

$$\left| e^{it} - \left(1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^n}{n!} \right) \right| \leq \frac{|t|^{n+1}}{(n+1)!}, \qquad (13.3.1)$$

$n = 0, 1, 2, \ldots$

Доказательство. Обозначим через $\rho_n(t)$ погрешность при замене e^{it} отрезком ряда Тейлора:

$$\rho_n(t) = e^{it} - \left(1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^n}{n!} \right),$$

$n = 0, 1, 2, \ldots$ В частности,

$$\rho_0(t) = e^{it} - 1.$$

Очевидно,

$$i \int\limits_0^t e^{ix} dx = e^{it} - 1 = \rho_0(t),$$

Непосредственной проверкой убеждаемся, что

$$\rho_n(t) = i\int\limits_0^t \rho_{n-1}(x)dx,\ n \geq 1.$$

Отсюда для $t > 0$ последовательно имеем:

$$|\rho_0(t)| = \left| i\int\limits_0^t e^{ix}dx \right| \leq \int\limits_0^t \left|e^{ix}\right| dx \leq t,$$

$$|\rho_1(t)| = \left| i\int\limits_0^t \rho_0(x)dx \right| \leq \int\limits_0^t |\rho_0(x)|\, dx \leq \int\limits_0^t x\, dx \leq \frac{t^2}{2}$$

и т. д.

$$|\rho_n(t)| \leq \frac{t^{n+1}}{(n+1)!} = \frac{|t|^{n+1}}{(n+1)!}.$$

Последнее неравенство доказано для $t > 0$, аналогично оно доказывается при $t < 0$.

Так что для целых $n \geq 0$

$$|\rho_n(t)| \leq \frac{|t|^{n+1}}{(n+1)!}.$$

Лемма доказана.

Следствие. *Для целых* $n \geq 0$

$$e^{it} = 1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^n}{n!} + \frac{(it)^n}{n!}\varepsilon(t), \tag{13.3.2}$$

причем $\varepsilon(t) \to 0$ *при* $t \to 0$ *и* $|\varepsilon(t)| \leq 2$, *в частности, при* $t \to 0$

$$e^{it} = 1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^n}{n!} + o(t^n).$$

Доказательство. В силу леммы

$$e^{it} = 1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^n}{n!} + \rho_n(t), \tag{13.3.3}$$

причем

$$|\rho_n(t)| \le \frac{|t|^{n+1}}{(n+1)!}.$$

Разделив почленно последнее неравенство на $\left|\dfrac{(it)^n}{n!}\right|$, получим:

$$\left|\rho_n(t)\Big/\frac{(it)^n}{n!}\right| \le \frac{|t|}{n+1}. \tag{13.3.4}$$

Если ввести обозначение

$$\varepsilon(t) = \rho_n(t)\Big/\frac{(it)^n}{n!},$$

то для $\rho_n(t)$ получим представление

$$\rho_n(t) = \frac{(it)^n}{n!}\varepsilon(t),$$

а неравенство (13.3.4) запишется в виде

$$|\varepsilon(t)| \le |t|/(n+1).$$

Отсюда следует, что $\varepsilon(t) \to 0$ при $t \to 0$ и

$$\rho_n(t) = o\left(t^n\right).$$

Так что равенство (13.3.3) можно записать в виде:

$$e^{it} = 1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^{n-1}}{(n-1)!} + \frac{(it)^n}{n!} + \frac{(it)^n}{n!}\varepsilon(t), \tag{13.3.5}$$

причем $\varepsilon(t) \to 0$ при $t \to 0$, в частности, при $t \to 0$

$$e^{it} = 1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^{n-1}}{(n-1)!} + \frac{(it)^n}{n!} + o(t^n).$$

Осталось показать, что имеет место оценка $|\varepsilon(t)| \le 2$. Перепишем равенство (13.3.5) так:

$$e^{it} = 1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^{n-1}}{(n-1)!} + \frac{(it)^n}{n!}(1 + \varepsilon(t)).$$

С другой стороны,

$$e^{it} = 1 + \frac{it}{1!} + \frac{(it)^2}{2!} + \ldots + \frac{(it)^{n-1}}{(n-1)!} + \rho_{n-1}(t).$$

Из двух последних равенств имеем:

$$\rho_{n-1}(t) = \frac{(it)^n}{n!}(1 + \varepsilon(t)).$$

Отсюда, учитывая, что

$$\left| \frac{(it)^n}{n!}(1 + \varepsilon(t)) \right| = |\rho_{n-1}(t)| \leq \left| \frac{(it)^n}{n!} \right|,$$

последовательно получаем:

$$|1 + \varepsilon(t)| \leq 1, \ |\varepsilon(t)| = |\varepsilon(t) + 1 - 1| \leq |\varepsilon(t) + 1| + 1 \leq 2.$$

Так что

$$|\varepsilon(t)| \leq 2.$$

Тем самым следствие доказано.

Пусть F — распределение вероятностей и φ — его характеристическая функция. Далее будем использовать обозначения

$$m_n = \int\limits_{\mathbb{R}^1} x^n \mathsf{F}(dx), \ M_n = \int\limits_{\mathbb{R}^1} |x|^n \mathsf{F}(dx).$$

Напомним, что функция интегрируема по Лебегу вместе со своим модулем.

Теорема 13.3.1 (о дифференцируемости характеристической функции). *Если n-й момент распределения F конечен, то у его характеристической функции*

$$\varphi(t) = \int\limits_{\mathbb{R}^1} e^{itx} \mathsf{F}(dx)$$

существует n-я производная, она непрерывна, и может быть получена дифференцированием под знаком интеграла:

$$\varphi^{(n)}(t) = i^n \int\limits_{\mathbb{R}^1} e^{itx} x^n \mathsf{F}(dx).$$

Доказательство. Рассмотрим разностное отношение

$$\frac{\varphi(t+h)-\varphi(t)}{h} = \int\limits_{\mathbb{R}^1} \frac{e^{i(t+h)x}-e^{itx}}{h}\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} e^{itx}\frac{e^{ihx}-1}{h}\mathsf{F}(dx). \tag{13.3.6}$$

Подынтегральная функция $e^{itx}\dfrac{e^{ihx}-1}{h}$ (зависящая от параметра h):

1° мажорируема интегрируемой функцией $|x|$

$$\left|e^{itx}\frac{e^{ihx}-1}{h}\right| = \left|e^{itx}\right|\left|\frac{e^{ihx}-1}{h}\right| \le \frac{|ihx|}{|h|} = |x|$$

(функция $|x|$ интегрируема, так как по условию $|x|^n$ интегрируема, см. неравенство Ляпунова (9.3.2));

2° сходится к ixe^{itx} при $h\to 0$, так как

$$\left|e^{itx}\frac{e^{ihx}-1}{h} - ixe^{itx}\right| = \left|e^{itx}\right|\left|\frac{e^{ihx}-(1+ihx)}{h}\right| \le$$

$$\le \frac{|ihx|^2/2!}{|h|} = \frac{|h|}{2}|x|^2$$

(мы воспользовались леммой об оценке погрешности (лемма 13.3.1)). Поэтому в силу теоремы Лебега о мажорируемой сходимости

$$\lim_{h\to 0}\int\limits_{\mathbb{R}^1} e^{itx}\frac{e^{ihx}-1}{h}\mathsf{F}(dx) =$$

$$= \int\limits_{\mathbb{R}^1} \lim_{h\to 0} e^{itx}\frac{e^{ihx}-1}{h}\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} ixe^{itx}\mathsf{F}(dx).$$

Отсюда и из равенства (13.3.6) следует, во-первых, существование

$$\lim_{h\to 0}\frac{\varphi(t+h)-\varphi(t)}{h},$$

а во-вторых, равенство

$$\varphi'(t) = i\int\limits_{\mathbb{R}^1} e^{itx}x\mathsf{F}(dx).$$

По индукции получаем:

$$\varphi^{(n)}(t) = i^n \int\limits_{\mathbb{R}^1} e^{itx} x^n \mathsf{F}(dx).$$

Непрерывность $\varphi^{(n)}(t)$ следует из оценки

$$\left|\varphi^{(n)}(t+h) - \varphi^{(n)}(t)\right| \leq \int\limits_{\mathbb{R}^1} \left|i^n x^n e^{i(t+h)x} - i^n x^n e^{itx}\right| \mathsf{F}(dx) =$$

$$= \int\limits_{\mathbb{R}^1} |x|^n \left|e^{ihx} - 1\right| \mathsf{F}(dx).$$

Подынтегральная функция $|x|^n \left|e^{ihx} - 1\right|$ мажорируема интегрируемой по распределению F функцией $2|x|^n$ (по условию $2|x|^n$ интегрируема) и при $h \to 0$ сходится к нулю, поскольку

$$|x|^n \left|e^{ihx} - 1\right| \leq |x|^n |hx|.$$

Поэтому в силу теоремы Лебега о мажорируемой сходимости при $h \to 0$

$$\int\limits_{\mathbb{R}^1} |x|^n \left|e^{ihx} - 1\right| \mathsf{F}(dx) \to 0$$

а, следовательно, и

$$|\varphi^{(n)}(t+h) - \varphi^{(n)}(t)| \to 0.$$

Следствие. *Если n-й момент случайной величины ξ (распределения F) конечен, то*

$$\varphi^{(n)}(0) = i^n m_n = i^n \mathsf{M}\xi^n. \tag{13.3.7}$$

Достаточно в равенстве

$$\varphi^{(n)}(t) = i^n \int\limits_{\mathbb{R}^1} e^{itx} x^n \mathsf{F}(dx)$$

положить $t = 0$.

Теорема 13.3.2 (о разложении Тейлора для характеристической функции). *Если n-й момент вероятностного распределения F конечен, то для его характеристической функции φ в окрестности нуля имеет место разложение Тейлора:*

$$\varphi(t) = 1 + \frac{\varphi'(0)}{1!}t + \frac{\varphi''(0)}{2!}t^2 + \cdots + \frac{\varphi^{(n)}(0)}{n!}t^n + \frac{t^n}{n!}\delta(t),$$

причем $\delta(t) \to 0$ при $t \to 0$ и $|\delta(t)| \le 2\mathsf{M}_n$.

Доказательство. В силу следствия из леммы 13.3.1 об оценке погрешности

$$e^{itx} = 1 + \frac{itx}{1!} + \frac{(itx)^2}{2!} + \ldots + \frac{(itx)^n}{n!} + \frac{(itx)^n}{n!}\varepsilon(tx),$$

причем $\varepsilon(tx) \to 0$ при $tx \to 0$ и $|\varepsilon(tx)| \le 2$. Почленное интегрирование по распределению F последнего равенства, с учетом того, что

$$\mathsf{M}_n = \int\limits_{\mathbb{R}^1} |x|^n \mathsf{F}(dx) < \infty \quad \text{и} \quad \varphi^{(k)}(0) = i^k m_k = i^k \int\limits_{\mathbb{R}^1} x^k \mathsf{F}(dx),$$

$k = 1, 2, \ldots, n$, дает:

$$\varphi(t) = 1 + \frac{\varphi'(0)}{1!}t + \frac{\varphi''(0)}{2!}t^2 + \cdots + \frac{\varphi^{(n)}(0)}{n!}t^n + \frac{t^n}{n!}\delta(t),$$

где

$$\delta(t) = i^n \int\limits_{\mathbb{R}^1} x^n \varepsilon(tx) \mathsf{F}(dx).$$

Далее, $x^n\varepsilon(tx) \to 0$ при $t \to 0$ и $|x^n\varepsilon(tx)| \le 2|x|^n$, причем функция $|x|^n$ интегрируема по распределению F. Поэтому в силу теоремы Лебега о мажорируемой сходимости $\delta(t) \to 0$ при $t \to 0$. Осталось заметить, что

$$|\delta(t)| \le \int\limits_{\mathbb{R}^1} |x^n\varepsilon(tx)| \mathsf{F}(dx) \le 2\int\limits_{\mathbb{R}^1} |x|^n \mathsf{F}(dx) = 2\mathsf{M}_n.$$

Тем самым теорема доказана.

Следствие. *В условиях теоремы при $t \to 0$*

$$\varphi(t) = 1 + \frac{\varphi'(0)}{1!}t + \frac{\varphi''(0)}{2!}t^2 + \cdots + \frac{\varphi^{(n)}(0)}{n!}t^n + o(t^n).$$

Теорема 13.3.3. *Характеристическая функция нормального распределения с параметрами* $(0;1)$ *равна*

$$e^{-t^2/2}.$$

Доказательство. По определению

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \cos tx \cdot e^{-x^2/2} dx + i \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sin tx \cdot e^{-x^2/2} dx.$$

Последний интеграл равен нулю как интеграл от нечетной функции по симметричному промежутку:

$$\int_{-\infty}^{+\infty} \sin tx \cdot e^{-x^2/2} dx = \lim_{n} \int_{-n}^{n} \sin tx \cdot e^{-x^2/2} dx = 0.$$

Поэтому

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \cos tx \cdot e^{-x^2/2} dx. \tag{13.3.8}$$

У нормального распределения конечен первый момент (как, впрочем, все остальные) поэтому последнее равенство можно продифференцировать (см. теорему 13.3.1), причем в правой части под знаком интеграла:

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x(-\sin tx) e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sin tx \, de^{-x^2/2} =$$

$$= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t \cos tx \cdot e^{-x^2/2} dx = -t\varphi(t).$$

Так что $\varphi(t)$ удовлетворяет дифференциальному уравнению

$$\varphi'(t)/\varphi(t) = -t.$$

Решая его с учетом начального условия $\varphi(0) = 1$, получаем:

$$\varphi(t) = e^{-t^2/2}.$$

Следствие. *Характеристическая функция нормального распределения с параметрами* $(a;\sigma^2)$ *равна*

$$e^{ita - t^2\sigma^2/2}.$$

Доказательство. Если η — нормально распределенная с параметрами $(a;\sigma^2)$ случайная величина, то случайная величина $\xi = (\eta - a)/\sigma$ распределена $\mathsf{N}_{0;1}$ (в чем убеждаемся непосредственной проверкой). Поэтому $\mathsf{N}_{a;\sigma^2}$-распределенную случайную величину η можно представить в виде:

$$\eta = a + \sigma\xi,$$

где ξ распределена $\mathsf{N}_{0;1}$. Отсюда

$$\mathsf{M}e^{it\eta} = \mathsf{M}e^{it(a+\sigma\xi)} = e^{ita}\mathsf{M}e^{i(t\sigma)\xi} = e^{ita}e^{-t^2\sigma^2/2} = e^{ita - t^2\sigma^2/2}.$$

13.4 Теорема единственности, формула обращения

Равенство Парсеваля. *Пусть* F *и* G *— вероятностные распределения с характеристическими функциями* φ *и* γ *соответственно, тогда имеет место равенство Парсеваля:*

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{G}(ds) = \int\limits_{\mathbb{R}^1} \gamma(x-t)\mathsf{F}(dx). \tag{13.4.1}$$

По определению характеристической функции распределения F

$$\varphi(s) = \int\limits_{\mathbb{R}^1} e^{isx}\mathsf{F}(dx).$$

Умножив правую и левую части равенства на e^{-ist}:

$$e^{-ist}\varphi(s) = \int\limits_{\mathbb{R}^1} e^{is(x-t)}\mathsf{F}(dx)$$

и проинтегрировав по распределению G, получим:

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{G}(ds) = \int\limits_{\mathbb{R}^1}\left(\int\limits_{\mathbb{R}^1} e^{is(x-t)}\mathsf{F}(dx)\right)\mathsf{G}(ds) =$$

$$= \int\limits_{\mathbb{R}^1}\left(\int\limits_{\mathbb{R}^1} e^{is(x-t)}\mathsf{G}(ds)\right)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1}\gamma(x-t)\mathsf{F}(dx).$$

Теорема 13.4.1 (единственности). *Различные вероятностные распределения имеют различные характеристические функции.*

Доказательство. Для доказательства теоремы достаточно доказать, что распределение F однозначно выражается через свою характеристическую функцию φ. Чем мы и займемся.

Введем распределение

$$\mathsf{F}_\sigma = \mathsf{F} * \mathsf{N}_{0;\sigma^2} = \mathsf{N}_{0;\sigma^2} * \mathsf{F}$$

и установим, что

1° семейство распределений F_σ сходится к распределению F при $\sigma \to 0$;

2° плотность f_σ распределения F_σ, а вместе с ним и само распределение, однозначно выражается через характеристическую функцию φ распределения F.

Сначала докажем, что

$$\lim_{\sigma\to 0}\mathsf{F}_\sigma = \mathsf{F}.$$

Для этого покажем, что в каждой точке непрерывности y функции распределения $F(x)$

$$\lim_{\sigma\to 0}F_\sigma(y) = F(y).$$

Равенство $\mathsf{F}_\sigma = \mathsf{N}_{0;\sigma^2} * \mathsf{F}$ означает, что

$$F_\sigma(y) = \int\limits_{-\infty}^{+\infty} F(y-x)\mathsf{N}_{0;\sigma^2}(dx) =$$

$$= \int_{-\infty}^{+\infty} F(y-x) \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} dx.$$

Сделав замену $x/\sigma = t$, получим

$$F_\sigma(y) = \int_{-\infty}^{+\infty} F(y-\sigma t) \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt =$$

$$= \int_{-\infty}^{+\infty} F(y-\sigma t) \mathsf{N}_{0,1}(dt).$$

Подынтегральная функция $F(y-\sigma t)$ мажорируема интегрируемой по $\mathsf{N}_{0,1}$-распределению функцией $f(t) = 1$ и в каждой точке непрерывности y функции $F(x)$

$$F(y-\sigma t) \to F(y)$$

при $\sigma \to 0$. Поэтому в силу теоремы Лебега о мажорируемой сходимости

$$\lim_{\sigma\to 0} F_\sigma(y) = \lim_{\sigma\to 0} \int_{-\infty}^{+\infty} F(y-\sigma t) \mathsf{N}_{0,1}(dt) =$$

$$= \int_{-\infty}^{+\infty} (\lim_{\sigma\to 0} F(y-\sigma t)) \mathsf{N}_{0,1}(dt) =$$

$$= \int_{-\infty}^{+\infty} F(y) \mathsf{N}_{0,1}(dt) = F(y) \int_{-\infty}^{+\infty} \mathsf{N}_{0,1}(dt) = F(y).$$

Тем самым мы установили, что в каждой точке непрерывности y функции распределения $F(x)$

$$F(y) = \lim_{\sigma\to 0} F_\sigma(y). \tag{13.4.2}$$

Далее докажем, что плотность f_σ абсолютно непрерывного распределения $\mathsf{F}_\sigma = \mathsf{F} * \mathsf{N}_{0;\sigma^2}$, а вместе с ней и само распределение F_σ, однозначно выражается через характеристическую функцию φ распределения F. Плотность f_σ свертки $\mathsf{F}_\sigma = \mathsf{F} * \mathsf{N}_{0;\sigma^2}$ равна свертке плотности

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{s^2}{2\sigma^2}\right\}$$

распределения $\mathsf{N}_{0;\sigma^2}$ с распределением F:

$$f_\sigma(t) = \int\limits_{\mathbb{R}^1} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(t-x)^2}{2\sigma^2}\right\}\mathsf{F}(dx) \tag{13.4.3}$$

(см. теорему 11.1.3).

Чтобы выразить f_σ через характеристическую функцию φ распределения F, воспользуемся равенством Парсеваля:

$$\int\limits_{-\infty}^{+\infty} e^{-ist}\varphi(s)\mathsf{G}(ds) = \int\limits_{-\infty}^{+\infty} \gamma(x-t)\mathsf{F}(dx).$$

Рассмотрим в качестве распределения G нормальное распределение $\mathsf{N}_{0;1/\sigma^2}$, его плотность

$$g(s) = \frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{s^2\sigma^2}{2}\right\},$$

а характеристическая функция

$$\gamma(s) = \exp\left\{-\frac{s^2}{2\sigma^2}\right\}.$$

Получим

$$\int\limits_{-\infty}^{+\infty} e^{-ist}\varphi(s)\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{s^2\sigma^2}{2}\right\}ds = \int\limits_{-\infty}^{+\infty} \exp\left\{-\frac{(x-t)^2}{2\sigma^2}\right\}\mathsf{F}(dx)$$

или, почленно умножив последнее равенство на $\dfrac{1}{\sqrt{2\pi}\sigma}$,

$$\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} e^{-ist}\varphi(s)\exp\left\{-\frac{s^2\sigma^2}{2}\right\}ds =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(t-x)^2}{2\sigma^2}\right\} \mathsf{F}(dx) = f_\sigma(t)$$

(см. также равенство (13.4.3)). Так что плотность $f_\sigma(t)$ распределения F_σ однозначно выражается через характеристическую функцию φ распределения F:

$$f_\sigma(t) = \frac{1}{2\pi} \int_{\mathbb{R}^1} e^{-ist} \varphi(s) \exp\left\{-\frac{s^2\sigma^2}{2}\right\} ds. \tag{13.4.4}$$

Отсюда, учитывая равенства (13.4.2) имеем:

$$F(y) = \lim_{\sigma\to 0} F_\sigma(y) = \lim_{\sigma\to 0} \int_{-\infty}^{y} f_\sigma(t)dt =$$

$$= \lim_{\sigma\to 0} \int_{-\infty}^{y} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ist} \varphi(s) \exp\left\{-\frac{s^2\sigma^2}{2}\right\} ds \right) dt.$$

Итак, функция распределения $F(y)$, а следовательно, и само распределение F, однозначно определяется по φ. Более того, мы получили представление для $F(y)$ через характеристическую функцию φ распределения F.

Пример 13.4.1 (свертка нормальных распределений). *Пусть* F *и* Q *— нормальные распределения с параметрами* $(a_1; \sigma_1^2)$ *и* $(a_2; \sigma_2^2)$ *соответственно. Найти свертку* $\mathsf{F} * \mathsf{Q}$ *распределений* Q *и* F.

Решение. Воспользовавшись мультипликативным свойством характеристических функций (характеристическая функция свертки вероятностных распределений равна произведению их характеристических функций), по известным характеристическим функциям

$$\varphi_1(t) = \exp\left\{ita_1 - \frac{t^2\sigma_1^2}{2}\right\} \text{ и } \varphi_2(t) = \exp\left\{ita_2 - \frac{t^2\sigma_2^2}{2}\right\}$$

нормальных распределений F и Q с параметрами соответственно $(a_1; \sigma_1^2)$ и $(a_2; \sigma_2^2)$ получим характеристическую функцию свертки $\mathsf{F} * \mathsf{Q}$:

$$\varphi(t) = \varphi_1(t)\varphi_2(t) = \exp\left\{it(a_1+a_2) - \frac{t^2(\sigma_1^2+\sigma_2^2)}{2}\right\}.$$

С другой стороны

$$\varphi(t) = \exp\left\{it(a_1 + a_2) - \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}\right\}$$

— характеристическая функция нормального распределения с параметрами $(a_1+a_2, \sigma_1^2+\sigma_2^2)$. Поэтому в силу теоремы единственности распределение $\mathsf{F} * \mathsf{Q}$ совпадает с нормальным распределением с параметрами $(a_1 + a_2; \sigma_1^2 + \sigma_2^2)$.

Так что *свертка нормальных распределений с параметрами* $(a_1; \sigma_1^2)$ *и* $(a_2; \sigma_2^2)$ *является нормальным распределением с параметрами* $(a_1 + a_2; \sigma_1^2 + \sigma_2^2)$.

Фактически получен следующий результат: *класс нормальных распределений замкнут относительно операции свертки.*

Пример 13.4.2 (свертка пуассоновских распределений). *Пусть* F *и* Q *— пуассоновские распределения с параметрами соответственно* λ_1 *и* λ_2. *Найти свертку* $\mathsf{F} * \mathsf{Q}$ *распределений* Q *и* F.

Решение. Характеристические функции пуассоновских распределений с параметрами λ_1 и λ_2 соответственно равны

$$\varphi_1(t) = \exp\{\lambda_1(e^{it} - 1)\} \text{ и } \varphi_2(t) = \exp\{\lambda_2(e^{it} - 1)\}$$

(см. пример 13.2.1), а характеристическая функция $\varphi(t)$ свертки $\mathsf{F} * \mathsf{Q}$ в силу мультипликативного свойства равна

$$\varphi(t) = \varphi_1(t)\varphi_2(t) = \exp\{(\lambda_1 + \lambda_2)(e^{it} - 1)\}.$$

Но $\exp\{(\lambda_1 + \lambda_2)(e^{it} - 1)\}$ — характеристическая функция пуассоновского распределения с параметром $\lambda_1 + \lambda_2$. Поэтому в силу теоремы единственности распределение $\mathsf{F} * \mathsf{Q}$ совпадает с пуассоновским распределением с параметром $\lambda_1 + \lambda_2$.

Следовательно, *свертка пуассоновских распределений с параметрами* λ_1 *и* λ_2 *является пуассоновским распределением с параметром* $\lambda_1 + \lambda_2$.

Так что получен следующий результат: *класс пуассоновских распределений замкнут относительно операции свертки.*

Теорема 13.4.2 (формула обращения для плотности). *Если характеристическая функция* φ *вероятностного распределения* F *интегрируема по мере Лебега* $\left(\int\limits_{\mathbb{R}^1} |\varphi(s)|ds < \infty\right)$, *то распределение* F *абсолютно непрерывно, его плотность* $f(t)$ *представима в виде*

$$f(t) = \frac{1}{2\pi} \int\limits_{\mathbb{R}^1} e^{-ist} \varphi(s) ds \tag{13.4.5}$$

и является ограниченной и непрерывной.

Доказательство. При доказательстве теоремы единственности было введено абсолютно непрерывное распределение

$$\mathsf{F}_\sigma = \mathsf{F} * \mathsf{N}_{0;\sigma^2},$$

и установлено, что, во-первых, при $\sigma \to 0$

$$\mathsf{F}_\sigma \to \mathsf{F},$$

а, во-вторых, плотность распределения F_σ

$$f_\sigma(t) = \frac{1}{2\pi} \int\limits_{\mathbb{R}^1} e^{-ist} \varphi(s) \exp\left\{-\frac{s^2\sigma^2}{2}\right\} ds. \tag{13.4.6}$$

Мы докажем, что при $\sigma \to 0$ плотность $f_\sigma(t)$ имеет предел

$$\lim_{\sigma \to 0} f_\sigma(t) dt = f(t),$$

и предельная функция $f(t)$ является плотностью распределения F.

Чтобы установить существование $\lim\limits_{\sigma \to 0} f_\sigma(t) dt$ достаточно заметить, что подынтегральная функция в правой части равенства (13.4.6) мажорируема интегрируемой по мере Лебега функцией $|\varphi(s)|$:

$$\left| e^{-ist} \varphi(s) \exp\left\{-\frac{s^2\sigma^2}{2}\right\} \right| \leq |\varphi(s)|,$$

при $\sigma \to 0$ сходится

$$e^{-ist} \varphi(s) \exp\left\{-\frac{s^2\sigma^2}{2}\right\} \to e^{-ist} \varphi(s),$$

и воспользоваться теоремой Лебега о мажорируемой сходимости:

$$\lim_{\sigma\to 0} f_\sigma(t)dt = \lim_n \frac{1}{2\pi}\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\exp\left\{-\frac{s^2\sigma^2}{2}\right\}ds =$$

$$= \frac{1}{2\pi}\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)ds = f(t). \qquad (13.4.7)$$

Далее, поскольку $\mathsf{F}_\sigma \to \mathsf{F}$, то для каждого промежутка непрерывности $[a,b)$ распределения F

$$\mathsf{F}([a,b)) = \lim_{\sigma\to 0}\mathsf{F}_\sigma([a,b)),$$

а, учитывая, что $f_\sigma(t)$ — плотность распределения F_σ

$$\mathsf{F}([a,b)) = \lim_{\sigma\to 0}\int\limits_{[a,b)} f_\sigma(t)dt. \qquad (13.4.8)$$

В правой части (13.4.8) возможен предельный переход под знаком интеграла при $\sigma \to 0$. Достаточно заметить, что подынтегральная функция

$$f_\sigma(t) = \frac{1}{2\pi}\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\exp\left\{-\frac{s^2\sigma^2}{2}\right\}ds.$$

при $\sigma \to 0$ имеет предел (см. (13.4.7)) и ограничена константой:

$$|f_\sigma(t)| \le \frac{1}{2\pi}\int\limits_{\mathbb{R}^1}\left|e^{-ist}\varphi(s)\exp\left\{-\frac{s^2\sigma^2}{2}\right\}\right|ds \le \int\limits_{\mathbb{R}^1}|\varphi(s)|\,ds = C < \infty$$

— интегрируемой по Лебегу функцией на каждом конечном промежутке, и воспользоваться теоремой Лебега о мажорируемой сходимости:

$$\mathsf{F}([a,b)) = \lim_{\sigma\to 0}\int\limits_{[a,b)} f_\sigma(t)dt = \int\limits_{[a,b)} f(t)dt.$$

Последнее равенство обозначает, что функция

$$f(t) = \frac{1}{2\pi} \int\limits_{\mathbb{R}^1} e^{-ist} \varphi(s) ds$$

является плотностью распределения F.

Осталось заметить, что плотность

$$f(t) = \frac{1}{2\pi} \int\limits_{\mathbb{R}^1} e^{-ist} \varphi(s) ds$$

ограничена (это следует из интегрируемости $\varphi(s)$) и непрерывна, последнее следует из неравенств

$$|f(t+h) - f(t)| \leq \frac{1}{2\pi} \int\limits_{\mathbb{R}^1} \left| e^{-is(t+h)} \varphi(s) - e^{-ist} \varphi(s) \right| ds \leq$$

$$\leq \frac{1}{2\pi} \int\limits_{\mathbb{R}^1} \left| e^{-ish} - 1 \right| |\varphi(s)| ds$$

и теоремы Лебега о мажорируемой сходимости.

Пример 13.4.3 (характеристическая функция распределения Коши). *Вычислить характеристическую функцию распределения* $\mathsf{C}_{a,0}$ *— распределения Коши с параметрами* $(a, 0)$, *его плотность:*

$$p(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}, \; x \in \mathbb{R}^1.$$

Решение. Сначала найдем характеристическую функцию двустороннего показательного распределения, с плотностью $f(t) = \frac{a}{2} \exp\{-a|t|\}$:

$$\varphi(s) = \int\limits_{-\infty}^{+\infty} e^{ist} f(t) dt = \int\limits_{-\infty}^{+\infty} e^{ist} \frac{a}{2} e^{-a|t|} dt = \frac{a^2}{a^2 + s^2}.$$

Характеристическая функция

$$\varphi(s) = \frac{a^2}{a^2 + s^2}$$

интегрируема по мере Лебега

$$\int\limits_{\mathbb{R}^1} |\varphi(s)| ds < \infty.$$

Поэтому в силу теоремы обращения для плотности имеем:

$$\frac{a}{2} e^{-a|t|} = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} e^{-its} \frac{a^2}{a^2+s^2} ds,$$

что можно переписать так:

$$e^{-a|t|} = \int\limits_{-\infty}^{+\infty} e^{its} \frac{1}{\pi} \frac{a}{a^2+s^2} ds.$$

Остается заметить, что последний интеграл — не что иное, как характеристическая функция распределения Коши с параметрами $(a, 0)$.

Следствие. *Характеристическая функция распределения Коши с параметрами* (a, b) (*распределения* $\mathsf{C}_{a,b}$)

$$\varphi(t) = e^{-a|t|+itb}.$$

Достаточно заметить, что если случайная величина ξ имеет распределение $\mathsf{C}_{a,0}$, то случайная величина $\eta = \xi + b$ имеет распределение $\mathsf{C}_{a,b}$, в последнем убеждаемся непосредственной проверкой .

Пример 13.4.4 (свертка распределений Коши). *Пусть* F *и* Q — *распределения Коши с соответственно с параметрами* $(a_1, 0)$ *и* $(a_2, 0)$. *Найти свертку* $\mathsf{F} * \mathsf{Q}$ *распределений* Q *и* F.

Решение. Характеристические функции распределений Коши с параметрами $(a_1, 0)$ и $(a_2, 0)$ соответственно равны

$$\varphi_1(t) = e^{-a_1|t|} \text{ и } \varphi_2(t) = e^{-a_2|t|}$$

(см. пример 13.4.3), а характеристическая функция $\varphi(t)$ свертки $\mathsf{F} * \mathsf{Q}$ в силу мультипликативного свойства характеристических функций равна

$$\varphi(t) = \varphi_1(t)\varphi_2(t) = e^{-(a_1+a_2)|t|}.$$

Но $e^{-(a_1+a_2)|t|}$ — характеристическая функция распределения Коши с параметрами $(a_1 + a_2, 0)$. Поэтому в силу теоремы единственности распределение $\mathsf{F} * \mathsf{Q}$ совпадает с распределением Коши с параметрами $(a_1 + a_2, 0)$.

Следовательно, *свертка распределений Коши с параметрами* $(a_1, 0)$ *и* $(a_2, 0)$ *является распределением Коши с параметрами* $(a_1 + a_2, 0)$.

Аналогично получаем, что

$$\mathsf{C}_{a_1,b_1} * \mathsf{C}_{a_2,b_2} = \mathsf{C}_{a_1+a_2,\, b_1+b_2}.$$

Так что получен следующий результат: *класс распределений Коши замкнут относительно операции свертки.*

Теорема 13.4.3 (о сохранении вида распределения).

1° *Сумма независимых нормально распределенных случайных величин с параметрами* (a_1, σ_1^2) *и* (a_2, σ_2^2) *является нормально распределенной случайной величиной с параметрами* $(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

В терминах сверток — свертка нормальных распределений с параметрами (a_1, σ_1^2) *и* (a_2, σ_2^2) *является нормальным распределением с параметрами* $(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$:

$$\mathsf{N}_{a_1,\sigma_1^2} * \mathsf{N}_{a_2,\sigma_2^2} = \mathsf{N}_{a_1+a_2,\, \sigma_1^2+\sigma_2^2}.$$

2° *Сумма независимых случайных величин, имеющих распределения Коши с параметрами* (a_1, b_1) *и* (a_2, b_2), *имеет распределение Коши с параметрами* $(a_1 + a_2, b_1 + b_2)$.

В терминах сверток — свертка распределений Коши с параметрами (a_1, b_1) *и* (a_2, b_2) *является распределением Коши с параметрами* $(a_1 + a_2, b_1 + b_2)$:

$$\mathsf{C}_{a_1,b_1} * \mathsf{C}_{a_2,b_2} = \mathsf{C}_{a_1+a_2,\, b_1+b_2}.$$

3° *Сумма независимых случайных величин, имеющих гамма-распределения с параметрами* $(\nu_1; \theta)$ *и* (ν_2, θ), *имеет гамма-распределение с параметрами* $(\nu_1 + \nu_2, \theta)$.

В терминах сверток — свертка гамма-распределений с параметрами (ν_1, θ) *и* (ν_2, θ) *является гамма-распределением с параметрами* $(\nu_1 + \nu_2, \theta)$:

$$\mathsf{G}_{\nu_1,\theta} * \mathsf{G}_{\nu_2,\theta} = \mathsf{G}_{\nu_1+\nu_2,\, \theta}.$$

4° *Сумма независимых пуассоновских случайных величин соответственно с параметрами* θ_1 *и* θ_2 *является пуассоновской случайной величиной с параметром* $\theta_1 + \theta_2$.

В терминах сверток — свертка пуассоновских распределений с параметрами θ_1 и θ_2 является пуассоновским распределением с параметром $\theta_1 + \theta_2$:

$$\mathsf{P}_{\theta_1} * \mathsf{P}_{\theta_2} = \mathsf{P}_{\theta_1+\theta_2}.$$

Доказательство. Доказательства пунктов 1°, 2° и 4° см. соответственно в примерах 13.4.1, 13.4.4 и 13.4.2, доказательство пункта 3° — см. теорему 11.1.5.

13.5 Теорема Леви

Теорема 13.5.1 (Леви). *Сходимость последовательности $\varphi_n(t)$ характеристических функций распределений F_n к непрерывной функции $\varphi(t)$ в каждой точке $t \in \mathbb{R}^1$ влечет собственную сходимость последовательности распределений F_n и непрерывный предел $\varphi(t)$ последовательности характеристических функций $\varphi_n(t)$ является характеристической функцией предельного распределения F последовательности F_n.*

Справедливо и обратное. Собственная сходимость последовательности вероятностных распределений F_n влечет сходимость последовательности их характеристических функций $\varphi_n(t)$ к характеристической функции $\varphi(t)$ предельного распределения F причем сходимость $\varphi_n(t) \to \varphi(t)$ равномерная на каждом конечном промежутке.

Доказательство. Сначала докажем, что из собственной сходимости $\mathsf{F}_n \to \mathsf{F}$ последовательности распределений F_n следует сходимость последовательности $\varphi_n(t)$ характеристических функций распределений F_n к характеристической функции $\varphi(t)$ предельного распределения F.

Поскольку для каждого t функции $u_t = u_t(x) = e^{itx}$, $x \in \mathbb{R}^1$, принадлежат классу $C(-\infty, +\infty)$, то в силу теоремы о слабой сходимости распределений (см. теорему 12.3.1) из собственной сходимости распределений F_n к распределению F для каждого $t \in \mathbb{R}^1$ следует сходимость их характеристических функций

$$\varphi_n(t) = \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}_n(dx) \to \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}(dx) = \varphi(t),$$

а чтобы установить, что эта сходимость равномерная на ограниченных промежутках, согласно следствию из теоремы о слабой сходимости распределений (теорема 12.3.1) достаточно проверить, что семейство функций

$$u_t = u_t(x) = e^{itx},\ x \in \mathbb{R}^1,\ |t| \le a < \infty,$$

зависящих от параметра t, равномерно ограничено и равностепенно непрерывно. Равномерная ограниченность этого семейства очевидна, а равностепенная непрерывность следует из неравенств

$$\begin{aligned}|u_t(x+h) - u_t(x)| &= |e^{it(x+h)} - e^{itx}| = \\ &= |e^{itx}||e^{ith} - 1| \le |ith| = |t||h| \le a|h|.\end{aligned}$$

Покажем теперь, что из сходимости последовательности

$$\varphi_n(t) = \int\limits_{\mathbb{R}^1} e^{itx} \mathsf{F}_n(dx)$$

характеристических функций распределений F_n к некоторой непрерывной функции $\varphi(t)$ следует собственная сходимость последовательности вероятностных распределений $\{\mathsf{F}_n\}$.

В силу теоремы Хелли о выборе из последовательности распределений $\{\mathsf{F}_n\}$ можно выбрать подпоследовательность $\{\mathsf{F}_{n_k}\}$, сходящуюся к некоторому вероятностному распределению F (в собственном или несобственном смысле). Далее покажем, что

1° сходимость подпоследовательности $\{\mathsf{F}_{n_k}\}$ к F собственная (т. е. F — собственное вероятностное распределение) и

2° не только подпоследовательность $\{\mathsf{F}_{n_k}\}$, но и сама последовательность $\{\mathsf{F}_n\}$ сходится к F.

Для доказательства собственной сходимости $\mathsf{F}_{n_k} \to \mathsf{F}$ воспользуемся равенством Парсеваля:

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{G}(ds) = \int\limits_{\mathbb{R}^1} \gamma(x-t)\mathsf{F}(dx)$$

(см. (13.5.1)), рассмотрев в качестве распределения F распределение F_{n_k} (с характеристической функцией $\varphi_{n_k}(s)$), а в качестве распределения G — нормальное распределение $\mathsf{N}_{0;\sigma^2}$ (его характеристическая функция $\gamma(s) = \exp\left\{-s^2\sigma^2/2\right\}$):

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi_{n_k}(s)\mathsf{N}_{0;\sigma^2}(ds) = \int\limits_{\mathbb{R}^1} \exp\left\{-\frac{(t-x)^2\sigma^2}{2}\right\} \mathsf{F}_{n_k}(dx). \tag{13.5.2}$$

Убедимся, что в последнем равенстве можно перейти к пределу при $n_k \to \infty$, причем под знаком интеграла.

Для каждого фиксированного t подынтегральная функция $e^{-ist}\varphi_{n_k}(s)$ в левой части равенства (13.5.2) сходится (при $n_k \to \to \infty$) и мажорируема интегрируемой функцией $u(s) = 1$ ($u(s) = 1$ интегрируема по распределению $\mathsf{N}_{0;\sigma^2}$). Поэтому в силу теоремы Лебега о мажорируемой сходимости существует предел левой части равенства (13.5.2), а следовательно существует предел и правой части равенства. И в силу теоремы о слабой сходимости распределений возможен предельный переход под знаком интеграла (сходимость $\mathsf{F}_{n_k} \to \mathsf{F}$ эквивалентна слабой сходимости F_{n_k} к F относительно класса $C_0[-\infty, +\infty]$, а функция $\exp\left\{-\dfrac{(t-x)^2\sigma^2}{2}\right\}$ при каждом фиксированном t принадлежит $C_0[-\infty, +\infty]$). В результате предельного перехода при $n_k \to \infty$ имеем:

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{N}_{0;\sigma^2}(ds) = \int\limits_{\mathbb{R}^1} \exp\left\{-\frac{(t-x)^2\sigma^2}{2}\right\}\mathsf{F}(dx) \leq$$

$$\leq \int\limits_{\mathbb{R}^1} \mathsf{F}(dx) = \mathsf{F}(\mathbb{R}^1).$$

Так что

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{N}_{0;\sigma^2}(ds) \leq \mathsf{F}(\mathbb{R}^1). \tag{13.5.3}$$

Далее, при $\sigma \to 0$ распределение $\mathsf{N}_{0;\sigma^2}$ сходится в собственном смысле к атомическому распределению $\mathsf{N}_{0;0}$, сосредоточенному в точке 0, а функция $e^{-ist}\varphi(s)$ принадлежит классу $C(-\infty, +\infty)$ непрерывных ограниченных функций ($\varphi(s)$ непрерывна по условию, и ограничена: $|\varphi(s)| \leq 1$ — поскольку $\varphi(s) = \lim\limits_n \varphi_n(s)$, а $|\varphi_n(s)| \leq 1$). Поэтому в силу теоремы 12.3.1 существует

$$\lim_{\sigma\to 0} \int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{N}_{0;\sigma^2}(ds)$$

и в неравенстве (13.5.3) возможен предельный переход под знаком интеграла:

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{N}_{0;0}(ds) \le \mathsf{F}(\mathbb{R}^1).$$

Осталось заметить, что

$$\int\limits_{\mathbb{R}^1} e^{-ist}\varphi(s)\mathsf{N}_{0;0}(ds) = \varphi(0) = 1,$$

$\varphi(0) = 1$, как предел последовательности $\varphi_n(0) = 1, n = 1, 2, \ldots$ Так что

$$1 \le \mathsf{F}(\mathbb{R}^1),$$

т. е. F — собственное вероятностное распределение и, следовательно, сходимость $\mathsf{F}_{n_k} \to \mathsf{F}$ собственная.

Далее покажем, что не только $\mathsf{F}_{n_k} \to \mathsf{F}$, но и $\mathsf{F}_n \to \mathsf{F}$. Если бы $\{\mathsf{F}_n\}$ не сходилась к F, то в силу теоремы о выборе Хелли нашлась бы подпоследовательность $\{\mathsf{F}_{n'_l}\}$ последовательности $\{\mathsf{F}_n\}$, сходящаяся в собственном смысле к некоторому распределению $\tilde{\mathsf{F}}$, отличному от F (собственная сходимость устанавливается так же, как и ранее). Но тогда, с одной стороны, в силу теоремы о слабой сходимости распределений

$$\lim_{n_k} \varphi_{n_k}(t) = \lim_{n_k} \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}_{n_k}(dx) = \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}(dx);$$

$$\lim_{n'_l} \varphi_{n'_l}(t) = \lim_{n'_l} \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}_{n'_l}(dx) = \int\limits_{\mathbb{R}^1} e^{itx}\tilde{\mathsf{F}}(dx).$$

С другой стороны, $\{\varphi_{n_k}(t)\}$ и $\{\varphi_{n'_l}(t)\}$ как подпоследовательности последовательности $\{\varphi_n(t)\}$, сходящейся к $\varphi(t)$, сходятся к $\varphi(t)$. Поэтому их пределы совпадают и равны пределу $\varphi(t)$ последовательности $\varphi_n(t)$:

$$\int\limits_{\mathbb{R}^1} e^{itx}\tilde{\mathsf{F}}(dx) = \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}(dx) = \varphi(t), t \in \mathbb{R}^1, \tag{13.5.4}$$

т. е. совпадают характеристические функции распределений $\tilde{\mathsf{F}}$ и F. А следовательно, в силу теоремы единственности, совпадают и сами распределения: $\mathsf{F} = \tilde{\mathsf{F}}$, что противоречит предположению $\mathsf{F} \neq \tilde{\mathsf{F}}$. Это противоречие доказывает сходимость последовательности $\{\mathsf{F}_n\}$ к F.

Равенство
$$\int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}(dx) = \varphi(t),$$
(см. (13.5.4)) означает, что предельная функция $\varphi(t)$ является характеристической функцией предельного распределения F.

Тем самым теорема доказана.

Теорема 13.5.2 (пуассоновское распределение в схеме серий). *Пусть* $\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,k}, \ldots, \xi_{n,n}$ *— независимые случайные величины, каждая с распределением*
$$\begin{pmatrix} 1 & 0 \\ p_n & 1-p_n \end{pmatrix},$$
$$S_n = \sum_{k=1}^{n} \xi_{n,k}.$$
Если при $n \to \infty$ *значение* $p_n \to 0$ *так, что*
$$np_n = \mathsf{M}S_n \to \lambda > 0,$$
то распределение суммы S_n *сходится к пуассоновскому распределению с параметром* λ*:*
$$\mathsf{P}\{S_n = m\} \to \frac{\lambda^m}{m!}e^{-\lambda},\ m = 0, 1, \ldots$$

Доказательство. Исследуем поведение характеристической функции случайной величины S_n при $n \to \infty$:
$$\psi_n(t) = \mathsf{M}\exp\{itS_n\} = \mathsf{M}\exp\left\{it\sum_{k=1}^{n}\xi_{n,k}\right\} =$$
$$= \prod_{k=1}^{n}\mathsf{M}\exp\{it\xi_{n,k}\} = \prod_{k=1}^{n}\big(e^{it}p_n + (1-p_n)\big) = \big(1 + p_n(e^{it}-1)\big)^n.$$

При $n \to \infty$ значение $p_n \to 0$ так, что $np_n \to \lambda$, поэтому для каждого фиксированного t

$$\ln \psi_n(t) = n \ln\left(1 + p_n\left(e^{it} - 1\right)\right) \sim np_n\left(e^{it} - 1\right) \to \lambda\left(e^{it} - 1\right),$$

отсюда

$$\psi_n(t) \to \exp\left\{\lambda\left(e^{it} - 1\right)\right\}$$

— характеристическая функция $\psi_n(t)$ случайной величины S_n сходится к характеристической функции $\exp\left\{\lambda\left(e^{it} - 1\right)\right.$ пуассоновского распределения с параметром λ. Поэтому в силу теоремы Леви распределение суммы $S_n = \sum\limits_{k=1}^{n} \xi_{n,k}$ сходится к распределению Пуассона с параметром λ.

13.6 Характеристическая функция смеси распределений

Смесь распределений. Пусть $\mathsf{F}_0, \mathsf{F}_1, \ldots$ — собственные вероятностные распределения на $\mathbb{R}^1$, $\{p_k\}$ — собственное вероятностное распределение на $\{0, 1, 2, \ldots\}$ $\left(\sum\limits_{k=0}^{\infty} p_k = 1\right)$. И пусть

$$Q_n(x) = \sum_{k=0}^{n} p_k F_k(x),\ x \in \mathbb{R}^1,$$

$$Q(x) = \sum_{k=0}^{\infty} p_k F_k(x) = \lim_n \sum_{k=0}^{n} p_k F_k(x) = \lim_n Q_n(x),\ x \in \mathbb{R}^1,$$

$Q(x)$ — собственная вероятностная функция распределения (в справедливости последнего убеждаемся непосредственной проверкой), Q и Q_n — распределения, соответствующие функциям распределения $Q(x)$ и $Q_n(x)$.

Поскольку для любого фиксированного x

$$Q(x) = \lim_n Q_n(x),\ x \in \mathbb{R}^1,$$

то в силу достаточного условия сходимости распределений

$$\mathsf{Q} = \lim_n \mathsf{Q}_n,$$

причем если $\mathsf{F}_0, \mathsf{F}_1, \ldots$ — собственные вероятностные распределения, то сходимость собственная.

Определение. Распределение Q, определенное равенством

$$\mathsf{Q} = \lim_n \mathsf{Q}_n = \lim_n \sum_{k=0}^{n} p_k \mathsf{F}_k, \tag{13.6.1}$$

называется *смесью распределений* $\mathsf{F}_k, k = 0, 1, \ldots$, и обозначается

$$\mathsf{Q} = \sum_{k=0}^{\infty} p_k \mathsf{F}_k.$$

Теорема 13.6.1 (о характеристической функции смеси). *Пусть* $\mathsf{F}_0, \mathsf{F}_1, \ldots$ *— собственные вероятностные распределения на* $\mathbb{R}^1$, $\varphi_0, \varphi_1, \ldots$ *— их характеристические функции,* $\{p_k\}$ *— собственное вероятностное распределение на* $\{0, 1, 2, \ldots\}$. *Смесь*

$$\mathsf{Q} = \sum_{k=0}^{\infty} p_k \mathsf{F}_k$$

распределений $\mathsf{F}_0, \mathsf{F}_1, \ldots$ *имеет своей характеристической функцией*

$$\omega(t) = \sum_{k=0}^{\infty} p_k \varphi_k(t).$$

Доказательство. По определению смеси Q распределений $\mathsf{F}_0, \mathsf{F}_1, \ldots$

$$\mathsf{Q} = \lim_n \sum_{k=0}^{n} p_k \mathsf{F}_k = \lim_n \mathsf{Q}_n,$$

и поскольку распределения $\mathsf{F}_0, \mathsf{F}_1, \ldots$ собственные, то сходимость

$$\mathsf{Q}_n \to \mathsf{Q}$$

собственная. Поэтому в силу теоремы о слабой сходимости распределений (см. теорему 12.3.1) имеем

$$\omega(t) = \int\limits_{\mathbb{R}^1} e^{itx} \mathsf{Q}(dx) = \lim_n \int\limits_{\mathbb{R}^1} e^{itx} \mathsf{Q}_n(dx) =$$

$$= \lim_n \sum_{k=0}^{n} p_k \int\limits_{\mathbb{R}^1} e^{itx} F_k(dx) = \lim_n \sum_{k=0}^{n} p_k \varphi_k(t) = \sum_{k=0}^{\infty} p_k \varphi_k(t).$$

Сумма случайного числа слагаемых. Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины, $\xi_0 = 0$,

$$S_k = \xi_0 + \xi_1 + \ldots + \xi_k,\ k = 0, 1, 2, \ldots,$$

ν — неотрицательная целочисленная случайная величина, причем случайная величина ν и случайные величины $\xi_1, \xi_2, \ldots$ независимые.

Случайную величину S_ν, определенную равенством

$$S_\nu = \sum_{k=0}^{\infty} S_k I_{\{\nu = k\}},$$

называют *суммой случайного числа случайных величин*, ее еще обозначают так:

$$S_\nu = \sum_{k=0}^{\nu} \xi_k.$$

Теорема 13.6.2 (о распределении суммы случайного числа слагаемых). *Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины, каждая с распределением F и характеристической функцией φ, ν — неотрицательная целочисленная случайная величина с распределением*

$$\mathsf{P}\{\nu = k\} = p_k,\ k = 0, 1, \ldots,$$

не зависящая от случайных величин $\xi_1, \xi_2, \ldots$ Тогда распределение Q суммы

$$S_\nu = \sum_{k=0}^{\nu} \xi_k,\ \xi_0 = 0,$$

равно смеси распределений $\mathsf{F}^{0*}, \mathsf{F}^{1*}, \mathsf{F}^{2*}, \ldots$, а именно

$$\mathsf{Q} = \sum_{k=0}^{\infty} p_k \mathsf{F}^{k*}$$

(F^{0*} — *атомическое распределение, сосредоточенное в точке* 0, $\mathsf{F}^{1*} = \mathsf{F}$, $\mathsf{F}^{k*} = \mathsf{F}^{1*} * \mathsf{F}^{(k-1)*}, k = 2, 3, \ldots$), *а характеристическая функция суммы* S_ν

$$\omega(t) = \sum_{k=0}^{\infty} p_k \varphi^k(t) = \mathsf{P}(\varphi(t)),$$

где $P(s) = \sum\limits_{k=0}^{\infty} p_k s^k$.

Доказательство. Убедимся, что характеристическая функция смеси

$$\mathsf{Q} = \sum_{k=0}^{\infty} p_k \mathsf{F}^{k*}$$

и характеристическая функция случайной величины S_ν совпадают.

Характеристическая функция смеси

$$\mathsf{Q} = \sum_{k=0}^{\infty} p_k \mathsf{F}^{k*}$$

распределений $\mathsf{F}^{k*}, k = 0, 1, 2, \ldots$, в силу теорему 13.6.1 равна

$$\omega(t) = \sum_{k=0}^{\infty} p_k \varphi^k(t) = P(\varphi(t))$$

(характеристическая функция F^{k*} равна $\varphi^k(t)$ в силу мультипликативного свойства характеристических функций).

Характеристическую функцию случайной величины S_ν вычислим, воспользовавшись свойством счетной аддитивности математического ожидания и мультипликативным свойством характеристических функций:

$$\mathsf{M}e^{itS_\nu} = \sum_{k=0}^{\infty} \mathsf{M}e^{itS_\nu} I_{\{\nu=k\}} = \sum_{k=0}^{\infty} \mathsf{M}e^{itS_k} I_{\{\nu=k\}} =$$

$$= \sum_{k=0}^{\infty} \mathsf{M}e^{itS_k} \mathsf{M}I_{\{\nu=k\}} = \sum_{k=0}^{\infty} \mathsf{M}e^{itS_k} \mathsf{P}\{\nu = k\} =$$

$$= \sum_{k=0}^{\infty} p_k \mathsf{M} e^{itS_k} = \sum_{k=0}^{\infty} p_k \varphi^k(t) = P(\varphi(t))$$

($\mathsf{M}e^{itS_k} = \varphi^k(t)$ — как характеристическая функция суммы S_k независимых случайных величин).

Характеристическая функция $\omega(t)$ смеси

$$\mathsf{Q} = \sum_{k=0}^{\infty} p_k \mathsf{F}^{k*}$$

совпадает с характеристической функцией случайной величины S_ν, поэтому в силу теоремы единственности распределение случайной величины S_ν совпадает с распределением Q.

13.7 Примеры и задачи

Примеры

Пример 13.7.1 . *Пусть $\eta = \xi + a$ (a — константа), $\varphi(t)$ — характеристическая функция случайной величины ξ. Вычислить характеристическую функцию η.*

Решение. Характеристическая функция константы a равна e^{ita}, а $\varphi_\eta(t) = e^{ita}\varphi(t)$, т. е. сомножитель e^{ita} “отвечает” за сдвиг ξ на величину a.

Пример 13.7.2. *Вычислить характеристическую функцию случайной величины ξ, имеющей биномиальное распределение с параметрами (n, p):*

$$\mathsf{P}\{\xi = k\} = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

Решение. Случайную величину ξ, имеющую биномиальное распределение с параметрами (n, p), можно представить в виде суммы

$$\xi = \sum_{k=1}^{n} \mu_k$$

независимых случайных величин $\mu_1, \mu_2, \ldots, \mu_n$, каждая с распределением

$$\mathsf{P}\{\mu_k = s\} = p^s (1-p)^{1-s},\ s = 0, 1;\ k = 1, 2, \ldots, n,$$

и, следовательно, каждая с характеристической функцией

$$\varphi(t) = \mathsf{M}e^{it\mu_k} = e^{it\cdot 1}p + e^{it\cdot 0}(1-p) = pe^{it} + 1 - p = 1 + p(e^{it} - 1),$$

$k = 1, 2, \ldots, n$. Отсюда, пользуясь мультипликативным свойством характеристических функций, получаем:

$$\varphi_\xi(t) = \prod_{k=1}^{n} \varphi(t) = \left(1 + p(e^{it} - 1)\right)^n.$$

Задачи

13.1. Вычислить характеристическую функцию случайной величины ξ, имеющей геометрическое распределение с параметром p:

$$\mathsf{P}\{\xi = k\} = p(1-p)^k,\ k = 0, 1, \ldots$$

13.2. Вычислить характеристическую функцию случайной величины, равномерно распределенной на $[-a; a]$.

13.3. Пусть ξ_{n,p_n} — биномиально распределенная с параметрами (n, p_n) случайная величина. Доказать, что если при $n \to \infty$, значение $p_n \to 0$, так, что среднее $np_n \to \lambda$, то распределение случайной величины ξ_{n,p_n} сходится к пуассоновскому распределению с параметром λ (см. также теорему 13.5.2).

13.4. В теореме 13.5.2 исследовать поведение распределения суммы S_n, если при $n \to \infty$ значение $np_n \to 0$.

13.5. Пусть случайная величина ξ_λ имеет распределение Пуассона с параметром λ.

Доказать, что при $\lambda \to \infty$

$$\mathsf{P}\left\{\frac{\xi_\lambda - \lambda}{\sqrt{\lambda}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt.$$

13.6. В каждой точке $x_j = j/n, j = 0, 1, \ldots, n$, отрезка $[0; 1]$ сосредоточена масса $1/(n+1)$. Обозначим через F_n так определенное вероятностное распределение.

1° Найти характеристическую функцию $\varphi_n(t)$ распределения F_n.

2° Вычислить предел $\varphi_n(t)$ при $n \to \infty$.

3° Доказать, что при $n \to \infty$ последовательность распределений $\{\mathsf{F}_n\}$ сходится к равномерному на $[0; 1]$ распределению.

13.7. Доказать, что атомическое распределение

$$\mathsf{D} = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

можно представить в виде предела абсолютно непрерывных распределений.

Указание. Достаточно установить, что характеристическая функция распределения $\mathsf{F}_\sigma = \mathsf{D} * \mathsf{N}_{0,\sigma^2}$ сходится к характеристической функции распределения D.

13.8. Пусть P_θ — пуассоновское распределение с параметром θ.

Исследовать P_θ на сходимость при $\theta \to 0$.

13.9. Пусть G_p — геометрическое распределение с параметром p.

Исследовать G_p на сходимость при $p \to 1$.

13.10 (центральная предельная теорема). Пусть $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин со средним 0 и конечной дисперсией $\mathsf{D}\xi_k = \sigma^2$. Доказать, что при $n \to \infty$, распределение случайной величины

$$S_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} \xi_k$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Указание. Обозначим через $\varphi(t)$ характеристическую функцию случайной величины ξ_k $(k = 1, 2, \ldots)$, а через $\psi_n(t)$ — случайной величины $\frac{1}{\sigma\sqrt{n}} \sum\limits_{k=1}^{n} \xi_k$.

Последовательно решите такие задачи:

1° Выписать характеристическую функцию $\psi_n(t)$ через характеристическую функцию $\varphi(t)$.

2° Выписать разложение Тейлора для $\varphi(t)$ в окрестности точки $t = 0$, содержащее t^2.

3° Воспользовавшись результатами пунктов 1° и 2°, найти предел характеристической функции $\psi_n(t)$, при $n \to \infty$.

4° Воспользовавшись результатом пункта 3° и теоремой Леви, доказать, что при $n \to \infty$, распределение случайной величины $\frac{1}{\sigma\sqrt{n}} \sum\limits_{k=1}^{n} \xi_k$ сходится к нормальному распределению с параметрами $(0; 1)$.

13.11. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые одинаково распределенные случайные величины со значениями в $\mathbb{R}^1$ (каждая с распределением F),

$$\mathbb{R}^1 = \bigcup_{i=1}^{r} X_i,\ X_i \cap X_j = \emptyset,\ i \neq j,$$

ν_i — число случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, попавших в X_i, $p_i = \mathsf{F}(X_i) = \mathsf{P}\{\xi_k \in X_i\}$ — вероятность, того, что случайная величина ξ_k попадет в X_i, $i = 1, 2, \ldots, r$.

Доказать, что при $n \to \infty$, распределение

$$\frac{\nu_i - np_i}{\sqrt{np_i(1-p_i)}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

13.12. Пусть $\{\xi_k\}$ — последовательность независимых одинаково распределенных случайных величин.

Если

$$\mathsf{D}\xi_k = \sigma^2 < \infty,$$

то из неравенства Чебышёва

$$\mathsf{P}\{|\zeta - \mathsf{M}\zeta| \geq \varepsilon\} \leq \mathsf{D}\zeta/\varepsilon^2$$

следует, что при $n \to \infty$, последовательность

$$S_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$$

сходится по вероятности к $a = \mathsf{M}\xi_1$, т. е. для любого $\varepsilon > 0$

$$\mathsf{P}\{|S_n - a| \geq \varepsilon\} \to 0.$$

Теперь не будем требовать, чтобы $\sigma^2 < \infty$, но пусть

$$\mathsf{M}\xi_k = a \neq \infty.$$

Доказать, что при $n \to \infty$ последовательность S_n сходится по вероятности к a (закон больших чисел в форме Хинчина).

Глава 14

Закон больших чисел и его приложения

14.1 Закон больших чисел

Закон больших чисел в форме Чебышёва. В приложениях случайная величина обычно возникает как результат измерений (наблюдений). Любые измерения неизбежно сопровождаются погрешностями (измерений без погрешностей не бывает), при этом объекты измерений зачастую выбираются случайным образом, а чаще всего имеет место и то, и другое. Давно было замечено, что в то время как результаты $\xi_1, \xi_2, \ldots, \xi_n$ отдельных измерений (наблюдений) ведут себя довольно "нерегулярно", их среднее арифметическое $(\xi_1+\xi_2+\ldots+\xi_n)/n$ сравнительно устойчиво. Этот экспериментальный факт, в частности, наблюдается при изучении частоты

$$\nu_n(A) = \frac{1}{n}k_n(A) = \frac{1}{n}\left(I_A^{(1)} + I_A^{(2)} + \cdots + I_A^{(n)}\right)$$

события в последовательности экспериментов ($I_A^{(k)}$ — индикатор события A в k-м эксперименте) — если $I_A^{(k)}$ ведут себя "нерегулярно" — принимают значения 1 или 0, то частота $\nu_n(A)$ ведет себя устойчиво — заметные уклонения $\nu_n(A)$ от некоторого числа наблюдаются изредка.

Поэтому естественно ожидать существования результатов, соответствующих эмпирическому факту устойчивости средних. Одним из таких результатов является закон больших чисел.

Теорема 14.1.1 (закон больших чисел в форме Чебышева). *Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины с дисперсиями, ограниченными одной и той же константой C. Тогда для любого $\varepsilon > 0$ при $n \to \infty$*

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}\mathsf{M}\xi_i\right| > \varepsilon\right\} \to 0.$$

Доказательство повторяет доказательство теоремы 5.4.1.

Следствие. *Пусть $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с математическим ожиданием $\mathsf{M}\xi_i = a$ и конечной дисперсией ($\mathsf{D}\xi_i = \sigma^2 < \infty$). Тогда для любого $\varepsilon > 0$ при $n \to \infty$*

$$\mathsf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| > \varepsilon\right\} \to 0.$$

Доказательство. Достаточно заметить, что для случайных величин $\xi_1, \xi_2, \ldots$ выполняются условия теоремы и к тому же

$$\frac{1}{n}\sum_{i=1}^{n}\mathsf{M}\xi_i = \frac{1}{n}\sum_{i=1}^{n}a = a.$$

Закон больших чисел в форме Хинчина. Если $\xi_1, \xi_2, \ldots$ — независимые одинаково распределенные случайные величины с конечными математическими ожиданиями, то закон больших чисел имеет место и без предположения об ограниченности дисперсии. Для формулировки теоремы нам понадобится понятие сходимости по вероятности.

Определение. Будем говорить, что последовательность $\{\zeta_n\}$ случайных величин *сходится по вероятности* к случайной величине ζ, если для каждого $\varepsilon > 0$ при $n \to \infty$

$$\mathsf{P}\{\omega : \ |\zeta_n(\omega) - \zeta(\omega)| > \varepsilon\} \to 0.$$

Определение. Будем говорить, что последовательность $\{\zeta_n\}$ случайных величин *сходится по вероятности* к ∞, если для любого $N > 0$ при $n \to \infty$

$$\mathsf{P}\{\omega : \ |\zeta_n(\omega)| > N\} \to 1.$$

Обозначать сходимость по вероятности последовательности $\{\zeta_n\}$ к ζ будем так:

$$\zeta_n \xrightarrow{\mathsf{P}} \zeta.$$

Теорема 14.1.2 (Хинчин). *Для независимых одинаково распределенных случайных величин* ξ_k, $k = 1, 2, \ldots$, *с конечным математическим ожиданием* $\mathsf{M}\xi_k = a$ *при* $n \to \infty$

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{\mathsf{P}} a.$$

Д о к а з а т е л ь с т в о. Обозначим $\sum\limits_{k=1}^{n}\xi_k$ через S_n. Сначала докажем, что при $n \to \infty$ распределение F_n случайной величины S_n/n сходится к собственному атомическому распределению F_0, сосредоточенному в точке a. Для этого воспользуемся теоремой Леви (теорема 13.5.1), согласно которой достаточно доказать, что последовательность характеристических функций $\{\psi_n(t)\}$ случайных величин S_n/n сходится к характеристической функции e^{ita} атомического распределения, сосредоточенного в точке a, для каждого $t \in \mathbb{R}^1$.

Обозначим через $\varphi(t)$ характеристическую функцию случайной величины ξ_k, $k = 1, 2, \ldots$ В силу мультипликативного свойства характеристических функций

$$\psi_n(t) = \mathsf{M}\exp\left\{it\frac{1}{n}S_n\right\} = \mathsf{M}\exp\left\{i\frac{t}{n}\sum_{k=1}^{n}\xi_k\right\} = \left(\varphi\left(\frac{t}{n}\right)\right)^n.$$

Так как $\mathsf{M}\xi_k = a$ конечно, то для $\varphi(s)$ в окрестности нуля имеет место разложение

$$\varphi(s) = 1 + s\varphi'(0) + o(s) = 1 + isa + o(s)$$

(см. следствие из теоремы 13.3.2). Отсюда, положив $s = t/n$, для каждого фиксированного t, при $n \to \infty$ имеем:

$$\varphi\left(\frac{t}{n}\right) = 1 + i\frac{t}{n}a + o\left(\frac{1}{n}\right),$$

а для $\psi_n(t)$

$$\psi_n(t) = \left(1 + i\frac{t}{n}a + o\left(\frac{1}{n}\right)\right)^n.$$

Из последнего равенства при $n \to \infty$ получаем:

$$\ln \psi_n(t) = n \ln \left(1 + i\frac{t}{n}a + o\left(\frac{1}{n}\right)\right) \sim$$

$$\sim n\left(i\frac{t}{n}a + o\left(\frac{1}{n}\right)\right) = ita + o(1),$$

$$\psi_n(t) \to e^{ita}.$$

Поэтому в силу теоремы Леви распределение F_n случайной величины S_n/n при $n \to \infty$ сходится к собственному атомическому распределению F_0, сосредоточенному в точке a.

Далее установим, что из собственной сходимости распределений F_n случайных величин S_n/n к атомическому распределению F_0, сосредоточенному в точке a, следует сходимость S_n/n к a по вероятности.

Собственная сходимость последовательности $\{\mathsf{F}_n\}$ вероятностных распределений к F_0 эквивалентна сходимости в основном последовательности $\{F_n(x)\}$ функций распределений к функции распределения $F_0(x)$ атомического распределения, сосредоточенного в точке a. Поэтому при $n \to \infty$

$$\mathsf{P}\left\{\frac{S_n}{n} < x\right\} = F_n(x) \to F_0(x) = \begin{cases} 0, & \text{если } x \le a \\ 1, & \text{если } x > a. \end{cases} \qquad (14.1.1)$$

Далее, для любого $\varepsilon > 0$ имеем:

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - a\right| \ge \varepsilon\right\} = \mathsf{P}\left(\left\{\frac{S_n}{n} \le a - \varepsilon\right\} \cup \left\{\frac{S_n}{n} \ge a + \varepsilon\right\}\right) =$$

$$= \mathsf{P}\left\{\frac{S_n}{n} \le a - \varepsilon\right\} + \mathsf{P}\left\{\frac{S_n}{n} \ge a + \varepsilon\right\} \le$$

$$\le \mathsf{P}\left\{\frac{S_n}{n} < a - \frac{\varepsilon}{2}\right\} + \left(1 - \mathsf{P}\left\{\frac{S_n}{n} < a + \varepsilon\right\}\right) =$$

$$= F_n(a - \varepsilon/2) + (1 - F_n(a + \varepsilon)).$$

Отсюда, учитывая, что при $n \to \infty$

$$F_n(a - \varepsilon/2) \to F_0(a - \varepsilon/2) = 0,$$

$$1 - F_n(a + \varepsilon) \to 1 - F_0(a + \varepsilon) = 0$$

(см. (14.1.1)), получаем, что при $n \to \infty$

$$\mathsf{P}\left\{\left|\frac{S_n}{n} - a\right| \geq \varepsilon\right\} \to 0.$$

Метод Монте-Карло вычисления интегралов. Этот метод еще называют методом статистических испытаний.

Теорема. *Пусть $\xi_1, \xi_2, \ldots$ — независимые случайные величины, равномерно распределенные на $[0;1]$, $g(x)$ — функция на $[0;1]$ со значениями в $\mathbb{R}^1$, такая, что интеграл $\int_0^1 g(x)dx$ конечен, тогда при $n \to \infty$*

$$\frac{1}{n}\sum_{i=1}^{n} g(\xi_i) \xrightarrow{\mathsf{P}} \int_0^1 g(x)dx.$$

Доказательство. Вместе с последовательностью $\xi_1, \xi_2, \ldots$ рассмотрим последовательность независимых одинаково распределенных случайных величин $g(\xi_1), g(\xi_2), \ldots$ Для $g(\xi_i)$ имеем:

$$\mathsf{M}g(\xi_i) = \int_{\mathbb{R}^1} g(x)p_{\xi_i}(x)dx = \int_0^1 g(x)dx, \ i = 1, 2, \ldots$$

В силу теоремы Хинчина при $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} g(\xi_i) \xrightarrow{\mathsf{P}} \mathsf{M}g(\xi_k) = \int_0^1 g(x)dx.$$

Последнее утверждение означает, что при больших n значительные уклонения величины $\frac{1}{n}\sum_{i=1}^{n} g(\xi_i)$ от $\int_0^1 g(x)dx$ встречаются изредка и поэтому в качестве приближенного значения интеграла $\int_0^1 g(x)dx$ можно рассматривать $\frac{1}{n}\sum_{i=1}^{n} g(\xi_i)$.

Закон больших чисел дает теоретическое обоснование следующего алгоритма приближенного вычисления интеграла $\int_0^1 g(x)dx$.

1. Моделируем последовательность $\xi_1, \xi_2, \ldots, \xi_n$ независимых равномерно распределенных на $[0;1]$ случайных величин.

2. Полагаем $\int\limits_0^1 g(x)dx$ равным сумме $\frac{1}{n}\sum\limits_{i=1}^{n} g(\xi_i)$.

Заметим, что вычисление интеграла по любому промежутку сводится к вычислению по промежутку $[0;1]$.

Описанный метод вычисления интегралов называется методом Монте-Карло (методом статистических испытаний). Он особенно эффективен при вычислении интегралов большой кратности, когда другие методы приближенного анализа непригодны.

14.2 Слабый закон больших чисел

Слабый закон больших чисел является непосредственным следствием неравенства Чебышёва. Неравенство Чебышёва, несмотря на свою простоту, является сильным утверждением. Из него, в частности, следует теорема С. Н. Бернштейна об аппроксимации непрерывных функций полиномами Бернштейна (теорема 14.2.2). Заметим, что теорема Вейерштрасса об аппроксимации непрерывных функций полиномами устанавливает только существование аппроксимирующих полиномов.

Пусть $\{\eta_{n;\theta}\}$ — последовательность случайных величин с одним и тем же математическим ожиданием

$$\mathsf{M}\eta_{n;\theta} = \theta, n = 1, 2, \ldots,$$

и дисперсиями $\mathsf{D}\eta_{n;\theta} = \sigma_n^2(\theta)$, стремящимися к нулю при $n \to \infty$ (θ — параметр, принимающий значения из некоторого конечного или бесконечного промежутка $\Theta \subset \mathbb{R}^1$). Поскольку $\sigma_n^2(\theta)$ при больших n мала, то в силу неравенства Чебышёва значения случайной величины $\eta_{n;\theta}$ близки к θ и лишь изредка значительно уклоняются от своего среднего θ. Поэтому естественно ожидать, что для непрерывной функции $u(x)$ математическое ожидание $\mathsf{M}u(\eta_{n;\theta})$ будет близко к $u(\theta)$. Формально это утверждение записывается в виде следующей теоремы.

Теорема 14.2.1 (слабый закон больших чисел). *Пусть $\{\eta_{n;\theta}\}$ — последовательность случайных величин с математическим ожиданием $\mathsf{M}\eta_{n;\theta} = \theta$ и дисперсиями $\mathsf{D}\eta_{n;\theta} = \sigma_n^2(\theta)$, сходящимися к нулю при $n \to \infty$ для всех $\theta \in \Theta$. Тогда для* *ограниченной непрерывной функции $u = u(x)$ на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$ при $n \to \infty$*

$$\mathsf{M}u(\eta_{n;\theta}) \to u(\theta),\ \theta \in \Theta,$$

причем сходимость равномерная, если функция $u(\theta)$ равномерно непрерывна на Θ и дисперсии $\sigma_n^2(\theta)$ сходится к нулю равномерно.

Доказательство. Оценим $|\mathsf{M}u(\eta_{n;\theta}) - u(\theta)|$ сверху, воспользовавшись неравенством Чебышёва.

Обозначим через $\mathsf{P}_{n;\theta}$ распределение случайной величины $\eta_{n;\theta}$. Тогда

$$|\mathsf{M}u(\eta_{n;\theta}) - u(\theta)| \le \mathsf{M}\,|u(\eta_{n;\theta}) - u(\theta)| = \int\limits_{\mathbb{R}^1} |u(x) - u(\theta)|\mathsf{P}_{n;\theta}(dx).$$

Поскольку функция $u(x)$ непрерывна, то для данных $\theta \in \Theta$ и $\varepsilon > 0$ найдется такое $\delta(\varepsilon, \theta)$, что на множестве точек

$$\{x : |x - \theta| < \delta(\varepsilon, \theta)\}$$

имеет место неравенство $|u(x) - u(\theta)| < \varepsilon$, при этом область интегрирования $\mathbb{R}^1$ можно записать в виде

$$\{x : |x - \theta| < \delta(\varepsilon, \theta)\} \cup \{x : |x - \theta| \ge \delta(\varepsilon, \theta)\}.$$

Тогда, учитывая ограниченность $u(x)$ ($|u(x)| \le C < \infty$), имеем

$$\int\limits_{\mathbb{R}^1} |u(x) - u(\theta)|\mathsf{P}_{n;\theta}(dx) = \int\limits_{\{x:|x-\theta|<\delta(\varepsilon,\theta)\}} |u(x) - u(\theta)|\mathsf{P}_{n;\theta}(dx) +$$

$$+ \int\limits_{\{x:|x-\theta|\ge\delta(\varepsilon,\theta)\}} |u(x) - u(\theta)|\mathsf{P}_{n;\theta}(dx) \le$$

$$\le \int\limits_{\{x:|x-\theta|<\delta(\varepsilon,\theta)\}} \varepsilon\mathsf{P}_{n;\theta}(dx) + \int\limits_{\{x:|x-\theta|\ge\delta(\varepsilon,\theta)\}} 2C\mathsf{P}_{n;\theta}(dx) \le$$

$$\le \varepsilon + 2C\mathsf{P}_{n;\theta}\{x : |x - \theta| \ge \delta(\varepsilon, \theta)\}.$$

Последнее слагаемое оценим, воспользовавшись неравенством Чебышёва (10.2.1) для распределений:

$$\mathsf{P}_{n;\theta}\{x : |x-\theta| \geq \delta(\varepsilon,\theta)\} \leq \frac{\sigma_n^2(\theta)}{\delta^2(\varepsilon,\theta)},$$

первый момент распределения $\mathsf{P}_{n;\theta}$ равен θ, а дисперсия равна $\sigma_n^2(\theta)$. Так что

$$|\mathsf{M}u(\eta_{n;\theta}) - u(\theta)| \leq \varepsilon + 2C\frac{\sigma_n^2(\theta)}{\delta^2(\varepsilon,\theta)}.$$

И так как $\sigma_n^2(\theta)$ сходится к нулю при $n \to \infty$, то, начиная с некоторого N, имеем:

$$|\mathsf{M}u(\eta_{n;\theta}) - u(\theta)| \leq 2\varepsilon.$$

Для равномерно непрерывной на Θ функции $u(\theta)$ значение $\delta(\varepsilon,\theta)$ является общим для всех θ из Θ, т. е. $\delta(\varepsilon,\theta) = \delta(\varepsilon)$, если к тому же $\sigma_n^2(\theta)$ сходится к нулю равномерно по $\theta \in \Theta$, то

$$2C\frac{\sigma_n^2(\theta)}{\delta^2(\varepsilon,\theta)} = 2C\frac{\sigma_n^2(\theta)}{\delta^2(\varepsilon)} \leq \varepsilon,$$

начиная с некоторого N для всех $\theta \in \Theta$. Так что $\mathsf{M}u(\eta_{n;\theta})$ при $n \to \infty$ сходится к $u(\theta)$ равномерно по $\theta \in \Theta$.

Тем самым теорема доказана.

Следствие. *Пусть $\{\xi_i\}$ — последовательность независимых одинаково распределенных случайных величин с математическими ожиданиями $\mathsf{M}\xi_i = \theta$ и конечными дисперсиями $\mathsf{D}\xi_i = \sigma^2(\theta) < \infty$, $i = 1, 2, \ldots$, $\theta \in \Theta$. Тогда для последовательности случайных величин*

$$\eta_{n;\theta} = \frac{1}{n}\sum_{i=1}^{n}\xi_i,\ n = 1, 2, \ldots,$$

имеет место слабый закон больших чисел: если u — ограниченная непрерывная функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$, то при $n \to \infty$

$$\mathsf{M}u(\eta_{n;\theta}) \to u(\theta),\ \theta \in \Theta,$$

причем сходимость равномерная на Θ, если на Θ функция $u(\theta)$ равномерно непрерывна и $\mathsf{D}\eta_{n;\theta}$ сходится к нулю равномерно.

Для доказательства следствия достаточно заметить, что последовательность случайных величин

$$\eta_{n;\theta} = \frac{1}{n}\sum_{i=1}^{n}\xi_i, \ n = 1, 2, \ldots,$$

построенная по последовательности $\xi_1, \xi_2, \ldots$, удовлетворяет условиям теоремы:

$$\mathsf{M}\eta_{n;\theta} = \theta$$

и при $n \to \infty$

$$\mathsf{D}\eta_{n;\theta} = \sigma^2(\theta)/n \to 0.$$

Приложения к анализу. Из слабого закона больших чисел, фактически как его следствие, получаем теорему С. Н. Бернштейна об аппроксимации непрерывной функции полиномами.

Определение. *Полиномом Бернштейна* степени n функции $u(\theta)$, заданной на $[0;1]$, называется полином

$$B_{n,u}(\theta) = \sum_{k=0}^{n} u\left(\frac{k}{n}\right) C_n^k \theta^k (1-\theta)^{n-k}.$$

Теорема 14.2.2 (С.Н. Бернштейн). *Для непрерывной на замкнутом промежутке $[0;1]$ функции $u(\theta)$ последовательность ее полиномов Бернштейна*

$$B_{n,u}(\theta) = \sum_{k=0}^{n} u\left(\frac{k}{n}\right) C_n^k \theta^k (1-\theta)^{n-k}, \ n = 1, 2, \ldots,$$

при $n \to \infty$ равномерно сходится к $u(\theta)$, $\theta \in [0;1]$.

Доказательство. Пусть $\{\xi_i\}$ — последовательность независимых случайных величин, каждая из которых имеет распределение

$$\mathsf{P}_{\xi_i}(k) = \theta^k (1-\theta)^{1-k}, \ k = 0, 1.$$

Заметим, что

$$\mathsf{M}\xi_i = \theta, \ \mathsf{D}\xi_i = \sigma^2(\theta) = \theta(1-\theta), \ \theta \in [0;1].$$

У случайных величин

$$\eta_{n;\theta} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

$$\mathsf{M}\eta_{n;\theta} = \theta,\ \mathsf{D}\eta_{n;\theta} = \mathsf{D}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n}\sigma^2(\theta) = \frac{\theta(1-\theta)}{n}.$$

Поэтому в силу слабого закона больших чисел (теорема 14.2.1) для непрерывной на $[0;1]$ функции $u(\theta)$ при $n \to \infty$

$$\mathsf{M}u(\eta_{n;\theta}) \to u(\theta),$$

причем равномерно на $[0;1]$, поскольку $u(\theta)$ равномерно непрерывна на $[0,1]$ (как непрерывная функция на замкнутом промежутке), а дисперсия $\mathsf{D}\eta_{n;\theta} = \theta(1-\theta)/n$ при $n \to \infty$ сходится к нулю равномерно на $[0;1]$.

Вычислим

$$\mathsf{M}u(\eta_{n;\theta}) = \mathsf{M}u\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \mathsf{M}u\left(\frac{1}{n}\zeta_n\right)$$

как математическое ожидание функции $u\left(\frac{1}{n}\zeta_n\right)$ от случайной величины $\zeta_n = \sum\limits_{i=1}^{n}\xi_i$. Случайная величина $\zeta_n = \sum\limits_{i=1}^{n}\xi_i$ как сумма независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, каждая с распределением

$$\mathsf{P}_{\xi_i}(k) = \theta^k(1-\theta)^{1-k},\ k = 0, 1,$$

имеет биномиальное распределение:

$$\mathsf{P}\left\{\zeta_n = k\right\} = C_n^k\theta^k(1-\theta)^{n-k},\ k = 0, 1, \ldots, n.$$

Поэтому

$$\mathsf{M}u(\eta_{n;\theta}) = \mathsf{M}u\left(\frac{1}{n}\zeta_n\right) = \sum_{k=0}^{n}u\left(\frac{k}{n}\right)C_n^k\theta^k(1-\theta)^{n-k}.$$

Итак, для любой непрерывной на замкнутом промежутке $[0;1]$ функции $u(\theta)$ последовательность ее полиномов Бернштейна $B_{n,u}(\theta)$ равномерно сходится к $u(\theta)$.

Тем самым теорема доказана.

14.3 Примеры и задачи

Примеры

Пример 14.3.1. *Пусть u — функция со значениям в $\mathbb{R}^1$, заданная на положительной полуоси. Если $u = u(\theta)$ непрерывна на $[0; l]$, то последовательность функций*

$$\sum_{k=0}^{\infty} u\left(\frac{k}{n}\right) \frac{(n\theta)^k}{k!} e^{-n\theta}$$

при $n \to \infty$ равномерно сходится к $u(\theta)$ на $[0, l]$.

Решение. Пусть $\{\xi_i\}$ — последовательность независимых одинаково распределенных случайных величин, каждая из которых имеет распределение Пуассона с параметром θ:

$$\mathsf{P}_{\xi_i}(k) = \frac{\theta^k}{k!} e^{-\theta}, \; k = 0, 1, \ldots, \; \theta \in (0, \infty).$$

Так как $\mathsf{M}\xi_i = \theta$, $\mathsf{D}\xi_i = \sigma^2(\theta) = \theta$, $i = 1, 2, \ldots$, то для последовательности случайных величин $\eta_{n;\theta} = \frac{1}{n} \sum\limits_{i=1}^{n} \xi_i$ имеем $\mathsf{M}\eta_{n;\theta} = \theta$, $\mathsf{D}\eta_{n;\theta} = \theta/n$.

На каждом конечном промежутке $[0; l]$ изменения θ последовательность $\mathsf{D}\eta_{n;\theta} = \sigma^2(\theta)/n = \theta/n$ при $n \to \infty$ сходится к нулю равномерно, а непрерывная на $[0; l]$ функция $u = u(\theta)$ является равномерно непрерывной, поэтому при $n \to \infty$ сходимость

$$\mathsf{M}u(\eta_{n;\theta}) \to u(\theta)$$

равномерная по θ на $[0; l]$.

Вычислим

$$\mathsf{M}u(\eta_{n;\theta}) = \mathsf{M}u\left(\frac{1}{n} \sum_{i=1}^{n} \xi_i\right) = \mathsf{M}u\left(\frac{1}{n}\zeta_n\right)$$

как математическое ожидание функции от случайной величины $\zeta_n = \sum\limits_{i=1}^{n} \xi_i$, которая, как сумма n независимых пуассоновских случайных величин, имеет распределение Пуассона с параметром $n\theta$:

$$\mathsf{P}\{\zeta_n = k\} = \frac{(n\theta)^k}{k!}e^{-n\theta},\ k = 0, 1, \ldots$$

Имеем:

$$\mathsf{M}u(\eta_{n;\theta}) = \mathsf{M}u\left(\frac{1}{n}\zeta_n\right) = \sum_{k=0}^{\infty} u\left(\frac{k}{n}\right)\frac{(n\theta)^k}{k!}e^{-n\theta}.$$

Итак, для непрерывной на $[0; l]$ функции $u = u(\theta)$

$$\sum_{k=0}^{\infty} u\left(\frac{k}{n}\right)\frac{(n\theta)^k}{k!}e^{-n\theta}$$

сходится равномерно к $u(\theta)$ при $n \to \infty$.

Задачи

14.1. Пусть $\{\xi_k\}$ — последовательность независимых одинаково распределенных случайных величин с распределением F. Для целого $n \geq 1$ определим на $\mathbb{R}^1$ функцию

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{(-\infty,x)}(\xi_k)$$

($\hat{F}_n(x) = \hat{F}_n(x, \omega)$ при каждом фиксированном ω — функция распределения).

Доказать, что при $n \to \infty$ для каждого $x \in \mathbb{R}^1$ случайная величина $\hat{F}_n(x)$ сходится по вероятности к $F(x)$.

14.2. Для действительной непрерывной на $[0; 1]$ функции f вычислить предел:

$$\lim_{n\to\infty} \underbrace{\int_0^1 \cdots \int_0^1}_{n} f\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) dx_1 dx_2 \ldots dx_n.$$

Указание. Воспользуйтесь следствием из теоремы 14.2.1. Рассмотрите в качестве последовательности $\{\xi_i\}$ последовательность независимых случайных величин, каждая из которых равномерно распределена на $[0; 1]$.

Ответ: $f(1/2)$.

Глава 15

Центральная предельная теорема

Мы изучаем стохастический эксперимент, исследуя его математическую модель — вероятностное пространство $\{\Omega, \mathfrak{F}, \mathsf{P}\}$. Вместе с $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ были введены основные понятия теории вероятностей: событие, вероятность, случайная величина, математическое ожидание, характеристическая функция. Взаимосвязи и соотношения между ними выражаются рядом теорем. Анализ содержания введенных понятий обнаруживает, что они не являются новыми в математике. Эти понятия хорошо известны в анализе:

событие — измеримое множество,

вероятность — нормированная мера,

вероятностное пространство — пространство с мерой,

случайная величина — измеримая функция,

математическое ожидание — интеграл Лебега,

характеристическая функция — преобразование Фурье.

Однако за каждой наукой признается право на самостоятельность, если в ней имеются результаты, которые не могут быть получены вне ее рамок. Для теории вероятностей таким результатом, занимающим в ней исключительное место, является центральная предельная теорема (ЦПТ), утверждающая, что при достаточно общих условиях сумма большого числа независимых случайных величин, каждая из которых мала по сравнению со всей суммой, имеет распределение, близкое к нормальному. Значение этой теоремы выходит далеко за рамки теории вероятностей. Она, в частности, объясняет особую роль нормального распределения при рассмотрении многих практических задач — всегда, когда можно считать, что рассматриваемая случайная величина является суммой большого числа малых независимых слагаемых, ее распределение несущественно отличается от нормального. Такими случайными величинами являются, например, погрешности измерений, продолжительность исправной работы прибора (когда выход его из строя является следствием постоянно действующих усталостных изменений), скорости движения молекул газа и т. д.

Первый вариант центральной предельной теоремы — интегральная теорема Муавра—Лапласа (1812) — утверждает, что для числа успехов $\mu = \sum\limits_{k=1}^{n} \mu_k$ в n испытаниях Бернулли при $n \to \infty$

$$\mathsf{P}\left\{x_1 < \frac{\mu - \mathsf{M}\mu}{\sqrt{\mathsf{D}\mu}} < x_2\right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{x_1}^{x_2} e^{-t^2/2} dt.$$

для любых действительных чисел x_1 и x_2 (говорят, что для случайных величин $\mu_1, \mu_2, \ldots$ имеет место ЦПТ).

Далее естественно возникает задача — для каких случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ (помимо упомянутых) имеет место ЦПТ, т. е. при $n \to \infty$ для $S_n = \sum\limits_{k=1}^{n} \xi_k$

$$\mathsf{P}\left\{x_1 < \frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}} < x_2\right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{x_1}^{x_2} e^{-t^2/2} dt.$$

Впервые достаточные условия применимости ЦПТ для широкого класса случайных величин были получены П. Л. Чебышёвым (1887). В доказательстве были обнаружены пробелы, восполненные А. А. Марковым (1898). Решение, близкое к окончательному, было получено А. М. Ляпуновым (1901). Окончательное решение — достаточные условия применимости ЦПТ, в некотором смысле близкие к необходимым, — получены Линдебергом. Необходимые условия найдены Феллером.

Сначала рассмотрим простой, но важный и часто встречающийся вариант ЦПТ, — центральную предельную теорему для одинаково распределенных случайных величин. Мы выделяем случай одинаково распределенных случайных величин как в силу его важности, так и возможности (и необходимости) продемонстрировать сущность метода характеристических функций, когда он свободен от трудностей, ему не свойственных.

15.1 ЦПТ для одинаково распределенных случайных величин

В главе 15 сумму $\sum\limits_{k=1}^{n} \xi_k$ будем обозначать через S_n, не оговаривая это специально каждый раз.

Далее нам понадобится следующая оценка:

Лемма 15.1.1. *При* $|z| \le 1/2$

$$|\ln(1+z) - z| \le |z|^2. \tag{15.1.1}$$

Действительно,

$$\ln(1+z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} z^{k+1} = z + \sum_{k=1}^{\infty} \frac{(-1)^k}{k+1} z^{k+1}, \ |z| < 1.$$

Отсюда

$$|\ln(1+z) - z| = \left| \sum_{k=1}^{\infty} \frac{(-1)^k}{k+1} z^{k+1} \right| \le \sum_{k=1}^{\infty} \frac{1}{k+1} |z|^{k+1} \le$$

$$\le \frac{1}{2} \sum_{k=1}^{\infty} |z|^{k+1} = \frac{|z|^2}{2(1-|z|)}.$$

При $|z| \le 1/2$ значение $1 - |z| \ge 1/2$, поэтому для таких z

$$|\ln(1+z) - z| \le |z|^2.$$

Тем самым искомая оценка получена. Из нее, в частности, следует, что при $z \to 0$

$$\frac{\ln(1+z)}{z} \to 1. \tag{15.1.2}$$

Следствие. *При* $|z_k| \le \varepsilon \le 1/2, \ k = 1, 2, \ldots, n,$

$$\sum_{k=1}^{n} |\ln(1+z_k) - z_k| \le \varepsilon \sum_{k=1}^{n} |z_k|.$$

Поскольку

$$|\ln(1+z_k) - z_k| \le |z_k|^2 \le \varepsilon |z_k|,$$

то

$$\sum_{k=1}^{n} |\ln(1+z_k) - z_k| \le \varepsilon \sum_{k=1}^{n} |z_k|.$$

Теорема 15.1.1 (ЦПТ для одинаково распределенных случайных величин). *Пусть* ξ_k, $k = 1, 2, \ldots,$ *— независимые одинаково распределенные случайные величины со средними* $m_1 = \mathsf{M}\xi_k$ *и конечными дисперсиями* $\mathsf{D}\xi_k = \sigma^2 > 0$, *тогда при* $n \to \infty$ *распределение нормированной суммы*

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Доказательство. Без ограничения общности будем считать, что $m_1 = \mathsf{M}\xi_k = 0$. Воспользуемся теоремой Леви (теорема 13.5.1), согласно которой для сходимости распределений случайных величин

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}} = \frac{1}{\sigma\sqrt{n}}\, S_n$$

к нормальному распределению с параметрами $(0; 1)$ достаточно доказать, что для каждого $t \in \mathbb{R}^1$ последовательность характеристических функций

$$\psi_n(t) = \mathsf{M}\exp\left\{it\frac{1}{\sigma\sqrt{n}}S_n\right\} = \mathsf{M}\exp\left\{it\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}\xi_k\right\}$$

случайных величин $\dfrac{1}{\sigma\sqrt{n}}\, S_n$ сходится к $e^{-t^2/2}$ — характеристической функции нормального распределения с параметрами $(0; 1)$.

Обозначим через $\varphi(t)$ характеристические функции случайных величин ξ_k:

$$\varphi(t) = \mathsf{M}\exp\{it\xi_k\},\ k = 1, 2, \ldots,$$

(они совпадают, поскольку ξ_k, $k = 1, 2, \ldots,$ одинаково распределены). В силу мультипликативного свойства характеристических функций (ξ_k, $k = 1, 2, \ldots,$ независимы)

$$\psi_n(t) = \mathsf{M}\exp\left\{i\frac{t}{\sigma\sqrt{n}}\sum_{k=1}^{n}\xi_k\right\} = \prod_{k=1}^{n}\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n,$$

т. е.

$$\psi_n(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n. \tag{15.1.3}$$

Далее, второй момент $m_2 = \sigma^2$ случайной величины ξ_k конечен, поэтому для характеристической функции $\varphi(s)$ случайной величины ξ_k при $s \to 0$ имеет место разложение

$$\varphi(s) = 1 + \varphi'(0)s + \frac{\varphi''(0)}{2!}s^2 + o(s^2)$$

(теорема 13.3.2), или, с учетом того, что

$$\varphi^{(k)}(0) = i^k m_k$$

(см. равенство (13.3.7)) и $m_1 = 0$, $m_2 = \sigma^2$,

$$\varphi(s) = 1 - \frac{s^2\sigma^2}{2!} + o(s^2).$$

Отсюда, положив $s = t/(\sigma\sqrt{n})$, для каждого фиксированного t при $n \to \infty$ имеем

$$\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{1}{2}\sigma^2\frac{t^2}{\sigma^2 n} + o\left(\frac{1}{n}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

Так что для $\psi_n(t)$ при $n \to \infty$ имеет место представление

$$\psi_n(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n$$

(см. равенство (15.1.3)). Далее, воспользовавшись тем, что при $z \to 0$

$$\ln(1+z) \sim z,$$

получим, что при $n \to \infty$

$$\ln\psi_n(t) = n\ln\left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right) \sim$$

$$\sim n\left(-\frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right) = -\frac{t^2}{2} + o(1)$$

и, следовательно,

$$\ln \psi_n(t) \to -\frac{t^2}{2}, \text{ а } \quad \psi_n(t) \to e^{-t^2/2}.$$

Отсюда и следует утверждение теоремы.

Замечание. Мы доказали теорему в предположении, что $m_1 = \mathsf{M}\xi_k = 0$. Если $m_1 = \mathsf{M}\xi_k \neq 0$, рассмотрим последовательность случайных величин $\eta_k = \xi_k - m_1,\ k = 1, 2, \ldots$ Случайные величины $\eta_k,\ k = 1, 2, \ldots$ независимы и одинаково распределены со средним $\mathsf{M}\eta_k = 0$ и конечными дисперсиями $\mathsf{D}\eta_k = \sigma^2 > 0$. Поэтому распределение случайной величины

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} \eta_k = \frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}$$

сходится к нормальному распределению с параметрами (0;1), где $S_n = \sum\limits_{k=1}^{n} \xi_k$.

Следствие (интегральная теорема Муавра—Лапласа). *Пусть μ — число успехов в n испытаниях Бернулли с вероятностью успеха p $(0 < p < 1)$ в одном испытании. При $n \to \infty$ распределение случайной величины*

$$\frac{\mu - np}{\sqrt{npq}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Доказательство. Как известно, число успехов μ можно представить в виде суммы независимых одинаково распределенных случайных величин μ_k:

$$\mu = \sum_{k=1}^{n} \mu_k,$$

каждая из которых имеет распределение

$$\mathsf{P}\{\mu_k = s\} = p^s (1-p)^{1-s}, s = 0; 1,$$

при этом

$$\mathsf{M}\mu_k = p,\ \mathsf{D}\mu_k = pq.$$

В силу теоремы 15.1.1 при $n \to \infty$ распределение случайной величины

$$\frac{\sum\limits_{k=1}^{n} \mu_k - \mathsf{M} \sum\limits_{k=1}^{n} \mu_k}{\sqrt{\mathsf{D} \sum\limits_{k=1}^{n} \mu_k}} = \frac{\mu - np}{\sqrt{npq}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

15.2 ЦПТ в форме Линдеберга

Обозначения. Пусть $\xi_k, \ k = 1, 2, \ldots,$ — последовательность независимых случайных величин соответственно с распределениями F_k. Не уменьшая общности, будем считать, что $\mathsf{M}\xi_k = 0, \ k = 1, 2, \ldots$ Дисперсию $\mathsf{D}\xi_k$ случайной величины ξ_k будем обозначать σ_k^2, характеристическую функцию — $\varphi_k(t)$. Напомним, что $\mathsf{M}\xi_k, \sigma_k^2, \varphi_k(t)$ по распределению F_k вычисляются так:

$$\mathsf{M}\xi_k = \int\limits_{\mathbb{R}^1} x\mathsf{F}_k(dx), \ \mathsf{D}\xi_k = \sigma_k^2 = \int\limits_{\mathbb{R}^1} x^2\mathsf{F}_k(dx), \ \varphi_k(t) = \int\limits_{\mathbb{R}^1} e^{itx}\mathsf{F}_k(dx).$$

Обозначим еще

$$S_n = \sum_{k=1}^{n} \xi_k, \ \mathsf{D}S_n = \mathsf{D}\sum_{k=1}^{n} \xi_k = \sum_{k=1}^{n} \sigma_k^2 = s_n^2.$$

Условие Линдеберга. Будем говорить, что для независимых случайных величин ξ_k с математическими ожиданиями $\mathsf{M}\xi_k = 0$ и конечными дисперсиями σ_k^2, $k = 1, 2, \ldots,$ выполняется условие Линдеберга, если для любого (сколь угодно малого) $\tau > 0$ при $n \to \infty$

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x| \geq \tau s_n} x^2\mathsf{F}_k(dx) \to 0.$$

Из условия Линдеберга, в частности, следует, что при больших n дисперсии σ_k^2 отдельных слагаемых ξ_k, $k = 1, 2, \ldots, n,$

малы по сравнению с дисперсией s_n^2 суммы $S_n = \sum\limits_{k=1}^{n} \xi_k$ в следующем смысле: для любого $\varepsilon > 0$ найдется N, такое, что при $n \geq N$
$$\frac{\sigma_k^2}{s_n^2} \leq \varepsilon,\ k = 1, 2, \ldots, n.$$

Действительно,
$$\frac{\sigma_k^2}{s_n^2} = \frac{1}{s_n^2} \int\limits_{|x|<\tau s_n} x^2 \mathsf{F}_k(dx) + \frac{1}{s_n^2} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx) \leq$$
$$\leq \frac{1}{s_n^2} \int\limits_{|x|<\tau s_n} \tau^2 s_n^2 \mathsf{F}_k(dx) + \frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx) =$$
$$= \tau^2 \int\limits_{|x|<\tau s_n} \mathsf{F}_k(dx) + \frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx) \leq$$
$$\leq \tau^2 + \frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx).$$

Выбирая τ так, чтобы $\tau^2 < \varepsilon/2$, а затем N настолько большим, чтобы при $n > N$
$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx) < \varepsilon/2,$$
получим:
$$\frac{\sigma_k^2}{s_n^2} < \varepsilon,\ k = 1, 2, \ldots, n.$$

Теорема 15.2.1 (Линдеберга). *Пусть для независимых случайных величин ξ_k, $k = 1, 2, \ldots$, со средними $\mathsf{M}\xi_k = 0$ и конечными дисперсиями σ_k^2 выполняется условие Линдеберга: для любого $\tau > 0$ при $n \to \infty$*
$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx) \to 0,$$

тогда распределение нормированной суммы

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}} = \frac{S_n}{s_n}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Доказательство. Воспользуемся теоремой Леви, согласно которой для доказательства сходимости распределений случайных величин S_n/s_n к нормальному распределению с параметрами $(0; 1)$ достаточно доказать, что для каждого $t \in \mathbb{R}^1$ последовательность характеристических функций

$$\psi_n(t) = \mathsf{M} \exp\left\{it\frac{S_n}{s_n}\right\}$$

случайных величин S_n/s_n сходится к характеристической функции $\exp\{-t^2/2\}$ нормального распределения с параметрами $(0; 1)$. Этим и займемся.

Выпишем представление для характеристической функции $\psi_n(t)$ суммы S_n/s_n через характеристические функции $\varphi_k(t)$ слагаемых ξ_k, $k = 1, 2, \ldots, n$. Воспользуемся мультипликативным свойством характеристических функций:

$$\psi_n(t) = \mathsf{M} \exp\left\{it\frac{S_n}{s_n}\right\} = \mathsf{M} \exp\left\{i\frac{t}{s_n}\sum_{k=1}^{n}\xi_k\right\} = \prod_{k=1}^{n}\varphi_k\left(\frac{t}{s_n}\right).$$

Докажем, что для каждого фиксированного $t \in \mathbb{R}^1$ при $n \to \infty$

$$\ln \psi_n(t) = \ln \prod_{k=1}^{n}\varphi_k\left(\frac{t}{s_n}\right) = \sum_{k=1}^{n}\ln\varphi_k\left(\frac{t}{s_n}\right) \to -\frac{t^2}{2}.$$

Для этого оценим сверху

$$\left|\ln\psi_n(t) - \left(-\frac{t^2}{2}\right)\right| = \left|\sum_{k=1}^{n}\ln\varphi_k\left(\frac{t}{s_n}\right) + \frac{t^2}{2}\right|.$$

Из условия Линдеберга следует, что при $n \to \infty$ неизбежно $s_n \to +\infty$. Отсюда имеем, что для каждого фиксированного t при $n \to \infty$ дробь $t/s_n \to 0$, а следовательно,

$$\varphi_k(t/s_n) - 1 \to 0.$$

Поэтому для $\ln\varphi_k(t/s_n)$ при $n\to\infty$ имеем:

$$\ln\varphi_k\left(\frac{t}{s_n}\right)=\ln\left(1+\left(\varphi_k\left(\frac{t}{s_n}\right)-1\right)\right)\sim\varphi_k\left(\frac{t}{s_n}\right)-1,$$

т. е.

$$\ln\varphi_k\left(\frac{t}{s_n}\right)\sim\varphi_k\left(\frac{t}{s_n}\right)-1,$$

Это соотношение мотивирует такую оценку:

$$\left|\ln\psi_n(t)+\frac{t^2}{2}\right|=\left|\sum_{k=1}^{n}\ln\varphi_k\left(\frac{t}{s_n}\right)+\frac{t^2}{2}\right|=$$

$$=\left|\sum_{k=1}^{n}\ln\varphi_k\left(\frac{t}{s_n}\right)-\sum_{k=1}^{n}\left(\varphi_k\left(\frac{t}{s_n}\right)-1\right)+\right.$$

$$\left.+\sum_{k=1}^{n}\left(\varphi_k\left(\frac{t}{s_n}\right)-1\right)+\frac{t^2}{2s_n^2}\sum_{k=1}^{n}\sigma_k^2\right|\le$$

$$\le\sum_{k=1}^{n}\left|\ln\left(1+\left(\varphi_k\left(\frac{t}{s_n}\right)-1\right)\right)-\left(\varphi_k\left(\frac{t}{s_n}\right)-1\right)\right|+$$

$$+\sum_{k=1}^{n}\left|\varphi_k\left(\frac{t}{s_n}\right)-1+\frac{t^2}{2s_n^2}\sigma_k^2\right|=A_n+B_n. \tag{15.2.1}$$

Покажем, что при $n\to\infty$ к нулю сходится каждое из слагаемых A_n и B_n в правой части последнего неравенства. Для этого оценим их сверху. Основным “инструментом” при этом будет оценка погрешности при замене экспоненты отрезком ряда Тейлора (см. оценку (13.3.1)). Заметим, что поскольку гарантирована конечность только вторых моментов случайных величин ξ_k, то в оценках могут участвовать только моменты случайных величин не выше второго порядка.

Сначала оценим

$$A_n=\sum_{k=1}^{n}\left|\ln\left(1+\left(\varphi_k\left(\frac{t}{s_n}\right)-1\right)\right)-\left(\varphi_k\left(\frac{t}{s_n}\right)-1\right)\right|.$$

Мы предполагаем воспользоваться неравенством

$$\sum_{k=1}^{n} |\ln(1+z_k) - z_k| \le \varepsilon \sum_{k=1}^{n} |z_k|,\ |z_k| < \varepsilon \le \frac{1}{2},\ k = 1, 2, \dots, n, \tag{15.2.2}$$

с $(\varphi_k(t/s_n) - 1)$ в качестве z_k. Для этого оценим $|\varphi_k\,(t/s_n) - 1|$. Воспользуемся оценкой погрешности при замене $\exp\left\{i\dfrac{t}{s_n}x\right\}$ отрезком ряда Тейлора (напомним, что $\mathsf{M}\xi_k = 0$):

$$\left|\varphi_k\left(\frac{t}{s_n}\right) - 1\right| = \left|\int\limits_{\mathbb{R}^1} \left(\exp\left\{i\frac{t}{s_n}x\right\} - \left(1 + i\frac{t}{s_n}x\right)\right)\mathsf{F}_k(dx)\right| \le$$

$$\le \int\limits_{\mathbb{R}^1} \left|\exp\left\{i\frac{t}{s_n}x\right\} - \left(1 + i\frac{t}{s_n}x\right)\right| \mathsf{F}_k(dx) \le \int\limits_{\mathbb{R}^1} \left|\frac{1}{2}\left(\frac{itx}{s_n}\right)^2\right| \mathsf{F}_k(dx) =$$

$$= \frac{t^2}{2s_n^2}\int\limits_{\mathbb{R}^1} x^2\mathsf{F}_k(dx) = \frac{t^2}{2s_n^2}\sigma_k^2.$$

Так что

$$\left|\varphi_k\left(\frac{t}{s_n}\right) - 1\right| \le \frac{t^2}{2s_n^2}\sigma_k^2. \tag{15.2.3}$$

Из последнего неравенства, в частности, следует, что для данного ε $(\varepsilon \le 1/2)$ при достаточно больших n

$$\left|\varphi_k\left(\frac{t}{s_n}\right) - 1\right| \le \varepsilon \le \frac{1}{2},\ k = 1, 2, \dots, n,$$

поскольку при достаточно больших n $(n \ge N)$ в силу условия Линдеберга

$$\frac{t^2}{2} \cdot \frac{\sigma_k^2}{s_n^2} < \varepsilon,\ k = 1, 2, \dots, n.$$

Далее, пользуясь сначала неравенством (15.2.2) с $(\varphi_k(t/s_n) - 1)$ в качестве z_k, а затем оценкой (15.2.3), получаем:

$$A_n = \sum_{k=1}^{n} \left|\ln\left(1 + \left(\varphi_k\left(\frac{t}{s_n}\right) - 1\right)\right) - \left(\varphi_k\left(\frac{t}{s_n}\right) - 1\right)\right| \le$$

$$\leq \varepsilon \sum_{k=1}^{n} \left| \varphi_k \left(\frac{t}{s_n} \right) - 1 \right| \leq \varepsilon \sum_{k=1}^{n} \frac{t^2}{2s_n^2} \sigma_k^2 = \varepsilon \frac{t^2}{2s_n^2} \sum_{k=1}^{n} \sigma_k^2 = \frac{1}{2} \varepsilon t^2.$$

Так что для фиксированного $t \in \mathbb{R}^1$ при $n \to \infty$

$$A_n \to 0.$$

Далее оценим сверху

$$B_n = \sum_{k=1}^{n} \left| \varphi_k \left(\frac{t}{s_n} \right) - 1 + \frac{t^2}{2s_n^2} \sigma_k^2 \right|.$$

Имеем

$$B_n = \sum_{k=1}^{n} \left| \varphi_k \left(\frac{t}{s_n} \right) - 1 + \frac{t^2}{2s_n^2} \sigma_k^2 \right| =$$

$$= \sum_{k=1}^{n} \left| \int\limits_{\mathbb{R}^1} \left(\exp\left\{ i\frac{t}{s_n}x \right\} - \left(1 + i\frac{t}{s_n}x + \frac{1}{2} \left(i\frac{t}{s_n}x \right)^2 \right) \right) \mathsf{F}_k(dx) \right| \leq$$

$$\leq \sum_{k=1}^{n} \int\limits_{\mathbb{R}^1} \left| \exp\left\{ i\frac{t}{s_n}x \right\} - \left(1 + i\frac{t}{s_n}x + \frac{1}{2} \left(i\frac{t}{s_n}x \right)^2 \right) \right| \mathsf{F}_k(dx) =$$

$$= \sum_{k=1}^{n} \int\limits_{|x|<\tau s_n} \left| \exp\left\{ i\frac{t}{s_n}x \right\} - \left(1 + i\frac{t}{s_n}x + \frac{1}{2} \left(i\frac{t}{s_n}x \right)^2 \right) \right| \mathsf{F}_k(dx) +$$

$$+ \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} \left| \exp\left\{ i\frac{t}{s_n}x \right\} - \left(1 + i\frac{t}{s_n}x + \frac{1}{2} \left(i\frac{t}{s_n}x \right)^2 \right) \right| \mathsf{F}_k(dx).$$

Первую из двух последних сумм, пользуясь оценкой погрешности при замене $\exp\left\{ i\frac{t}{s_n}x \right\}$ отрезком ряда Тейлора, оценим так:

$$\sum_{k=1}^{n} \int\limits_{|x|<\tau s_n} \left| \exp\left\{ i\frac{t}{s_n}x \right\} - \left(1 + i\frac{t}{s_n}x + \frac{1}{2} \left(i\frac{t}{s_n}x \right)^2 \right) \right| \mathsf{F}_k(dx) \leq$$

$$\leq \sum_{k=1}^{n} \int\limits_{|x|<\tau s_n} \frac{1}{3!}\frac{|t|^3}{s_n^3}|x|^3 \mathsf{F}_k(dx) \leq \frac{1}{3!}\frac{|t|^3}{s_n^3}\sum_{k=1}^{n} \int\limits_{|x|<\tau s_n} |x||x|^2 \mathsf{F}_k(dx) \leq$$

$$\leq \frac{1}{6}\frac{|t|^3}{s_n^3}\sum_{k=1}^{n} \int\limits_{|x|<\tau s_n} \tau s_n x^2 \mathsf{F}_k(dx) \leq \frac{1}{6}\frac{|t|^3}{s_n^3}\tau s_n \sum_{k=1}^{n} \int\limits_{\mathbb{R}^1} x^2 \mathsf{F}_k(dx) =$$

$$= \frac{1}{6}\tau\frac{|t|^3}{s_n^2}\sum_{k=1}^{n}\sigma_k^2 = \frac{1}{6}\tau|t|^3\frac{1}{s_n^2}s_n^2 = \frac{1}{6}\tau|t|^3.$$

Вторую сумму оценим так:

$$\sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} \left|\exp\left\{i\frac{t}{s_n}x\right\} - \left(1 + i\frac{t}{s_n}x + \frac{1}{2}\left(i\frac{t}{s_n}x\right)^2\right)\right| \mathsf{F}_k(dx) \leq$$

$$\leq \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} \left(\left|\exp\left\{i\frac{t}{s_n}x\right\} - \left(1 + i\frac{t}{s_n}x\right)\right| + \frac{1}{2}\frac{t^2}{s_n^2}x^2\right) \mathsf{F}_k(dx) \leq$$

$$\leq \sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} \left(\frac{1}{2}\frac{t^2}{s_n^2}x^2 + \frac{1}{2}\frac{t^2}{s_n^2}x^2\right) \mathsf{F}_k(dx) = t^2\frac{1}{s_n^2}\sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx).$$

Так что для B_n имеем оценку:

$$B_n = \sum_{k=1}^{n}\left|\varphi_k\left(\frac{t}{s_n}\right) - 1 + \frac{t^2\sigma_k^2}{2s_n^2}\right| \leq \frac{\tau|t|^3}{6} + \frac{t^2}{s_n^2}\sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx).$$

Выбирая τ так, чтобы $\tau|t|^3/6 < \varepsilon/2$, а затем n настолько большим, чтобы

$$t^2\frac{1}{s_n^2}\sum_{k=1}^{n} \int\limits_{|x|\geq\tau s_n} x^2 \mathsf{F}_k(dx) \leq \frac{1}{2}\varepsilon,$$

получаем, что при достаточно больших n:

$$B_n = \sum_{k=1}^{n}\left|\varphi_k\left(\frac{t}{s_n}\right) - 1 + \frac{t^2}{2s_n^2}\sigma_k^2\right| \leq \varepsilon.$$

Поэтому из (15.2.1), учитывая, что и $A_n \to 0$ при $n \to \infty$, следует, что

$$\ln \psi_n(t) = \sum_{k=1}^{n} \ln \varphi_k \left(\frac{t}{s_n}\right) \to -\frac{t^2}{2}.$$

Тем самым центральная предельная теорема в форме Линдеберга доказана.

Замечание. Центральная предельная теорема для независимых одинаково распределенных случайных величин может быть получена как следствие теоремы Линдеберга поскольку для таких случайных величин условие Линдеберга выполняется.

В самом деле, если ξ_k, $k = 1, 2, \ldots,$ — независимые одинаково распределенные случайные величины с конечными дисперсиями $\sigma_k^2 = \sigma^2$, $k = 1, 2, \ldots,$ то

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x| \geq \tau s_n} x^2 \mathsf{F}_k(dx) = \frac{1}{\sigma^2 n} \sum_{k=1}^{n} \int\limits_{|x| \geq \tau\sigma\sqrt{n}} x^2 \mathsf{F}(dx) =$$

$$= \frac{1}{\sigma^2 n} n \int\limits_{|x| \geq \tau\sigma\sqrt{n}} x^2 \mathsf{F}(dx) = \frac{1}{\sigma^2} \int\limits_{|x| \geq \tau\sigma\sqrt{n}} x^2 \mathsf{F}(dx).$$

Функция множества

$$\mathsf{Q}(A) = \frac{1}{\sigma^2} \int\limits_{A} x^2 \mathsf{F}(dx), \; A \in \mathfrak{B}^1,$$

является вероятностным распределением на $\mathbb{R}^1$. Последовательность множеств $A_n = \{x : |x| \geq \tau\sigma\sqrt{n}\}$, $n = 1, 2, \ldots,$ монотонно убывающая ($A_{n+1} \subset A_n, n = 1, 2, \ldots$) и $B = \bigcap\limits_{n=1}^{\infty} A_n = \emptyset$. Поэтому в силу свойства непрерывности вероятности (распределения) при $n \to \infty$

$$\mathsf{Q}(A_n) = \frac{1}{\sigma^2} \int\limits_{A_n} x^2 \mathsf{F}(dx) \to \frac{1}{\sigma^2} \int\limits_{B} x^2 \mathsf{F}(dx) = 0,$$

а следовательно, при $n \to \infty$ и

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{|x| \geq \tau s_n} x^2 \mathsf{F}_k(dx) = \frac{1}{\sigma^2} \int\limits_{|x| \geq \tau\sigma\sqrt{n}} x^2 \mathsf{F}(dx) \to 0.$$

Центральная предельная теорема, разумеется, имеет место и для случайных величин ξ_k с $\mathsf{M}\xi_k = a_k \neq 0$.

Теорема. *Если для независимых случайных величин ξ_k, $k = 1, 2, \ldots$, со средними $\mathsf{M}\xi_k = a_k$ и конечными дисперсиями σ_k^2 выполняется условие Линдеберга: для каждого $\tau > 0$ при $n \to \infty$*

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{y:|y-a_k|\geq \tau s_n} (x - a_k)^2 \mathsf{F}_k(dx) \to 0,$$

то распределение нормированной суммы

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Пусть

$$\eta_k = \xi_k - a_k, \ k = 1, 2, \ldots$$

У случайных величин η_k

$$\mathsf{M}\eta_k = 0, \ \mathsf{D}\eta_k = \sigma_k^2.$$

Обозначим распределение η_k через G_k. Убедимся, что для независимых случайных величин $\eta_1, \eta_2, \ldots$ со средним $\mathsf{M}\eta_k = 0$ выполняется условие Линдеберга: для каждого $\tau > 0$ при $n \to \infty$

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{y:|y|\geq \tau s_n} x^2 \mathsf{G}_k(dx) \to 0.$$

В самом деле,

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{y:|y|\geq \tau s_n} x^2 \mathsf{G}_k(dx) =$$

$$= \frac{1}{s_n^2} \sum_{k=1}^{n} \int\limits_{\mathbb{R}^1} x^2 I_{\{y:|y|\geq \tau s_n\}}(x) \mathsf{G}_k(dx) = \frac{1}{s_n^2} \sum_{k=1}^{n} \mathsf{M}\eta_k^2 I_{\{y:|y|\geq \tau s_n\}}(\eta_k) =$$

$$= \frac{1}{s_n^2} \sum_{k=1}^{n} \mathsf{M}(\xi_k - a_k)^2 I_{\{y:|y|\geq \tau s_n\}}(\xi_k - a_k) =$$

$$= \frac{1}{s_n^2} \sum_{k=1}^{n} \mathsf{M}(\xi_k - a_k)^2 I_{\{y:|y-a_k| \geq \tau s_n\}}(\xi_k) =$$

$$= \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\mathbb{R}^1} (x - a_k)^2 I_{\{y:|y-a_k| \geq \tau s_n\}}(x) \mathsf{F}_k(dx) =$$

$$= \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{y:|y-a_k| \geq \tau s_n} (x - a_k)^2 \mathsf{F}_k(dx) \to 0$$

при $n \to \infty$. Поэтому для случайных величин η_k имеет место ЦПТ: распределение случайной величины $\frac{1}{s_n} \sum\limits_{k=1}^{n} \eta_k$ сходится к нормальному распределению с параметрами $(0; 1)$. А так как

$$\frac{1}{s_n} \sum_{k=1}^{n} \eta_k = \frac{1}{s_n} \sum_{k=1}^{n} (\xi_k - a_k) = \frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}},$$

то и распределение нормированной суммы

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Теорема 15.2.2 (Ляпунова). *Если для независимых случайных величин ξ_k, $k = 1, 2, \ldots$, со средними $\mathsf{M}\xi_k = a_k$ и конечными дисперсиями σ_k^2 выполняется условие Ляпунова: существует такое $\delta > 0$, что при $n \to \infty$*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \int_{\mathbb{R}^1} |x - a_k|^{2+\delta} \mathsf{F}_k(dx) \to 0,$$

то распределение нормированной суммы

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Для доказательства теоремы убедимся, что из условия Ляпунова следует условие Линдеберга.

Действительно,

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int\limits_{x:|x-a_k|\geq\tau s_n}(x-a_k)^2\mathsf{F}_k(dx)=\frac{1}{s_n^2}\sum_{k=1}^{n}\int\limits_{x:\frac{|x-a_k|}{\tau s_n}\geq 1}(x-a_k)^2\mathsf{F}_k(dx)\leq$$

$$\leq\frac{1}{s_n^2}\sum_{k=1}^{n}\int\limits_{x:\frac{|x-a_k|}{\tau s_n}\geq 1}(x-a_k)^2\left(\frac{|x-a_k|}{\tau s_n}\right)^{\delta}\mathsf{F}_k(dx)=$$

$$=\frac{1}{\tau^{\delta}s_n^{2+\delta}}\sum_{k=1}^{n}\int\limits_{x:\frac{|x-a_k|}{\tau s_n}\geq 1}|x-a_k|^{2+\delta}\mathsf{F}_k(dx)\leq$$

$$\leq\frac{1}{\tau^{\delta}}\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n}\int\limits_{\mathbb{R}^1}|x-a_k|^{2+\delta}\mathsf{F}_k(dx),$$

что для каждого $\tau>0$ стремится к нулю при $n\to\infty$.

О другой формулировке ЦПТ. В рассмотренных предельных теоремах в тех или иных предположениях относительно независимых случайных величин $\xi_k, k=1,2,\ldots,$ с конечными дисперсиями σ_k^2 утверждается, что распределение нормированной суммы

$$\frac{S_n-\mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}$$

сходится к распределению $\mathsf{N}_{0;1}$ (в собственном смысле), а собственная сходимость, как известно, равносильна сходимости функций распределения (см. гл. 13), поэтому можно утверждать, что при $n\to\infty$

$$\mathsf{P}\left\{\frac{S_n-\mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}}<x\right\}\to\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-t^2/2}dt.$$

Часто ЦПТ именно в таком виде и формулируется.

15.3 Примеры и задачи

Примеры

Пример 15.3.1. *Известно, что вероятность p рождения мальчика составляет* 0,515. *Найти вероятность того, что из* $n = 10\,000$ *новорожденных детей мальчиков будет меньше, чем девочек.*

Решение. Обозначим через ξ количество мальчиков среди $n = 10\,000$ новорожденных, тогда девочек будет $10\,000 - \xi$. Необходимо вычислить $\mathsf{P}\{\xi < 10\,000 - \xi\} = \mathsf{P}\{\xi < 5000\}$.

Количество ξ мальчиков можно представить в виде $\sum\limits_{i=1}^{n} \mu_i$, где μ_i принимает значение 1, если i-й новорожденный — мальчик и 0 — если девочка. Случайные величины μ_i, $i = 1, 2, \ldots, n$, независимы, каждая с распределением

$$\mathsf{P}\{\mu_i = 1\} = p,\ \mathsf{P}\{\mu_i = 0\} = q.$$

Далее воспользуемся интегральной теоремой Муавра—Лапласа (см. следствие из теоремы 15.1.1). Поскольку число независимых слагаемых в сумме $\xi = \sum\limits_{i=1}^{n} \mu_i$ большое ($n = 10\,000$), то можно считать, что

$$\mathsf{P}\left\{\frac{\xi - np}{\sqrt{npq}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

И следовательно,

$$\mathsf{P}\{\xi < 5000\} =$$

$$= \mathsf{P}\left\{\frac{\xi - 5150}{\sqrt{10000 \cdot 0{,}515 \cdot 0{,}485}} < \frac{-150}{\sqrt{10000 \cdot 0{,}515 \cdot 0{,}485}}\right\} =$$

$$= \mathsf{P}\left\{\frac{\xi - 5150}{49{,}98} < -3{,}0\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{-3{,}0} \exp\left\{-\frac{t^2}{2}\right\} dt = 0{,}0013$$

(значение интеграла получено по таблице 27.1.1 нормального распределения).

Так что вероятность того, что среди 10 000 новорожденных детей мальчиков будет меньше, чем девочек, равна 0,0013.

Пример 15.3.2. *Театр, вмещающий* 1000 *человек, имеет два разных входа. Около каждого из входов имеется свой гардероб (количество мест в гардеробах одинаково). Сколько мест должно быть в гардеробах для того, чтобы в среднем в* 90 *случаях из* 100 *все зрители могли раздеться в гардеробе того входа, через который они вошли? Рассмотреть два случая: а) зрители приходят парами; б) зрители приходят поодиночке. Предположить, что входы зрители выбирают с равными вероятностями.*

Решение. Рассмотрим случай, когда зрители приходят поодиночке. Обозначим через k количество мест в каждом гардеробе. Пусть μ_i — случайная величина, принимающая значение 1, если i-й зритель выбрал вход 1, и 0 — если вход 2, $i = 1, 2, \ldots, 1000$. Случайные величины μ_i, $i = 1, 2, \ldots, 1000$, независимые, каждая с распределением

$$\mathsf{P}\{\mu_i = s\} = 1/2,\ s = 0, 1.$$

Тогда $\mu = \sum\limits_{i=1}^{1000} \mu_i$ — количество зрителей, которые вошли через вход 1, а $(1000 - \mu)$ — через вход 2.

Событие “все зрители имеют возможность оставить одежду в гардеробе” в терминах случайной величины μ запишется следующим образом:

$$\{\mu < k, 1000 - \mu < k\} = \{1000 - k < \mu < k\}.$$

Число мест k (в каждом гардеробе) находится как минимальное, удовлетворяющее условию

$$\mathsf{P}\{1000 - k < \mu < k\} \geq 0,90.$$

Чтобы определить k, воспользуемся интегральной теоремой Муавра—Лапласа (см. следствие из теоремы 15.1.1, а также пример 15.3.1).

Ответ: $k = 527$.

Задачи

15.1. Пусть случайная величина ξ_λ имеет распределение Пуассона с параметром λ.

Доказать, что при $\lambda \to \infty$

$$\mathsf{P}\left\{\frac{\xi_\lambda - \lambda}{\sqrt{\lambda}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt.$$

15.2. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые одинаково распределенные (каждая с распределением F) случайные величины со значениями в $\mathbb{R}^s$ и пусть

$$\mathbb{R}^s = \bigcup_{j=1}^{r} X_j,\ X_i \cap X_j = \emptyset,\ i \neq j;$$

$p_j = \mathsf{F}(X_j) = \mathsf{P}\{\xi_k \in X_j\}$ — вероятность, того, что случайная величина ξ_k попадет в X_j, $j = 1, 2, \ldots, r$, ν_j — число случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, попавших в X_j:

$$\nu_j = \sum_{k=1}^{n} I_{X_j}(\xi_k),\ j = 1, 2, \ldots, r.$$

Доказать, что при $n \to \infty$ распределение случайной величины

$$\sqrt{\frac{n}{p_j(1-p_j)}}\left(\frac{\nu_j}{n} - p_j\right)$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Указание. Воспользуйтесь центральной предельной теоремой.

Глава 16

Задачи математической статистики

Изложение основ математической статистики и ее связей с теорией вероятностей начнем с постановки типичных задач теории вероятностей и математической статистики на примере биномиального распределения.

Напомним, что биномиальное распределение с параметрами (n, p) определяется равенством

$$B_{n,p}(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

Оно, в частности, описывает распределение числа успехов в схеме испытаний Бернулли. Например, число выпадений герба в n независимых подбрасываниях монеты, вероятность выпадения герба которой равна p, имеет биномиальное распределение с параметрами (n, p).

Типичная задача теории вероятностей. Известна математическая модель $\{\Omega, \mathsf{P}\}$:

$$\Omega = \{0, 1, \ldots, n\},\ \mathsf{P}(k) = B_{n,p}(k),\ k = 0, 1, \ldots, n,$$

стохастического эксперимента, состоящего в подбрасывании n раз монеты, вероятность выпадения герба которой равна p. Вычислить вероятность выпадения четного числа гербов, нечетного, числа гербов, не превосходящего данного m — вычислить вероятность того или иного события по известной модели стохастического эксперимента — типичная задача теории вероятностей.

Типичная задача математической статистики. Стохастический эксперимент состоит в подбрасывании независимым образом n раз монеты, вероятность выпадения герба p которой неизвестна. Результат стохастического эксперимента известен — герб выпал μ раз, или, что то же, частота выпадения герба равна μ/n. Выбрать модель стохастического эксперимента по его исходу — по числу выпадения герба — типичная задача математической статистики применительно к биномиальному распределению.

В общем виде типичная задача математической статистики формулируется так — по исходу стохастического эксперимента предложить (выбрать) его математическую модель. К её решению возможны по меньшей мере два подхода, приводящие соответственно к задаче оценивания параметров распределения и к задаче проверки статистических гипотез.

Задача оценивания параметров распределения. Применительно к биномиальному распределению задача оценивания параметров распределения формулируется следующим образом. Монету, вероятность выпадения герба p которой неизвестна, независимым образом подбрасывают n раз и регистрируют число μ выпавших гербов. Случайная величина μ имеет биномиальное распределение:

$$\mathsf{P}\{\mu = k\} = B_{n,p}(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

По результату эксперимента — числу выпавших гербов μ (частоте μ/n появления герба) необходимо оценить (определить) неизвестный параметр p распределения $B_{n,p}$ (неизвестную вероятность p выпадения герба).

В общем виде задача оценивания параметров распределения $\mathsf{F}(\cdot;\theta)$ формулируется так. Функция распределения $F(x;\theta)$ (а вместе с ней и распределение $\mathsf{F}(\cdot;\theta)$) случайной величины ξ со значениями в $\mathbb{R}^n$ зависит от параметра θ (скалярного или векторного), значение которого неизвестно. По результату эксперимента, состоящего в наблюдении значений случайной величины ξ, имеющей распределение $\mathsf{F}(\cdot;\theta)$, необходимо определить (оценить) неизвестный параметр θ распределения.

Различают задачи точечного оценивания параметров и задачи интервального оценивания параметров. Если по результату наблюдений необходимо получить значение $\hat{\theta}$, которое будет использоваться в качестве параметра θ, то мы имеем дело с задачей точечного оценивания параметра, а если требуется получить интервал, который с той или иной вероятностью содержит неизвестное истинное значение параметра θ, это задача интервального оценивания параметра.

Задача проверки статистических гипотез. Другой подход к решению задачи выбора модели стохастического эксперимента по его исходу приводит к задаче проверки статистических гипотез. Применительно к биномиальному распределению эта задача формулируется так. Монету, вероятность выпадения герба p которой неизвестна, независимым образом подбрасывают n раз и регистрируют результат эксперимента — число μ выпавших гербов. Относительно неизвестного параметра p биномиального распределения

$$B_{n,p}(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n,$$

согласно которому распределена случайная величина μ, выдвигается гипотеза (предположение) — параметр p равен данному числу p_0, например, $1/2$. Задача проверки статистических гипотез — выяснить (проверить), согласуется ли выдвинутая гипотеза с исходом эксперимента "число успехов равно μ" или, что то же, "частота выпадения герба равна μ/n".

В общем виде задача проверки статистических гипотез формулируется следующим образом. Наблюдается случайная величина ξ со значениями в $\mathbb{R}^n$, распределение F которой неизвестно. Относительно распределения F выдвигается гипотеза: F совпадает с данным распределением G или, например, F принадлежит данному классу распределений, скажем, является нормальным распределением. Необходимо проверить, согласуется ли выдвинутая гипотеза с результатами наблюдений случайной величины.

З а м е ч а н и е. К задаче проверки статистических гипотез, помимо выбора модели стохастического эксперимента, сводится задача проверки адекватности модели стохастическому эксперименту. Эта задача естественным образом возникает на практике: теория вероятностей изучает стохастические эксперименты, исследуя их математические модели, предложенные исследователем из тех или иных соображений (рекомендаций по выбору моделей теория вероятностей не дает). Чтобы получить правильные выводы о стохастическом эксперименте, изучая его математическую модель, сначала необходимо убедиться, адекватно ли (правильно ли) модель его описывает.

Задача проверки адекватности модели стохастическому эксперименту сводится к задаче проверки статистических гипотез.

Например, рассматривается стохастический эксперимент, состоящий в подбрасывании монеты. Экспериментатор считает, что вероятность выпадения герба равна $\mathsf{P}(\Gamma) = 1/2$. Это, вообще говоря, ниоткуда не следует, и, скажем, гипотеза $\mathsf{P}(\Gamma) = 2/3$ ничем не хуже (как, впрочем, и не лучше) первой. Но чтобы модель можно было использовать для изучения эксперимента, она должна адекватно его описывать. Поэтому необходима проверка адекватности модели эксперименту. Для этого достаточно большое число раз проводят эксперимент и смотрят, согласуется ли предложенная модель с его результатами.

Резюме. В задачах теории вероятностей по известной математической модели стохастического эксперимента мы вычисляем вероятности тех или иных событий.

В задачах же математической статистики по результату эксперимента, сформулированному, как правило, в терминах частот, необходимо предложить (выбрать) математическую модель $\{\Omega, \mathfrak{F}, \mathsf{P}\}$ стохастического эксперимента.

У нас не будет трудностей с выбором множества исходов эксперимента и класса наблюдаемых событий — ими, как правило, будут $\mathbb{R}^n$ и $\mathfrak{B}^n$ (или подмножество $B \subset \mathbb{R}^n$ и $\mathfrak{B}^n_B$), чего не скажешь о выборе распределения на $\mathfrak{B}^n$. Так что выбор модели фактически сводится к выбору распределения вероятностей на $\mathbb{R}^n$.

В приведенном описании задач математической статистики слово “частота” упоминалось неоднократно. Поэтому при изложении математической статистики естественным математическим аппаратом будет дисциплина, изучающая частоты событий в стохастическом эксперименте, т. е. теория вероятностей.

Глава 17

Оценивание параметров распределения

17.1 Основные понятия и определения

Выборка. В математической статистике приняты следующие определения.

Определение. Случайный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ со значениями в $\mathbb{R}^n$ будем называть *выборкой* (*выборочным вектором*).

Выборку $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, образованную последовательностью независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n$, каждая с распределением G, будем называть *выборкой объемом n из распределения* (*закона*) G.

Значение $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ будем называть ее реализацией.

Множество X всех возможных значений (реализаций) выборки будем называть *выборочным пространством* (X — это $\mathbb{R}^n$ или его часть).

Мы будем иметь дело с выборками $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, распределения $\mathsf{F}(\cdot; \theta)$ (функции распределения $F(x; \theta)$) которых зависят от параметра θ, принимающего значения из некоторого множества Θ возможных значений, как правило, Θ — подмножество конечномерного пространства $\mathbb{R}^s$. Распределениями, зависящими от параметра, в частности, являются: нормальное распределение, его плотность

$$f(x; \theta) = f(x; a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\},$$

параметр $\theta = (a; \sigma^2)$ принимает значения из множества

$$\Theta = \{(a; \sigma^2); a \in \mathbb{R}^1, \sigma^2 > 0\} \subset \mathbb{R}^2;$$

равномерное на промежутке $[a, b]$ распределение, его плотность

$$f(x; \theta) = f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & \text{если } x \in [a, b]; \\ 0, & \text{если } x \notin [a, b], \end{cases}$$

параметр $\theta = (a; b)$ принимает значения из множества

$$\Theta = \{(a, b); a, b \in \mathbb{R}^1, a < b\} \subset \mathbb{R}^2;$$

пуассоновское распределение:

$$\mathsf{P}(k; \theta) = \frac{\theta^k}{k!} e^{-\theta}, \ k = 0, 1, \ldots,$$

параметр θ принимает значения из множества

$$\Theta = \{\theta : \theta > 0\} \subset \mathbb{R}^1.$$

Задача оценивания параметров распределения. Имеется реализация $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, распределение $\mathsf{F}(\cdot; \theta)$ которой зависит от параметра $\theta \in \Theta$. Параметр θ неизвестен. По реализации $\xi(\omega)$ выборки ξ (по выборке ξ) необходимо оценить (определить) параметр θ. Эта задача известна как задача оценивания параметра распределения.

Единственное, что у нас имеется (что нам известно) для определения неизвестного параметра θ, — это реализация $\xi(\omega)$ выборки ξ (кроме $\xi(\omega)$ мы не имеем ничего, что бы несло информацию о параметре θ). Поэтому "оценить (определить) θ по $\xi(\omega)$", точно или хотя бы приближенно, означает поставить в соответствие реализации $\xi(\omega)$ значение $\hat{\theta}$ из множества Θ возможных значений θ, которое мы будем использовать в качестве неизвестного нам параметра θ. Формально, для оценивания параметра θ на выборочном пространстве X — множестве реализаций выборки — необходимо определить (построить) функцию $h(\cdot)$ со значениями в Θ — множестве возможных значений параметра θ — такую, что значение

$$\hat{\theta} = h(\xi(\omega))$$

мы и будем использовать в качестве неизвестного параметра θ.

Определение. Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка, распределение $\mathsf{F}(\cdot; \theta)$ которой зависит от параметра $\theta \in \Theta$. Борелевскую функцию $h(\cdot)$, заданную на выборочном пространстве $X \subset \mathbb{R}^n$, со значениями в множестве Θ возможных значений параметра θ, будем называть *статистикой*, а случайную величину

$$\hat{\theta} = h(\xi) = h(\xi_1, \xi_2, \ldots, \xi_n)$$

— функцию от выборки — *оценкой.*

Получать (строить) статистики $h(\cdot)$, а вместе с ними и оценки $\hat{\theta} = h(\xi)$ параметра θ, точно (однозначно) определяющие неизвестный параметр θ:

$$\theta = \hat{\theta} = h(\xi),$$

заведомо не удастся уже хотя бы потому, что θ — константа, а оценка $\hat{\theta} = h(\xi)$, как функция от выборки — случайная величина (для каждой реализации $\xi(\omega)$ значение $\hat{\theta} = h(\xi(\omega))$, используемое в качестве θ, свое). Поэтому, нравится нам это или нет, но оценку $\hat{\theta} = h(\xi)$ параметра θ мы можем рассматривать только как его приближенное значение.

Далее оборот "$\hat{\theta} = h(\xi)$ — оценка параметра θ" означает, что $\hat{\theta} = h(\xi)$ используется в качестве приближенного значения параметра θ; оборот "оценить параметр θ" означает получить приближенное значение $\hat{\theta} = h(\xi)$ параметра θ; оборот "$\hat{\theta} = h(\xi)$ — оценка" означает, что $\hat{\theta} = h(\xi)$ — функция (борелевская) от выборки.

Для одного и того же параметра θ можно предложить не одну оценку (функций $\hat{\theta} = h(\xi)$ от выборки ξ со значениями в Θ много). Например, для оценивания параметра θ равномерного на промежутке $[\theta - 1, \theta + 1]$ распределения можно предложить следующие оценки

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} \xi_i; \;\; \hat{\theta}_2 = \frac{\xi_1 + \xi_n}{2}; \;\; \hat{\theta}_3 = \frac{1}{2}(\max\{\xi_i\} + \min\{\xi_i\}).$$

При этом (и не только здесь), естественно, возникают вопросы: "Какая из оценок лучше?", "Как строить хорошие оценки?", "Существует ли наилучшая оценка?"

Погрешности оценивания параметра. В связи с задачей оценивания неизвестного параметра θ распределения $\mathsf{F}(\cdot;\theta)$ по выборке ξ, как задачей получения приближенного значения $\hat\theta = h(\xi)$ (оценки) параметра θ, возникает вопрос о погрешности оценивания параметра θ оценкой $\hat\theta$ — о погрешности при замене θ его приближенным значением $\hat\theta = h(\xi)$. Говорить о $\hat\theta = h(\xi)$ как о приближенном значении параметра θ имеет смысл только тогда, когда мы имеем возможность оценить погрешность при замене θ на $\hat\theta$.

Пусть $\hat\theta$ — оценка параметра θ, т. е. $\hat\theta = h(\xi)$ рассматривается как приближенное значение параметра θ. В качестве погрешности при замене параметра θ его оценкой $\hat\theta = h(\xi)$ (в качестве количественной меры разброса значений оценки $\hat\theta$ относительно оцениваемого параметра θ) будем рассматривать величину

$$\mathsf{M}(\hat\theta - \theta)^2,$$

а если параметр θ векторный: $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$ — величину

$$\mathsf{M}\|\hat\theta - \theta\|^2 = \mathsf{M}\sum_{i=1}^{s}(\hat\theta_i - \theta_i)^2 = \sum_{i=1}^{s}\mathsf{M}(\hat\theta_i - \theta_i)^2.$$

Перепишем $\mathsf{M}(\hat\theta - \theta)^2$ так:

$$\mathsf{M}(\hat\theta - \theta)^2 = \mathsf{M}\left((\hat\theta - \mathsf{M}\hat\theta) + (\mathsf{M}\hat\theta - \theta)\right)^2 =$$

$$= \mathsf{M}((\hat\theta - \mathsf{M}\hat\theta)^2 + 2(\hat\theta - \mathsf{M}\hat\theta)(\mathsf{M}\hat\theta - \theta) + (\mathsf{M}\hat\theta - \theta)^2) =$$

$$= \mathsf{D}\hat\theta + (\mathsf{M}\hat\theta - \theta)^2,$$

т. е.

$$\mathsf{M}(\hat\theta - \theta)^2 = \mathsf{D}\hat\theta + (\mathsf{M}\hat\theta - \theta)^2. \tag{17.1.1}$$

Таким образом, величина разброса $\mathsf{M}(\hat\theta - \theta)^2$ оценки $\hat\theta$ относительно оцениваемого параметра θ (погрешность при замене θ на $\hat\theta$) складывается из величины разброса самой оценки $\mathsf{M}(\hat\theta - \mathsf{M}\hat\theta)^2 = \mathsf{D}\hat\theta$ и величины $(\mathsf{M}\hat\theta - \theta)^2$.

Для оценивания параметра θ естественно использовать оценки $\hat\theta$, разброс $\mathsf{M}(\hat\theta - \theta)^2$ которых относительно θ (погрешность при замене θ на $\hat\theta$) как можно меньше. Из равенства

$$\mathsf{M}(\hat\theta - \theta)^2 = \mathsf{D}\hat\theta + (\mathsf{M}\hat\theta - \theta)^2$$

следует, что для этого от оценки $\hat{\theta}$ параметра θ необходимо потребовать, чтобы $(\mathsf{M}\hat{\theta} - \theta)^2 = 0$ или, что то же, $\mathsf{M}\hat{\theta} = \theta$, и чтобы как можно меньшим был разброс $\mathsf{D}\hat{\theta}$ самой оценки $\hat{\theta}$.

В связи с этим вводятся следующие определения.

Определение. Оценку $\hat{\theta}$ будем называть *несмещенной оценкой параметра* θ, если

$$\mathsf{M}\hat{\theta} = \theta.$$

Наглядное содержание свойства несмещенности оценки $\hat{\theta}$ параметра θ состоит в том, что при многократном использовании $\hat{\theta}$ в качестве значения θ среднее значение погрешности $\hat{\theta} - \theta$ равно нулю.

Часто мы имеем возможность "наращивать" объем выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ и в связи с этим рассматривать не одну оценку $\hat{\theta} = h(\xi) = h(\xi_1, \xi_2, \ldots, \xi_n)$, а последовательность оценок $\hat{\theta}_n = h_n(\xi_1, \xi_2, \ldots, \xi_n)$, $n = 1, 2, \ldots$ В этом случае естественно говорить об их *асимптотическом поведении.*

Определение. Последовательность

$$\hat{\theta}_n = h_n(\xi_1, \xi_2, \ldots, \xi_n),\ n = 1, 2, \ldots,$$

будем называть *асимптотически несмещенной последовательностью оценок параметра* θ, если при $n \to \infty$

$$\mathsf{M}\hat{\theta}_n \to \theta.$$

Определение. Последовательность

$$\hat{\theta}_n = h_n(\xi_1, \xi_2, \ldots, \xi_n),\ n = 1, 2, \ldots,$$

будем называть *состоятельной последовательностью оценок параметра* θ, если для каждого $\varepsilon > 0$ при $n \to \infty$

$$\mathsf{P}\{|\hat{\theta}_n - \theta| > \varepsilon\} \to 0,$$

другими словами, если последовательность $\hat{\theta}_n$ сходится по вероятности к θ.

Сходимость по вероятности $\hat{\theta}_n$ к θ будем обозначать так

$$\hat{\theta}_n \xrightarrow{\mathsf{P}} \theta.$$

Если последовательность несмещенных оценок $\hat{\theta}_n$ параметра θ такая, что $\mathsf{D}\hat{\theta}_n \to 0$, то $\hat{\theta}_n$ является состоятельной последовательностью оценок параметра θ. Это следует из неравенства Чебышёва:

$$\mathsf{P}\{|\hat{\theta}_n - \theta| > \varepsilon\} \leq \frac{\mathsf{D}\hat{\theta}_n}{\varepsilon^2}.$$

Замечание 1. К задаче оценивания параметров распределения сводится более общая задача — определение неизвестного распределения F. Основной подход к ее решению — подбор F из некоторого параметрического семейства распределений. Предполагается, что функция распределения $F(x)$ определяется небольшим числом (обычно не более трех-четырех) параметров, которые подлежат оцениванию. Например, нормальное распределение определяется двумя параметрами: математическим ожиданием и средним квадратическим отклонением, пуассоновское — одним. Чаще всего в качестве параметров выбирают моменты

$$a_k = \int\limits_{-\infty}^{+\infty} x^k \mathsf{F}(dx),\ k = 1, 2, \ldots$$

Замечание 2. Согласно сложившейся в математической статистике традиции, выборкой называют не только случайный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, но и его реализацию, употребляя к тому же одни и те же обозначения как для случайного вектора ξ, так и для его реализации.

С интуитивной точки зрения, до проведения эксперимента, выборка — это случайный вектор, после проведения — реализация случайного вектора, хотя по-прежнему мы будем называть его выборкой.

Например, в эксперименте девять раз независимым образом измеряется точка плавления железа. “Будущий результат” эксперимента — случайный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_9)$, результат эксперимента (проведенного) — последовательность чисел: 1493, 1519, 1518, 1514, 1489, 1508, 1485, 1503, 1497 — реализация $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_9(\omega))$ случайного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_9)$.

Пример. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью*

$$p(x) = \begin{cases} 0, & \text{если} \quad x \leq \alpha, \\ e^{-(x-\alpha)}, & \text{если} \quad x > \alpha. \end{cases}$$

Является ли $\hat{\alpha} = -1/n + \min\{\xi_1, \xi_2, \ldots, \xi_n\}$ *несмещенной и состоятельной последовательностью оценок параметра* α?

Решение. По определению $\hat{\alpha}$ является несмещенной оценкой параметра α, если $\mathsf{M}\hat{\alpha} = \alpha$. Убедимся, что $\mathsf{M}\hat{\alpha} = \alpha$.

Обозначим $\zeta = \min\{\xi_1, \xi_2, \ldots, \xi_n\}$. Тогда

$$\mathsf{M}\hat{\alpha} = \mathsf{M}\left(-\frac{1}{n} + \zeta\right) = -\frac{1}{n} + \mathsf{M}\zeta.$$

Вычислим $\mathsf{M}\zeta$. Для этого найдем плотность распределения случайной величины $\zeta = \min\{\xi_1, \xi_2, \ldots, \xi_n\}$. Сначала найдем функцию распределения $F_\zeta(x)$. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и их плотности известны:

$$p_{\xi_k}(x) = p(x) = \begin{cases} 0, & \text{если} \quad x \leq \alpha, \\ e^{-(x-\alpha)}, & \text{если} \quad x > \alpha, \end{cases}$$

$k = 1, 2, \ldots, n$. Отсюда функция распределения случайной величины ξ_k

$$F(x) = \mathsf{P}\{\xi_k < x\} = \int\limits_{-\infty}^{x} p(t)dt = \begin{cases} 0, & \text{если} \quad x \leq \alpha, \\ 1 - e^{-(x-\alpha)}, & \text{если} \quad x > \alpha. \end{cases}$$

Функция распределения $F_\zeta(x)$ случайной величины ζ:

$$F_\zeta(x) = 1 - \mathsf{P}\{\zeta \geq x\} = 1 - \mathsf{P}\{\min\{\xi_1, \xi_2, \ldots, \xi_n\} \geq x\} =$$

$$= 1 - \mathsf{P}\left(\bigcap_{i=1}^{n}\{\xi_i \geq x\}\right) = 1 - \prod_{i=1}^{n} \mathsf{P}\{\xi_i \geq x\} =$$

$$= 1 - \prod_{i=1}^{n}(1 - F(x)) = 1 - (1 - F(x))^n =$$

$$= \begin{cases} 0, & \text{если} \quad x \leq \alpha, \\ 1 - e^{-n(x-\alpha)}, & \text{если} \quad x > \alpha; \end{cases}$$

$F_\zeta(x)$ — абсолютно непрерывна, поэтому случайная величина ζ имеет плотность распределения и она может быть получена дифференцированием функции распределения:

$$p_\zeta(x) = \frac{d}{dx}F_\zeta(x) = \begin{cases} 0, & \text{если} \quad x \leq \alpha, \\ ne^{-n(x-\alpha)}, & \text{если} \quad x > \alpha. \end{cases}$$

Отсюда

$$\mathsf{M}\zeta = \int\limits_{\alpha}^{\infty} x p_\zeta(x) dx = \int\limits_{\alpha}^{\infty} x n e^{-n(x-\alpha)} dx = \alpha + \frac{1}{n},$$

$$\mathsf{M}\hat{\alpha} = -\frac{1}{n} + \mathsf{M}\zeta = -\frac{1}{n} + \alpha + \frac{1}{n} = \alpha,$$

т. е. $\hat{\alpha}$ является несмещенной оценкой α.

Выясним теперь, является ли $\hat{\alpha}$ состоятельной оценкой параметра α, т. е. будет ли $\hat{\alpha}$ сходиться по вероятности к α при $n \to \infty$.

Пусть $\varepsilon > 0$. Вычислим $\mathsf{P}\{|\hat{\alpha} - \alpha| \geq \varepsilon\}$. При достаточно больших n

$$\mathsf{P}\{|\hat{\alpha} - \alpha| \geq \varepsilon\} = \mathsf{P}\{|\zeta - (\alpha + 1/n)| \geq \varepsilon\} = \int\limits_{|x-(\alpha+1/n)| \geq \varepsilon} p_\zeta(x) dx =$$

$$= \int\limits_{x \leq \alpha+1/n-\varepsilon} p_\zeta(x) dx + \int\limits_{x \geq \alpha+1/n+\varepsilon} p_\zeta(x) dx = \int\limits_{\alpha+1/n+\varepsilon}^{\infty} n e^{-n(x-\alpha)} dx = e^{-n\varepsilon-1}.$$

Отсюда следует, что $\mathsf{P}\{|\hat{\alpha} - \alpha| \geq \varepsilon\}$ стремится к нулю при $n \to \infty$, т. е. $\hat{\alpha}$ является состоятельной оценкой параметра α.

17.2 Примеры и задачи

Примеры

Пример 17.2.1. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью*

$$f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & \text{если } x \in [a, b]; \\ 0, & \text{если } x \notin [a, b]. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \min\{\xi_1, \xi_2, \ldots, \xi_n\}, \ \ \hat{\theta}_2 = \max\{\xi_1, \xi_2, \ldots, \xi_n\},$$

$$\hat{\theta}_3 = \frac{1}{n}\sum_{i=1}^{n} \xi_i, \ \ \hat{\theta}_4 = \frac{\xi_{n-1} + \xi_n}{2}.$$

Какие из оценок $\hat\theta_1, \hat\theta_2, \hat\theta_3, \hat\theta_4$ и для каких параметров (возможно, отличных от a и b) являются несмещенными оценками? Состоятельными оценками?

Решение. Рассмотрим каждую из оценок.

Оценка $\hat\theta_2 = \max\{\xi_1, \xi_2, \ldots, \xi_n\}$. Сначала найдем распределение $\hat\theta_2$. По условию $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из равномерного на отрезке $[a, b]$ распределения, т. е. $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, каждая из которых распределена равномерно на отрезке $[a, b]$. Отсюда, учитывая, что случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ абсолютно непрерывны с плотностями $f(x; a, b)$, имеем:

$$F_{\hat\theta_2}(x) = \mathsf{P}\{\hat\theta_2 < x\} = \mathsf{P}\{\max\{\xi_1, \xi_2, \ldots, \xi_n\} < x\} =$$

$$= \mathsf{P}\{\xi_1 < x, \xi_2 < x, \ldots, \xi_n < x\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i < x\} =$$

$$= \prod_{i=1}^{n} \int\limits_{-\infty}^{x} f(y; a, b)\, dy = \left(\int\limits_{-\infty}^{x} f(y; a, b)\, dy \right)^{n}.$$

Поэтому $\hat\theta_2$ — абсолютно непрерывная случайная величина и ее плотность распределения

$$f_{\hat\theta_2}(x) = \frac{d}{dx} F_{\hat\theta_2}(x) = n f(x; a, b) \left(\int\limits_{-\infty}^{x} f(y; a, b)\, dy \right)^{n-1}.$$

А так как

$$\int\limits_{-\infty}^{x} f(y; a, b)\, dy = \begin{cases} 0, & \text{если } x < a; \\ \dfrac{x-a}{b-a}, & \text{если } a \le x \le b; \\ 1, & \text{если } x > b, \end{cases}$$

то

$$f_{\hat\theta_2}(x) = \begin{cases} 0, & \text{если } x \notin [a, b]; \\ \dfrac{n}{(b-a)^n}(x-a)^{n-1}, & \text{если } x \in [a, b]. \end{cases}$$

График плотности $f_{\hat\theta_2}(x)$ изображен на рис. 17.2.1.

По известной плотности распределения оценки $\hat{\theta}_2$ вычисляем ее математическое ожидание:

$$\mathsf{M}\hat{\theta}_2 = \int\limits_{-\infty}^{\infty} x f_{\hat{\theta}_2}(x)\,dx =$$

$$= \int\limits_a^b x\frac{n}{(b-a)^n}(x-a)^{n-1}\,dx = \frac{n}{n+1}b + \frac{a}{n+1}.$$

Так что оценка $\hat{\theta}_2$ не является несмещенной оценкой ни параметра a, ни параметра b. Но при $n \to \infty$

$$\mathsf{M}\hat{\theta}_2 = \frac{n}{n+1}b + \frac{a}{n+1} \to b.$$

Последнее означает, что $\hat{\theta}_2$ является асимптотически несмещенной оценкой параметра b.

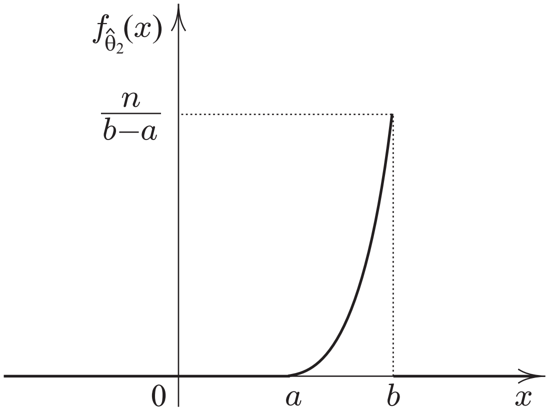

Рис. 17.2.1: Плотность распределения оценки $\hat{\theta}_2$

Выясним, является ли $\hat{\theta}_2$ состоятельной оценкой параметра b, т. е. сходится ли $\hat{\theta}_2$ по вероятности к b. Для достаточно малых ε имеем:

$$\mathsf{P}\{|\hat{\theta}_2 - b| > \varepsilon\} = \mathsf{P}\{\hat{\theta}_2 \in (-\infty, b-\varepsilon) \cup (b+\varepsilon, +\infty)\} =$$

$$= \int\limits_{-\infty}^{b-\varepsilon} f_{\hat{\theta}_2}(x)\,dx + \int\limits_{b+\varepsilon}^{+\infty} f_{\hat{\theta}_2}(x)\,dx = \int\limits_{a}^{b-\varepsilon} f_{\hat{\theta}_2}(x)\,dx =$$

$$= \int\limits_{a}^{b-\varepsilon} \frac{n}{(b-a)^n}(x-a)^{n-1}\,dx = \left(1 - \frac{\varepsilon}{b-a}\right)^n,$$

что сходится к 0 при $n \to \infty$. Следовательно, при $n \to \infty$

$$\mathsf{P}\{|\hat{\theta}_2 - b| > \varepsilon\} \to 0.$$

Последнее означает, что $\hat{\theta}_2$ является состоятельной оценкой параметра b.

Оценка $\hat{\theta}_1 = \min\{\xi_1, \xi_2, \ldots, \xi_n\}$. Оценка $\hat{\theta}_1$ исследуется аналогично оценке $\hat{\theta}_2$.

Распределение оценки $\hat{\theta}_1$ абсолютно непрерывное с плотностью

$$f_{\hat{\theta}_1}(x) = \begin{cases} 0, & \text{если } x \notin [a,b]; \\ \dfrac{n}{(b-a)^n}(b-x)^{n-1}, & \text{если } x \in [a,b]. \end{cases}$$

Отсюда

$$\mathsf{M}\hat{\theta}_1 = \frac{n}{n+1}a + \frac{b}{n+1}.$$

Поэтому $\hat{\theta}_1$ не является несмещенной оценкой ни параметра a, ни параметра b, но она является асимптотически несмещенной оценкой параметра a.

Оценка $\hat{\theta}_1$ является состоятельной оценкой параметра a.

Оценка $\hat{\theta}_3 = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$. Поскольку ξ_i, $i = 1, 2, \ldots, n$, — равномерно распределенные на отрезке $[a,b]$ случайные величины, а для них математическое ожидание m_1, как известно, равно $(a+b)/2$, то

$$\mathsf{M}\hat{\theta}_3 = \mathsf{M}\frac{1}{n}\sum_{i=1}^{n}\xi_i = m_1 = \frac{a+b}{2}.$$

Следовательно, $\hat{\theta}_3$ не является несмещенной оценкой ни параметра a, ни параметра b, но $\hat{\theta}_3$ — несмещенная оценка параметра

$m_1 = (a+b)/2$ — математического ожидания равномерного на отрезке $[a,b]$ распределения.

Далее, согласно закону больших чисел, $\hat\theta_3 = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ сходится по вероятности к $m_1 = (a+b)/2$, т. е. $\hat\theta_3$ — состоятельная оценка параметра $m_1 = (a+b)/2$.

Оценка $\hat\theta_4 = (\xi_{n-1}+\xi_n)/2$. Для оценки $\hat\theta_4$

$$\mathsf{M}\hat\theta_4 = \mathsf{M}\frac{\xi_{n-1}+\xi_n}{2} = \frac{1}{2}(\mathsf{M}\xi_{n-1}+\mathsf{M}\xi_n) = \frac{a+b}{2}.$$

Следовательно, $\hat\theta_4$ — несмещенная оценка параметра $m_1 = (a+b)/2$.

Оценка $\hat\theta_4$ не сходится по вероятности ни к одной константе (параметру). В самом деле, плотность распределения случайной величины $\hat\theta_4 = (\xi_{n-1}+\xi_n)/2$

$$f_{\hat\theta_4}(x) = \begin{cases} 0, & \text{если } x \notin [a,b]; \\ \dfrac{4}{(b-a)^2}(x-a), & \text{если } x \in [a, \dfrac{a+b}{2}]; \\ \dfrac{4}{(b-a)^2}(b-x), & \text{если } x \in [\dfrac{a+b}{2}, b] \end{cases}$$

(сначала получена плотность $p_\zeta(x)$ суммы $\zeta = \xi_{n-1}+\xi_n$, как свертка плотностей равномерных на отрезке $[a,b]$ распределений, а затем плотность $f_{\hat\theta_4}(x)$ случайной величины $\hat\theta_4 = \zeta/2$, график $f_{\hat\theta_4}(x)$ изображен на рис. 17.2.2).

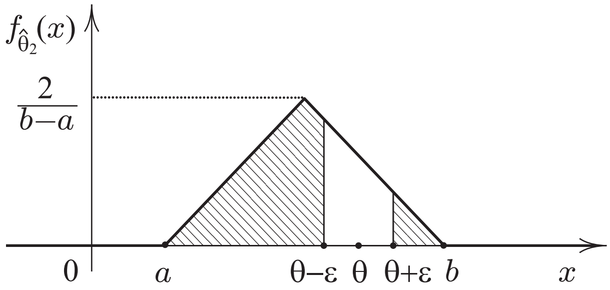

Рис. 17.2.2: Плотность распределения оценки $\hat\theta_4$

И если предположить, что $\hat{\theta}_4$ сходится по вероятности к некоторой константе θ (см. рис. 17.2.2), то для каждого достаточно малого $\varepsilon > 0$ при $n \to \infty$ вероятность

$$\mathsf{P}\{|\hat{\theta}_4 - \theta| > \varepsilon\} = \int\limits_{\{x:|x-\theta|>\varepsilon\}} f_{\hat{\theta}_4}(x)dx$$

должна стремится к нулю, но $\mathsf{P}\{|\hat{\theta}_4 - \theta| > \varepsilon\} > 0$ и не завит от n (численно эта вероятность равна заштрихованной площади, см. рис. 17.2.2) и, следовательно, не сходится к нулю при $n \to \infty$.

Задачи

17.1. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;\theta,b) = \begin{cases} \dfrac{1}{\theta}\exp\{-\dfrac{1}{\theta}(x-b)\}, & \text{если } x > b; \\ 0, & \text{если } x \le b. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n}\xi_i; \;\; \hat{\theta}_2 = \min\{\xi_i\}; \;\; \hat{\theta}_3 = \hat{\theta}_1 - \hat{\theta}_2.$$

Имеются ли среди $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ несмещенные, состоятельные оценки параметров θ и b? Возможно, среди них имеются несмещенные и состоятельные оценки других параметров? Каких именно?

17.2. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;\theta) = \begin{cases} \dfrac{1}{2h_0}, & \text{если } x \in [\theta - h_0, \theta + h_0]; \\ 0, & \text{если } x \notin [\theta - h_0, \theta + h_0], \end{cases}$$

h_0 — известно.

Рассматриваются оценки

$$\hat{\theta}_1 = \min\{\xi_i\}; \; \hat{\theta}_2 = \max\{\xi_i\}; \; \hat{\theta}_3 = \frac{1}{n}\sum_{i=1}^{n}\xi_i;$$

$$\hat{\theta}_\lambda = \lambda(\max\{\xi_i\} - h_0) + (1-\lambda)(\min\{\xi_i\} + h_0), \;\; \lambda \in [0;1].$$

Выяснить, какие из оценок $\hat{\theta}_i$ $(i = 1, 2, 3)$, $\hat{\theta}_\lambda$ несмещенные, состоятельные оценки параметра θ. Возможно, среди них имеются несмещенные, состоятельные оценки других параметров? Каких именно?

17.3. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, h) = \begin{cases} \dfrac{1}{2h}, & \text{если } x \in [\theta - h, \theta + h]; \\ 0, & \text{если } x \notin [\theta - h, \theta + h]. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \frac{1}{2}\left(\max\{\xi_i\} - \min\{\xi_i\}\right); \ \hat{\theta}_2 = \frac{1}{2}(\max\{\xi_i\} + \min\{\xi_i\});$$

$$\hat{\theta}_3 = \min\{\xi_i\} - \frac{1}{n-1}(\max\{\xi_i\} - \min\{\xi_i\});$$

$$\hat{\theta}_4 = \max\{\xi_i\} + \frac{1}{n-1}(\max\{\xi_i\} - \min\{\xi_i\}).$$

Выяснить, какие из оценок $\hat{\theta}_i$, $i = 1, 2, 3, 4$, несмещенные, состоятельные оценки параметров θ и h. Возможно, среди них имеются состоятельные оценки других параметров? Каких именно?

17.4. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из равномерного на отрезке $[a; a+1]$ распределения. Рассматриваются две оценки параметра a:

$$\hat{a}_1 = \frac{1}{n}\sum_{i=1}^{n} \xi_i - \frac{1}{2}; \ \hat{a}_2 = \max\{\xi_i\} - \frac{n}{n+1}.$$

Доказать, что $\hat{a}_1$ и $\hat{a}_2$ являются несмещенными оценками параметра a.

Найти $\mathsf{D}\hat{a}_1$ и $\mathsf{D}\hat{a}_2$. Доказать, что

$$\mathsf{D}\hat{a}_2 = o(\mathsf{D}\hat{a}_1),$$

при $n \to \infty$.

Глава 18

Оценки с минимальной дисперсией

Распределение $\mathsf{F}(\cdot;\theta)$, выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ зависит от неизвестного параметра $\theta \in \Theta$. Пусть

$$\hat{\theta} = h(\xi)$$

— оценка параметра θ — приближенное значение, полученное по выборке $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$. Погрешность при замене параметра θ его оценкой $\hat{\theta}$ мы измеряем величиной $\mathsf{M}(\hat{\theta} - \theta)^2$, ее можно представить в виде

$$\mathsf{M}(\hat{\theta} - \theta)^2 = \mathsf{D}\hat{\theta} + (\mathsf{M}\hat{\theta} - \theta)^2,$$

см. (17.1.1). Отсюда следует, что для оценивания параметра θ, во-первых, необходимо использовать несмещенные оценки параметра θ, т. е. такие оценки, для которых

$$\mathsf{M}\hat{\theta} = \theta$$

(если $\mathsf{M}\hat{\theta} \neq \theta$, то погрешность при замене θ на $\hat{\theta}$ заведомо больше чем у несмещенной оценки), и, во-вторых, среди несмещенных оценок $\hat{\theta}$ параметра θ следует выбирать те, которые имеют по возможности меньшую дисперсию $\mathsf{D}\hat{\theta}$, а в идеале минимально возможную.

Здесь, естественно, сразу же возникает задача — а насколько малыми могут быть дисперсии оценок $\hat{\theta}$ параметра θ?

Оказывается, что если распределение $\mathsf{F}(\cdot;\theta)$ выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ достаточно регулярно зависит от параметра θ, то можно указать нижнюю границу дисперсий всех несмещенных оценок $\hat{\theta} = h(\xi)$ параметра θ. Эта граница определяется распределением $\mathsf{F}(\cdot;\theta)$ выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ и дается в форме *неравенства Крамера—Рао.*

18.1 Неравенство Крамера—Рао

Напомним условие дифференцируемости функции $f(x;\theta)$, $(x,\theta) \in \mathbb{R}^n \times (a,b)$, по параметру θ под знаком интеграла.

Теорема 18.1.1 *Если функция $f(x;\theta)$, $(x,\theta) \in \mathbb{R}^n \times (a,b)$, дифференцируема по θ для почти всех $x \in \mathbb{R}^n$ и производная $\dfrac{\partial}{\partial\theta} f(x;\theta)$ мажорируема интегрируемой функцией:*

$$\left|\frac{\partial}{\partial\theta} f(x;\theta)\right| \leq Y(x), \ \int\limits_{\mathbb{R}^n} Y(x)dx < \infty,$$

то

$$\frac{\partial}{\partial\theta} \int\limits_{\mathbb{R}^n} f(x;\theta)dx = \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(x;\theta)dx.$$

Действительно,

$$\frac{1}{h}\left(\int\limits_{\mathbb{R}^n} f(x;\theta+h)dx - \int\limits_{\mathbb{R}^n} f(x;\theta)dx\right) = \int\limits_{\mathbb{R}^n} \frac{f(x;\theta+h) - f(x;\theta)}{h} dx. \tag{18.1.1}$$

В силу теоремы о среднем

$$\left|\frac{f(x;\theta+h) - f(x;\theta)}{h}\right| = \left|\frac{\partial}{\partial t} f(x;t)|_{t=\theta+sh}\right| \leq Y(x), \ 0 \leq s \leq 1,$$

поэтому при $h \to 0$ согласно теореме Лебега о мажорируемой сходимости существует предел правой части равенства (18.1.1):

$$\lim_{h\to 0} \int\limits_{\mathbb{R}^n} \frac{f(x;\theta+h) - f(x;\theta)}{h} dx = \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(x;\theta)dx,$$

а следовательно, существует предел и левой части. Переходя в равенстве (18.1.1) к пределу при $h \to 0$, получим:

$$\frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} f(x;\theta)dx = \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(x;\theta)dx.$$

Следствие. *Если функция $f(x;\theta)$, $(x,\theta) \in \mathbb{R}^n \times (a,b)$, m раз дифференцируема по параметру θ для почти всех $x \in \mathbb{R}^n$ и производные $\frac{\partial^k}{\partial\theta^k} f(x;\theta)$ мажорируемы интегрируемыми функциями:*

$$\left|\frac{\partial^k}{\partial\theta^k} f(x;\theta)\right| \le Y_k(x),\ \int\limits_{\mathbb{R}^n} Y_k(x)dx < \infty, x \in \mathbb{R}^n,\ k = 1,2,\ldots,m,$$

то

$$\frac{\partial^k}{\partial\theta^k}\int\limits_{\mathbb{R}^n} f(x;\theta)dx = \int\limits_{\mathbb{R}^n} \frac{\partial^k}{\partial\theta^k} f(x;\theta)dx,\ k = 1,2,\ldots,m.$$

Далее мы будем рассматривать выборку $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ фиксированного объема n и будем предполагать, что её распределение абсолютно непрерывно, плотность распределения выборки будем обозначать через $f(x;\theta) = f(x_1, x_2, \ldots, x_n;\theta)$, параметр θ будем считать одномерным, а в качестве множества его возможных значений Θ будем рассматривать конечный интервал.

Определение. Функцию

$$I(\theta) = \mathsf{D}\frac{\partial}{\partial\theta}\ln f(\xi;\theta),\ \theta \in \Theta,$$

будем называть *количеством информации по Фишеру.*

Теорема 18.1.2 (неравенство Крамера—Рао). *Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — абсолютно непрерывный выборочный вектор с плотностью $f(x;\theta)$, $x \in \mathbb{R}^n$, $\theta \in \Theta$, и $\hat{\theta} = h(\xi)$ — несмещенная оценка параметра θ.*

Если функции $f(x;\theta)$ и $h(x)f(x;\theta)$ удовлетворяют условиям дифференцируемости по параметру $\theta \in \Theta$ под знаком интеграла, причем $f(x;\theta)$ — дифференцируемости дважды (см. теорему 18.1.1), и

$$\mathsf{M}\left|\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta)\right| < \infty,$$

то $I(\theta) < \infty$ *и*

$$\mathsf{D}\hat{\theta} \geq \frac{1}{I(\theta)}, \tag{18.1.2}$$

причем равенство достигается тогда и только тогда, когда

$$\frac{\partial}{\partial\theta} \ln f(x;\theta) = C(\theta)(h(x) - \theta).$$

Доказательство. Сначала установим, что в условиях теоремы при всех $\theta \in \Theta$

$$\mathsf{M}\frac{\partial}{\partial\theta} \ln f(\xi;\theta) = 0,$$

$$I(\theta) = \mathsf{D}\frac{\partial}{\partial\theta} \ln f(\xi;\theta) = -\mathsf{M}\frac{\partial^2}{\partial\theta^2} \ln f(\xi;\theta) < \infty. \tag{18.1.3}$$

Для плотности вероятностного распределения $f(x;\theta) = f(x_1, x_2, \ldots, x_n;\theta)$, $x \in \mathbb{R}^n$, $\theta \in \Theta$, всегда имеет место равенство

$$\int\limits_{\mathbb{R}^n} f(x;\theta)dx = 1.$$

И коль скоро $f(x;\theta)$ удовлетворяет условиям дифференцируемости по параметру $\theta \in \Theta$ под знаком интеграла (см. теорему 18.1.1), то имеет место равенство

$$\int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f(x;\theta)dx = 0, \tag{18.1.4}$$

а вместе с ним и равенство

$$\int\limits_{\mathbb{R}^n} \left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right) f(x;\theta)dx = 0,$$

которое в терминах математического ожидания можно записать так

$$\mathsf{M}\frac{\partial}{\partial\theta} \ln f(\xi;\theta) = 0. \tag{18.1.5}$$

Так что последнее равенство всегда имеет место, лишь бы $f(x;\theta)$ можно было дифференцировать по параметру под знаком интеграла.

Плотность $f(x;\theta)$ удовлетворяет условию дифференцируемости дважды по параметру θ под знаком интеграла, поэтому из (18.1.4) следует

$$\int\limits_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta^2} f(x;\theta)dx = 0.$$

По условию теоремы $\mathsf{M}\dfrac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta)$ конечно, поэтому в силу теоремы о вычислении математического ожидания функций от случайной величины $\xi = (\xi_1, \xi_2 \ldots, \xi_n)$ (см. теорему 10.1.4)

$$\mathsf{M}\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta) = \int\limits_{\mathbb{R}^n}\left(\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta)\right) f(x;\theta)dx =$$

$$= \int\limits_{\mathbb{R}^n}\left(\frac{\partial}{\partial\theta}\left(\frac{f'_\theta(x;\theta)}{f(x;\theta)}\right)\right) f(x;\theta)dx =$$

$$= \int\limits_{\mathbb{R}^n} \frac{f''_{\theta\theta}(x;\theta)f(x;\theta) - \left(f'_\theta(x;\theta)\right)^2}{f^2(x;\theta)} f(x;\theta)dx =$$

$$= \int\limits_{\mathbb{R}^n} f''_{\theta\theta}(x;\theta)dx - \int\limits_{\mathbb{R}^n}\left(\frac{f'_\theta(x;\theta)}{f(x;\theta)}\right)^2 f(x;\theta)dx =$$

$$= -\int\limits_{\mathbb{R}^n}\left(\frac{f'_\theta(x;\theta)}{f(x;\theta)}\right)^2 f(x;\theta)dx =$$

$$= -\int\limits_{\mathbb{R}^n}\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)^2 f(x;\theta)dx = -\mathsf{M}\left(\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right)^2.$$

Так что

$$I(\theta) = \mathsf{D}\frac{\partial}{\partial\theta}\ln f(\xi;\theta) =$$

$$= \mathsf{M}\left(\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right)^2 = -\mathsf{M}\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta) < \infty.$$

Далее, по условию теоремы $\hat{\theta} = h(\xi)$ — несмещенная оценка параметра θ, т. е.

$$\theta = \mathsf{M}\hat{\theta} = \mathsf{M}h(\xi)$$

или

$$\theta = \int\limits_{\mathbb{R}^n} h(x)f(x;\theta)dx. \tag{18.1.6}$$

Почленно продифференцируем это равенство (условия теоремы обеспечивают дифференцирование в правой части под знаком интеграла):

$$1 = \int\limits_{\mathbb{R}^n} h(x)f'_\theta(x;\theta)dx.$$

Преобразуем правую часть полученного равенства:

$$\int\limits_{\mathbb{R}^n} h(x)f'_\theta(x;\theta)dx = \int\limits_{\mathbb{R}^n} \left(h(x)\frac{\partial}{\partial\theta}\ln f(x;\theta)\right) f(x;\theta)dx =$$

$$= \mathsf{M}h(\xi)\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = \mathsf{M}\hat{\theta}\frac{\partial}{\partial\theta}\ln f(\xi;\theta).$$

Так что из (18.1.6) имеем

$$\mathsf{M}\hat{\theta}\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = 1, \tag{18.1.7}$$

а из (18.1.5) —

$$\mathsf{M}\theta\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = 0. \tag{18.1.8}$$

Вычитая почленно из равенства (18.1.7) равенство (18.1.8), получаем:

$$\mathsf{M}(\hat{\theta}-\theta)\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = 1.$$

Отсюда, воспользовавшись неравенством Коши—Буняковского, имеем:

$$1 = \left(\mathsf{M}(\hat{\theta}-\theta)\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right)^2 \le$$

$$\le \mathsf{M}(\hat{\theta}-\theta)^2\mathsf{M}\left(\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right)^2 = \mathsf{D}\hat{\theta}I(\theta), \tag{18.1.9}$$

или

$$\mathsf{D}\hat{\theta} \geq \frac{1}{I(\theta)}.$$

Заметим, что неравенство (18.1.9) обращается в равенство тогда и только тогда, когда

$$\frac{\partial}{\partial\theta} \ln f(x;\theta) = C(\theta)(h(x) - \theta).$$

На неформальном языке неравенство Крамера—Рао означает, что погрешность при оценивании параметра θ по выборке не может быть меньше величины $1/I(\theta)$ (какой бы несмещенной оценкой для оценивания θ мы не пользовались).

Замечание. При доказательстве леммы неявно предполагалось, что $f(x;\theta) > 0$ на $\mathbb{R}^n$. Если $f(x;\theta)$ может обращаться в нуль, то интегрирование ведется по множеству точек X, где $f(x;\theta) > 0$.

Определение. Несмещенную оценку θ^* параметра θ будем называть *эффективной* или *наилучшей несмещенной оценкой*, если

$$\mathsf{D}\theta^* = \inf_{\hat{\theta}:\mathsf{M}\hat{\theta}=\theta} \mathsf{D}\hat{\theta}.$$

Следствие 1. *Если для несмещенной оценки θ^* параметра θ неравенство Крамера—Рао обращается в равенство*

$$\frac{1}{I(\theta)} = \mathsf{D}\theta^*,$$

то θ^* является эффективной оценкой параметра θ.

В условиях следствия для любой несмещенной оценки $\hat{\theta}$ параметра θ

$$\mathsf{D}\theta^* = \frac{1}{I(\theta)} \leq \mathsf{D}\hat{\theta}.$$

Отсюда

$$\mathsf{D}\theta^* \leq \inf_{\hat{\theta}:\mathsf{M}\hat{\theta}=\theta} \mathsf{D}\hat{\theta}.$$

Но θ^* — одна из несмещенных оценок, поэтому

$$\inf_{\hat{\theta}:\mathsf{M}\hat{\theta}=\theta} \mathsf{D}\hat{\theta} \leq \mathsf{D}\theta^*,$$

Так что
$$\mathsf{D}\theta^* = \inf_{\hat\theta:\mathsf{M}\hat\theta=\theta} \mathsf{D}\hat\theta,$$

т. е. θ^* — эффективная оценка параметра θ.

Следствие 2. *Пусть* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ *— абсолютно непрерывный выборочный вектор с плотностью* $f(x;\theta), x \in \mathbb{R}^n$, $\theta \in \Theta$, $\hat\theta = h(\xi)$ *— несмещенная оценка параметра* θ.

Если плотность $f(x;\theta)$ *и статистика* $h(x)$ *связаны соотношением*
$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = C(\theta)(h(x)-\theta),\ x \in \mathbb{R}^n, \tag{18.1.10}$$

то $\hat\theta = h(\xi)$ *— эффективная оценка параметра* θ.

Действительно, равенство (18.1.10) гарантирует обращение неравенства Крамера—Рао в равенство, а вместе с ним и эффективность оценки.

Заметим, что если для плотности $f(x;\theta)$ и статистики $h(x)$ имеет место равенство (18.1.10), то
$$I(\theta) = \mathsf{M}\left(\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right)^2 = C^2(\theta)\mathsf{M}(h(\xi)-\theta)^2 = C^2(\theta)\mathsf{D}\hat\theta$$

и
$$\frac{1}{I(\theta)} = \mathsf{D}\hat\theta.$$

Из двух последних равенств получаем соотношение $|C(\theta)| = I(\theta)$.

Следствие 3. *Пусть* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ *— выборка из распределения с плотностью* $g(x;\theta)$, $\hat\theta = h(\xi)$ *— несмещенная оценка параметра* θ, *причем статистика* $h(x)$ *и совместная плотность*
$$f(x;\theta) = f(x_1, x_2, \ldots, x_n;\theta) = \prod_{i=1}^n g(x_i;\theta)$$

удовлетворяют условиям теоремы. Тогда информация
$$I(\theta) = n\mathsf{D}\frac{\partial}{\partial\theta}\ln g(\xi_1;\theta), \tag{18.1.11}$$

а неравенство Крамера—Рао можно записать в виде

$$\mathsf{D}\hat{\theta} \geq \frac{1}{n\mathsf{D}\frac{\partial}{\partial\theta}\ln g(\xi_1;\theta)}. \tag{18.1.12}$$

Достаточно заметить, что в предположениях следствия для количества информации $I(\theta)$ имеет место представление:

$$I(\theta) = \mathsf{D}\frac{\partial}{\partial\theta}\ln f(\xi;\theta) =$$

$$= -\mathsf{M}\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta) = -\mathsf{M}\frac{\partial^2}{\partial\theta^2}\ln\prod_{i=1}^{n} g(\xi_i;\theta) =$$

$$= -\mathsf{M}\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n}\ln g(\xi_i;\theta) = -\mathsf{M}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta) =$$

$$= -\sum_{i=1}^{n}\mathsf{M}\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta) = -n\mathsf{M}\frac{\partial^2}{\partial\theta^2}\ln g(\xi_1;\theta) =$$

$$= n\mathsf{M}\left(\frac{\partial}{\partial\theta}\ln g(\xi_1;\theta)\right)^2 = n\mathsf{D}\frac{\partial}{\partial\theta}\ln g(\xi_1;\theta).$$

Если $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из данного распределения, то из неравенства Крамера—Рао (18.1.12) *следует, что нижняя граница дисперсий несмещенных оценок при $n \to \infty$ имеет порядок $1/n$.*

Неравенство Крамера—Рао (дискретное распределение). Неравенство Крамера—Рао, разумеется, имеет место и для дискретно распределенных выборочных векторов. Напомним, что распределение P_ξ вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ дискретно, если существует не более чем счетное множество $X \subset \mathbb{R}^n$ такое, что

$$\sum_{x\in X}\mathsf{P}_\xi(x) = 1.$$

Напомним еще условие дифференцируемости ряда по параметру.

Теорема 18.1.3. *Если функция* $\mathsf{P}(x;\theta)$, $x \in X \subset \mathbb{R}^n$, $\theta \in \Theta$, *при всех* x *дифференцируема по параметру* θ *и ряд*

$$\sum_{x \in X} \frac{\partial}{\partial \theta} \mathsf{P}(x;\theta)$$

сходится абсолютно и равномерно относительно $\theta \in \Theta$, *то ряд* $\sum\limits_{x \in X} \mathsf{P}(x;\theta)$ *можно почленно дифференцировать по параметру:*

$$\frac{\partial}{\partial \theta} \sum_{x \in X} \mathsf{P}(x;\theta) = \sum_{x \in X} \frac{\partial}{\partial \theta} \mathsf{P}(x;\theta).$$

Теорема 18.1.4 (неравенство Крамера—Рао, дискретное распределение). *Пусть* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — *дискретный выборочный вектор с распределением* $\mathsf{P}_\xi(x) = \mathsf{P}(x;\theta)$, $x \in X$, $\theta \in \Theta$ *и* $\hat{\theta} = h(\xi)$ — *несмещенная оценка параметра* θ.

Если ряды

$$\sum_{x \in X} \mathsf{P}(x;\theta) \ \text{и} \ \sum_{x \in X} h(x) \mathsf{P}(x;\theta)$$

удовлетворяют условиям дифференцируемости по параметру $\theta \in \Theta$, *причем ряд* $\sum\limits_{x \in X} \mathsf{P}(x;\theta)$ — *дифференцируемости дважды* (*см. теорему* 18.1.3), *и*

$$\mathsf{M} \left| \frac{\partial^2}{\partial \theta^2} \ln \mathsf{P}(\xi;\theta) \right| < \infty,$$

то $I(\theta) < \infty$ *и*

$$\mathsf{D}\hat{\theta} \geq \frac{1}{I(\theta)},$$

причем равенство достигается тогда и только тогда, когда

$$\frac{\partial}{\partial \theta} \ln \mathsf{P}(x;\theta) = C(\theta)(h(x) - \theta).$$

Доказательство теоремы 18.1.4 аналогично доказательству теоремы 18.1.2.

Из теоремы 18.1.4 получаются следствия аналогичные следствиям 1, 2, 3 из теоремы 18.1.2. Их формулировки совпадают с формулировками следствий 1, 2, 3 из теоремы 18.1.2 с заменой плотности $f(x,\theta)$ на распределение $\mathsf{P}(x,\theta)$.

18.2 Примеры и задачи

Примеры

Пример 18.2.1. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из нормального распределения с параметрами $(\theta; 1)$, θ принадлежит отрезку $[a, b]$. Является ли*

$$\hat{\theta} = \overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

эффективной оценкой параметра θ?

Решение. Для того чтобы выяснить, является ли несмещенная оценка $\hat{\theta} = \overline{\xi}$ параметра θ его эффективной оценкой, воспользуемся неравенством Крамера—Рао (см. следствие из теоремы 18.1.2): если

$$\mathsf{D}\hat{\theta} = \frac{1}{I(\theta)},$$

то оценка $\hat{\theta}$ — эффективная.

Сначала убедимся, что для выборки из нормального распределения с параметрами $(\theta; 1)$ и функции

$$h(x) = \frac{1}{n}\sum_{k=1}^{n}x_k$$

выполняются условия теоремы 18.1.2.

Обозначим через $g(t; \theta)$ плотность

$$\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(t-\theta)^2\right\}$$

нормального распределения с параметрами $(\theta; 1)$. Плотность

$$f(x; \theta) = f(x_1, x_2, \ldots, x_n; \theta) =$$

$$= \prod_{i=1}^{n}g(x_i; \theta) = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

распределения вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ дважды дифференцируема по θ и

$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = \frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} g(x_i;\theta) =$$

$$= \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln g(x_i;\theta) = \sum_{i=1}^{n}(x_i-\theta);$$

$$\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}(x_i-\theta) = -n;$$

$$\mathsf{M}\left|\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta)\right| = \mathsf{M}|-n| = n.$$

Далее, функция $h(x)f(x;\theta)$ дифференцируема по θ, причем

$$\frac{\partial}{\partial\theta}h(x)f(x,\theta) = \frac{\partial}{\partial\theta}h(x_1,x_2,\ldots,x_n)f(x_1,x_2,\ldots,x_n;\theta) =$$

$$= \frac{\partial}{\partial\theta}\overline{x}\prod_{j=1}^{n} g(x_j;\theta) = \sum_{i=1}^{n}\overline{x}\left(\frac{\partial}{\partial\theta}g(x_i;\theta)\right)\prod_{j=1,j\neq i}^{n} g(x_j;\theta) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}x_k\left(\frac{\partial}{\partial\theta}g(x_i;\theta)\right)\prod_{j=1,j\neq i}^{n} g(x_j;\theta) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}x_k(x_i-\theta)\prod_{j=1}^{n} g(x_j;\theta).$$

Мы воспользовались тем, что

$$\frac{\partial}{\partial\theta}g(x_i;\theta) = \frac{\partial}{\partial\theta}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_i-\theta)^2}{2}\right\} =$$

$$= (x_i-\theta)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_i-\theta)^2}{2}\right\} = (x_i-\theta)g(x_i;\theta).$$

Чтобы убедиться в существовании интегрируемой мажоранты для функции $h(x)\dfrac{\partial}{\partial\theta}f(x;\theta)$, достаточно установить, что она существует для каждой из функций

$$x_k x_i \prod_{j=1}^{n} g(x_j;\theta),\ \ k,i=1,2,\ldots,n;\ \ \theta\in[a,b].$$

Определим функцию $v(t)$ так:

$$v(t)=\begin{cases} g(t;b), & \text{если } t>b;\\ g(t;a), & \text{если } t<a;\\ c, & \text{если } a\le t\le b,\end{cases}$$

где $c=\max\limits_{t,\theta\in[a,b]} g(t;\theta)$.

Функция $|t|v(t)$ является интегрируемой мажорантой для $tg(t;\theta)$, $\theta\in[a,b]$. В последнем убеждаемся непосредственной проверкой. Поэтому функция

$$|x_k|v(x_k)|x_i|v(x_i)\prod_{j=1,j\ne i,j\ne k}^{n} v(x_j)$$

является интегрируемой мажорантой для функции

$$x_k x_i \prod_{j=1}^{n} g(x_j;\theta).$$

В последнем легко убедиться, воспользовавшись теоремой Фубини.

Следовательно, условия теоремы 18.1.2 выполняются и неравенство Крамера—Рао имеет место:

$$\frac{1}{I(\theta)}\le \mathsf{D}\hat{\theta}.$$

При этом

$$I(\theta)=\mathsf{M}\left(\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right)^2=-\mathsf{M}\left(\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta)\right)=n;$$

$$\mathsf{D}\hat{\theta} = \mathsf{D}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n^2}n\mathsf{D}\xi_i = \frac{1}{n}1 = \frac{1}{n},$$

т. е. неравенство Крамера—Рао обращается в равенство, что является достаточным условием эффективности оценки $\hat{\theta}$ параметра θ.

Другой способ решения примера состоит в проверке достаточного условия обращения неравенства Крамера—Рао в равенство, а именно:

$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = c(\theta)(h(x) - \theta);\ x \in \mathbb{R}^n.$$

Имеем:

$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = \frac{\partial}{\partial\theta}\ln f(x_1, x_2, \ldots, x_n;\theta) =$$

$$= \frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} g(x_i;\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\ln g(x_i;\theta) =$$

$$= \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_i-\theta)^2}{2}\right\} =$$

$$= \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left(-\frac{(x_i-\theta)^2}{2}\right) = \sum_{i=1}^{n}x_i - n\theta =$$

$$= n\left(\frac{1}{n}\sum_{i=1}^{n}x_i - \theta\right) = n(h(x) - \theta).$$

Отсюда следует, что $\hat{\theta} = \overline{\xi}$ — эффективная оценка параметра θ.

Задачи

18.1. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}\exp\left\{-\frac{1}{\theta}x\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0, \end{cases}$$

θ принадлежит конечному отрезку $[a, b]$.

Является ли

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

эффективной оценкой параметра θ?

18.2. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из биномиального распределения

$$\mathsf{P}(k;\theta) = C_N^k \theta^k (1-\theta)^{N-k},\ k = 0, 1, \ldots, N,$$

где N — известное целое число, θ принадлежит конечному отрезку $[a, b]$.

Является ли

$$\hat{\theta} = \frac{1}{nN}\sum_{i=1}^{n}\xi_i$$

эффективной оценкой параметра θ?

18.3. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;\theta) = \begin{cases} \dfrac{x}{\theta}\ \exp\left\{-\dfrac{x^2}{2\theta}\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0 \end{cases}$$

($f(x;\theta)$ — плотность распределения Рэлея), параметр θ принадлежит конечному отрезку $[a; b]$.

Является ли

$$\hat{\theta} = \frac{1}{2n}\sum_{i=1}^{n}{\xi_i}^2$$

эффективной оценкой параметра θ?

18.4. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;\theta) = \frac{1}{2\theta}\ \exp\left\{-\frac{1}{\theta}|x|\right\},\ \ \theta > 0.$$

($f(x;\theta)$ — плотность двустороннего показательного распределения).

Является ли

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}|\xi_i|$$

эффективной оценкой параметра θ?

18.5. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения

$$\mathsf{P}(k;\theta) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^k, \; k = 0, 1, \ldots,$$

где $\theta > 0$ и принадлежит конечному отрезку $[a, b]$.

Является ли

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

эффективной оценкой параметра θ?

Замечание. Если обозначить $\dfrac{1}{1+\theta} = p$, то распределение

$$\mathsf{P}(k;\theta) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^k, \; k = 0, 1, \ldots,$$

принимает вид хорошо известного геометрического распределения

$$\mathsf{P}(k;p) = p(1-p)^k, \; k = 0, 1, \ldots$$

18.6. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$p(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma^2}x}\exp\left\{-\dfrac{(\ln x - \mu_0)^2}{2\sigma^2}\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0, \end{cases}$$

($p(x)$ — плотность логарифмически нормального распределения), μ_0 — известно, $\sigma^2 > 0$ и принадлежит конечному отрезку $[a, b]$, $a > 0$.

Является ли

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln\xi_i - \mu_0)^2$$

эффективной оценкой параметра σ^2?

18.7. Доказать теорему 18.1.4.

Глава 19

Эмпирические характеристики

19.1 Оценивание вероятности события

Пусть A — наблюдаемое событие стохастического эксперимента, вероятность $p = \mathsf{P}(A)$ которого неизвестна. Известно число μ наступлений события A в последовательности n независимых экспериментов. Другими словами, известно число μ успехов в n испытаниях Бернулли, успех — наступление события A. Необходимо оценить неизвестную вероятность p. Поскольку частота μ/n события A близка к его вероятности $p = \mathsf{P}(A)$, то её и естественно рассматривать в качестве оценки $\hat{p}$ вероятности p.

Теорема 19.1.1. *Частота $\hat{p} = \mu/n$ успеха в последовательности n испытаний Бернулли является несмещенной и состоятельной оценкой вероятности успеха p.*

Доказательство. Число μ успехов в n независимых испытаниях имеет биномиальное распределение с параметрами $(n; p)$. Поэтому
$$\mathsf{M}\mu = np,\ \ \mathsf{D}\mu = npq.$$
Отсюда
$$\mathsf{M}\hat{p} = \mathsf{M}\frac{\mu}{n} = \frac{1}{n}\mathsf{M}\mu = \frac{1}{n}np = p,$$
т. е. $\hat{p}$ — несмещенная оценка параметра p.

В силу неравенства Чебышева
$$\mathsf{P}\left\{|\hat{p} - p| > \varepsilon\right\} = \mathsf{P}\left\{|\hat{p} - \mathsf{M}\hat{p}| > \varepsilon\right\} \le$$
$$\le \frac{\mathsf{D}\hat{p}}{\varepsilon^2} = \frac{\mathsf{D}(\mu/n)}{\varepsilon^2} = \frac{pq}{n\varepsilon^2}.$$

Последняя дробь стремится к нулю при $n \to \infty$ и, следовательно, $\hat{p}$ — состоятельная оценка параметра p.

Замечание. Сходимость $\hat{p} \xrightarrow{\mathsf{P}} p$ также следует из закона больших чисел.

19.2 Эмпирическая функция распределения

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из неизвестного распределения F, по которой его необходимо оценить. Поскольку распределение F однозначно определяется своей функцией распределения $F(x)$, то достаточно оценить функцию распределения $F(x)$, т. е. оценить значение $F(x)$ в каждой точке $x \in \mathbb{R}^1$.

Для каждого фиксированного $x \in \mathbb{R}^1$ выборка $\xi_1, \xi_2, \ldots, \xi_n$ из распределения F (последовательность независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, каждая с распределением F) задает последовательность независимых событий

$$\{\xi_k \in (-\infty, x)\}, \ k = 1, 2, \ldots, n,$$

вероятность появления каждого из которых

$$p_x = \mathsf{P}\{\xi_k \in (-\infty, x)\} = F(x),$$

а вместе с ними задает последовательность n испытаний Бернулли: успех в k-м испытании — произошло событие $\{\xi_k \in (-\infty, x)\}$, вероятность успеха $p_x = F(x)$. А, как известно (см. теорему 19.1.1), несмещенной и состоятельной оценкой вероятности $p_x = F(x)$ успеха является его частота в последовательности независимых испытаний:

$$\hat{p}_x = \frac{\mu_x}{n},$$

где μ_x — число успехов в n испытаниях или, что то же, число выборочных значений $\xi_1, \xi_2, \ldots, \xi_n$ меньших x (попавших в $(-\infty, x)$). Очевидно, μ_x можно представить в виде:

$$\mu_x = \sum_{k=1}^{n} I_{(-\infty, x)}(\xi_k).$$

Поэтому для оценки $\hat{p}_x$ вероятности $p_x = F(x)$ имеет место представление

$$\hat{p}_x = \frac{\mu_x}{n} = \frac{1}{n}\sum_{k=1}^{n} I_{(-\infty,x)}(\xi_k) = \hat{F}_n(x).$$

Определение. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения F. Функция $\hat{F}_n(x)$, определенная на $\mathbb{R}^1$ равенством

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{(-\infty,x)}(\xi_k),$$

или, что то же,

$$\hat{F}_n(x) = \frac{\mu_x}{n},$$

называется *эмпирической* функцией распределения.

Учитывая, что частота $\hat{p} = \mu/n$ события в n независимых наблюдениях является несмещенной и состоятельной оценкой вероятности p его появления (см. теорему 19.1.1), а $\hat{F}_n(x)$ является частотой события $\{\xi_k \in (-\infty, x)\}$, вероятность которого равна $F(x)$, имеем:

Теорема 19.2.1 . *Для каждого фиксированного x эмпирическая функция распределения $\hat{F}_n(x)$ является несмещенной и состоятельной оценкой $F(x)$:*

$$\mathsf{M}\hat{F}_n(x) = F(x),$$

$$\hat{F}_n(x) \xrightarrow{\mathsf{P}} F(x)$$

при $n \to \infty$.

В этом смысле эмпирическая функция распределения $\hat{F}_n(x)$ является хорошей оценкой функции распределения $F(x)$.

Эмпирическую функцию распределения $\hat{F}_n(x)$ часто называют *статистическим аналогом функции распределения* $F(x)$ (теоретической функции распределения).

Далее будем предполагать, что $F(x)$ непрерывна, при этом вероятность совпадения выборочных значений равна нулю.

Распределение $\hat{F}_n(x)$. Отметим, что при каждом фиксированном $x \in \mathbb{R}^1$, эмпирическая функция распределения

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{(-\infty,x)}(\xi_k),$$

как функция случайного вектора $\xi_1, \xi_2, \ldots, \xi_n$, является случайной величиной:

$$\hat{F}_n(x) = \hat{F}_n(x, \omega),\ \omega \in \Omega,$$

поэтому для каждого $x \in \mathbb{R}^1$ можно говорить о распределении $\hat{F}_n(x)$.

Теорема 19.2.2 (о распределении $\hat{F}_n(x)$). *При каждом фиксированном x*

$$\mathsf{P}\{\hat{F}_n(x) = k/n\} = C_n^k (F(x))^k (1 - F(x))^{n-k},\ k = 0, 1, \ldots, n,$$

$$\mathsf{M}(\hat{F}_n(x)) = F(x),$$

$$\mathsf{D}(\hat{F}_n(x)) = \frac{1}{n} F(x)(1 - F(x)).$$

Доказательство. При каждом фиксированном x случайная величина

$$n\hat{F}_n(x) = \sum_{j=1}^{n} I_{(-\infty,x)}(\xi_j),$$

как сумма независимых одинаково распределенных случайных величин $I_{(-\infty,x)}(\xi_j),\ j = 1, 2, \ldots, n$, каждая с распределением

$$\mathsf{P}\left\{I_{(-\infty,x)}(\xi_j) = 1\right\} = F(x),\ \mathsf{P}\left\{I_{(-\infty,x)}(\xi_j) = 0\right\} = 1 - F(x),$$

биномиально распределена с параметрами $(n, F(x))$:

$$\mathsf{P}\left\{n\hat{F}_n(x) = k\right\} = C_n^k (F(x))^k (1 - F(x))^{n-k},\ k = 0, 1, \ldots, n,$$

(см. теорему 6.1.4), поэтому

$$\mathsf{M}(n\hat{F}_n(x)) = nF(x),$$

$$\mathsf{D}(n\hat{F}_n(x)) = nF(x)(1 - F(x)),$$

$$\mathsf{P}\left\{\hat{F}_n(x) = k/n\right\} =$$

$$= \mathsf{P}\left\{n\hat{F}_n(x) = k\right\} = C_n^k (F(x))^k (1 - F(x))^{n-k},\ k = 0, 1, \ldots, n.$$

Также очевидно, что

$$\mathsf{M}\hat{F}_n(x) = F(x),$$

$$\mathsf{D}(\hat{F}_n(x)) = \frac{1}{n}F(x)(1 - F(x)).$$

Вариационный ряд. По случайным величинам $\xi_1, \xi_2, \ldots, \xi_n$ определим случайные величины $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ следующим образом: для каждого фиксированного $\omega \in \Omega$ упорядочим значения $\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega)$ в порядке возрастания и эти упорядоченные значения обозначим через $\xi_1^*(\omega)$, $\xi_2^*(\omega)$, $\ldots$, $\xi_n^*(\omega)$, $\omega \in \Omega$; заметим, что, в частности,

$$\xi_1^*(\omega) = \min\{\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega)\},$$

$$\xi_n^*(\omega) = \max\{\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega)\}.$$

При фиксированном ω значение $\xi_k^*(\omega)$ $(k = 1, 2, \ldots, n)$ совпадает с одним из значений $\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega)$, но как случайная величина $\xi_k^* = \xi_k^*(\omega)$, $\omega \in \Omega$, вообще говоря, не совпадает ни с одной из случайных величин $\xi_1, \xi_2, \ldots, \xi_n$.

Определение. Последовательность случайных величин $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ называется *вариационным рядом* последовательности $\xi_1, \xi_2, \ldots, \xi_n$, а случайные величины ξ_i^*, $i = 1, 2, \ldots, n$, называют *порядковыми статистиками.*

Индекс j в порядковой статистике ξ_j^* равен числу выборочных значений, лежащих левее ξ_j^* (включая ξ_j^*).

Реализация эмпирической функции распределения. Поведение реализации эмпирической функции распределения описывается следующим утверждением:

Теорема. *При каждом фиксированном ω реализация $\hat{F}_n(x) = \hat{F}_n(x, \omega)$ эмпирической функции распределения, построенной по выборке $\xi_1, \xi_2, \ldots, \xi_n$, является неотрицательной, монотонно неубывающей, непрерывной слева, кусочно-постоянной функцией со скачками, равными $1/n$ в точках ξ_k, $k = 1, 2, \ldots, n$, являющихся выборочными значениями.*

Доказательство. Число

$$\mu_x = \sum_{k=1}^{n} I_{(-\infty, x)}(\xi_k)$$

выборочных значений ξ_k, меньших x, равно числу

$$\mu_x^* = \sum_{k=1}^{n} I_{(-\infty, x)}(\xi_k^*)$$

значений ξ_k^*, меньших x,

$$\mu_x^* = \mu_x.$$

Поэтому эмпирическую функцию распределения $\hat{F}_n(x)$ можно представить в виде:

$$\hat{F}_n(x) = \hat{F}_n(x, \omega) = \frac{\mu_x^*}{n}. \tag{19.2.1}$$

Зафиксируем ω. Непосредственно из равенства (19.2.1) получаем, что $\hat{F}_n(x) = 0$ на промежутке $(-\infty, \xi_1^*]$, так как при $x \leq \xi_1^*$ число тех ξ_k^*, для которых $\xi_k^* < x$ равно нулю; $\hat{F}_n(x) = 1/n$ на промежутке $(\xi_1^*, \xi_2^*]$, поскольку при $x \in (\xi_1^*, \xi_2^*]$ число тех ξ_k^*, для которых $\xi_k^* < x$ равно 1 и т. д., и, наконец, $\hat{F}_n(x) = 1$ на промежутке $(\xi_n^*, +\infty)$, так как при $x \in (\xi_n^*, +\infty)$ число тех ξ_k^*, для которых $\xi_k^* < x$ равно n (см. рис. 19.2.1, а также рис. 19.2.2).

Из сказанного следует, что реализация эмпирической функции распределения $\hat{F}_n(x) = \hat{F}_n(x, \omega)$:

1° неотрицательна;

2° постоянна на каждом из промежутков $(-\infty, \xi_1^*], (\xi_k^*, \xi_{k+1}^*]$, $k = 1, 2, \ldots, n-1$, $(\xi_n^*, +\infty)$ (а, следовательно, непрерывна слева);

3° монотонно неубывающая, при этом $\hat{F}_n(x)$ "возрастает скачками" $1/n$, в точках ξ_k^*, $k = 1, 2, \ldots, n$.

Заметим, что при каждом фиксированном ω эмпирическая функция распределения $\hat{F}_n(x) = \hat{F}_n(x, \omega)$ как функция x обладает всеми свойствами функции распределения.

Типичный график реализации эмпирической функции распределения

$$\hat{F}_n(x) = \hat{F}_n(x, \omega) = \frac{1}{n} \sum_{k=1}^{n} I_{(-\infty, x)}(\xi_k^*(\omega)),$$

изображен на рис. 19.2.1.

Представление об эмпирической функции распределения $\hat{F}_n(x) = \hat{F}_n(x, \omega)$ как функции двух переменных x и ω дает рис. 19.2.2.

Эмпирическая функция распределения может иметь скачки величиной l/n (в случае совпадения l выборочных значений).

Совпадать выборочные значения могут как за счет того, что они фактически регистрируются с определенной точностью (часть десятичных знаков отбрасывается), так и за за счет того, что выборка может быть получена из дискретного распределения.

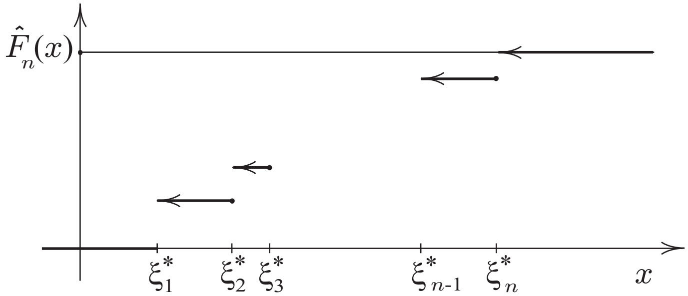

Рис. 19.2.1: Реализация эмпирической функции распределения

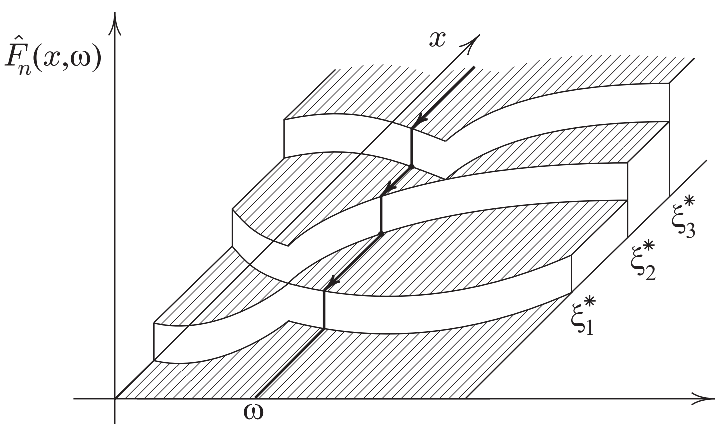

Рис. 19.2.2: К определению эмпирической функции распределения

Определение. При каждом фиксированном ω распределение $\hat{\mathsf{F}}_n$, соответствующее эмпирической функции распределения $\hat{F}_n(x)$, называется *эмпирическим.*

Эмпирическое распределение является атомическим распределением, сосредоточенным в атомах $\xi_k(\omega)$, $k = 1, 2, \ldots, n$, с массой $1/n$ в каждом атоме.

Об уклонениях эмпирической функции распределения $\hat{F}_n(x)$ от теоретической $F(x)$. Эмпирическая функция распределения $\hat{F}_n(x)$, будучи при каждом x несмещенной и состоятельной оценкой $F(x)$, дает хорошее приближение для $F(x)$. Но в ряде задач этого не достаточно — необходимо знать насколько эмпирическая функция распределения $\hat{F}_n(x)$, построенная по выборке $\xi_1, \xi_2, \ldots, \xi_n$ из F, уклоняется от $F(x)$ (какими могут быть уклонения эмпирической функции распределения от теоретической).

Меру уклонения эмпирической функции распределения $\hat{F}_n(x)$ от $F(x)$ можно вводить различными способами. Например, так:

$$\sup_{-\infty<x<\infty} \left| F(x) - \hat{F}_n(x) \right| \text{ (А. Н. Колмогоров).}$$

Или так:

$$\sum_{i=1}^{r} \frac{n}{p_i} \left(\frac{\nu_i}{n} - p_i \right)^2 \text{ (К. Пирсон),}$$

где p_i — вероятность попадания выборочного значения в промежуток $[a_i, b_i)$, вычисленная по функции распределения $F(x)$, ν_i/n — вероятность попадания в тот же промежуток, вычисленная по эмпирической функции распределения $\hat{F}_n(x)$ (вся прямая разбита на r непересекающихся промежутков $[a_j, b_j)$: $\bigcup\limits_{j=1}^{r} [a_j, b_j) = \mathbb{R}^1$, два крайних из них бесконечны). Оказывается, что введенные меры уклонения эмпирической функции распределения $\hat{F}_n(x)$ от теоретической $F(x)$ обладают замечательными свойствами — вне зависимости от распределения F, из которого получена выборка, при достаточно больших n распределение уклонения $\sup\limits_{-\infty<x<\infty} \left| F(x) - \hat{F}_n(x) \right| \Big/ \frac{1}{\sqrt{n}}$ близко к распределению А. Н. Колмогорова, его функция распределения

$$K(x) = \sum_{n=-\infty}^{+\infty} (-1)^n e^{-2n^2 x^2},$$

а распределение уклонения К. Пирсона близко к χ^2-распределению с $(r-1)$ степенями свободы (уклонение $\hat{F}_n(x)$ от $F(x)$ является случайной величиной, и когда мы говорим о таком уклонении, необходимо говорить о его распределении).

Выборочные квантили. Пусть F — непрерывное распределение на $\mathbb{R}^1$, $\alpha \in (0,1)$. Число x_α, являющееся решением уравнения

$$F(x) = \alpha,$$

называется *α-квантилью* распределения F; 1/2-квантиль называется *медианой распределения* F.

Решение x_α уравнения $F(x) = \alpha$, вообще говоря, не единственное, но если функция распределения $F(x)$ строго возрастающая, то решение x_α единственное.

α-Квантиль x_α распределения F — это число, "отсекающее левый хвост" распределения F с "массой" α:

$$F(x_\alpha) = \mathsf{F}((-\infty, x_\alpha)) = \alpha$$

(см. рис. 19.2.3).

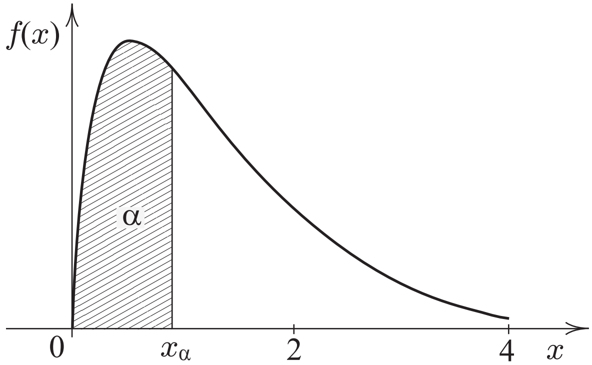

Рис. 19.2.3: К определению α-квантили распределения, $f(x)$ — плотность распределения F

Пусть ξ — случайная величина с непрерывным распределением F, α-квантиль x_α распределения F еще называют *α-квантилью случайной величины ξ*. Другими словами, α-квантиль x_α случайной величины ξ — это число x_α, для которого

$$\mathsf{P}\{\xi < x_\alpha\} = \alpha.$$

В частности, *медиана* $x_{1/2}$ случайной величины ξ — это число, для которого

$$\mathsf{P}\left\{\xi < x_{1/2}\right\} = 1/2.$$

Заметим, что

$$\mathsf{P}\left\{\xi < x_{1/2}\right\} = \mathsf{P}\left\{\xi \geq x_{1/2}\right\},$$

т. е. с равными вероятностями случайная величина ξ принимает значения большие $x_{1/2}$ и меньшие $x_{1/2}$.

Для данных $\alpha \in (0,1)$ и целого n $(n = 1, 2, \ldots)$ определим целое число $k(\alpha)$ равенством

$$k(\alpha) = \begin{cases} n\alpha, & \text{если } n\alpha \text{ целое;} \\ [n\alpha]+1, & \text{если } n\alpha \text{ не целое.} \end{cases}$$

Определение. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения F. *Выборочной* (*эмпирической*) *α-квантилью* распределения F будем называть порядковую статистику

$$\zeta_\alpha = \xi^*_{k(\alpha)},$$

в частности, порядковую статистику

$$\zeta_{1/2} = \xi^*_{k(1/2)}$$

называют *выборочной медианой.*

Теорема. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения F с непрерывной строго монотонной функцией распределения $F(x)$, тогда порядковая статистика*

$$\zeta_\alpha = \xi^*_{k(\alpha)}$$

является состоятельной оценкой α-квантили x_α распределения F.

Доказательство. Для произвольного фиксированного x

$$\mu(x) = \sum_{j=1}^{n} I_{(-\infty,x)}(\xi_j)$$

— количество выборочных значений меньших x, следовательно $\xi^*_{\mu(x)}$ — наибольшее из них и

$$\xi^*_1 \leq \xi^*_2 \leq \cdots \leq \xi^*_{\mu(x)}.$$

Поэтому, если для индекса $k(\alpha)$ значение $\xi^*_{k(\alpha)} < x$, то $k(\alpha) \le \mu(x)$, т. е.

$$\{\xi^*_{k(\alpha)} < x\} \subset \{k(\alpha) \le \mu(x)\}$$

Обратно, если для индекса $k(\alpha)$ имеет место неравенство $k(\alpha) \le \mu(x)$, то $\xi^*_{k(\alpha)} < \xi^*_{\mu(x)} < x$, т. е.

$$\{\xi^*_{k(\alpha)} < x\} \subset \{k(\alpha) \le \mu(x)\}.$$

Так что

$$\{\xi^*_{k(\alpha)} < x\} = \{k(\alpha) \le \mu(x)\}. \tag{19.2.2}$$

Пусть $\varepsilon > 0$ произвольное, фиксированное число. Из соотношения (19.2.2) при $x = x_\alpha + \varepsilon$ имеем:

$$\Big\{\omega : \xi^*_{k(\alpha)}(\omega) < x_\alpha + \varepsilon\Big\} = \Big\{\omega : k(\alpha) \le \mu(x_\alpha + \varepsilon)\Big\} =$$

$$= \Big\{\omega : \frac{k(\alpha)}{n} \le \frac{\mu(x_\alpha + \varepsilon)}{n}\Big\} = \Big\{\omega : \frac{k(\alpha)}{n} \le \hat{F}_n(x_\alpha + \varepsilon)\Big\}.$$

При $n \to \infty$

$$\hat{F}_n(x_\alpha + \varepsilon) \to F(x_\alpha + \varepsilon) > F(x_\alpha) = \alpha,$$

т. е. $\hat{F}_n(x_\alpha + \varepsilon)$ стремится к величине $F(x_\alpha + \varepsilon)$ большей α, а отношение $k(\alpha)/n$ сходится к α, поскольку

$$n\alpha \le k(\alpha) \le n\alpha + 1.$$

Поэтому при достаточно больших n вероятность события

$$\Big\{\omega : \frac{k(\alpha)}{n} \le \hat{F}_n(x_\alpha + \varepsilon)\Big\}$$

близка к 1 и, следовательно, близка к 1 вероятность события

$$\{\omega :\ \xi^*_{k(\alpha)}(\omega) < x_\alpha + \varepsilon\}.$$

Аналогично, из равенства

$$\{\omega :\ \xi^*_{k(\alpha)}(\omega) \ge x\} = \{\omega : k(\alpha) \ge \mu(x)\},$$

которое является другой формой записи (19.2.2), при $x = x_\alpha - \varepsilon$ получаем, что для достаточно больших n вероятность события

$$\{\omega :\ \xi^*_{k(\alpha)}(\omega) \geq x_\alpha - \varepsilon\}$$

близка к 1.

Таким образом, для любого $\varepsilon > 0$ за счет выбора n достаточно большим $\mathsf{P}\{x_\alpha - \varepsilon \leq \xi^*_{k(\alpha)} < x_\alpha + \varepsilon\} = \mathsf{P}\{|\xi^*_{k(\alpha)} - x_\alpha| \leq \varepsilon\}$ можно сделать сколь угодно близкой к 1. Последнее означает, что $\xi^*_{k(\alpha)}$ сходится по вероятности к x_α, другими словами,

$$\zeta_\alpha = \xi^*_{k(\alpha)}$$

является состоятельной оценкой для x_α.

19.3 Эмпирические значения параметров

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения $\mathsf{F}(\cdot;\theta)$, зависящего от параметра θ, причем параметр θ однозначно определяется распределением $\mathsf{F}(\cdot;\theta)$ (функцией распределения $F(\cdot;\theta)$), т. е. является функционалом

$$\theta = \varPhi(F(\cdot;\theta)),$$

заданным на некотором множестве распределений F (функций распределения $F(x)$). Например, среднее $a = \int\limits_{\mathbb{R}^1} x\mathsf{F}(dx)$, моменты $m_r = \int\limits_{\mathbb{R}^1} x^r\mathsf{F}(dx), r = 1, 2, \ldots$, дисперсия $\sigma^2 = \int\limits_{\mathbb{R}^1}\Bigl(x - \int\limits_{\mathbb{R}^1} t\mathsf{F}(dt)\Bigr)^2\mathsf{F}(dx)$ являются функционалами на множестве функций распределения $F(x)$.

По эмпирическому распределению $\hat{\mathsf{F}}_n$ (по эмпирической функции распределения $\hat{F}_n(x)$) строятся так называемые эмпирические (выборочные) оценки параметров $\theta = \varPhi(F(\cdot;\theta))$.

Выборка $\xi_1, \xi_2, \ldots, \xi_n$ из распределения $\mathsf{F}(\cdot;\theta)$ задает эмпирическую функцию распределения $\hat{F}_n(x)$. И поскольку $\hat{F}_n(x)$ "близка" к $F(x;\theta)$, а $\theta = \varPhi(F(\cdot;\theta))$, то в качестве оценки параметра θ естественно рассматривать

$$\hat{\theta} = \varPhi(\hat{F}_n(\cdot)).$$

Определение. Оценка $\hat{\theta}$ параметра $\theta = \varPhi(F(\cdot;\theta))$, определенная равенством

$$\hat{\theta} = \varPhi\left(\hat{F}_n(\cdot)\right),$$

называется *эмпирической оценкой* θ или *эмпирическим* (*выборочным*) значением параметра θ.

Например, эмпирическим (выборочным) значением среднего

$$a = \int\limits_{\mathbb{R}^1} x\mathsf{F}(dx) = \varPhi_1(F(\cdot))$$

является

$$\hat{a} = \int\limits_{\mathbb{R}^1} x\hat{\mathsf{F}}_n(dx) = \frac{1}{n}\sum_{i=1}^{n}\xi_i,$$

эмпирическим (выборочным) значением момента

$$m_r = \int\limits_{\mathbb{R}^1} x^r\mathsf{F}(dx) = \varPhi_r(F(\cdot))$$

является

$$\hat{m}_r = \int\limits_{\mathbb{R}^1} x^r\hat{\mathsf{F}}_n(dx) = \frac{1}{n}\sum_{i=1}^{n}\xi_i^r,$$

эмпирическим (выборочным) значением дисперсии

$$\sigma^2 = \int\limits_{\mathbb{R}^1}\Bigl(x - \int\limits_{\mathbb{R}^1} t\mathsf{F}(dt)\Bigr)^2\mathsf{F}(dx) = \varPhi(F(\cdot))$$

является

$$\hat{\sigma^2} = \int\limits_{\mathbb{R}^1}\Bigl(x - \int\limits_{\mathbb{R}^1} t\hat{\mathsf{F}}_n(dt)\Bigr)^2\hat{\mathsf{F}}_n(dx) = \int\limits_{\mathbb{R}^1}\Bigl(x - \frac{1}{n}\sum_{k=1}^{n}\xi_k\Bigr)^2\hat{\mathsf{F}}_n(dx) =$$

$$= \int\limits_{\mathbb{R}^1}(x-\hat{a})^2\hat{\mathsf{F}}_n(dx) = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \hat{a})^2.$$

Интегралы вычислены как интегралы Лебега по дискретному (атомическому) распределению $\hat{\mathsf{F}}_n$, сосредоточенному в точках $\xi_1, \xi_2, \ldots, \xi_n$ с массой $1/n$ в каждой:

$$\int\limits_{\mathbb{R}^1} g(x)\hat{\mathsf{F}}_n(dx) = \sum_{k=1}^{n} g(\xi_k)\hat{\mathsf{F}}_n\left(\{\xi_k\}\right) = \sum_{k=1}^{n} g(\xi_k)\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} g(\xi_k).$$

Эмпирическое значение параметра $\boldsymbol{G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx)}$.

О распределении F можно судить по значениям интеграла

$$G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx)$$

на некоторых классах функций g. В частности, если, скажем, $g = g(x) = x$, то значение интеграла

$$G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} x\,\mathsf{F}(dx) = m_1$$

равно первому моменту распределения F и дает представление о среднем распределения F, если $g = g(x) = (x - m_1)^2$, то значение

$$G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx) = \int\limits_{\mathbb{R}^1} (x - m_1)^2\mathsf{F}(dx) = \sigma^2$$

равно второму центральному моменту распределения F. Во многих случаях значения интеграла $G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx)$ на сравнительно узком классе функций полностью определяют распределение F. Например, нормальное распределение определяется значениями интеграла

$$G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx)$$

на классе, состоящем из двух функций: $g_1 = g_1(x) = x$ и $g_2 = g_2(x) = (x - m_1)^2$.

На $G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx)$ можно смотреть как на параметр распределения F.

Теорема 19.3.1 (о выборочном значении для параметра $G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx)$). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения F, $g(x)$ — борелевская функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$ и*

$$G = \int\limits_{\mathbb{R}^1} g(x)\mathsf{F}(dx).$$

Выборочное значение

$$\hat{G}_n = \int\limits_{\mathbb{R}^1} g(x)\hat{\mathsf{F}}_n(dx) = \frac{1}{n}\sum_{k=1}^{n} g(\xi_k)$$

параметра G является его несмещенной и состоятельной оценкой (в предположении, что G конечно).

Доказательство.

$$\mathsf{M}\hat{G}_n = \mathsf{M}\frac{1}{n}\sum_{k=1}^{n} g(\xi_k) = \frac{1}{n}\sum_{k=1}^{n} \mathsf{M}g(\xi_k) = \frac{1}{n}\sum_{k=1}^{n} G = G,$$

т. е. $\hat{G}_n$ является несмещенной оценкой параметра G.

Далее, $g(\xi_1), g(\xi_2), \ldots$ — независимые одинаково распределенные случайные величины (как функции от независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots$) и

$$\mathsf{M}g(\xi_k) = G \neq \infty,$$

поэтому в силу закона больших чисел в форме Хинчина при $n \to \infty$

$$\frac{1}{n}\sum_{k=1}^{n} g(\xi_k) \xrightarrow{\mathsf{P}} G,$$

т. е. $\hat{G}_n$ — состоятельная оценка G.

Следствие. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения F, r — целое положительное число. Если r-й момент распределения F конечен:*

$$m_r = \int\limits_{\mathbb{R}^1} x^r\mathsf{F}(dx) \neq \infty,$$

то r-й выборочный момент

$$\hat{m}_r = \frac{1}{n} \sum_{k=1}^{n} \xi_k^r$$

является несмещенной и состоятельной оценкой m_r.

Достаточно в теореме в качестве $g(x)$ рассмотреть $g(x) = x^r$.

Теорема 19.3.2. *Для непрерывной функции $\varphi(t_1, t_2, \ldots, t_s)$ (на $\mathbb{R}^s$ со значениями в $\mathbb{R}^1$), из сходимости при $n \to \infty$ последовательностей случайных величин*

$$\xi_n^{(k)} \xrightarrow{\mathsf{P}} \xi^{(k)}, k = 1, 2, \ldots, s,$$

следует сходимость

$$\varphi(\xi_n^{(1)}, \xi_n^{(2)}, \ldots, \xi_n^{(s)}) \xrightarrow{\mathsf{P}} \varphi(\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(s)}).$$

В частности, если $\lim\limits_n \xi_n \overset{\mathsf{P}}{=} \xi$, $\lim\limits_n \eta_n \overset{\mathsf{P}}{=} \eta$, то

$$\lim_n (\alpha\xi_n + \beta\eta_n) \overset{\mathsf{P}}{=} \alpha\xi + \beta\eta, \ \lim_n \xi_n\eta_n \overset{\mathsf{P}}{=} \xi\eta.$$

Замечание. Сумму вида $\frac{1}{n} \sum\limits_{k=1}^{n} a_k$ принято обозначать $\overline{a}$.

Теорема 19.3.3 (об оценках $\hat{\sigma^2}$, s^2). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения F с конечной дисперсией σ^2. Выборочная дисперсия*

$$\hat{\sigma^2} = \frac{1}{n} \sum_{k=1}^{n} (\xi_k - \overline{\xi})^2$$

является состоятельной и асимптотически несмещенной оценкой дисперсии σ^2, а оценка

$$s^2 = \frac{1}{n-1} \sum_{k=1}^{n} (\xi_k - \overline{\xi})^2$$

— состоятельной и несмещенной.

Доказательство. Вычислим

$$\mathsf{M}\hat{\sigma^2} = \mathsf{M}\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^2 = \frac{1}{n}\sum_{k=1}^{n}\mathsf{M}(\xi_k - \overline{\xi})^2.$$

Если для случайной величины η математическое ожидание $\mathsf{M}\eta = 0$, то $\mathsf{D}\eta = \mathsf{M}\eta^2$, поэтому, учитывая что $\mathsf{M}(\xi_k - \overline{\xi}) = 0$, имеем:

$$\mathsf{M}(\xi_k - \overline{\xi})^2 = \mathsf{D}(\xi_k - \overline{\xi}) = \mathsf{D}\left(\xi_k - \frac{1}{n}\sum_{j=1}^{n}\xi_j\right) =$$

$$= \mathsf{D}\left(\left(1 - \frac{1}{n}\right)\xi_k - \frac{1}{n}\sum_{j=1,\, j\neq k}^{n}\xi_j\right) = \left(\frac{n-1}{n}\right)^2\mathsf{D}\xi_k +$$

$$+\frac{1}{n^2}\sum_{j=1,\, j\neq k}^{n}\mathsf{D}\xi_j = \left(\frac{n-1}{n}\right)^2\sigma^2 + \frac{1}{n^2}(n-1)\sigma^2 = \frac{n-1}{n}\sigma^2.$$

И следовательно

$$\mathsf{M}\hat{\sigma^2} = \frac{1}{n}\sum_{k=1}^{n}\mathsf{M}(\xi_k - \overline{\xi})^2 = \frac{1}{n}\,n\left(\frac{n-1}{n}\,\sigma^2\right) = \frac{n-1}{n}\,\sigma^2$$

(отсюда, в частности, следует, что $\hat{\sigma^2}$ — асимптотически несмещенная оценка σ^2).

Оценка s^2 выражается через $\hat{\sigma^2}$ так:

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^2 = \frac{n}{n-1}\,\hat{\sigma^2}.$$

Поэтому

$$\mathsf{M}s^2 = \mathsf{M}\frac{n}{n-1}\,\hat{\sigma^2} = \frac{n}{n-1}\cdot\frac{n-1}{n}\,\sigma^2 = \sigma^2.$$

Так что s^2 — несмещенная оценка σ^2.

Покажем, что $\hat{\sigma^2}$ — состоятельная оценка параметра σ^2. Выборочную дисперсию $\hat{\sigma^2}$ можно записать в виде:

$$\hat{\sigma^2} = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^2 = \frac{1}{n}\left(\sum_{k=1}^{n}\xi_k^2 - 2\overline{\xi}\sum_{k=1}^{n}\xi_k + n\overline{\xi}^2\right) =$$

$$= \frac{1}{n}\sum_{k=1}^{n}\xi_k^2 - \overline{\xi}^2 = \hat{m}_2 - (\hat{m}_1)^2.$$

Выборочные моменты $\hat{m}_1$ и $\hat{m}_2$ сходятся по вероятности соответственно к m_1 и m_2 (см. следствие из теоремы 19.3.1), функция $\varphi(t_2, t_1) = t_2 - t_1^2$ непрерывна, поэтому в силу теоремы 19.3.2 случайная величина $\varphi(\hat{m}_2, \hat{m}_1) = \hat{m}_2 - (\hat{m}_1)^2 = \hat{\sigma^2}$ сходится по вероятности к $\varphi(m_2, m_1) = m_2 - m_1^2 = \sigma^2$. Так что $\hat{\sigma^2}$ — состоятельная оценка σ^2.

Состоятельность оценки s^2 для параметра σ^2 следует из соотношения

$$s^2 = \frac{n}{n-1}\,\hat{\sigma^2}$$

и состоятельности оценки $\hat{\sigma^2}$ для σ^2.

19.4 Примеры и задачи

Примеры

Пример 19.4.1. *На отрезок* $[0; 2]$ *наудачу бросают точку, делящую его на две части, и фиксируют координату* ξ *точки деления.*

Эксперимент проведен 10 *раз. При этом получено* 10 *значений независимых наблюдений случайной величины* ξ*:* 0,00; 0,31; 0,70; 1,40; 0,08; 1,93; 0,79; 1,43; 1,42; 1,69.

Оценить по выборке математическое ожидание объема куба с ребром, равным большей части отрезка.

Решение. Ребро куба равно $\max\{\xi,\ 2-\xi\}$. Объем куба $g(\xi) = (\max\{\xi,\ 2-\xi\})^3$ — функция случайной величины ξ. Значение $G = \mathsf{M}g(\xi)$ конечно, поскольку распределение случайной величины ξ сосредоточено на конечном промежутке [0;2], а функция $g(x)$ на этом промежутке ограничена. Поэтому несмещенной и состоятельной оценкой математического ожидания объема куба $G = \mathsf{M}g(\xi) = \mathsf{M}(\max\{\xi,\, 2-\xi\})^3$ является

$$\hat{G}_n = \frac{1}{n}\sum_{i=1}^{n} g(\xi_i) = \frac{1}{n}\sum_{i=1}^{n} (\max\{\xi_i,\, 2-\xi_i\})^3$$

(см. теорему 19.3.1). В рассматриваемом примере $n = 10$, а выборочные значения $\xi_1, \xi_2, \ldots, \xi_n$ приведены в условии. Значения $\hat{G}_{10} = 4{,}44$.

Задачи

19.1. С помощью таблицы случайных чисел (табл. 27.10.1) оценить значение интеграла Пуассона

$$\int\limits_{-\infty}^{+\infty} \exp\left\{-\frac{x^2}{2}\right\} dx.$$

Указание 1. Сделать в интеграле замену

$$y = \frac{1}{\pi}\operatorname{arctg} x + \frac{1}{2}.$$

Указание 2. По таблице случайных чисел (табл. 27.10.1) получить выборку из равномерного на отрезке $[0;1]$ распределения и воспользоваться теоремой 19.3.1.

19.2. С помощью таблицы случайных чисел получить выборку объемом $n = 20$ из равномерного на отрезке [0;1] распределения. По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$, $n = 20$. Построить график функции $F(x)$ равномерного на отрезке [0;1] распределения. Вычислить

$$\sup_x \left| F(x) - \hat{F}_n(x) \right|.$$

19.3. Воспользовавшись таблицей случайных чисел (см. табл. 27.10.1), получить 10 независимых выборок объемом 12 из равномерного на отрезке [0;1] распределения. Обозначим их

$$\xi_{k,1}, \xi_{k,2}, \ldots, \xi_{k,12},\ k = 1, 2, \ldots, 10.$$

Пусть

$$\eta_k = \sum_{i=1}^{12} \xi_{k,i} - 6, \ k = 1, 2, \ldots, 10.$$

Последовательность $\eta_1, \eta_2, \ldots, \eta_{10}$ является выборкой из некоторого распределения. По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_{10}(x)$. Построить график функции $N_{0;1}(x)$ нормального распределения с параметрами (0;1). Вычислить

$$\sup_x \left| N_{0;1}(x) - \hat{F}_{10}(x) \right|.$$

19.4. Продолжительность работы элемента до первого выхода из строя является показательно распределенной случайной величиной. Наблюдалась работа 20 элементов и фиксировалась продолжительность их работы до первого выхода из строя (в часах): 11, 149, 846, 563, 384, 950, 864, 63, 990, 77, 685, 158, 348, 318, 25, 278, 1803, 83, 1544, 380.

По выборке найти оценку математического ожидания продолжительности безотказной работы элемента, если его меняют после 200 часов непрерывной работы, даже если элемент не вышел из строя.

19.5. Наблюдаются цифровые части номеров 10 автомобилей (в номере пять цифр), проезжающих мимо вас. Пусть a_i, b_i, c_i, d_i, e_i — цифры, встречающиеся в номере i-го автомобиля, если его читать слева направо. Строится последовательность чисел

$$\xi_i = 10^{-1}a_i + 10^{-2}b_i + 10^{-3}c_i + 10^{-4}d_i + 10^{-5}e_i,$$

$i = 1, 2, \ldots, 10$. Их можно рассматривать как выборку из некоторого распределения.

По выборке $\xi_1, \xi_2, \ldots, \xi_{10}$ построить график эмпирической функции распределения $\hat{F}_n(x)$ $(n = 10)$. Построить график функции $F(x)$ равномерного на отрезке [0;1] распределения. Вычислить

$$\sup_x \left| F(x) - \hat{F}_n(x) \right|.$$

З а м е ч а н и е. Аналогично формулируется задача, если в цифровой части номера автомобиля четыре цифры, при этом рассматривается последовательность

$$\xi_i = 10^{-1}a_i + 10^{-2}b_i + 10^{-3}c_i + 10^{-4}d_i,$$

$i = 1, 2, \ldots, 10$.

Глава 20

Методы получения оценок

20.1 Метод моментов

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения $\mathsf{F}(\cdot; \theta_1, \theta_2, \ldots$ $\ldots, \theta_s)$, зависящего от параметров $\theta_1, \theta_2, \ldots, \theta_s$. Параметры $\theta_1, \theta_2, \ldots, \theta_s$ неизвестны и их необходимо оценить по выборке $\xi_1, \xi_2, \ldots, \xi_n$. Первым общим методом получения оценок неизвестных параметров по выборке $\xi_1, \xi_2, \ldots, \xi_n$ был метод моментов, предложенный К. Пирсоном. Этот метод состоит в приравнивании определенного количества выборочных моментов $\hat{m}_r$ соответствующим моментам

$$m_r(\theta_1, \theta_2, \ldots, \theta_s) = \int\limits_{\mathbb{R}^1} x^r \mathsf{F}(dx; \theta_1, \theta_2, \ldots, \theta_s) \tag{20.1.1}$$

распределения $\mathsf{F}(\cdot; \theta_1, \theta_2, \ldots, \theta_s)$, вычисленным при значениях параметров $\theta_1, \theta_2, \ldots, \theta_s$, равных соответственно $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s$ (моменты $m_r = m_r(\theta_1, \theta_2, \ldots, \theta_s)$ распределения $\mathsf{F}(\cdot; \theta_1, \theta_2, \ldots, \theta_s)$ являются функциями параметров $\theta_1, \theta_2, \ldots, \theta_s$, см. равенство (20.1.1)). А именно:

$$\hat{m}_1 = \int\limits_{\mathbb{R}^1} x \mathsf{F}(dx; \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s),$$

$$\hat{m}_2 = \int\limits_{\mathbb{R}^1} x^2 \mathsf{F}(dx; \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s),$$

$$\vdots$$

$$\hat{m}_s = \int\limits_{\mathbb{R}^1} x^s \mathsf{F}(dx; \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s).$$

Рассматривая количество моментов, равное количеству неизвестных параметров, подлежащих оцениванию, получаем такое же количество уравнений для определения неизвестных параметров. Решения этих уравнений относительно $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s$ дают искомые оценки.

Метод применим, если существуют все перечисленные моменты, и на практике приводит к сравнительно простым вычислениям.

Если оценивается s параметров, то, как правило, приравнивают моменты $\hat{m}_r$ и $m_r(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s)$ с 1-го по s-й.

Пример 20.1.1. *Получить оценки параметров a и σ^2 нормального распределения*

$$F(x; a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{x} \exp\left\{-\frac{(t-a)^2}{2\sigma^2}\right\} dt$$

по методу моментов.

Решение.

$$m_1(a, \sigma^2) = \int\limits_{\mathbb{R}^1} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = a,$$

$$m_2(a, \sigma^2) = \int\limits_{\mathbb{R}^1} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = \sigma^2 + a^2,$$

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} \xi_i, \;\; \hat{m}_2 = \frac{1}{n} \sum_{i=1}^{n} \xi_i^2.$$

Приравнивая моменты $m_1(a, \sigma^2)$ и $m_2(a, \sigma^2)$, вычисленные при значениях параметров a и σ^2, равных соответственно $\hat{a}$ и $\widehat{\sigma^2}$, соответствующим эмпирическим моментам, получаем уравнения относительно $\hat{a}$ и $\widehat{\sigma^2}$:

$$m_1(\hat{a}, \widehat{\sigma^2}) = \hat{m}_1, \;\; m_2(\hat{a}, \widehat{\sigma^2}) = \hat{m}_2,$$

подробнее:

$$\hat{a} = \hat{m}_1, \ \hat{\sigma^2} + \hat{a}^2 = \hat{m}_2.$$

Решения этих уравнений и будут оценками соответственно параметров a и σ^2, полученными по методу моментов. Оценкой для a является

$$\hat{a} = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n}\xi_i,$$

для σ^2 :

$$\hat{\sigma^2} = \hat{m}_2 - (\hat{m}_1)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \overline{\xi}\right)^2.$$

20.2 Метод максимального правдоподобия

Пусть дана реализация $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ выборочного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, распределение $\mathsf{F}(\cdot;\theta)$ которого зависит от параметра θ ($\theta \in \Theta$). Параметр θ неизвестен и его необходимо оценить по реализации $\xi(\omega)$ выборки ξ, другими словами, по реализации $\xi(\omega)$ выборки ξ необходимо из множества возможных значений Θ параметра θ выбрать $\hat{\theta}$, которое будем использовать в качестве приближенного значения неизвестного параметра θ.

Мы будем рассматривать ситуацию, когда распределение $\mathsf{F}(\cdot;\theta)$, однозначно определяется параметром θ и наоборот, распределение однозначно задает параметр. Например, нормальное распределение $N_{a;\sigma^2}$ однозначно определяется параметром $\theta = (a;\sigma^2)$ и наоборот.

Общим методом получения оценок, наиболее важным как в теоретическом отношении, так и в приложениях, является метод максимального правдоподобия, предложенный Фишером. Метод максимального правдоподобия основан на следующих неформальных соображениях (для наглядности предположим, что распределение $\mathsf{F}(\cdot;\theta)$ выборочного вектора ξ абсолютно непрерывно, его плотность обозначим через $f(x;\theta)$).

По плотности распределения f случайного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ всегда можно вычислить вероятность попадания значения вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ в любое борелевское множество B:

$$\mathsf{P}\{\xi \in B\} = \int\limits_B f(x)dx, \ x \in \mathbb{R}^n.$$

Из последнего равенства следует, что случайный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ чаще (по сравнению с другими значениями) принимает значения из окрестности точки x_m, в которой плотность $f(x)$ достигает наибольшего значения. Этот факт иллюстрирует рис. 20.2.1 (для наглядности $n = 1$, при этом $\xi = (\xi_1, \xi_2, \ldots, \xi_n) = \xi_1$). Вероятность попадания значения ξ в окрестность (a, b) точки x_m больше вероятности попадания ξ в промежутки (c, d) вне окрестности точки x_m (численно эти вероятности равны площадям соответствующих криволинейных трапеций, см. рис. 20.2.1). Поэтому, *если $\xi(\omega)$ — реализация случайного вектора ξ с плотностью $f(x)$, то "чаще всего" значение $f(\xi(\omega))$ плотности $f(x)$ в точке $\xi(\omega)$ и ее наибольшее значение $f(x_m)$ близки между собой.*

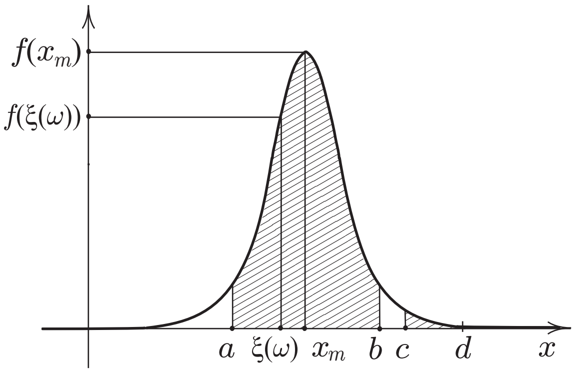

Рис. 20.2.1: К вычислению вероятности $\mathsf{P}\{\xi \in (a, b)\}$

Отсюда следует, что реализация $\xi(\omega)$ случайного вектора ξ из семейства $f(x; \theta)$, $\theta \in \Theta$, его возможных плотностей ("кандидатур" на плотность вектора ξ) выделяет плотность $f(x, \hat{\theta})$ вектора ξ (а вместе с ней и значение $\hat{\theta}$ параметра θ) тем, что значение

$f(\xi(\omega), \hat{\theta})$ плотности $f(x, \hat{\theta})$ вектора ξ в точке $\xi(\omega)$ наибольшее по сравнению со значениями $f(\xi(\omega), \theta)$ в точке $\xi(\omega)$ других возможных плотностей, а именно:

$$f(\xi(\omega), \hat{\theta}) = \max_{\theta \in \Theta} f(\xi(\omega); \theta). \qquad (20.2.1)$$

Поэтому *в качестве оценки $\hat{\theta}$ параметра θ естественно выбрать точку $\hat{\theta}$, в которой функция $f(\xi(\omega), \theta)$, от $\theta \in \Theta$, достигает наибольшего значения, т.е. точку $\hat{\theta}$, являющуюся решением уравнения* (20.2.1).

Проиллюстрируем сказанное. Пусть $\xi(\omega)$ — реализация выборки ξ объемом 1 из распределения $N_{\theta;1}$, $\theta \in \mathbb{R}^1$, и $f(x;\theta)$ — его плотность (объем выборки 1 — для наглядности). Семейство возможных плотностей

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta)^2}{2}\right\}, \ \theta \in \mathbb{R}^1, \qquad (20.2.2)$$

выборки ξ и ее реализация $\xi(\omega)$ — значение, принятое случайной величиной ξ, изображено на рис. 20.2.2.

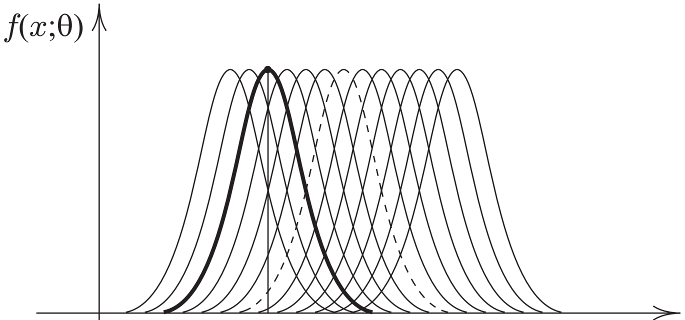

Рис. 20.2.2: К выбору плотности ξ
из семейства $f(x;\theta)$ по значению $\xi(\omega)$

Из всех плотностей семейства (20.2.2) точка $\xi(\omega)$ выделяет как плотность распределения случайной величины ξ ту плотность, которая в точке $\xi(\omega)$ достигает наибольшего значения (плотность, выделенная на рис. 20.2.2 “жирно”, в точке $\xi(\omega)$ принимает наибольшее значение, а, например, плотность, выделенная пунктиром, — нет). Вместе с плотностью случайной величины ξ мы получаем значение параметра θ, точнее, его оценку. Здесь оценкой параметра θ является $\hat{\theta} = \xi(\omega)$.

Приведенные выше соображения можно повторить с незначительными изменениями и для дискретно распределенного вектора ξ.

Далее изложенный подход получения оценок приводится в формализованном виде.

Определение. *Функцией максимального правдоподобия* выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ с распределением $\mathsf{F}(\cdot;\theta)$, зависящим от параметра $\theta \in \Theta \subset \mathbb{R}^s$, называется функция $L(\theta)$ параметра $\theta \in \Theta$, определяемая равенством

$$L(\theta) = f(\xi;\theta),\ \theta \in \Theta,$$

если выборочный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ абсолютно непрерывный с плотностью $f(x;\theta) = f(x_1, x_2, \ldots, x_n;\theta)$, $x \in \mathbb{R}^n$, и равенством

$$L(\theta) = \mathsf{F}(\xi;\theta),\ \theta \in \Theta,$$

если выборочный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ дискретный с распределением $\mathsf{F}(x;\theta) = \mathsf{F}(x_1, x_2, \ldots, x_n;\theta)$.

Определение. *Оценкой максимального правдоподобия* параметра θ называется точка $\hat{\theta}$, в которой функция максимального правдоподобия достигает наибольшего значения, другими словами, оценка максимального правдоподобия параметра θ — это решение уравнения

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta).$$

Поскольку функции $L(\theta)$ и $\ln L(\theta)$ достигают наибольшего значения в одних и тех же точках (функция $y = \ln x$ монотонно возрастает), а находить точку, в которой $\ln L(\theta)$ достигает наибольшего значения, зачастую проще, то оценку максимального правдоподобия можно получить и как решение уравнения

$$\ln L(\hat{\theta}) = \max_{\theta \in \Theta} \ln L(\theta).$$

Функцию $\ln L(\theta)$ еще называют *логарифмической функцией максимального правдоподобия.*

Если функция максимального правдоподобия $L(\theta_1, \theta_2, \ldots, \theta_s)$ непрерывно дифференцируема по $\theta_1, \theta_2, \ldots, \theta_s$, то для решения уравнения

$$\ln L(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s) = \max_{(\theta_1, \theta_2, \ldots, \theta_s) \in \Theta} \ln L(\theta_1, \theta_2, \ldots, \theta_s) \qquad (20.2.3)$$

достаточно найти стационарные точки функции $\ln L(\theta_1, \theta_2, \ldots, \theta_s)$, решая уравнения

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \theta_2, \ldots, \theta_s) = 0, \; i = 1, 2, \ldots, s,$$

и, сравнив значения функции $\ln L(\theta_1, \theta_2, \ldots, \theta_s)$ в ее стационарных точках и граничных точках множества Θ, выбрать точку $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_s)$, в которой функция $\ln L(\theta_1, \theta_2, \ldots, \theta_s)$ достигает наибольшего значения. Эта точка и будет решением уравнения (20.2.3).

Уравнения

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \theta_2, \ldots, \theta_s) = 0, \; i = 1, 2, \ldots, s,$$

называют *уравнениями максимального правдоподобия.*

З а м е ч а н и е. Получая оценку параметра θ как решение уравнения $L(\hat{\theta}) = \max\limits_{\theta \in \Theta} L(\theta)$, следует отбросить все решения вида $\theta = const$. Оценки, не зависящие от выборки, для нас интереса не представляют.

Пример 20.2.1 (оценивание вероятности события). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения*

$$\mathsf{Q}(x; p) = p^x (1-p)^{1-x}, \; x = 0, 1; \; p \in (0; 1).$$

Получить оценку максимального правдоподобия параметра p.

Р е ш е н и е. Оценкой максимального правдоподобия $\hat{p}$ параметра p является точка, в которой функция максимального правдоподобия достигает наибольшего значения.

Выпишем функцию максимального правдоподобия. Совместное распределение независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$

$$\mathsf{F}(x_1, x_2, \ldots, x_n; p) = \prod_{i=1}^{n} \mathsf{Q}(x_i; p) =$$

$$= \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum\limits_{i=1}^{n} x_i} (1-p)^{n - \sum\limits_{i=1}^{n} x_i},$$

а, следовательно, по определению функция максимального правдоподобия

$$L(p) = p^{\sum\limits_{i=1}^{n} \xi_i} (1-p)^{n-\sum\limits_{i=1}^{n} \xi_i}.$$

А логарифмическая функция правдоподобия

$$\ln L(p) = \sum_{i=1}^{n} \xi_i \ln p + (n - \sum_{i=1}^{n} \xi_i) \ln(1-p) =$$

$$= n(\overline{\xi} \ln p + (1-\overline{\xi}) \ln(1-p)).$$

Функция $\ln L(p)$ дифференцируема по p, поэтому можно выписать уравнение правдоподобия

$$\frac{d}{dp} \ln L(p) = 0,$$

или, после дифференцирования,

$$n(\overline{\xi}/p - (1-\overline{\xi})/(1-p)) = 0.$$

Решением уравнения правдоподобия является $p = \overline{\xi}$. На границе области допустимых значений для p значение $L(p) = 0$, поэтому функция максимального правдоподобия $L(p)$ достигает наибольшего значения в точке $p = \overline{\xi}$, и, следовательно,

$$\hat{p} = \overline{\xi}$$

является оценкой максимального правдоподобия параметра p.

Пример 20.2.2. *По выборке $\xi_1, \xi_2, \ldots, \xi_n$ из показательного распределения с плотностью*

$$p(x;\theta) = \begin{cases} \theta e^{-\theta x}, & \text{при} \quad x > 0, \\ 0, & \text{при} \quad x \le 0 \end{cases}$$

получить оценку максимального правдоподобия для параметра θ.

Решение. Сначала выпишем функцию максимального правдоподобия. Совместная плотность $p(x_1, x_2, \ldots, x_n; \theta)$ распределения независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ равна произведению плотностей $p(x_i; \theta)$ компонент:

$$p(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum\limits_{i=1}^{n} x_i},$$

если все $x_i > 0$, а если хотя бы одно $x_i \leq 0$, то

$$p(x_1, x_2, \ldots, x_n; \theta) = 0.$$

По определению функция максимального правдоподобия $L(\theta)$ выборки $\xi_1, \xi_2, \ldots, \xi_n$

$$L(\theta) = \theta^n \exp\left\{-\theta \sum_{i=1}^{n} \xi_i\right\} = \theta^n \exp\left\{-\theta n\overline{\xi}\right\},$$

если все $\xi_i > 0$ (и равна 0, если хотя бы одно $\xi_j \leq 0$,) а логарифмическая функция правдоподобия

$$\ln L(\theta) = n \ln \theta - n\theta\overline{\xi}$$

(если все $\xi_i > 0$). Логарифмическая функция максимального правдоподобия дифференцируема по θ, поэтому можно выписать уравнение правдоподобия

$$\frac{d}{d\theta} \ln L(\theta) = 0$$

или, после дифференцирования,

$$n/\theta - n\overline{\xi} = 0.$$

Решением уравнения правдоподобия является $\theta = 1/\overline{\xi}$. Функция максимального правдоподобия $L(\theta)$ на границе области допустимых значений параметра θ равна 0. Поэтому в точке $\hat{\theta} = 1/\overline{\xi}$ функция максимального правдоподобия $L(\theta)$ достигает наибольшего значения, и, следовательно,

$$\hat{\theta} = 1/\overline{\xi}$$

является оценкой максимального правдоподобия параметра θ.

Асимптотические свойства оценок максимального правдоподобия. Мы рассмотрели метод максимального правдоподобия получения оценок. Оказывается, что оценки, полученные по методу максимального правдоподобия, “хорошие”, а именно, они являются состоятельными, асимптотически нормальными и асимптотически эффективными.

Напомним, что несмещенная оценка θ^* параметра θ называется эффективной, если

$$\mathsf{D}\theta^* = \inf_{\hat\theta\colon \mathsf{M}\hat\theta=\theta} \mathsf{D}\hat\theta.$$

Определение. Величина

$$e(\hat\theta) = \frac{\inf\limits_{\hat\theta\colon \mathsf{M}\hat\theta=\theta} \mathsf{D}\hat\theta}{\mathsf{D}\hat\theta}$$

называется *эффективностью* несмещенной оценки $\hat\theta$ параметра θ.

Ясно, что для эффективной оценки θ^* параметра θ значение $e(\theta^*) = 1$, и наоборот, если для несмещенной оценки θ^* параметра θ значение $e(\theta^*) = 1$, то θ^* — эффективная оценка параметра θ.

Пусть объем выборки $\xi_1, \xi_2, \ldots, \xi_n$ может неограниченно возрастать, и пусть $\hat\theta_n$ — несмещенная оценка параметра θ, а $e(\hat\theta_n)$ — эффективность оценки $\hat\theta_n$, $n = 1, 2, \ldots$ Если при $n \to \infty$

$$e(\hat\theta_n) = \frac{\inf\limits_{\hat\theta\colon \mathsf{M}\hat\theta=\theta} \mathsf{D}\hat\theta}{\mathsf{D}\hat\theta_n} \to 1,$$

то последовательность оценок $\{\hat\theta_n\}$ параметра θ называют *асимптотически эффективной.*

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из абсолютно непрерывного распределения с плотностью $g(x;\theta)$, $\hat\theta_n$ — несмещенная оценка параметра θ, причем $g(x;\theta)$ и $\hat\theta_n$ удовлетворяют условиям, при которых имеет место неравенство Крамера—Рао (см. теорему 18.1.2), тогда

$$\mathsf{D}\hat\theta_n \geq \frac{1}{n\mathsf{D}\left(\dfrac{\partial}{\partial\theta} \ln g(\xi_1;\theta)\right)}$$

(см. равенство (18.1.12)). Поэтому при каждом фиксированном n

$$\inf_{\hat\theta_n\colon \mathsf{M}\hat\theta_n=\theta} \mathsf{D}\hat\theta_n \geq \frac{1}{n\mathsf{D}\left(\dfrac{\partial}{\partial\theta} \ln g(\xi_1;\theta)\right)},$$

и, следовательно,

$$e(\hat{\theta}_n) = \frac{\inf\limits_{\hat{\theta}\colon \mathsf{M}\hat{\theta}=\theta} \mathsf{D}\hat{\theta}}{\mathsf{D}\hat{\theta}_n} \geq \frac{1}{n\mathsf{D}\left(\dfrac{\partial}{\partial\theta}\ln g(\xi_1;\theta)\right)\mathsf{D}\hat{\theta}_n}.$$

Если при этом

$$\frac{1}{n\mathsf{D}\left(\dfrac{\partial}{\partial\theta}\ln g(\xi_1;\theta)\right)\mathsf{D}\hat{\theta}_n} \to 1,$$

то последовательность оценок $\{\hat{\theta}_n\}$ является асимптотически эффективной.

Часто бывает известно, что последовательность оценок $\{\hat{\theta}_n\}$ данного параметра θ асимптотически нормальна с параметрами $(\theta; 1/(c^2n))$ $(c > 0)$, т. е. функция распределения случайной величины

$$(\hat{\theta}_n - \theta) \Big/ \frac{1}{c\sqrt{n}}$$

сходится в каждой точке к функции распределения нормального распределения с параметрами $(0;1)$. При этом $\hat{\theta}_n$ могут быть как несмещенными, так и смещенными оценками θ, но если при этом

$$\frac{1}{n\mathsf{D}\left(\dfrac{\partial}{\partial\theta}\ln g(\xi_1;\theta)\right)\dfrac{1}{c^2n}} = \frac{1}{\dfrac{1}{c^2}\mathsf{D}\left(\dfrac{\partial}{\partial\theta}\ln g(\xi_1;\theta)\right)} = 1, \qquad (20.2.4)$$

то асимптотически нормальную с параметрами $(\theta; (1/(c^2n)))$ последовательность оценок $\{\hat{\theta}_n\}$ мы также будем называть асимптотически эффективной.

Теорема 20.2.1 (о решении уравнения максимального правдоподобия). *Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из абсолютно непрерывного распределения с плотностью $g(x;\theta)$, $\theta \in \Theta$. Если $g(x;\theta)$ трижды дифференцируема по параметру $\theta \in \Theta$, для нее выполняется условие дифференцируемости дважды по параметру θ под знаком интеграла (см. теорему 18.1.1) и*

$$\mathsf{M}\left|\frac{\partial^2}{\partial\theta^2}\ln g(\xi_1;\theta)\right| < +\infty,$$

$$\mathsf{M}\left|\frac{\partial^3}{\partial\theta^3}\ln g(\xi_1;\theta)\right| \le K$$

для всех $\theta \in \Theta$, то решение θ_n^* уравнение правдоподобия

$$\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} g(x_i;\theta) = 0$$

является состоятельной, асимптотически нормальной с параметрами $(\theta, 1/I_n(\theta))$ и асимптотически эффективной оценкой параметра θ, где

$$I_n(\theta) = n\mathsf{D}\left(\frac{\partial}{\partial\theta}\ln g(\xi_1;\theta)\right) = n\sigma^2$$

— информация по Фишеру.

Доказательство. Совместную плотность $\xi_1, \xi_2, \ldots, \xi_n$ обозначим через $f(x;\theta)$:

$$f(x;\theta) = f(x_1, x_2, \ldots, x_n;\theta) = \prod_{i=1}^{n} g(x_i;\theta).$$

Через θ_0 обозначим неизвестное истинное значение параметра θ. Будем полагать, что θ_0 — внутренняя точка интервала Θ.

Сначала покажем, что решение θ_n^* уравнения правдоподобия

$$\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = 0$$

является состоятельной оценкой θ_0, т. е. при $n \to \infty$ сходится по вероятности к θ_0.

Выпишем разложение Тейлора для функции

$$\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = \frac{\partial}{\partial\theta}\ln f(\xi_1, \xi_2, \ldots, \xi_n, ;\theta) =$$

$$= \frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} g(\xi_i;\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta)$$

в окрестности точки θ_0. Для каждой функции $\dfrac{\partial}{\partial\theta}\ln g(\xi_i;\theta)$, $i = 1, 2, \ldots, n$, в окрестности точки θ_0 имеем

$$\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta) = \frac{\partial}{\partial\theta}\ln g(\xi_i;\theta_0) + (\theta-\theta_0)\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta_0) +$$

$$+\frac{1}{2}\,(\theta-\theta_0)^2\,\frac{\partial^3}{\partial\theta^3}\ln g(\xi_i;\theta_0+t(\xi_i)),$$

$\theta_0+t(\xi_i)$ — точка из промежутка с концами θ_0 и $\theta, i=1,2,\ldots,n$. Просуммировав левую и правую части по i от 1 до n и умножив на $1/n$, получим:

$$\frac{1}{n}\frac{\partial}{\partial\theta}\ln f(\xi;\theta)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta_0)+$$

$$+(\theta-\theta_0)\,\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta_0)+$$

$$+\frac{1}{2}\,(\theta-\theta_0)^2\,\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta^3}\ln g(\xi_i;\theta_0+t(\xi_i))=$$

$$=B_0+(\theta-\theta_0)B_1+\frac{1}{2}(\theta-\theta_0)^2B_2,$$

где

$$B_0=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta_0),\;\; B_1=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta_0),$$

$$B_2=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta^3}\ln g(\xi_i;\theta_0+t(\xi_i)).$$

Тогда уравнение правдоподобия

$$\frac{\partial}{\partial\theta}\ln f(\xi;\theta)=0$$

можно записать в виде:

$$B_0+(\theta-\theta_0)B_1+\frac{1}{2}(\theta-\theta_0)^2B_2=0. \tag{20.2.5}$$

Оценим значения квадратного относительно θ трехчлена

$$\frac{1}{n}\frac{\partial}{\partial\theta}\ln f(\xi;\theta)=B_0+(\theta-\theta_0)B_1+\frac{1}{2}(\theta-\theta_0)^2B_2$$

на концах промежутка $[\theta_0-\delta,\theta_0+\delta]$ при достаточно малых δ и достаточно больших n. Для этого сначала выясним, как ведут себя коэффициенты B_0, B_1, B_2 при $n\to\infty$. Из условия теоремы следует, что

$$\mathsf{M}\frac{\partial^3}{\partial\theta^3}\ln g(\xi_i;\theta_0+t(\xi_i)) = C\neq\infty,$$

а в теореме 18.1.2 установлено, что

$$\mathsf{M}\left(\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta_0)\right)=0,$$

$$\mathsf{M}\left(\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta_0)\right)=-\mathsf{D}\left(\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta_0)\right)=-\sigma^2\neq\infty.$$

Поэтому при $n\to\infty$ согласно закону больших чисел в форме Хинчина (см. теорему 14.1.2) коэффициенты B_0, B_1, B_2 сходятся по вероятности, а именно

$$B_0=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta_0)\xrightarrow{\mathsf{P}}\mathsf{M}\frac{\partial}{\partial\theta}\ln g(\xi_i;\theta_0)=0,$$

$$B_1=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta_0)\xrightarrow{\mathsf{P}}\mathsf{M}\frac{\partial^2}{\partial\theta^2}\ln g(\xi_i;\theta_0)=-\sigma^2,$$

$$B_2=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta^3}\ln g(\xi_i;\theta_0+t(\xi_i))\xrightarrow{\mathsf{P}}\mathsf{M}\frac{\partial^3}{\partial\theta^3}\ln g(\xi_i;\theta_0+t(\xi_i))=C.$$

И, следовательно, для данных $\delta>0$ и $\varepsilon>0$, начиная с некоторого N $(n>N)$,

$$\mathsf{P}\left\{-\delta^2<B_0<\delta^2\right\}\geq 1-\varepsilon/3,\ \ \mathsf{P}\left\{-3\sigma^2/2<B_1<-\sigma^2/2\right\}\geq 1-\varepsilon/3,$$

$$\mathsf{P}\left\{-2|C|<B_2<2|C|\right\}\geq 1-\varepsilon/3,$$

а вероятность события

$$S=\{-\delta^2<B_0<\delta^2\}\cap$$

$$\cap\{-3\sigma^2/2<B_1<-\sigma^2/2\}\cap\{-2|C|<B_2<2|C|\}$$

больше $1-\varepsilon$:
$$\mathsf{P}(S) \geq 1-\varepsilon$$
(легко проверить, что $\mathsf{P}(\overline{S}) \leq \varepsilon$).

Пусть $\omega \in S$ и $n > N$. Определим знак многочлена
$$\frac{1}{n}\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = B_0 + (\theta-\theta_0)B_1 + \frac{1}{2}(\theta-\theta_0)^2 B_2$$
на концах промежутка $[\theta_0-\delta, \theta_0+\delta]$. В точке $\theta_0+\delta$ значение
$$\frac{1}{n}\frac{\partial}{\partial\theta}\ln f(\xi;\theta_0+\delta) = B_0 + \delta B_1 + \frac{1}{2}\delta^2 B_2 \leq$$
$$\leq \delta^2 - \delta\sigma^2/2 + \frac{1}{2}\delta^2 2|C| = \delta\left(\delta(1+|C|) - \sigma^2/2\right)$$
меньше нуля при достаточно малых δ. В точке $\theta_0-\delta$ значение
$$\frac{1}{n}\frac{\partial}{\partial\theta}\ln f(\xi;\theta_0-\delta) = B_0 - \delta B_1 + \frac{1}{2}\delta^2 B_2 \geq$$
$$\geq -\delta^2 + \delta\sigma^2/2 + \frac{1}{2}\delta^2(-2|C|) = \delta\left(\sigma^2/2 - \delta(1+|C|)\right),$$
положительно при достаточно малых δ.

Так что при достаточно малых $\delta > 0$ и $\omega \in S$ непрерывная по θ функция
$$\frac{1}{n}\frac{\partial}{\partial\theta}\ln f(\xi;\theta) = B_0 + (\theta-\theta_0)B_1 + \frac{1}{2}(\theta-\theta_0)^2 B_2$$
на концах промежутка $[\theta_0-\delta, \theta_0+\delta]$ принимает значения разных знаков. Поэтому при $\omega \in S$ на промежутке $[\theta_0-\delta, \theta_0+\delta]$ существует точка θ_n^*, для которой
$$\frac{\partial}{\partial\theta}\ln f(\xi;\theta_n^*) = 0,$$
т. е. на промежутке $[\theta_0-\delta, \theta_0+\delta]$ уравнение правдоподобия имеет корень. Фактически установлено, что при достаточно больших n
$$S \subset \{\omega:\ |\theta_n^* - \theta_0| < \delta\}$$
и, следовательно,
$$\mathsf{P}\left\{|\theta_n^* - \theta_0| < \delta\right\} \geq \mathsf{P}(S) \geq 1-\varepsilon.$$

Последнее означает, что решение θ_n^* уравнения правдоподобия при $n \to \infty$ сходится по вероятности к θ_0, т. е. θ_n^* — состоятельная оценка θ_0.

Докажем теперь, что θ_n^* асимптотически нормальна с параметрами $(\theta_0; (1/I_n(\theta))$, т. е. функция распределения случайной величины

$$(\theta_n^* - \theta_0) \Big/ \frac{1}{\sqrt{I_n(\theta)}} = (\theta_n^* - \theta_0) \Big/ \frac{1}{\sigma\sqrt{n}}$$

сходится к функции $N_{0;1}(x)$ нормального распределения с параметрами $(0; 1)$.

Для θ_n^* имеет место равенство:

$$B_0 + (\theta_n^* - \theta_0)B_1 + \frac{1}{2}\,(\theta_n^* - \theta_0)^2 B_2 = 0.$$

Отсюда

$$\theta_n^* - \theta_0 = \frac{-B_0}{B_1 + \frac{1}{2}\,(\theta_n^* - \theta_0)B_2} = \frac{\sum\limits_{i=1}^{n} \frac{\partial}{\partial\theta} \ln g(\xi_i; \theta_0)}{-n(B_1 + \frac{1}{2}\,(\theta_n^* - \theta_0)B_2)},$$

$$(\theta_n^* - \theta_0) \Big/ \frac{1}{\sigma\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} \frac{\partial}{\partial\theta} \ln g(\xi_i; \theta_0)}{-\frac{\sqrt{n}}{\sigma}\left(B_1 + \frac{1}{2}(\theta_n^* - \theta_0)B_2\right)} =$$

$$= \frac{\left(\sum\limits_{i=1}^{n} \frac{\partial}{\partial\theta} \ln g(\xi_i; \theta_0)\right) \Big/ \sigma\sqrt{n}}{-\frac{1}{\sigma^2}\left(B_1 + \frac{1}{2}\,(\theta_n^* - \theta_0)B_2\right)}. \tag{20.2.6}$$

Функция распределения числителя правой части (20.2.6) сходится к $N_{0;1}(x)$ (в силу центральной предельной теоремы для одинаково распределенных случайных величин), а знаменатель сходится по вероятности к 1, что следует из сходимости

$$B_1 \xrightarrow{\mathsf{P}} -\sigma^2,\ B_2 \xrightarrow{\mathsf{P}} C,\ \theta_n^* \xrightarrow{\mathsf{P}} \theta_0$$

при $n \to \infty$. Поэтому функция распределения дроби в правой части равенства (20.2.6) сходится к функции распределения $N_{0;1}(x)$ (см., например, Г. Крамер "Математические методы статистики", стр. 281).

Так что функция распределения случайной величины

$$(\theta_n^* - \theta_0) \Big/ \frac{1}{\sigma\sqrt{n}} = (\theta_n^* - \theta_0) \Big/ \frac{1}{\sqrt{I_n(\theta)}}$$

сходится к $\mathsf{N}_{0;1}(x)$, т. е. случайная величина θ_n^* является асимптотически нормальной с параметрами $(\theta_0, 1/I_n(\theta))$.

Далее, так как θ_n^* асимптотически нормальна с параметрами $(\theta_0; 1/I_n(\theta))$, а

$$\frac{1}{n\mathsf{D}\left(\frac{\partial}{\partial\theta}\ln g(\xi_1;\theta)\right)\frac{1}{I_n(\theta)}} = \frac{1}{n\sigma^2\frac{1}{\sigma^2 n}} = 1,$$

то θ_n^* — асимптотически эффективна (см. равенство (20.2.4)).

Тем самым теорема полностью доказана.

20.3 Примеры и задачи

Примеры

Следующий пример иллюстрирует получение оценки максимального правдоподобия в ситуации, когда функция максимального правдоподобия не дифференцируема.

Пример 20.3.1. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью*

$$f(x;\alpha,\mu) = \frac{1}{2\alpha}\exp\left\{-\frac{|x-\mu|}{\alpha}\right\},\ x \in \mathbb{R}^1,\ \alpha > 0,$$

($f(x;\alpha,\mu)$ — плотность распределения Лапласа). Найти оценки максимального правдоподобия параметров α и μ.

Р е ш е н и е. Совместная плотность распределения случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ равна

$$f(x_1, x_2, \ldots, x_n;\alpha,\mu) = \prod_{i=1}^{n} f(x_i;\alpha,\mu) =$$

$$= \prod_{i=1}^{n} \frac{1}{2\alpha} \exp\left\{-\frac{|x_i-\mu|}{\alpha}\right\} = \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}\sum_{i=1}^{n}|x_i-\mu|\right\}.$$

Поэтому функция правдоподобия

$$L(\alpha,\mu) = f(\xi_1,\xi_2,\ldots,\xi_n;\alpha,\mu) = \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}\sum_{i=1}^{n}|\xi_i-\mu|\right\}.$$

Оценкой максимального правдоподобия параметра $(\alpha;\mu)$ является точка $(\hat\alpha;\hat\mu)$, в которой $L(\alpha;\mu)$ достигает наибольшего значения. Найдем эту точку. Заметим, что функция правдоподобия по μ не дифференцируема.

Пусть

$$Q(\mu) = \sum_{i=1}^{n}|\xi_i-\mu|.$$

Через Q^* обозначим минимальное значение функции $Q(\mu)$. Пусть оно достигается при $\mu=\hat\mu$, т. е.

$$Q^* = Q(\hat\mu)$$

(при $n \geq 2$ значение $Q^* > 0$). Тогда для каждого α

$$L(\alpha,\mu) = \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}Q(\mu)\right\} \leq \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}Q^*\right\}.$$

Далее, пусть $\hat\alpha$ — точка, в которой функция

$$L(\alpha,\hat\mu) = \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}Q^*\right\}$$

достигает наибольшего значения. Для произвольной точки $(\alpha;\mu)$ и точки $(\hat\alpha;\hat\mu)$ имеем:

$$L(\alpha,\mu) \leq \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}Q^*\right\} \leq \left(\frac{1}{2\hat\alpha}\right)^n \exp\left\{-\frac{1}{\hat\alpha}Q^*\right\}.$$

Поэтому $(\hat\alpha;\hat\mu)$ — точка, в которой $L(\alpha;\mu)$ достигает наибольшего значения. И ее можно получить так: 1° находим точку $\hat\mu$, в которой

$$Q(\mu) = \sum_{i=1}^{n}|\xi_i-\mu|$$

достигает наименьшего значения; 2° находим точку $\hat{\alpha}$, в которой

$$L(\alpha, \hat{\mu}) = \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}Q^*\right\}$$

достигает наибольшего значения. Точка $(\hat{\alpha}; \hat{\mu})$ и будет искомой.

Найдем точку $\hat{\mu}$, в которой функция $Q(\mu) = \sum\limits_{i=1}^{n} |\xi_i - \mu|$ достигает минимального значения.

Пусть $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ — вариационный ряд последовательности $\xi_1, \xi_2, \ldots, \xi_n$ — выборочные значения, расположенные в порядке возрастания. Очевидно

$$Q(\mu) = \sum_{i=1}^{n} |\xi_i - \mu| = \sum_{i=1}^{n} |\xi_i^* - \mu|.$$

Рассмотрим два случая: 1) n — нечетно и 2) n — четно.

1) Пусть $n = 2k - 1$. На каждом из промежутков $(-\infty, \xi_1^*]$, $(\xi_{i-1}^*, \xi_i^*]$, $i = 2, 3, \ldots, n$, $(\xi_n^*, +\infty)$ функция $Q(\mu) = \sum\limits_{i=1}^{n} |\xi_i^* - \mu|$ линейна. При этом на промежутках $(-\infty, \xi_1^*]$, $(\xi_{i-1}^*, \xi_i^*]$, $i = 2, 3, \ldots, k$, убывает, так как угловой коэффициент — коэффициент при μ — отрицателен, а на каждом из промежутков $(\xi_{i-1}^*, \xi_i^*]$, $i = k+1, \ldots, 2k-1$, (ξ_{2k-1}^*, ∞) возрастает (коэффициент при μ положителен). Поэтому наименьшего значения непрерывная функция $Q(\mu)$ достигает в точке ξ_k^* (см. рис. 20.3.1).

2) Пусть теперь $n = 2k$. Тогда на каждом из промежутков $(-\infty, \xi_1^*]$, $(\xi_{i-1}^*, \xi_i^*]$, $i = 2, 3, \ldots, k$, функция $Q(\mu)$ убывает, на промежутке $(\xi_k^*, \xi_{k+1}^*]$ постоянна, на промежутках $(\xi_{i-1}^*, \xi_i^*]$, $i = k+2, k+3, \ldots, 2k$, $(\xi_n^*, +\infty)$ возрастает и, следовательно, наименьшего значения функция $Q(\mu)$ достигает в каждой точке промежутка $(\xi_k^*, \xi_{k+1}^*]$ (см. рис. 20.3.2).

Далее,

$$L(\alpha, \hat{\mu}) = \left(\frac{1}{2\alpha}\right)^n \exp\left\{-\frac{1}{\alpha}Q^*\right\},$$

$$\ln L(\alpha, \hat{\mu}) = -n\ln(2\alpha) - \frac{Q^*}{\alpha},$$

$$\frac{d}{d\alpha}\ln L(\alpha, \hat{\mu}) = -\frac{n}{\alpha} + \frac{1}{\alpha^2}Q^*.$$

Решая уравнение

$$-\frac{n}{\alpha}+\frac{1}{\alpha^2}Q^*=0,$$

получаем точку $\hat{\alpha}=Q^*/n$, в которой функция $L(\alpha,\hat{\mu})$ достигает наибольшего значения.

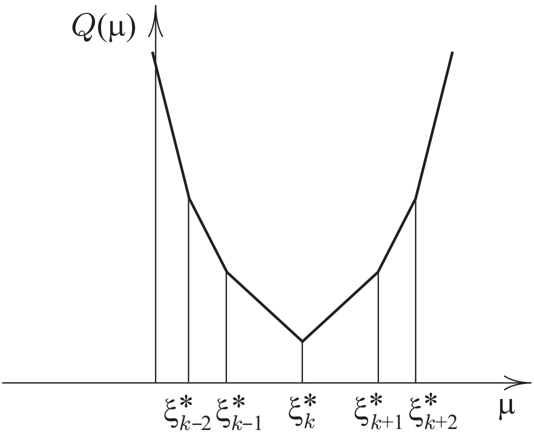

Рис. 20.3.1: График функции $Q(\mu)$ при $n=2k-1$

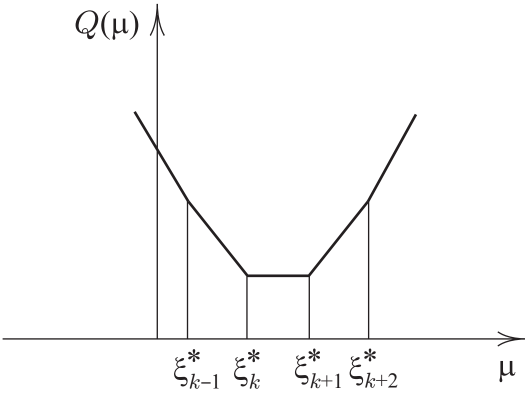

Рис. 20.3.2: График функции $Q(\mu)$ при $n=2k$

Итак, при $n=2k-1$ оценкой максимального правдоподобия параметра (α, μ) является

$$(\hat{\alpha}, \hat{\mu}) = \left(\frac{1}{n} Q^*, \xi_k^*\right),$$

при $n = 2k$ —

$$(\hat{\alpha}, \hat{\mu}) = \left(\frac{1}{n} Q^*, \hat{\mu}\right),$$

где $\hat{\mu}$ — произвольная точка промежутка $(\xi_k^*, \xi_{k+1}^*]$.

Задачи

20.1. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & \text{если } x \in [a, b]; \\ 0, & \text{если } x \notin [a, b]. \end{cases}$$

Найти оценки параметров a и b по методу максимального правдоподобия.

Выяснить, являются ли оценки максимального правдоподобия $\hat{a}$ и $\hat{b}$ параметров a и b их несмещенными оценками? Состоятельными оценками?

20.2. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, h) = \begin{cases} \dfrac{1}{2h}, & \text{если } x \in [\theta - h, \theta + h]; \\ 0, & \text{если } x \notin [\theta - h, \theta + h] \end{cases}$$

($f(x; \theta, h)$ — плотность равномерного на отрезке $[\theta - h, \theta + h]$ распределения).

Найти оценки параметров θ и h по методу максимального правдоподобия.

Выяснить, являются ли оценки максимального правдоподобия $\hat{\theta}$ и $\hat{h}$ параметров θ и h несмещенными оценками? Состоятельными оценками?

20.3. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \dfrac{1}{2h_0}, & \text{если } x \in [\theta - h_0, \theta + h_0]; \\ 0, & \text{если } x \notin [\theta - h_0, \theta + h_0]. \end{cases}$$

Найти оценку $\hat{\theta}$ параметра θ по методу максимального правдоподобия.

Выяснить, будет ли $\hat{\theta}$ несмещенной, состоятельной оценкой параметра θ.

20.4. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;h) = \begin{cases} \dfrac{1}{2h}, & \text{если } x \in [\theta_0 - h, \theta_0 + h]; \\ 0, & \text{если } x \notin [\theta_0 - h, \theta_0 + h]. \end{cases}$$

Найти оценку параметра h по методу максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра h его несмещенной и состоятельной оценкой.

20.5. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из геометрического распределения с параметром p:

$$\mathsf{P}(k;p) = (1-p)^k p, \ k = 0, 1, \ldots$$

Оценить параметр p по методу моментов.

20.6. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;a,b) = \begin{cases} \dfrac{1}{a} \exp\left\{-\dfrac{1}{a}(x-b)\right\}, & \text{если } x > b; \\ 0, & \text{если } x \le b \end{cases}$$

($f(x;a,b)$ — плотность смещенного показательного распределения).

Найти оценки параметров a и b по методу максимального правдоподобия.

Выяснить, являются ли оценки максимального правдоподобия параметров a и b несмещенными и состоятельными оценками.

20.7. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения

$$\mathsf{P}(k;\theta) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^k, k = 0, 1, \ldots, \ \theta > 0.$$

Найти оценку параметра θ по методу максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра θ его несмещенной, состоятельной, эффективной оценкой.

20.8. Обозначим через ξ случайную величину — число неудач до r-го успеха в неограниченной последовательности независимых испытаний с вероятностью успеха p в одном испытании. Случайная величина ξ имеет отрицательное биномиальное распределение (распределение Паскаля) с параметрами (r, p):

$$\mathsf{P}\{\xi = k\} = C_{r-1+k}^{k} p^r (1-p)^k, \;\; k = 0, 1, \ldots$$

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из отрицательного биномиального распределения с параметрами (r, p), r — известно. Найти оценку параметра p по методу моментов.

У к а з а н и е. Общее количество неудач до r-го успеха можно представить в виде суммы:

$$\xi = \eta_1 + \eta_2 + \ldots + \eta_r,$$

где $\eta_1, \eta_2, \ldots, \eta_r$ — независимые случайные величины, каждая из которых имеет геометрическое распределение с параметром p.

20.9. С целью определения количества N рыбин в озере выловили K рыбин, пометили их и выпустили в озеро. Через некоторое время в озере выловили n рыбин. При этом k из них оказались меченными.

Какое наиболее вероятное количество рыбин в озере?

У к а з а н и е. Задачу решить в предположении, что количество N рыбин в озере велико по сравнению с n. В этом предположении случайную величину ξ — количество меченных рыбин среди n выловленных — можно считать биномиально распределенной.

20.10. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \dfrac{1}{a} \exp\left\{-\dfrac{1}{a}(x-b)\right\}, & \text{если } x > b; \\ 0, & \text{если } x \le b. \end{cases}$$

Найти оценки параметров a и b по методу моментов.

20.11. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из биномиального распределения

$$\mathsf{P}(k; m) = C_m^k p^k (1-p)^{m-k}, \;\; k = 0, 1, \ldots, m.$$

(m известно).

Найти оценку $\hat{p}$ параметра p по методу моментов. Выяснить, является ли оценка $\hat{p}$ несмещенной, состоятельной и эффективной оценкой p.

20.12. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения Пуассона

$$\mathsf{P}(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, \;\; k = 0, 1, \ldots$$

Найти оценку параметра λ по методу максимального правдоподобия.

Выяснить, является оценка максимального правдоподобия параметра λ его несмещенной, состоятельной, эффективной оценкой.

20.13. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения с плотностью

$$f(x;\theta,h) = \begin{cases} \dfrac{1}{2\sqrt{3h}}, & \text{если } x \in [\theta - \sqrt{3h}, \theta + \sqrt{3h}]; \\ 0, & \text{если } x \notin [\theta - \sqrt{3h}, \theta + \sqrt{3h}]. \end{cases}$$

Найти оценки $\hat{\theta}$ и $\hat{h}$ параметров θ и h по методу моментов.

Выяснить, являются ли $\hat{\theta}$, $\hat{h}$ несмещенными, состоятельными оценками параметров θ, h.

Глава 21

Нормальная выборка

21.1 Многомерное нормальное распределение

Здесь нам будет удобно рассматривать вектор $(\xi_1, \xi_2, \ldots, \xi_n)$ как вектор-столбец $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ (штрих обозначает операцию транспонирования).

Определение. *Математическим ожиданием случайного вектора* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ называется вектор

$$\mathsf{M}\xi = (\mathsf{M}\xi_1, \mathsf{M}\xi_2, \ldots, \mathsf{M}\xi_n)'.$$

Определение. *Математическим ожиданием случайной матрицы* $V^{(n\times m)} = [v_{kj}],\ k = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m,$ называется матрица

$$\mathsf{M}V = [\mathsf{M}v_{kj}],\ k = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m.$$

Свойство линейности математического ожидания. *Для случайных матриц* $V^{(n\times m)}$ *и* $W^{(n\times m)}$

$$\mathsf{M}(V + W) = \mathsf{M}V + \mathsf{M}W.$$

Для случайной матрицы $V^{(n\times m)}$ *и постоянных матриц* $A^{(p\times n)}$, $B^{(m\times q)}$

$$\mathsf{M}(AVB) = A(\mathsf{M}V)B.$$

Равенства очевидным образом получаются из свойства линейности математического ожидания для скалярных случайных величин.

Ковариационная матрица случайного вектора. *Ковариационной матрицей случайного вектора* $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ называется матрица

$$\Gamma = \mathsf{M}(\xi - \mathsf{M}\xi)(\xi - \mathsf{M}\xi)',$$

другими словами, матрица $\Gamma = [\Gamma_{kj}]$, $k, j = 1, 2, \ldots, n$, элементами Γ_{kj} которой являются

$$\Gamma_{kj} = \mathsf{M}(\xi_k - \mathsf{M}\xi_k)(\xi_j - \mathsf{M}\xi_j),\ k, j = 1, 2, \ldots, n.$$

Заметим, что диагональные элементы $\Gamma_{jj}, j = 1, 2, \ldots, n$, ковариационной матрицы $\Gamma = [\Gamma_{kj}], k, j = 1, 2, \ldots, n$, равны соответственно $\mathsf{D}\xi_j$:

$$\Gamma_{jj} = \mathsf{M}(\xi_j - \mathsf{M}\xi_j)(\xi_j - \mathsf{M}\xi_j) = \mathsf{M}(\xi_j - \mathsf{M}\xi_j)^2 = \mathsf{D}\xi_j.$$

Лемма 21.1.1. *Ковариационная матрица случайного вектора*

$$\eta = A\xi,$$

полученного из вектора ξ линейным преобразованием A, равна

$$\Gamma_\eta = A\Gamma_\xi A'.$$

Доказательство.

$$\Gamma_\eta = \mathsf{M}(\eta - \mathsf{M}\eta)(\eta - \mathsf{M}\eta)' = \mathsf{M}(A\xi - \mathsf{M}A\xi)(A\xi - \mathsf{M}A\xi)' =$$

$$= \mathsf{M}A(\xi - \mathsf{M}\xi)(\xi - \mathsf{M}\xi)'A' = A\mathsf{M}(\xi - \mathsf{M}\xi)(\xi - \mathsf{M}\xi)'A' = A\Gamma_\xi A'.$$

Замечание. *Ковариационные матрицы векторов ξ и $\xi + a$ совпадают.*

Теорема. *Ковариационная матрица является симметричной и положительно полуопределенной.*

Доказательство. Пусть Γ — ковариационная матрица случайного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$. Ее элементами являются

$$\Gamma_{kj} = \mathsf{M}(\xi_k - \mathsf{M}\xi_k)(\xi_j - \mathsf{M}\xi_j),\ k, j = 1, 2, \ldots, n.$$

Очевидно, что

$$\Gamma_{kj} = \Gamma_{jk},\ k, j = 1, 2, \ldots, n,$$

т. е. ковариационная матрица является симметричной.

Напомним, что матрица $\Gamma = [\Gamma_{kj}]$ положительно полуопределена, если для любого вектора $u = (u_1, u_2, \ldots, u_n)'$

$$(\Gamma u, u) \geq 0,$$

или, что то же

$$\sum_{k,j=1}^{n} \Gamma_{kj} u_k u_j \geq 0.$$

Для ковариационной матрицы $\Gamma = [\Gamma_{kj}]$ последнее неравенство имеет место поскольку

$$\Gamma_{kj} = \mathsf{M}(\xi_k - \mathsf{M}\xi_k)(\xi_j - \mathsf{M}\xi_j), \ k, j = 1, 2, \ldots, n,$$

$$\sum_{k,j=1}^{n} \Gamma_{kj} u_k u_j = \sum_{k,j=1}^{n} \mathsf{M}(\xi_k - \mathsf{M}\xi_k)(\xi_j - \mathsf{M}\xi_j) u_k u_j =$$

$$= \mathsf{M} \sum_{k,j=1}^{n} ((\xi_k - \mathsf{M}\xi_k) u_k)((\xi_j - \mathsf{M}\xi_j) u_j) =$$

$$= \mathsf{M} \left(\sum_{k=1}^{n} (\xi_k - \mathsf{M}\xi_k) u_k \right)^2 \geq 0.$$

Гауссовские векторы. Случайный вектор $\xi = (\xi_1, \xi_2, \ldots$ $\ldots, \xi_n)'$ называется *стандартным гауссовским вектором* (*стандартным нормальным вектором*), если его компоненты независимы и каждая имеет распределение $\mathsf{N}_{0;1}$.

Стандартный гауссовский вектор еще называют *вектором, распределенным* $\mathsf{N}_{0;I}$.

Ковариационная матрица стандартного гауссовского вектора равна единичной матрице I (в этом убеждаемся непосредственной проверкой).

Определение. Вектор η называется *гауссовским*, если он получен из стандартного гауссовского вектора ξ преобразованием

$$\eta = A\xi + a.$$

Гауссовский вектор

$$\eta = A\xi + a$$

имеет среднее $\mathsf{M}\eta = a$ *и ковариационную матрицу*

$$\Gamma_\eta = AA'.$$

В частности, если $A = \Lambda = [\delta_{k,j}\lambda_k],\ k, j = 1, 2, \ldots, n,$ — диагональная матрица, то гауссовский вектор

$$\eta = \Lambda\xi$$

имеет среднее $\mathsf{M}\eta = 0$ и ковариационную матрицу

$$\Gamma_\eta = [\delta_{k,j}\lambda_k^2],\ k, j = 1, 2, \ldots, n.$$

Теорема 21.1.1. *Линейным преобразованием гауссовский вектор преобразуется в гауссовский.*

Доказательство. Пусть η — гауссовский вектор, т. е. представим в виде $\eta = A\xi + a$, где ξ — стандартный гауссовский вектор. Вектор ζ получен из η линейным преобразованием B, т. е. $\zeta = B\eta$. Тогда

$$\zeta = B\eta = B(A\xi + a) = (BA)\xi + Ba.$$

Такое представление ζ по определению обозначает, что ζ — гауссовский вектор.

Характеристическая функция случайного вектора. Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ — случайный вектор и F — его распределение. *Характеристической функцией случайного вектора* ξ (*характеристической функцией распределения* F) называется комплекснозначная функция $\varphi(t)$, определяемая для действительных $t \in \mathbb{R}^n$ равенством

$$\varphi(t) = \mathsf{M}\exp\{i(t,\xi)\} = \int\limits_{\mathbb{R}^n} \exp\{i(t,x)\}\,\mathsf{F}(dx),$$

где (t,ξ) — скалярное произведение векторов $t = (t_1, t_2, \ldots, t_n)$ и $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$:

$$(t,\xi) = t_1\xi_1 + t_2\xi_2 + \cdots + t_n\xi_n.$$

В координатной форме,

$$\varphi(t) = \varphi(t_1, t_2, \ldots, t_n) = \mathsf{M}\exp\{i(t_1\xi_1 + t_2\xi_2 + \cdots + t_n\xi_n)\} =$$

$$= \int\limits_{\mathbb{R}^n} \exp\{i(t_1x_1 + t_2x_2 + \dots + t_nx_n)\} \, \mathsf{F}(d(x_1, x_2, \dots, x_n)).$$

Характеристическая функция

$$\varphi(t) = \mathsf{M} \exp\{i(t, \xi)\} = \int\limits_{\mathbb{R}^n} \exp\{i(t, x)\} \, \mathsf{F}(dx)$$

любого случайного вектора (любого вероятностного распределения F) определена, поскольку $|e^{i(t,\xi)}| \le 1$, а математическое ожидание ограниченной случайной величины существует и конечно.

Как и в одномерном случае, для случайных векторов (вероятностных распределений на $\mathbb{R}^n$) имеет место теорема единственности:

Теорема 21.1.2 (единственности). *Различным вероятностным распределениям на $\mathbb{R}^n$ соответствуют различные характеристические функции.*

Характеристическая функция гауссовского вектора. Сначала найдем характеристическую функцию стандартного гауссовского вектора.

Теорема 21.1.3. *Характеристическая функция стандартного гауссовского вектора равна*

$$\varphi(t) = \exp\left\{-\frac{1}{2}(t, t)\right\}, \ t \in \mathbb{R}^n.$$

Д о к а з а т е л ь с т в о. Компоненты стандартного гауссовского вектора $\xi = (\xi_1, \xi_2, \dots, \xi_n)'$ независимы, поэтому его характеристическая функция

$$\varphi(t) = \mathsf{M} \exp\{i(t, \xi)\} = \mathsf{M} \exp\left\{i \sum_{j=1}^{n} t_j \xi_j\right\} =$$

$$= \mathsf{M} \prod_{j=1}^{n} \exp\{it_j\xi_j\} = \prod_{j=1}^{n} \mathsf{M} \exp\{it_j\xi_j\}. \qquad (21.1.1)$$

А поскольку каждая компонента ξ_j, $j = 1, 2, \dots, n$, распределена $\mathsf{N}_{0;1}$, то

$$\varphi_{\xi_j}(t_j) = \mathsf{M} \exp\{it_j\xi_j\} = \exp\left\{-\frac{1}{2}t_j^2\right\}, \ j = 1, 2, \dots, n.$$

Из последних равенств и равенства (21.1.1) имеем

$$\varphi(t)=\prod_{j=1}^{n}\exp\left\{-\frac{1}{2}t_j^2\right\}=\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}t_j^2\right\}=\exp\left\{-\frac{1}{2}(t,t)\right\}.$$

Теорема 21.1.4. *Характеристическая функция гауссовского вектора с математическим ожиданием a и ковариационной матрицей* Γ

$$\varphi(t)=\exp\left\{i(t,a)-\frac{1}{2}(\Gamma t,t)\right\}.$$

Доказательство. Гауссовский вектор η получается из стандартного гауссовского вектора ξ преобразованием

$$\eta=A\xi+a,$$

при этом

$$\Gamma=\Gamma_\eta=A\Gamma_\xi A'=AIA'=AA'.$$

Характеристическая функция вектора η

$$\varphi(t)=\mathsf{M}\exp\{i(t,\eta)\}=\mathsf{M}\exp\{i(t,A\xi+a)\}=$$

$$=\mathsf{M}\exp\{i(t,a)+i(t,A\xi)\}=\exp\{i(t,a)\}\,\mathsf{M}\exp\{i(t,A\xi)\}=$$

$$=\exp\{i(t,a)\}\,\mathsf{M}\exp\left\{i(A't,\xi)\right\},$$

где $\mathsf{M}\exp\{i(A't,\xi)\}$ — характеристическая функция стандартного гауссовского вектора ξ, вычисленная в точке $s=A't$. И поскольку

$$\mathsf{M}\exp\{i(s,\xi)\}=\exp\{-(s,s)/2\})$$

(см. теорему 21.1.3), то

$$\mathsf{M}\exp\left\{i(A't,\xi)\right\}=\mathsf{M}\exp\{i(s,\xi)\}=\exp\left\{-\frac{1}{2}(s,s)\right\}=$$

$$=\exp\left\{-\frac{1}{2}(A't,A't)\right\}=\exp\left\{-\frac{1}{2}(AA't,t)\right\}=\exp\left\{-\frac{1}{2}(\Gamma t,t)\right\}.$$

Так что

$$\varphi(t)=\exp\{i(t,a)\}\,\mathsf{M}\exp\left\{i(A't,\xi)\right\}=$$

$$= \exp\{i(t,a)\}\exp\left\{-\frac{1}{2}(\Gamma t,t)\right\} = \exp\left\{i(t,a) - \frac{1}{2}(\Gamma t,t)\right\}.$$

Тем самым теорема доказана.

В координатной форме характеристическая функция $\varphi(t)$ гауссовского вектора записывается так:

$$\varphi(t) = \varphi(t_1, t_2, \ldots, t_n) = \exp\left\{i\sum_{j=1}^{n} t_j\, a_j - \frac{1}{2}\sum_{k,j=1}^{n}\Gamma_{kj}\, t_k\, t_j\right\}.$$

Следствие. *Распределение гауссовского вектора однозначно определяется его математическим ожиданием и ковариационной матрицей.*

Доказательство. Математическое ожидание a и ковариационная матрица Γ гауссовского вектора η однозначно определяют его характеристическую функцию

$$\varphi(t) = \mathsf{M}\exp\{i(t,\eta)\} = \exp\left\{i(t,a) - \frac{1}{2}(\Gamma t,t)\right\},$$

а вместе с ней, в силу теоремы единственности, и распределение.

Гауссовский вектор со средним a и ковариационной матрицей Γ еще называют *случайным вектором, распределенным* $\mathsf{N}_{a;\Gamma}$, в частности, стандартный гауссовский вектор называют вектором, распределенным $\mathsf{N}_{0;I}$.

Напомним, что преобразование C из $\mathbb{R}^n$ в $\mathbb{R}^n$ называется ортогональным, если $C'C = I$, при этом

$$(Cx, Cx) = (C'Cx, x) = (x, x).$$

Ортогональное преобразование сохраняет скалярное произведение, а вместе с ним и расстояние между точками в $\mathbb{R}^n$, поскольку $\|x\| = \sqrt{(x,x)}$.

Теорема 21.1.5 (об ортогональном преобразовании гауссовского вектора). *Ортогональным преобразованием стандартный гауссовский вектор преобразуется в стандартный гауссовский вектор.*

Доказательство. Пусть вектор

$$\zeta = C\xi$$

получен из стандартного гауссовского вектора ξ ортогональным преобразованием C. Вектор ζ — гауссовский, причем

$$\mathsf{M}\zeta = \mathsf{M}C\xi = C\mathsf{M}\xi = C0 = 0,$$

$$\Gamma_\zeta = C\Gamma_\xi C' = CIC' = CC' = I,$$

поэтому

$$\varphi_\zeta(t) = \exp\left\{-\frac{1}{2}(t,t)\right\}.$$

Так что характеристическая функция вектора ζ совпадает с характеристической функцией стандартного гауссовского вектора. Поэтому в силу теоремы единственности ζ — стандартный гауссовский вектор.

Некоррелированность и независимость компонент гауссовского вектора. Случайные величины ξ_k и ξ_j, $k \neq j$, называются *некоррелированными,* если

$$\mathsf{M}(\xi_k - \mathsf{M}\xi_k)(\xi_j - \mathsf{M}\xi_j) = 0.$$

Из независимости случайных величин следует их некоррелированность, обратное утверждение, вообще говоря, не имеет места. Но *у гауссовского вектора некоррелированность его компонент равносильна независимости.*

Теорема 21.1.6. *Для того чтобы компоненты гауссовского вектора были независимы, достаточно, чтобы они были некоррелированы.*

Доказательство. Сначала убедимся, что характеристическая функция гауссовского вектора с некоррелированными компонентами равна произведению характеристических функций его компонент.

Характеристическую функцию $\varphi_\xi(t_1, t_2, \ldots, t_n)$ гауссовского вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ со средним $\mathsf{M}\xi = (a_1, a_2, \ldots, a_n)'$ и некоррелированными компонентами: $\Gamma_{kj} = 0, k \neq j$, можно записать в виде:

$$\varphi_\xi(t_1, t_2, \ldots, t_n) = \mathsf{M}\exp\left\{i\sum_{j=1}^n t_j\xi_j\right\} =$$

$$= \exp\left\{i\sum_{j=1}^n t_j a_j - \frac{1}{2}\sum_{k,j=1}^n \Gamma_{kj}\, t_k\, t_j\right\} =$$

$$= \exp\left\{i\sum_{j=1}^n t_j a_j - \frac{1}{2}\sum_{j=1}^n \Gamma_{jj}\, t_j^2\right\}.$$

Так что для $\varphi_\xi(t_1, t_2, \ldots, t_n)$ имеем представление

$$\varphi_\xi(t_1, t_2, \ldots, t_n) = \mathsf{M} \exp\left\{ i \sum_{j=1}^{n} t_j \xi_j \right\} = \prod_{j=1}^{n} \exp\left\{ i\, t_j a_j - \frac{1}{2} \Gamma_{jj}\, t_j^2 \right\}. \tag{21.1.1}$$

В точке $(0, \ldots, 0, t_k, 0, \ldots, 0)$ характеристическая функция вектора ξ:

$$\varphi_\xi(0, \ldots, 0, t_k, 0, \ldots, 0) = \mathsf{M} \exp\{i\, t_k\, \xi_k\} = \exp\left\{ i\, t_k\, a_k - \frac{1}{2} \Gamma_{kk}\, t_k^2 \right\}$$

Из последнего равенства следует, что у гауссовского вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_k, \ldots, \xi_n)$ с некоррелированными компонентами k-я компонента ξ_k $(k = 1, 2, \ldots, n)$ распределена нормально, с параметрами $(a_k; \Gamma_{kk})$, а из равенства (21.1.1) следует, что характеристическая функция гауссовского вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_k, \ldots$ $\ldots, \xi_n)$ равна произведению характеристических функций его компонент:

$$\varphi_\xi(t_1, t_2, \ldots, t_n) = \varphi_{\xi_1}(t_1) \varphi_{\xi_2}(t_2) \ldots \varphi_{\xi_n}(t_n). \tag{21.1.2}$$

Переписав равенство (21.1.2) в интегральной форме и воспользовавшись теоремой Фубини, получим:

$$\int\limits_{\mathbb{R}^n} e^{i(t_1 x_1 + t_2 x_2 + \cdots + t_n x_n)} \mathsf{P}_\xi\left(d(x_1, x_2, \ldots, x_n)\right) =$$

$$= \int\limits_{\mathbb{R}^1} e^{i t_1 x_1} \mathsf{P}_{\xi_1}(dx_1) \int\limits_{\mathbb{R}^1} e^{i t_2 x_2} \mathsf{P}_{\xi_2}(dx_2) \ldots \int\limits_{\mathbb{R}^1} e^{i t_n x_n} \mathsf{P}_{\xi_n}(dx_n) =$$

$$= \int\limits_{\mathbb{R}^n} e^{i(t_1 x_1 + t_2 x_2 + \ldots + t_n x_n)} \mathsf{P}_{\xi_1} \times \mathsf{P}_{\xi_2} \times \ldots \times \mathsf{P}_{\xi_n}\left(d(x_1, \ldots, x_n)\right).$$

Так что характеристическая функция распределения P_ξ гауссовского вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ с некоррелированными компонентами совпадает с характеристической функцией распределения $\mathsf{P}_{\xi_1} \times \mathsf{P}_{\xi_2} \times \ldots \times \mathsf{P}_{\xi_n}$. А поэтому в силу теоремы единственности совпадают и сами распределения:

$$\mathsf{P}_\xi = \mathsf{P}_{\xi_1} \times \mathsf{P}_{\xi_2} \times \ldots \times \mathsf{P}_{\xi_n}. \tag{21.1.3}$$

Равенство же (21.1.3) обозначает, что случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы.

Так что из некоррелированности компонент $\xi_1, \xi_2, \dots, \xi_n$ гауссовского вектора следует их независимость.

Следствие. *Если компоненты гауссовского вектора* $\xi = (\xi_1, \xi_2, \dots, \xi_k, \dots, \xi_n)'$ *— со средним* $a = (a_1, a_2, \dots, a, \dots, a_n)'$ *и ковариационной матрицей* $\Gamma = [\Gamma_{kj}]$*, некоррелированы, то они независимы и распределены нормально: а именно,* ξ_k *распределена нормально со средним* a_k *и дисперсией* Γ_{kk}*,* $k = 1, 2, \dots, n$*.*

21.2 Распределения, связанные с нормальным распределением

Определение. χ^2*-распределением с* n *степенями свободы* будем называть распределение случайной величины

$$\chi_n^2 = \sum_{i=1}^{n} \xi_i^2, \tag{21.2.1}$$

где $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая с распределением $\mathsf{N}_{0;1}$.

Определение. *Распределением Стьюдента* или t*-распределением с* n *степенями свободы* будем называть распределение случайной величины

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}}, \tag{21.2.2}$$

где ξ и χ_n^2 — независимые случайные величины, ξ распределена $\mathsf{N}_{0;1}$, а χ_n^2 имеет χ^2-распределение с n степенями свободы.

Определение. *Распределением Фишера* или F*-распределением с* $(n; m)$ *степенями свободы* будем называть распределение случайной величины

$$\mathsf{F}_{n;m} = \frac{\frac{1}{n}\chi_n^2}{\frac{1}{m}\chi_m^2}, \tag{21.2.3}$$

где χ_n^2 и χ_m^2 — независимые случайные величины, имеющие χ^2-распределение с n и m степенями свободы соответственно.

Теорема 21.2.1. *Случайная величина* χ_n^2 *при* $n \to \infty$ *асимптотически нормальна со средним* n *и дисперсией* $2n$, *т. е.*

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{\chi_n^2 - n}{\sqrt{2n}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt$$

для каждого $x \in \mathbb{R}^1$.

Доказательство. В силу центральной предельной теоремы (см. теорему 15.1.1) сумма $\chi_n^2 = \sum\limits_{i=1}^{n} \xi_i^2$ независимых одинаково распределенных случайных величин ξ_i^2 с конечными дисперсиями (ξ_i распределены $\mathsf{N}_{0;1}$) распределена асимптотически нормально со средним $\mathsf{M}\chi_n^2$ и дисперсией $\mathsf{D}\chi_n^2$, т. е. при $n \to \infty$

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{\chi_n^2 - \mathsf{M}\chi_n^2}{\sqrt{\mathsf{D}\chi_n^2}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt,$$

$x \in \mathbb{R}^1$. Осталось заметить, что поскольку для $\mathsf{N}_{0;1}$-распределенной случайной величины ξ_i значения $\mathsf{M}\xi_i^{2k} = (2k-1)!!$ (см. теорему 10.2.2), то

$$\mathsf{M}\chi_n^2 = \mathsf{M}\sum_{i=1}^{n} \xi_i^2 = \sum_{i=1}^{n} \mathsf{M}\xi_i^2 = n,$$

$$\mathsf{D}\chi_n^2 = \mathsf{D}\sum_{i=1}^{n} \xi_i^2 = \sum_{i=1}^{n} \mathsf{D}\xi_i^2 = \sum_{i=1}^{n} \left(\mathsf{M}\xi_i^4 - (\mathsf{M}\xi_i^2)^2\right) = 2n.$$

Теорема 21.2.2. *Случайная величина* t_n *при* $n \to \infty$ *асимптотически нормальна с параметрами* $(0;1)$, *т. е.*

$$\lim_{n\to\infty} \mathsf{P}\{t_n < x\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt$$

для каждого $x \in \mathbb{R}^1$.

Доказательство. По определению

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}},$$

где случайная величина ξ имеет распределение $\mathsf{N}_{0;1}$, а $\chi_n^2 = \sum_{i=1}^{n} \xi_i^2$ имеет распределение χ^2 с n степенями свободы. В силу закона больших чисел случайная величина

$$\frac{1}{n}\chi_n^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2,$$

сходится по вероятности к $\mathsf{M}\xi_1^2 = 1$, и следовательно,

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}}$$

сходится по вероятности к ξ. Из сходимости по вероятности последовательности случайных величин t_n к случайной величине ξ (распределенной $\mathsf{N}_{0;1}$) следует сходимость их функций распределения, поэтому

$$\mathsf{P}\left\{\frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}} < x\right\} \to \mathsf{P}\{\xi < x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$$

для всех $x \in \mathbb{R}^1$.

Замечание. Распределения χ^2, t, F абсолютно непрерывны (см., например, Крамер Г., Математические методы статистики. — 2-е изд., перераб.— М.: Мир, 1975.— 648 с). Графики плотностей t-распределения, χ^2-распределения, F-распределения изображены на рис. 21.2.1 – 21.2.3.

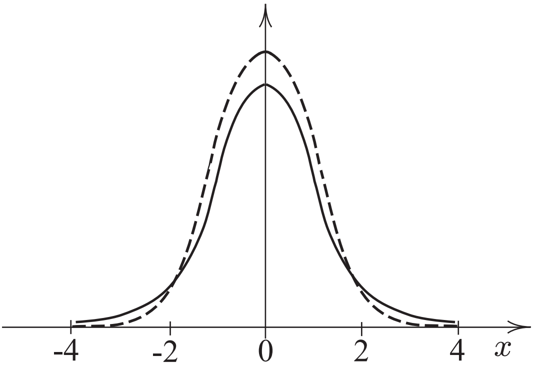

Рис. 21.2.1: Графики плотностей распределения Стьюдента с 2 степенями свободы (сплошная линия) и нормального распределения с параметрами $(0; 1)$ (пунктирная линия)

Распределение Стьюдента при больших значениях числа n степеней свободы близко к $\mathsf{N}_{0;1}$-распределению, для небольших значений n заметно отличается от нормального распределения. Вероятности больших отклонений от среднего значения больше для t-распределения, чем для нормального распределения.

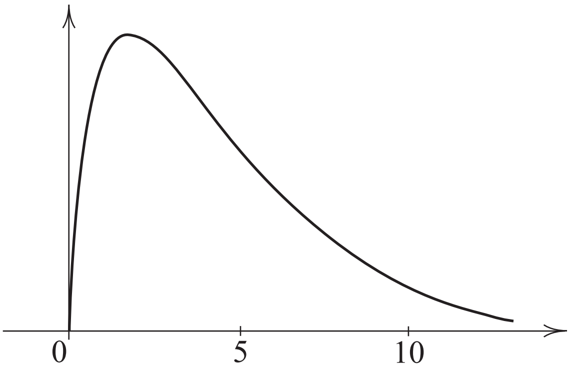

Рис. 21.2.2: График плотности χ^2-распределения с 4 степенями свободы

График плотности распределения χ^2 с числом степеней свободы большим 2 имеет качественно такой же вид, как и график плотности χ_4^2-распределения (см. рис. 21.2.2).

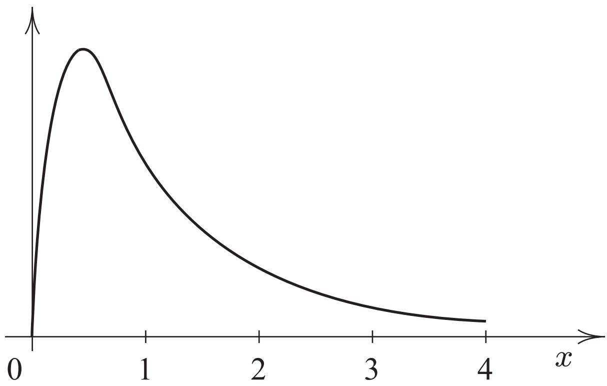

Рис. 21.2.3: График плотности F-распределения с $n = 4$, $m = 8$ степенями свободы

Верхний α-предел распределения. Пусть F — непрерывное вероятностное распределение, $0 < \alpha < 1$. *Верхним α-пределом* (*верхним* 100α*-процентным пределом,* 100α*-процентной точкой,* 100α*-процентным критическим значением*) распределения F будем называть число z_α, являющееся решением уравнения

$$\mathsf{F}\left([z_\alpha, +\infty)\right) = \alpha$$

или, что то же, уравнения

$$F(z_\alpha) = 1 - \alpha,$$

где $F(z_\alpha) = \mathsf{F}\left((-\infty, z_\alpha)\right).$

Интуитивно, верхний α-предел z_α распределения F — это точка, "отсекающая правый хвост" распределения F, на который приходится "масса" α (см. рис. 21.2.4).

Верхний α-предел z_α распределения F связан с его $(1-\alpha)$-квантилью $x_{1-\alpha}$ (точкой, для которой $F(x_{1-\alpha}) = 1 - \alpha$) равенством

$$z_\alpha = x_{1-\alpha},$$

что непосредственно следует из равенства $F(z_\alpha) = 1 - \alpha$.

Пусть ξ — случайная величина с непрерывным распределением F. *Верхним α-пределом случайной величины* ξ будем называть верхний α-предел z_α ее распределения F.

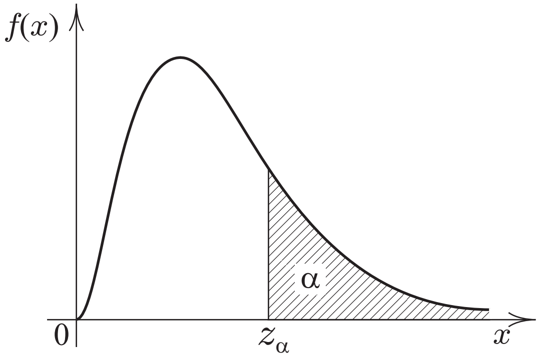

Рис. 21.2.4: К определению верхнего α-предела z_α распределения F

Для распределений Стьюдента, χ^2, Фишера имеем:

Определение. Пусть χ^2_n — случайная величина, имеющая распределение χ^2 с n степенями свободы, $0 < \alpha < 1$. Число $\chi^2_{\alpha;n}$, удовлетворяющее соотношению

$$\mathsf{P}\left\{\chi^2_n \geq \chi^2_{\alpha;n}\right\} = \alpha,$$

будем называть *верхним α-пределом* (*верхней* 100α*-процентной точкой*) *случайной величины χ^2_n, распределения χ^2 с n степенями свободы.*

Пусть t_n — случайная величина, имеющая распределение Стьюдента с n степенями свободы, $0 < \alpha < 1$. Число $t_{\alpha;n}$, удовлетворяющее соотношению

$$\mathsf{P}\left\{t_n \geq t_{\alpha;n}\right\} = \alpha,$$

будем называть *верхним α-пределом* (*верхней* 100α*-процентной точкой*) *случайной величины t_n, распределения Стьюдента с n степенями свободы.*

Пусть $F_{n;m}$ — случайная величина, имеющая F-распределение с n, m степенями свободы, $0 < \alpha < 1$. Число $F_{\alpha;n;m}$, удовлетворяющее соотношению

$$\mathsf{P}\left\{F_{n;m} \geq F_{\alpha;n;m}\right\} = \alpha,$$

будем называть *верхним α-пределом* (100α*-процентной точкой*) *случайной величины $F_{n;m}$, распределения Фишера с n, m степенями свободы.*

По заданным α и n верхние α-пределы (100α-процентные точки) $\chi^2_{\alpha;n}$, и $t_{\alpha;n}$ соответственно распределений χ^2 и Стьюдента табулированы в табл. 27.2.1 и 27.3.1, а по α, n, m верхние α-пределы $F_{\alpha;n,m}$ распределения Фишера табулированы в табл. 27.4.1, 27.4.2, точки $t_{\alpha;n}$, $\chi^2_{\alpha;n}$, $F_{\alpha;n,m}$ выбираются так, чтобы масса, приходящаяся на заштрихованные “хвосты” распределений, была равна α (см. рис. 27.2.1, 27.3.1, 27.4.1).

Верхний α-предел и p-value. Верхним α-пределом, $\alpha \in (0,1)$, абсолютно непрерывного распределения F с плотностью $f(t)$ ($f(t) > 0$) будем называть функцию

$$z_\alpha : \alpha \to z_\alpha$$

на $(0,1)$ со значениями в $\mathbb{R}^1$, значение z_α которой для каждого α получаются как решение уравнения

$$\int\limits_{z_\alpha}^{\infty} f(t)dt = \alpha,\ \alpha \in (0,1).$$

Определение. p-Value абсолютно непрерывного распределения F с плотностью $f(t)$ будем называть функцию

$$p_z : z \to \int\limits_{z}^{\infty} f(t)dt = p_z,$$

на $\mathbb{R}^1$ со значениями в $(0,1)$.

Если z_α — верхний α-предел распределения F — точка на $\mathbb{R}^1$, которая для данного α отсекает правый хвост распределения F с массой α, то p-value p_z — это масса распределения F, приходящаяся на данный хвост $[z, +\infty)$.

21.3 Нормальная выборка

Теорема 21.3.1 (о распределении вектора $(\overline{\xi}, s^2)$). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения $\mathsf{N}_{a;\sigma^2}$. Оценки*

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i \ \text{ и } \ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2$$

являются независимыми случайными величинами, при этом $\dfrac{n-1}{\sigma^2}s^2$ *имеет распределение* χ^2 *с* $n-1$ *степенями свободы, а* $\overline{\xi}$ *распределена* $\mathsf{N}_{a;\sigma^2/n}$.

Доказательство. Положим

$$\eta_i = \frac{\xi_i - a}{\sigma},\ i = 1, 2, \ldots, n.$$

Случайные величины $\eta_1, \eta_2, \ldots, \eta_n$ распределены $\mathsf{N}_{0;1}$, так как ξ_i, $i = 1, 2, \ldots, n$, распределены $\mathsf{N}_{a;\sigma^2}$ и независимы как функции от независимых случайных величин. Выразим $\overline{\xi}$ и s^2 через компоненты стандартного гауссовского вектора $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$:

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i = \frac{\sigma}{n}\sum_{i=1}^{n}\frac{\xi_i - a}{\sigma} + a = \sigma\frac{1}{n}\sum_{i=1}^{n}\eta_i + a = \sigma\overline{\eta} + a,$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i - (\sigma\overline{\eta} + a)\right)^2 =$$

$$= \frac{\sigma^2}{n-1}\sum_{i=1}^{n}\left(\frac{\xi_i - a}{\sigma} - \overline{\eta}\right)^2 = \frac{\sigma^2}{n-1}\sum_{i=1}^{n}(\eta_i - \overline{\eta})^2 =$$

$$= \frac{\sigma^2}{n-1}\left(\sum_{i=1}^{n}\eta_i^2 - n\overline{\eta}^2\right).$$

Далее, пусть C — матрица ортогонального преобразования из $\mathbb{R}^n$ в $\mathbb{R}^n$, элементы первой строки которой

$$c_{11} = c_{12} = \ldots = c_{1n} = 1/\sqrt{n}. \qquad (21.3.1)$$

Определим вектор

$$\zeta = C\eta.$$

Вектор ζ является стандартным гауссовским, поскольку получен ортогональным преобразованием из стандартного гауссовского вектора η. Выразим s^2 и $\overline{\xi}$ через вектор ζ. Из (21.3.1) следует, что ζ_1 — первая компонента вектора ζ — равна

$$\zeta_1 = \frac{1}{\sqrt{n}}\eta_1 + \frac{1}{\sqrt{n}}\eta_2 + \cdots + \frac{1}{\sqrt{n}}\eta_n = \sqrt{n}\overline{\eta}.$$

А из ортогональности преобразования C имеем:

$$\sum_{i=1}^{n} \eta_i^2 = (\eta, \eta) = (C\eta, C\eta) = (\zeta, \zeta) = \sum_{i=1}^{n} \zeta_i^2.$$

Поэтому

$$s^2 = \frac{\sigma^2}{n-1}\left(\sum_{i=1}^{n} \eta_i^2 - n\overline{\eta}^2\right) = \frac{\sigma^2}{n-1}\left(\sum_{i=1}^{n} \zeta_i^2 - \zeta_1^2\right) = \frac{\sigma^2}{n-1}\sum_{i=2}^{n} \zeta_i^2,$$

$$\overline{\xi} = \sigma\overline{\eta} + a = \sigma\frac{\zeta_1}{\sqrt{n}} + a.$$

Из этих представлений s^2 и $\overline{\xi}$ через $\zeta_1, \zeta_2, \ldots, \zeta_n$ следует, что s^2 и $\overline{\xi}$ — независимые случайные величины, как функции от независимых случайных величин ζ_1 и $(\zeta_2, \zeta_3, \ldots, \zeta_n)$.

Найдем распределения s^2 и $\overline{\xi}$. Так как

$$\frac{n-1}{\sigma^2}s^2 = \sum_{i=2}^{n} \zeta_i^2,$$

а ζ_i — независимые случайные величины, каждая с распределением $\mathsf{N}_{0;1}$, то, по определению, случайная величина $\sum\limits_{i=2}^{n} \zeta_i^2$, а вместе с ней и $\dfrac{n-1}{\sigma^2}s^2$, имеет распределение χ^2 с $n-1$ степенями свободы.

Случайная величина $\overline{\xi} = \dfrac{1}{n}\sum\limits_{i=1}^{n} \xi_i$ распределена нормально, как сумма независимых нормально распределенных случайных величин, при этом $\mathsf{M}\overline{\xi} = a$, $\mathsf{D}\overline{\xi} = \sigma^2/n$. Так что $\overline{\xi}$ распределена $\mathsf{N}_{a;\sigma^2/n}$.

Следствие 1. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения $\mathsf{N}_{a;\sigma^2}$. Случайная величина*

$$(\overline{\xi} - a) \Big/ \frac{s}{\sqrt{n}}$$

имеет распределение Стьюдента с $n-1$ степенями свободы.

Доказательство. Перепишем дробь $(\overline{\xi}-a) \Big/ \frac{s}{\sqrt{n}}$, разделив числитель и знаменатель на $\sigma/\sqrt{n}$:

$$(\overline{\xi}-a) \Big/ \frac{s}{\sqrt{n}} = \frac{(\overline{\xi}-a)/\frac{\sigma}{\sqrt{n}}}{\sqrt{s^2/\sigma^2}}.$$

В силу теоремы 21.3.1 о распределении вектора $(\overline{\xi}, s^2)$ числитель и знаменатель последней дроби являются независимыми случайными величинами (как функции от независимых случайных величин $\overline{\xi}$ и s^2), причем $(\overline{\xi}-a) \Big/ \frac{\sigma}{\sqrt{n}}$ распределена $\mathsf{N}_{0;1}$, а s^2/σ^2 распределена так же, как случайная величина $\frac{1}{n-1}\chi^2_{n-1}$. Поэтому по определению случайная величина $(\overline{\xi}-a) \Big/ \frac{s}{\sqrt{n}}$ имеет распределение Стьюдента с $n-1$ степенями свободы.

Следствие 2. *Несмещенной оценкой дисперсии*

$$\sigma^2_{\overline{\xi}} = \mathsf{D}\overline{\xi} = \sigma^2/n$$

выборочного среднего $\overline{\xi}$ *является*

$$\hat{\sigma}^2_{\overline{\xi}} = s^2/n.$$

Поскольку s^2 — несмещенная оценка σ^2, то

$$\mathsf{M}\hat{\sigma}^2_{\overline{\xi}} = \mathsf{M}\frac{1}{n}s^2 = \frac{1}{n}\sigma^2 = \sigma^2_{\overline{\xi}}$$

Доверительные интервалы для параметров нормального распределения. Помимо точечного оценивания параметров часто используется, так называемое, интервальное оценивание параметров.

Определение. Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборочный вектор с распределением, зависящим от параметра θ, и пусть $\gamma_1(x_1, x_2, \ldots, x_n)$ и $\gamma_2(x_1, x_2, \ldots, x_n)$ — борелевские функции на $\mathbb{R}^n$ со значениями в $\mathbb{R}^1$ такие, что

$$\gamma_1(x_1, x_2, \ldots, x_n) \leq \gamma_2(x_1, x_2, \ldots, x_n).$$

Если

$$\mathsf{P}\{\gamma_1(\xi_1, \xi_2, \ldots, \xi_n) \le \theta \le \gamma_2(\xi_1, \xi_2, \ldots, \xi_n)\} = 1 - \alpha,$$

то интервал $(\gamma_1(\xi_1, \xi_2, \ldots, \xi_n), \gamma_2(\xi_1, \xi_2, \ldots, \xi_n))$ называется *доверительным интервалом параметра θ с коэффициентом доверия* (*надежности*) $1 - \alpha$.

Теорема (о доверительном интервале для $\boldsymbol{\sigma^2}$). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения $\mathsf{N}_{a;\sigma^2}$, $\chi^2_{\alpha;n-1}$ — верхний α-предел, а $\chi^2_{1-\alpha;n-1}$ — верхний $(1-\alpha)$-предел χ^2-распределения с $n-1$ степенями свободы $(0 < \alpha < 1/2)$. Тогда*

$$\mathsf{P}\left\{\frac{(n-1)s^2}{\chi^2_{\alpha;n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha;n-1}}\right\} = 1 - 2\alpha. \tag{21.3.2}$$

Доказательство. Ясно, что

$$\mathsf{P}\{\chi^2_{1-\alpha;n-1} \le \chi^2_{n-1} \le \chi^2_{\alpha;n-1}\} = 1 - 2\alpha.$$

Случайная величина $\dfrac{n-1}{\sigma^2}s^2$ имеет распределение χ^2 с $n-1$ степенями свободы, поэтому

$$\mathsf{P}\left\{\chi^2_{1-\alpha;n-1} \le \frac{n-1}{\sigma^2}s^2 \le \chi^2_{\alpha;n-1}\right\} = 1 - 2\alpha$$

или

$$\mathsf{P}\left\{\frac{(n-1)s^2}{\chi^2_{\alpha;n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha;n-1}}\right\} = 1 - 2\alpha.$$

Последнее равенство обозначает, что интервал

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha;n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha;n-1}}\right)$$

является доверительным интервалом для σ^2 с коэффициентом доверия $1 - 2\alpha$.

Теорема (о доверительном интервале для a). *Пусть* $\xi_1, \xi_2, \ldots, \xi_n$ *— выборка из распределения* $\mathsf{N}_{a;\sigma^2}$, $t_{\alpha;n-1}$ *— верхний* α*-предел* t*-распределения с* $n-1$ *степенями свободы* $(0 < \alpha < 1/2)$. *Тогда*

$$\mathsf{P}\left\{\overline{\xi} - \frac{s}{\sqrt{n}}t_{\alpha;n-1} \le a \le \overline{\xi} + \frac{s}{\sqrt{n}}t_{\alpha;n-1}\right\} = 1 - 2\alpha. \qquad (21.3.3)$$

Доказательство. Из определения $t_{\alpha;n-1}$ имеем:

$$\mathsf{P}\{-t_{\alpha;n-1} \le t_{n-1} \le t_{\alpha;n-1}\} = 1 - 2\alpha.$$

Случайная величина $(\overline{\xi} - a)/\dfrac{s}{\sqrt{n}}$ имеет распределение Стьюдента с $n-1$ степенями свободы. Поэтому

$$\mathsf{P}\{-t_{\alpha;n-1} \le (\overline{\xi} - a)/\frac{s}{\sqrt{n}} \le t_{\alpha;n-1}\} = 1 - 2\alpha,$$

или

$$\mathsf{P}\left\{\overline{\xi} - \frac{s}{\sqrt{n}}t_{\alpha;n-1} \le a \le \overline{\xi} + \frac{s}{\sqrt{n}}t_{\alpha;n-1}\right\} = 1 - 2\alpha.$$

Последнее равенство обозначает, что интервал

$$\left(\overline{\xi} - \frac{s}{\sqrt{n}}t_{\alpha;n-1}, \overline{\xi} + \frac{s}{\sqrt{n}}t_{\alpha;n-1}\right)$$

является доверительным интервалом для a с коэффициентом доверия $1 - 2\alpha$.

Заметим, что поскольку $\widehat{\sigma^2_{\overline{\xi}}} = s^2/n$, то доверительный интервал для a можно записать в виде

$$\left(\overline{\xi} - \hat{\sigma}_{\overline{\xi}}t_{\alpha;n-1}, \overline{\xi} + \hat{\sigma}_{\overline{\xi}}t_{\alpha;n-1}\right),$$

где $\hat{\sigma}_{\overline{\xi}} = \sqrt{\widehat{\sigma^2_{\overline{\xi}}}}$.

Пример 21.3.1. *Ниже приведенные данные об измерениях неровностей поверхности одной и той же чистоты обработки при помощи двойного микроскопа:* 0,8; 1,9; 3,0; 3,5; 3,8; 2,5; 1,7; 0,9; 1,0; 2,3; 3,3; 3,4.

Найти доверительный интервал для показаний микроскопа с коэффициентом надежности: 1) 0,90; 2) 0,99.

Р е ш е н и е. Показания микроскопа естественно считать реализацией выборки из нормального распределения. Доверительный интервал для параметра a (математического ожидания) нормального распределения имеет вид (21.3.3). В рассматриваемом примере $n = 12$, $\overline{\xi} = 2{,}34$, $s^2 = 1{,}66$, $t_{0,05;11} = 1{,}796$, $t_{0,005;11} = 3{,}106$. Поэтому доверительным интервалом с коэффициентом надежности 0,90 является интервал $(1{,}68; 3)$, а с коэффициентом надежности 0,99 — интервал $(1{,}19; 3{,}49)$ (фактически найдены реализации доверительных интервалов, хотя мы говорим о доверительных интервалах).

Заметим, что за увеличение коэффициента надежности мы "расплачиваемся" расширением доверительного интервала, и наоборот — чем у́же доверительный интервал, тем меньше коэффициент надежности.

Глава 22

Проверка статистических гипотез

22.1 Основные определения и понятия

Задача проверки статистических гипотез. Многие задачи математической статистики формулируются так. Имеется результат $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))'$ стохастического эксперимента, состоящего в наблюдении случайного вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ со значениями в $\mathbb{R}^n$. Распределение ξ неизвестно. Что можно сказать о распределении P_ξ вектора ξ по его наблюдению $\xi(\omega)$? Например, "Можно ли считать, что распределением P_ξ является данное распределение G?" или, например, "Можно ли считать, что распределением P_ξ является нормальное распределение?"

Общий подход к решению этой задачи такой. Из класса возможных распределений $\mathcal{P}$ вектора ξ выбираем распределение G (или класс распределений $\mathcal{G}$) — кандидатуру на распределение вектора ξ, фактически выдвигаем гипотезу (предположение): распределением вектора ξ является распределение G (или распределение вектора ξ принадлежит классу $\mathcal{G}$). И далее следуем Сократу, когда он хотел опровергнуть оппонента — если выдвинутая гипотеза противоречит результату эксперимента, ее необходимо отклонить. Так и мы — выдвинутую гипотезу о распределении P_ξ вектора ξ будем отклонять, если она противоречит результату $\xi(\omega)$ эксперимента и не будем отклонять в противном случае. Но и отличие от Сократа у нас ситуация сложнее, поскольку мы имеем дело со стохастическим экспериментом и в наши выводы вмешивается случай.

Далее мы будем пользоваться следующими принятыми в теории проверки статистических гипотез определениями и понятиями.

Нулевая и альтернативные гипотезы. Гипотезы относительно распределений случайных величин будем называть *статистическими* гипотезами.

Выбор из класса $\mathcal{P}$ возможных распределений вектора ξ кандидатуры на его распределение, например, распределением P_ξ является G, называют выбором нулевой гипотезы (основной, проверяемой). Нулевую гипотезу обычно обозначают через H_0. Вместе с выбором нулевой гипотезы H_0 относительно распределения P_ξ (распределение $\mathsf{P}_\xi = G$) из класса возможных гипотез о распределении вектора ξ выбирают альтернативную (конкурирующую) к H_0 гипотезу H_1 — распределение P_ξ вектора ξ принадлежит классу $\mathcal{P}_1 \subset \mathcal{P}$. Далее нулевую гипотезу H_0 проверяют против альтернативной гипотезы — если H_0 противоречит результату $\xi(\omega)$ эксперимента, то H_0 отклоняют (в пользу альтернативной гипотезы), в противном случае — не отклоняют.

Альтернативная к H_0 гипотеза — это гипотеза, в пользу которой H_0 отклоняют (если H_0 отклоняют). Нулевая гипотеза всегда проверяется против альтернативной гипотезы.

В качестве класса $\mathcal{P}$ возможных распределений вектора $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ мы, как правило, будем рассматривать параметрический класс:

$$\mathcal{P} = \{\mathsf{P}_\theta : \theta \in \Theta \subset \mathbb{R}^s\}.$$

Проиллюстрируем введенные понятия на примере задачи проверки монеты на симметричность.

Монету, вероятность выпадения герба θ которой неизвестна, независимым образом подбрасывают n раз. Число выпавших гербов оказалось равным $\xi(\omega)$. Можно ли считать, что монета симметрична?

В терминах распределения наблюдаемой случайной величины ξ — числа выпавших гербов — задача о симметричности монеты формулируется так.

Распределение случайной величины ξ принадлежит классу вероятностных распределений

$$\mathcal{P} = \{B_{n,\theta} : B_{n,\theta}(k) = C_n^k \theta^k (1-\theta)^{n-k},\ k = 0, 1, \ldots, n;\ \theta \in (0,1)\}.$$

Относительно распределения P_ξ выдвигается гипотеза H_0: распределением P_ξ случайной величины ξ является биномиальное распределение $B_{n,1/2}$, другими словами,

$$H_0 : \mathsf{P}_\xi(k) = C_n^k (1/2)^k (1-1/2)^{n-k},\ k = 0, 1, \ldots, n.$$

Альтернативной (конкурирующей) к H_0 является гипотеза

$$H_1 : \mathsf{P}_\xi(k) = C_n^k \theta^k (1-\theta)^{n-k},\ k = 0, 1, \ldots, n,\ \theta \in (0, 1),\ \theta \neq 1/2.$$

Необходимо проверить гипотезу H_0 против альтернативы H_1 — отклонить H_0 или не отклонить. Если гипотеза H_0 противоречит результату эксперимента — число выпавших гербов равно $\xi(\omega)$ — мы ее отклоняем (в пользу альтернативы H_1), в противном случае — нет.

Гипотеза относительно распределения, однозначно его определяющая, называется *простой,* в противном случае — *сложной.* Например, гипотеза: случайная величина ξ имеет своим распределением

$$\mathsf{P}_{\theta_0}(k) = C_n^k \theta_0^k (1-\theta_0)^{n-k},\ k = 0, 1, \ldots, n,$$

(θ_0 и n — известны) — простая, а гипотеза: распределением случайной величины ξ является

$$\mathsf{P}_\theta(k) = C_n^k \theta^k (1-\theta)^{n-k},\ k = 0, 1, \ldots, n,\ \theta \in (1/2, 1)$$

— сложная.

Критерий. Пусть $\xi(\omega)$ — результат (исход) эксперимента, состоящего в наблюдении вектора ξ со значениями в $\mathbb{R}^n$. Распределение вектора ξ неизвестно, H_0 — нулевая гипотеза относительно распределения P_ξ, пусть к примеру $H_0 : \mathsf{P}_\xi = \mathsf{G}$, альтернативная гипотеза H_1. Необходимо проверить гипотезу H_0, а именно, вынести заключение: отклонить гипотезу H_0 или не отклонить. Согласно нашему общему подходу мы гипотезу H_0 отклоняем (в пользу альтернативы H_1), если она противоречит результату эксперимента, и не отклоняем в противном случае. Результатов эксперимента, которым гипотеза H_0 противоречит, разумеется, может быть много, обозначим это множество исходов — подмножество выборочного пространства $X \subset \mathbb{R}^n$ — через S. Тогда мы гипотезу H_0 отклоняем, если исход эксперимента $\xi(\omega)$ оказался в множестве S (множестве исходов, которым H_0 противоречит) и не отклоняем в противном случае.

Определение. Борелевское подмножество S выборочного пространства $X \subset \mathbb{R}^n$ такое, что при $\xi \in S$ гипотезу H_0 отклоняем, в противном случае — нет, называется *критерием* (*критическим множеством, критической областью*) для проверки гипотезы H_0.

Далее естественно возникает задача: *как выбрать* (*построить*) *критическое множество S для проверки данной гипотезы H_0?* Чтобы найти подходы к ее решению, сначала рассмотрим так называемые ошибки первого и второго рода, неизбежные при проверке статистических гипотез.

Ошибки первого и второго рода. Пусть S — критическое множество для проверки гипотезы H_0: если исход $\xi(\omega)$ эксперимента попадает в S, то гипотезу H_0 отклоняем, в противном случае — нет. Поскольку заключение об отклонении или неотклонении гипотезы H_0 мы делаем по исходу стохастического эксперимента, при проверке гипотезы H_0 возможна одна из четырех ситуаций: гипотеза H_0 верна или нет (априори мы этого не знаем), реализация $\xi(\omega)$ случайного вектора ξ попала в S или нет. Подробнее.

1° Гипотеза H_0 верна. Реализация $\xi(\omega)$ не попала в S, и согласно критерию S гипотезу H_0 не отклоняем.

2° Гипотеза H_0 верна. Реализация $\xi(\omega)$ попала в S, и согласно критерию S гипотезу H_0 отклоняем.

3° Гипотеза H_0 неверна. Реализация $\xi(\omega)$ не попала в S, и согласно критерию S гипотезу H_0 не отклоняем.

4° Гипотеза H_0 неверна. Реализация $\xi(\omega)$ попала в S, и согласно критерию S гипотезу H_0 отклоняем.

Из четырех ситуаций две: 1° и 4° удовлетворительны и две: 2° и 3° неудовлетворительны.

В случаях 2° и 3° мы говорим, что в результате использования критерия S допускается ошибка.

Ошибки "верную гипотезу H_0 отклоняем" и "неверную гипотезу H_0 не отклоняем" качественно различны. Проиллюстрируем это различие на примере исследования медицинского препарата на токсичность.

Медицинский препарат перед выпуском в продажу исследуют на токсичность биологическими методами: определенную дозу препарата вводят подопытным животным (мышам, кроликам) и, в зависимости от количества ξ летальных исходов (число летальных исходов ξ — случайная величина) делают выводы о токсичности препарата — "препарат токсичен" или "препарат нетоксичен".

Задачу исследования препарата на токсичность биологическими методами в терминах проверки статистических гипотез формулируют так: относительно токсичности препарата выдвигают гипотезу H_0: препарат токсичен (альтернативная гипотеза H_1 — препарат нетоксичен). Необходимо проверить H_0, т. е. отклонить H_0 или нет. Выбор между этими действиями осуществляется по реализации $\xi(\omega)$ случайной величины ξ (по числу летальных исходов в группе животных) — если $\xi(\omega)$ оказалось малым, гипотезу H_0 отклоняют, в противном случае — нет.

Как и в каждой задаче проверки статистических гипотез, здесь возможны ошибки: 1° гипотеза H_0 верна, но согласно критерию ее отклоняют, и 2° гипотеза H_0 неверна, но согласно критерию ее не отклоняют.

Посмотрим, каковы последствия (какова “цена”) этих ошибок в данной конкретной ситуации проверки препарата на токсичность (проверки гипотезы H_0: препарат токсичен).

1° Пусть допущена ошибка: гипотеза H_0 верна, но согласно критерию H_0 отклонили (число летальных исходов $\xi(\omega)$ оказалось малым). Утверждение “H_0 верна” означает, что препарат токсичен (опасен для здоровья пациентов, его принимающих). Утверждение “H_0 отклонили” означает, что препарат согласно критерию классифицирован как нетоксичный. В итоге — токсичный препарат (опасный для здоровья пациентов) классифицирован как нетоксичный (безопасный для здоровья пациентов) и отправлен в продажу для реализации. Последствия ошибки такого рода могут привести к летальным исходам пациентов, принимающих этот препарат (“цена” ошибки — летальные исходы).

2° Пусть допущена ошибка: гипотеза H_0 неверна, но согласно критерию H_0 не отклонили (число летальных исходов $\xi(\omega)$ оказалось большим). Утверждение “гипотеза H_0 неверна” в нашей задаче означает, что препарат нетоксичен (безопасен для здоровья пациентов). Утверждение “H_0 не отклонили” означает, что согласно критерию препарат классифицирован как токсичный. В итоге нетоксичный препарат классифицирован как токсичный, и, следовательно, отправлен поставщику (для переработки или уничтожения). Последствия ошибки такого рода выражаются в финансовых убытках, увеличении стоимости препарата (“цена” ошибки — финансовые потери).

Очевидно, из двух ошибок: верную гипотезу H_0 отклонить, (последствия ошибки — выпуск в продажу токсичного препарата), и неверную гипотезу H_0 не отклонить (последствия ошибки — возвращение поставщику нетоксичного препарата), важнее избежать первой.

Определение. Ошибка, состоящая в отклонении верной гипотезы H_0, называется ошибкой *первого рода.* Ошибка, состоящая в неотклонении неверной гипотезы H_0, называется ошибкой *второго рода.*

О выборе нулевой гипотезы. Испытуемую гипотезу из совокупности возможных гипотез мы выбираем сами. Ошибки, связанные с проверкой гипотез и состоящие в том, что данную гипотезу H_θ, $\theta \in \Theta$, отклоняют, когда она верна, можно рассматривать и не выбрав нулевую гипотезу. Так вот, в качестве проверяемой гипотезы, как правило, выбирают ту, для которой важней избежать ошибки, состоящей в ее отклонении, когда она верна. Ошибка первого рода — это ошибка, которой важней избежать ("цена" которой выше).

Относительно токсичности препарата имеется две гипотезы: H_1 — препарат нетоксичен, H_2 — токсичен (класс возможных гипотез состоит из двух гипотез: H_1 и H_2). Безусловно, важнее избежать ошибки, состоящей в классификации токсичного препарата как нетоксичного, т. е. в отклонении верной H_2. Поэтому в качестве проверяемой гипотезы H_0 выбираем гипотезу H_2 — препарат токсичен, тогда ошибка, состоящая в классификации токсичного препарата как нетоксичного, будет ошибкой первого рода.

Уровень значимости критерия. Мощность критерия. При проверке статистических гипотез ошибки неизбежны — каким бы ни было критическое множество S для проверки гипотезы H_0, значение $\xi(\omega)$ случайной величины ξ при верной H_0 может попасть в S, что приводит к ошибке первого рода, а при неверной гипотезе H_0 значение $\xi(\omega)$ может не попасть в S, что приводит к ошибке второго рода. И коль скоро невозможно (невозможно в принципе) строить критерии для проверки данной гипотезы, не приводящие к ошибкам, то естественно пытаться строить такие критерии, при использовании которых частота ошибок была бы по возможности меньшей (за ошибки надо "платить"). Посмотрим, как это можно делать.

Пусть H_0 — испытуемая гипотеза. Предположим для наглядности, что H_0 — простая и конкурирующая гипотеза только одна (обозначим ее H_1) и она также простая. Пусть S — критическое множество. Используя S для проверки гипотезы H_0, а именно, отклоняя H_0, если ξ попадает в S, и не отклоняя в противном случае, мы можем допустить ошибки первого и второго рода. Вероятность ошибки первого рода равна вероятности попадания ξ в критическое множество S при верной гипотезе H_0, т. е. равна $\mathsf{P}\{\xi \in S|H_0\}$ (коротко будем писать $\mathsf{P}(S|H_0)$). Вероятность ошибки второго рода равна вероятности попадания ξ в $\overline{S}$ при верной гипотезе H_1, т. е. $\mathsf{P}\{\xi \in \overline{S}|H_1\}$ (коротко будем писать $\mathsf{P}(\overline{S}|H_1)$). (Заметим, что $\mathsf{P}\{\xi \in S|H_0\}$ обозначает вероятность события $\{\xi \in S\}$, вычисленную при верной гипотезе H_0, и не имеет ничего общего с условной вероятностью.) Частоты ошибок первого и второго рода близки к вероятностям $\mathsf{P}(S|H_0)$ и $\mathsf{P}(\overline{S}|H_1)$ этих ошибок соответственно, поэтому, чтобы частоты ошибок были малыми, критерий S для проверки гипотезы H_0 следует выбирать так, чтобы $\mathsf{P}(S|H_0)$ и $\mathsf{P}(\overline{S}|H_1)$ были по возможности меньше. Обе вероятности $\mathsf{P}(S|H_0)$ и $\mathsf{P}(\overline{S}|H_1)$ однозначно задаются критическим множеством S. Поэтому, выбирая S, скажем, так, чтобы вероятность $\mathsf{P}(S|H_0)$ ошибки первого рода была малой, мы одновременно получаем вероятность ошибки второго рода. (Заметим, что критическое множество S можно выбрать и из условия малости вероятности ошибки второго рода, но при этом вероятность ошибки первого рода определится выбранным S.) Так что выбирать S так, чтобы можно было контролировались вероятности ошибок первого и второго рода “одновременно и независимо“ невозможно. Поэтому будем поступать следующим образом. Поскольку важнее избежать ошибки первого рода (ее “цена” выше), критерий S будем выбирать так, чтобы вероятность ошибки первого рода $\mathsf{P}(S|H_0)$ была малой. (Тогда, используя критерий S для проверки H_0, в длинной серии экспериментов верную гипотезу H_0 будем отклонять изредка.) Малую вероятность ошибки первого рода $\mathsf{P}(S|H_0)$ будем гарантировать следующим образом: фиксируем малое α и критическое множество S выберем так, чтобы вероятность ошибки первого рода не превосходила α:

$$\mathsf{P}(S|H_0) \leq \alpha.$$

Как правило, это требование может быть удовлетворено более чем одним способом. Поэтому окончательно критическое множество S выбираем так, чтобы вероятность ошибки второго рода $\mathsf{P}(\overline{S}|H_1)$ была минимальна. Для этого выбираем S “пошире”.

Определение. Число α, ограничивающее сверху вероятность ошибки первого рода, называется *уровнем значимости.*

Если критическое множество S удовлетворяет условию

$$\mathsf{P}(S|H_0) \leq \alpha,$$

то будем говорить, что оно соответствует уровню значимости α.

Определение. Вероятность $\mathsf{P}(S|H_1)$ отклонить нулевую гипотезу, когда верна альтернативная гипотеза H_1, называется *мощностью критерия* S.

Если альтернативная гипотеза сложная, причем когда она верна, верна одна из простых гипотез H_θ, $\theta \in \Theta$, то при каждом $\theta \in \Theta$ можно вычислить $\mathsf{P}(S|H_\theta)$.

Определение. Функция

$$\beta(\theta) = \mathsf{P}(S|H_\theta),\ \theta \in \Theta,$$

равная вероятности отклонить нулевую гипотезу H_0, если верна гипотеза H_θ, $\theta \in \Theta$, называется *функцией мощности критерия* S.

Мощность критерия $\mathsf{P}(S|H_1)$ и вероятность ошибки второго рода $\mathsf{P}(\overline{S}|H_1)$ связаны так:

$$\mathsf{P}(S|H_1) = 1 - \mathsf{P}(\overline{S}|H_1).$$

Чувствительность критерия и выбор уровня значимости. Задача критерия S для проверки гипотезы H_0 — не отклонять H_0, когда она верна, и отклонять, когда она неверна — мало того чтобы критерий не отклонял верную H_0, он должен отклонять H_0, когда H_0 неверна. Свойство критерия S отклонять H_0, кода H_0 неверна (когда верна альтернативная гипотеза H_1), называют чувствительностью критерия. Количественно чувствительность критерия S описывается мощностью $\mathsf{P}(S|H_1)$ критерия (функцией мощности $\beta(\theta) = \mathsf{P}(S|H_\theta),\ \theta \in \Theta$, критерия) — чем больше мощность критерия, тем критерий чувствительнее.

Процедура проверки гипотезы H_0 сопровождается ошибками (первого и второго рода). Выбором уровня значимости α мы не только контролируем (ограничиваем) вероятность ошибки первого рода (а именно, $\mathsf{P}(S|H_0) \leq \alpha$), но и определяем чувствительность $\mathsf{P}(S|H_1)$ критерия. При этом за уменьшение уровня значимости α (а вместе с ним и вероятности ошибки первого рода $\mathsf{P}(S|H_0)$) мы расплачиваемся потерей (уменьшением) чувствительности $\mathsf{P}(S|H_1)$ критерия — потерей свойства критерия отклонять гипотезу H_0, когда она неверна. Последнее необходимо иметь в виду при выборе уровня значимости — критерии, соответствующие слишком малым уровням значимости α, имеют малую чувствительность (гипотеза H_0 неверна, а критерий ее не отклоняет).

Замечание. Обычно уровень значимости критерия полагают равным 0,10; 0,05; 0,01 и т. д. Чем серьезнее последствия ошибки первого рода, тем меньше должен быть уровень значимости. Но за уменьшение уровня значимости критерия мы “расплачиваемся” уменьшением его чувствительности, в частности, когда альтернативная гипотеза простая, увеличением вероятности ошибки второго рода.

22.2 Пример. Диагностика туберкулеза

Рассмотрим задачу диагностики туберкулеза, связанную с открытием того факта, что метод рентгеновского анализа не является абсолютно надежным при диагностике туберкулеза. Открытие сделано известными специалистами в рентгенологии докторами Берклоу, Чемберленом, Фелпсом, Заксом и статистиком профессором Иерусалми (см. Journ. Am. Med. Assn., vol. 133, а также Public Health Report, vol. 62, 40).

Термин "рентгеновский анализ" употребляется для обозначения двух операций: (a) изготовление снимка; (b) толкование этого снимка рентгенологом. Таким образом, когда мы говорим, что проведено n рентгеновских анализов индивидуума, то имеем в виду, что сделано n снимков и по каждому из них поставлен диагноз.

Мы будем предполагать, что людей можно разделить на две категории: 1) не имеющие признаков туберкулеза в легких, их кратко будем называть "здоровыми"; 2) лица с признаками туберкулеза в легких, которых будем называть "больными". Здоров индивидуум или нет, исход его рентгеновского обследования на туберкулез может быть как положительным, так и отрицательным. (Термин "положительный анализ" означает, что обследуемому пациенту ставится диагноз "туберкулез легких".) Для каждой категории будем постулировать существование вероятностей положительного рентгеновского анализа. Если пациент страдает туберкулезом, то вероятность того, что отдельный рентгеновский анализ положителен, обозначим через p_0; если не имеет никаких признаков туберкулеза — p_1. Вероятности p_0 и p_1 зависят не только от обследуемых индивидуумов, но также от тщательности изготовления рентгеновских снимков и квалификации рентгенолога, исследующего снимки.

Предположим, что в некоторой клинике при обследовании состояния здоровья пациента делается одним и тем же способом несколько рентгеновских снимков с целью обнаружения возможных признаков туберкулеза. Допустим, что толкование нескольких снимков, принадлежащих одному и тому же пациенту, производится так, чтобы обеспечить независимость диагноза по каждому снимку от выводов, сделанных по предыдущим. Этого можно достигнуть, например, предлагая рентгенологу для просмотра снимки, принадлежащие разным людям, в случайном порядке.

Пусть наш предшествующий опыт выражается в следующем:

1. Если пациент страдает туберкулезом, то вероятность того, что отдельный рентгеновский анализ положителен, равна $p_0 = 0{,}80$.

2. Если пациент не несет никаких признаков туберкулеза, то вероятность положительного анализа равна $p_1 = 0{,}01$.

Задача. *Предположим, что для каждого пациента делается $n = 3$ независимых рентгеновских анализа, по результатам которых необходимо вынести заключение: болен человек туберкулезом или нет, а, значит, следует его лечить или нет.*

Интуитивно ясно, что если число положительных анализов "большое" (больше некоторого l), то необходимо вынести заключение, что пациент "болен" (заметим, что "большое" число положительных анализов может быть и у здорового пациента, при этом будет поставлен ошибочный диагноз — туберкулез легких); если число положительных анализов "малое" (меньше l), то заключаем, что пациент "здоров" (заметим, что "малое" число положительных анализов может быть и у "больного" пациента, при этом будет допущена ошибка — у больного "просмотрели" туберкулез). К таким естественным и правдоподобным выводам мы приходим и не пользуясь идеями проверки статистических гипотез. Но наши выводы не содержат количественной информации: во-первых, о границе l, разделяющей "большое" число положительных анализов и "малое", во-вторых, о частоте ошибок, состоящих в классификации больного как здорового и здорового как больного.

Попытаемся решить поставленную задачу диагностики туберкулеза, сформулировав ее в терминах проверки статистических гипотез.

1. Выбор нулевой гипотезы. Относительно состояния здоровья пациента естественно выдвинуть две гипотезы: H_1 — "пациент здоров" и H_2 — "пациент болен" (страдает туберкулезом). Из этих гипотез одну необходимо выбрать в качестве нулевой (испытуемой). Для этого рассмотрим последствия ошибок, состоящих в отклонении гипотез H_1 и H_2, когда они верны. Ошибка, состоящая в отклонении верной H_1, приводит к излишнему беспокойству здорового пациента, связанному с его лечением до тех пор, пока очередное обследование состояния здоровья пациента не установит, что его легкие в порядке. Такая излишняя предосторожность, проявленная клиникой, может повредить ее репутации. Однако ошибка, состоящая в отклонении H_2, когда она верна, чревата для пациента потерей драгоценной возможности избавиться от болезни в ранней стадии. К тому же для специалистов клиники "просмотреть" болезнь означало бы подорвать свою репутацию в большей мере, чем

неоправданным беспокойством здорового пациента. С этой точки зрения важнее избежать ошибки, состоящей в отклонении гипотезы H_2, когда она верна. Поэтому в качестве испытуемой (нулевой) гипотезы будем рассматривать гипотезу H_2 — пациент болен. Как обычно, обозначим ее H_0. Альтернативной к гипотезе H_0 является гипотеза H_1 — пациент здоров.

2. Эксперимент. Для того чтобы вынести заключение о состоянии здоровья человека относительно туберкулеза, проводим эксперимент, состоящий в получении реализации $\xi(\omega)$ случайной величины ξ — числа положительных рентгеновских анализов из трех проведенных (ξ принимает значения $0, 1, 2, 3$).

3. Семейство $\mathcal{P}$ возможных распределений случайной величины ξ и гипотезы H_0 и H_1. В условиях задачи семейство $\mathcal{P}$ возможных распределений случайной величины ξ состоит из двух распределений:

$$\mathsf{P}_0(k) = B_{3;p_0}(k) = C_3^k p_0^k (1-p_0)^{3-k},\ k = 0, 1, 2, 3; \qquad (22.2.1)$$

$$\mathsf{P}_1(k) = B_{3;p_1}(k) = C_3^k p_1^k (1-p_1)^{3-k},\ k = 0, 1, 2, 3. \qquad (22.2.2)$$

Гипотезы относительно здоровья пациента — это фактически гипотезы о распределении случайной величины ξ. А именно, при верной гипотезе H_0 случайная величина ξ имеет биномиальное распределение $B_{3;p_0}(k)$, а при верной гипотезе H_1 — биномиальное распределение $B_{3;p_1}(k)$.

4. Выбор критерия. Наша задача — по реализации $\xi(\omega)$ случайной величины ξ сделать заключение о ее распределении: P_0 может быть распределением ξ (гипотезу H_0 не отклоняем) или P_0 не может распределением ξ (гипотезу H_0 отклоняем). Для этого необходимо указать подмножество S множества $\{0, 1, 2, 3\}$ результатов $\xi(\omega)$ эксперимента, которым гипотеза H_0 противоречит — при попадании $\xi(\omega)$ в которое гипотезу отклоняем.

Если верна гипотеза H_1 (т. е. пациент здоров), случайная величина ξ имеет распределение P_1 (см. (22.2.2)). Табличная форма записи P_1:

k	0	1	2	3
$\mathsf{P}_1(k)$	0,9702	0,0294	0,0003	0,0000

Если H_0 верна (т. е. пациент болен), распределением случайной величины ξ является P_0 (см. 22.2.1). Табличная форма записи P_0:

k	0	1	2	3
$\mathsf{P}_0(k)$	0,0080	0,0960	0,3840	0,5120

Отсюда видно, что в качестве распределения ξ естественно рассматривать

$$\mathsf{P}_1(k) = \mathsf{P}_1\{\xi = k\} = C_3^k p_1^k (1-p_1)^{3-k},\ k = 0, 1, 2, 3,$$

если $\xi(\omega)$ принимает "малые" значения ($\xi \le l$), и распределение

$$\mathsf{P}_0(k) = \mathsf{P}_0\{\xi = k\} = C_3^k p_0^k (1-p_0)^{3-k}\ k = 0, 1, 2, 3,$$

если $\xi(\omega)$ принимает "большие" значения ($\xi > l$). Поэтому гипотезу H_0 будем отклонять, если $\xi \le l$, и не отклонять, если $\xi > l$. И следовательно, критическое множество для проверки гипотезы H_0: пациент болен, против альтернативы H_1: пациент здоров, естественно искать в виде $S = \{x : x \le l\}$. Окончательный выбор S определяется уровнем значимости и мощностью критерия.

5. Выбор уровня значимости. Уровень значимости α — это число, ограничивающее сверху вероятность ошибки первого рода, т. е. вероятность отклонения H_0, когда она верна. Вообще говоря, желательно, чтобы уровень значимости α был равен нулю. Для $\alpha = 0$ равенство

$$\mathsf{P}_0(S) = \mathsf{P}(S|H_0) = \mathsf{P}\{\xi \le l/H_0\} = \sum_{k=0}^{l} C_3^k p_0^k (1-p_0)^{3-k} \le \alpha = 0$$

имеет место только при $S = \emptyset$. Но для $S = \emptyset$ всегда $\xi(\omega) \notin S$, и, следовательно, вне зависимости от исхода $\xi(\omega)$ эксперимента гипотезу H_0 не отклоняем (всем ставим положительный диагноз). Такой критерий не учитывает результат эксперимента и, разумеется, интереса не представляет. Поэтому, хочется нам того или нет, но α необходимо выбрать отличным от нуля. Пусть $\alpha = 0{,}01$. По α определяем l как наибольшее натуральное, для которого

$$\sum_{k=0}^{l} C_3^k p_0^k (1-p_0)^{3-k} \le \alpha,$$

(число l определяется как наибольшее, чтобы обеспечить более широкое $S = \{x : x \le l\}$, а вместе с этим бóльшую мощность критерия S). Для $\alpha = 0{,}01$ получаем $l = 0$, так что критическое множество $S = \{x : x \le 0\} = \{x : x = 0\}$.

Таким образом, если число положительных анализов $\xi \in S$, т. е. ξ равно нулю, гипотезу H_0: пациент болен, отклоняем, если

$\xi \notin S$, т. е. хотя бы один анализ положителен, гипотезу H_0 не отклоняем — пациенту ставим положительный диагноз.

6. Вероятности ошибок первого и второго рода. Вероятность ошибки первого рода:

$$\mathsf{P}_0(S) = \mathsf{P}(S|H_0) = \mathsf{P}\{\xi \le 0|H_0\} = C_3^0(0{,}8)^0(0{,}2)^3 = 0{,}008.$$

Вероятность ошибки второго рода:

$$\mathsf{P}_1(\overline{S}) = \mathsf{P}(\overline{S}|H_1) = 1 - \mathsf{P}(S|H_1) = 1 - \mathsf{P}\{\xi \le 0|H_1\} =$$

$$= 1 - C_3^0 \cdot 0{,}01^0 \cdot 0{,}99^3 = 0{,}0297.$$

7. Мощность критерия. Мощность критерия равна вероятности отклонения H_0, когда H_0 неверна. В нашей задаче

$$\beta(1) = \mathsf{P}(S|H_1) = \mathsf{P}\{\xi \le 0|H_1\} = C_3^0 \cdot 0{,}01^0 \cdot 0{,}99^3 = 0{,}970.$$

8. Интерпретация результатов. Полученные результаты:

$$\mathsf{P}(S|H_0) = 0{,}008,\ \mathsf{P}(\overline{S}|H_1) = 0{,}0297,\ \mathsf{P}(S|H_1) = 0{,}970$$

можно интерпретировать следующим образом.

Если выполнены сделанные выше предположения и в клинике используется описанная процедура диагностики туберкулеза, то равенство $\mathsf{P}(S|H_0) = 0{,}008$ обозначает, что при верной гипотезе H_0 (пациент болен) вероятность отклонения H_0 (классификации пациента как здорового) равна 0,008. Этому результату можно дать такую частотную интерпретацию: из 100 больных туберкулезом в среднем одному будет ставиться отрицательный диагноз (больного пациента будут классифицировать как здорового). Из равенства $\mathsf{P}(S|H_0) = 0{,}008$ имеем $\mathsf{P}(\overline{S}|H_0) = 0{,}992$. Частотная интерпретация результата — из 100 больных туберкулезом в среднем 99 будет поставлен положительный диагноз (пациент болен).

Равенство $\mathsf{P}(\overline{S}|H_1) = 0{,}0297$ означает, что при верной гипотезе H_1 (пациент здоров) вероятность не отклонять гипотезу H_0 (пациент страдает туберкулезом) равна 0,0297. Частотная интерпретация результата — из 100 здоровых в среднем 3 будет поставлен положительный диагноз (из 100 здоровых пациентов 3 будут классифицированы как больные). Далее, $\mathsf{P}(S|H_1) = 0{,}970$. Последнее интерпретируется так: из 100 здоровых в среднем 97 будет поставлен отрицательный диагноз (из 100 здоровых пациентов 97 будут классифицированы как здоровые).

Замечание. Следует сказать об упрощениях, введенных при решении задачи диагностики туберкулеза.

Во-первых, мы предположили, что для всех пациентов, страдающих туберкулезом, вероятность p_0 положительного анализа одна и та же, что естественно, не так, поскольку p_0 зависит от тяжести болезни.

Во-вторых, мы предположили, что результаты анализов одного и того же пациента независимы. Однако, как правило, рентгенолог исследует сразу несколько снимков пациента и выносит решение о диагнозе, руководствуясь впечатлением от всех снимков.

22.3 Примеры и задачи

Примеры

Пример 22.3.1 (выборочный контроль). *Предназначенная для продажи партия изделий в количестве* 10 000 *штук проходит выборочный контроль на качество. Поставщик уверен, что доля дефектных изделий равна* 1 % *(или меньше), и хочет, чтобы всякий раз, когда эта доля составляет* 1 %*, партия выдерживала контроль с вероятностью* 0,90*. Покупатель считает, что партию целесообразно закупить даже в том случае, если доля дефектных изделий будет больше* 1 %*, но* 6 % *дефектных изделий он считает предельно допустимой долей и хочет, чтобы контроль обнаруживал партии с* 6 % *содержанием дефектных изделий с вероятностью* 0,95*. Поставщик и покупатель договорились осуществлять контроль следующим образом: из партии в* 10 000 *изделий извлекают случайную выборку объемом n, если при этом количество $\xi(\omega)$ дефектных изделий в выборке окажется малым (меньше некоторого l), то партия выдерживает контроль и закупается, в противном случае партия бракуется.*

Каким должен быть объем n выборки из партии и значение l?

Ответить на этот вопрос, сформулировав и решив поставленную задачу как задачу проверки статистических гипотез.

Решение. Изделия в выборку объемом n извлекают последовательно. Поскольку объем партии большой (10 000 изделий), то можно считать, что случайная величина ξ — число дефектных изделий в выборке — имеет биномиальное распределение с параметрами $(n;\theta)$:

$$\mathsf{P}\{\xi = k\} = B_{n,\theta}(k) = C_n^k \theta^k (1-\theta)^{n-k},\ k = 0, 1, \ldots, n,$$

параметр θ неизвестен (вероятность выбора дефектного изделия неизвестна). Семейством возможных распределений случайной величины ξ является $\mathcal{P} = \{B_{n,\theta}, \theta \in (0;1)\}$, через H_θ будем обозначать гипотезу — ξ имеет своим распределением $B_{n,\theta}$, $\theta \in (0;1)$.

Отметим, что распределение из семейства $\mathcal{P}$ однозначно определяется значением параметра θ и наоборот.

Для нас представляют интерес следующие значения параметра θ: $\theta_1 = 0{,}01$ и $\theta_2 = 0{,}06$ (им соответствует $1\,\%$ и $6\,\%$ содержание дефектных изделий в партии). Поэтому в качестве нулевой гипотезы естественно выбрать одну из гипотез: гипотезу $H_{0,01}$ — ξ имеет распределение $B_{n;0,01}$, альтернативная гипотеза $H_{(0,01;1)}$ — распределение ξ принадлежит классу распределений $\{B_{n,\theta},\ \theta \in (0{,}01;1)\}$, или гипотезу $H_{0,06}$ — ξ имеет распределение $B_{n;0,06}$, альтернативная гипотеза $H_{(0;0,06)}$ — распределение ξ принадлежит классу распределений $\{B_{n;\theta}, \theta \in (0;0{,}06)\}$. Остановимся на последней, а именно, $H_0 = H_{0,06}$, альтернативная гипотеза $H_1 = H_{(0;0,06)}$. Альтернатива $H_1 = H_{(0;0,06)}$ сложная, когда $H_{(0;0,06)}$ верна, верна одна из простых гипотез H_θ: ξ имеет распределение $B_{n,\theta}, \theta \in (0;0{,}06)$.

Неотклонение нулевой гипотезы $H_{0,06}$ означает классификацию партии как партии, содержащей $6\,\%$ дефектных изделий, с последующим отказом в закупке партии. Отклонение нулевой гипотезы в пользу альтернативы $H_1 = H_{(0;0,06)}$ означает классификацию партии как такой, которая содержит долю дефектных изделий меньшую 0,06, с последующей закупкой партии. Если при этом партия с $6\,\%$ брака (гипотеза $H_0 = H_{0,06}$ верна) классифицируется как партия, содержащая меньше $6\,\%$ дефектных изделий и, как следствие, закупается (допущена ошибка первого рода), то покупатель, разумеется, понесет убытки. Если какая-то из партий с малым процентом дефектных изделий — меньшим $6\,\%$ (верна гипотеза $H_1 = H_{(0;0,06)}$), классифицируется как партия, содержащая $6\,\%$ брака, то партия бракуется (допущена ошибка второго рода), и это уже проблема поставщика.

Для проверки гипотезы H_0 в качестве критического естественно выбрать множество вида $S = \{k : k \le l\}$. Это множество определяется числом l и объемом выборки n.

По договоренности между покупателем и поставщиком множество $S = \{k : k \le l\}$, а фактически числа n и l, должны быть такими, чтобы выполнялись приведенные ранее договоренности между покупателем и поставщиком, а именно.

1. Партия с 6 % дефектных изделий должна обнаруживаться с вероятностью 0,95, т. е.

$$\mathsf{P}(\overline{S}|H_{0,06}) \geq 0{,}95,$$

или, что то же,

$$\mathsf{P}(S|H_{0,06}) = \mathsf{P}\{\xi \in S|H_0\} = \mathsf{P}_{0,06}\{\xi \leq l\} =$$

$$= \sum_{k=0}^{l} C_n^k (0{,}06)^k (0{,}94)^{n-k} \leq \alpha = 0{,}05.$$

Последнее неравенство — формализованная запись требований покупателя, согласно которым партии с 6 % содержанием брака должны обнаруживаться с вероятностью не меньшей, чем 0,95.

2. Значение $\beta(\theta) = \mathsf{P}(S|H_\theta)$ функции мощности критерия S в точке $\theta = 0{,}01$ должно быть не меньшим 0,9:

$$\beta(0{,}01) = \mathsf{P}(S|H_{0,01}) = \mathsf{P}\{\xi \in S|H_{0,01}\} = \mathsf{P}\{\xi \leq l|H_{0,01}\} =$$

$$= \sum_{k=0}^{l} C_n^k (0{,}01)^k (0{,}99)^{n-k} \geq 0{,}9.$$

Последнее неравенство является формализованной записью требований поставщика, согласно которым партии с 1 % содержанием брака должны выдерживать контроль с вероятностью не меньшей, чем 0,9.

Таким образом, искомые n и l должны удовлетворять соотношениям:

$$\sum_{k=0}^{l} C_n^k (0{,}06)^k (0{,}94)^{n-k} \leq 0{,}05; \qquad (22.3.1)$$

$$\sum_{k=0}^{l} C_n^k (0{,}01)^k (0{,}99)^{n-k} \geq 0{,}9, \qquad (22.3.2)$$

разумеется, при этом n следует выбрать по возможности меньшим.

В рассматриваемой задаче биномиальное распределение

$$B_{n,\theta}(m) = C_n^m \theta^m (1-\theta)^{n-m},\ m = 0, 1, \ldots, n,$$

удобно аппроксимировать пуассоновским, поскольку параметр θ малый (вероятность θ выбора дефектного изделия мала), а объем n выборки — большой. Основанием для этого является теорема Пуассона (см. теорему 6.2.1). Чтобы погрешность аппроксимации биномиального распределения пуассоновским была незначительной, n должно быть не меньшим нескольких десятков (а лучше сотен), а произведение $n\theta$ должно лежать между 1 и 10. При этом

$$\mathsf{P}\{\xi = k | H_{0,06}\} = \mathsf{P}_{\lambda_0}(k) = \frac{\lambda_0^k}{k!} e^{-\lambda_0},\ k = 0, 1, 2, \ldots,\ \lambda_0 = n \cdot 0{,}06,$$

$$\mathsf{P}\{\xi = k | H_{0,01}\} = \mathsf{P}_{\lambda_1}(k) = \frac{\lambda_1^k}{k!} e^{-\lambda_1},\ k = 0, 1, 2, \ldots,\ \lambda_1 = n \cdot 0{,}01,$$

а n и l должны удовлетворять соотношениям:

$$\mathsf{P}\{\xi \le l | H_{0,06}\} = \sum_{k=0}^{l} \frac{\lambda_0^k}{k!} e^{-\lambda_0} \le 0{,}05,\ \ \lambda_0 = n \cdot 0{,}06; \tag{22.3.3}$$

$$\mathsf{P}\{\xi \le l | H_{0,01}\} = \sum_{k=0}^{l} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \ge 0{,}90,\ \ \lambda_1 = n \cdot 0{,}01 \tag{22.3.4}$$

(сравните с неравенствами (22.3.1) и (22.3.2)).

Определяя n и l, мы, естественно, будем стремиться выбрать n по возможности меньшим.

Посмотрим, можно ли, скажем, для $n = 50$ выбрать l так, чтобы выполнялись соотношения (22.3.3) и (22.3.4). Если нет, выбираем n бóльшим, и так будем поступать до тех пор, пока не найдем n и l, которые будут удовлетворять неравенствам (22.3.3) и (22.3.4).

При $n = 50$ имеем $\lambda_0 = 50 \cdot 0{,}06 = 3$; $\lambda_1 = 50 \cdot 0{,}01 = 0{,}5$. Пользуясь таблицей распределения Пуассона (см. табл. 27.6.1), находим, что неравенство (22.3.3) удовлетворяется при $l = 0$, а неравенство (22.3.4) — при $l \ge 1$. Поэтому для $n = 50$ не существует l, которое бы удовлетворяло неравенствам (22.3.3) и (22.3.4).

При $n = 100$ имеем $\lambda_0 = 100 \cdot 0{,}06 = 6$; $\lambda_1 = 100 \cdot 0{,}01 = 1$. Из таблицы распределения Пуассона находим, что неравенство (22.3.3) удовлетворяется, при $l \le 1$, а неравенство (22.3.4) — при $l \ge 2$. Поэтому для $n = 100$ не существует l, которое бы удовлетворяло неравенствам (22.3.3) и (22.3.4).

При $n = 150$ имеем $\lambda_0 = 150 \cdot 0{,}06 = 9$; $\lambda_1 = 150 \cdot 0{,}01 = 1{,}5$. Для таких λ_0 и λ_1 неравенство (22.3.3) удовлетворяется, при $l \leq 4$, а неравенство (22.3.4) — при $l \geq 3$. Попытаемся уменьшить n.

Пусть $n = 130$. Тогда $\lambda_0 = 130 \cdot 0{,}06 = 7{,}8$; $\lambda_1 = 130 \cdot 0{,}01 = 1{,}3$; неравенство (22.3.3) удовлетворяется, при $l \leq 3$, а неравенство (22.3.4) — при $l \geq 3$. Таким образом, для $n = 130$ и $l = 3$ оба неравенства (22.3.3) и (22.3.4) удовлетворяются. И, следовательно, критическое множество (критерий) для проверки гипотезы H_0 имеет вид

$$S = \{k\colon k \leq 3\}.$$

Для критерия $S = \{k : k \leq 3\}$ вероятность ошибки первого рода

$$\mathsf{P}\{\xi \leq 3 | H_{0,06}\} = \mathsf{P}_{0,06}(S) = \sum_{k=0}^{3} \frac{\lambda_0^k}{k!} e^{-\lambda_0} = \sum_{k=0}^{3} \frac{7{,}8^k}{k!} e^{-7{,}8} = 0{,}048.$$

Это можно интерпретировать так. Используя описанный критерий, на 100 партий изделий, содержащих 6 % брака, в среднем около 5 партий будут классифицироваться как содержащие меньше 6 % брака и, следовательно, будут закупаться.

Значения функции мощности

$$\beta(\theta) = \mathsf{P}(S|H_\theta) = \sum_{k=0}^{3} \frac{(130 \cdot \theta)^k}{k!} e^{-130\cdot\theta},\ \theta \in (0; 0{,}06),$$

критерия S в некоторых точках приведены в таблице 22.3.1.

Таблица 22.3.1

θ	0,005	0,01	0,02	0,04	0,05	0,06	0,07	0,08
$\beta(\theta)$	0,996	0,957	0,736	0,238	0,112	0,048	0,020	0,008

Заметим, что значение 0,048 функции мощности критерия $\beta(\theta)$ в точке $\theta = 0{,}06$ равно вероятности ошибки первого рода (напомним, что H_0: $\theta = 0{,}06$).

График функции мощности $\beta(\theta) = \mathsf{P}(S|H_\theta)$ критерия изображен на рис. 22.3.1.

В точке $\theta = 0{,}01$ значение функции мощности критерия

$$\beta(0{,}01) = \mathsf{P}\{\xi \leq 3 | H_{0,01}\} = \mathsf{P}_{0,01}(S) =$$

$$= \sum_{k=0}^{3} \frac{\lambda_1^k}{k!} e^{-\lambda_1} = \sum_{k=0}^{3} \frac{1{,}3^k}{k!} e^{-1{,}3} = 0{,}957.$$

Последнее обозначает, что на 100 партий изделий с 1 % содержанием брака в среднем около 96 партий будут выдерживать контроль.

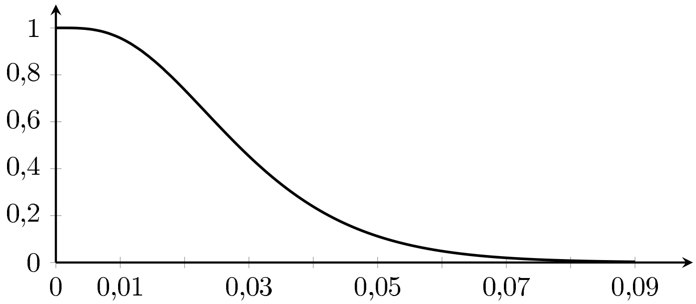

Рис. 22.3.1: График функции мощности критерия $\beta(\theta) = \mathsf{P}(S|H_\theta)$

В точке 0,005 значение

$$\beta(0{,}005) = \mathsf{P}\{\xi \leq 3|H_{0{,}005}\} = \mathsf{P}_{0{,}005}(S) = 0{,}995.$$

Это можно интерпретировать так: используя критерий $S = \{k \colon k \leq 3\}$, в среднем на 100 партий изделий с 0,5 % содержанием дефектных изделий около 99 партий будут выдерживать контроль.

Далее, хотя гипотеза $H_0 = H_{0{,}06}$ проверяется против альтернативы $H_1 = H_{(0;0{,}06)}$ — параметр $\theta \in (0; 0{,}06)$, формально мы можем вычислять значения функции мощности $\beta(\theta) = \mathsf{P}(S|H_\theta)$ и при значениях параметра $\theta > 0{,}06$, что соответствует гипотезам о содержании доли дефектных изделий в партии большем 0,06. В частности, например, в точке $\theta = 0{,}08$ значение функции мощности критерия

$$\beta(0{,}08) = \mathsf{P}\{\xi \leq 3|H_{0{,}08}\} = \mathsf{P}_{0{,}08}(S) = 0{,}008.$$

Последнее обозначает, что при верной гипотезе $H_{0{,}08}$ — партия содержит 8% дефектных изделий, гипотеза H_0 с вероятностью 0,008 будут отклоняется, и с вероятностью $1 - \beta(0{,}08) =$

$= 0{,}992$ не будет отклоняться и, как следствие, партия, содержащая 8% дефектных изделий, с вероятностью 0,992 будет браковаться. Так что критерий $\{k : k \leq 3\}$ "хорошо работает" и при значениях параметра θ больших 0,06.

Задачи

22.1 (о телепатах). Некоторые люди утверждают, что они могут читать мысли на расстоянии — являются телепатами. При этом телепат не претендует на то, что он может безошибочно читать мысли, но утверждает, что, иногда ошибаясь, он все-таки чаще читает мысли верно, чем неверно. Необходимо проверить способности телепата читать мысли на расстоянии. С этой целью предлагается провести такой эксперимент. Вы задумываете наудачу число (для простоты 0 или 1) и фиксируете его, например, на бумаге. Телепат читает вашу мысль — задуманное число и также фиксирует его, и так n раз. Если телепат действительно читает ваши мысли, то число правильно прочитанных нулей (нуль читается как нуль) и единиц (единица читается как единица), т. е. доля успехов (успех — правильно прочитанный символ) будет большой. Телепату предлагается прочитать n символов, $\xi(\omega)$ из них были прочитаны правильно. Свидетельствует ли это о способности телепата читать мысли на расстоянии?

Предложить правило (критерий), по которому можно обнаружить способности телепата читать мысли на расстоянии.

Сформулировать и решить поставленную задачу в терминах проверки статистических гипотез, а именно:

предложить нулевую (основную) и альтернативные гипотезы;

выяснить, в чем состоят ошибки первого и второго рода;

определить случайную величину, наблюдаемую в эксперименте (что означают выдвинутые гипотезы относительно распределения этой случайной величины?);

предложить критерий для проверки нулевой гипотезы;

выбрать уровень значимости критерия, обосновать выбор;

вычислить вероятность ошибки первого рода;

исследовать поведение функции мощности критерия;

дать частотную интерпретацию полученных результатов.

Указание. Нулевая гипотеза H_0: телепат мысли не читает.

22.2 (о статистике и экспериментаторе). Два студента (одного будем называть статистиком, другого — экспериментатором) имеют монеты: симметричную и несимметричную, вероятность выпадения "герба" последней θ $(\theta \neq 1/2)$. Экспериментатор и статистик договариваются о следующем: экспериментатор выходит в соседнюю комнату, выбирает одну из монет (какую — статистик не знает), подбрасывает ее n раз и результат (число выпадений “герба”) сообщает статистику. Последний утверждает, что может определить, какую из двух монет подбрасывал экспериментатор. Однако экспериментатор ставит под сомнение эти способности статистика и как аргумент выдвигает такие возражения: поскольку результат эксперимента (число выпадений “герба”) предсказать невозможно, причем как для симметричной монеты, так и для несимметричной это число может быть любым: $0, 1, \ldots, n$, то определить, какая из монет бросалась, невозможно.

Кто прав — статистик или экспериментатор?

Сформулируйте поставленную задачу с позиции статистика, а именно, как задачу проверки статистических гипотез (см. задачу 22.1). Что можно сказать относительно возможности различить симметричную и несимметричную монеты?

Пусть $n = 25$. Какими будут выводы, если герб выпал 21 раз?

У к а з а н и е. Нулевая гипотеза H_0: монета симметричная, альтернатива двусторонняя.

22.3 (о контроле препарата на токсичность). Процесс производства некоторого медицинского препарата достаточно сложный, так что несущественные на первый взгляд отклонения от технологии могут вызвать появление высокотоксичной побочной примеси. Токсичность последней может оказаться столь высокой, что даже незначительное ее количество, которое не может быть обнаружено при обычном химическом анализе, опасно для человека, принимающего препарат. Поэтому до реализации партии препарата его исследуют на токсичность биологическими методами. Определенная доза препарата вводится подопытным животным и регистрируется результат (количество летальных исходов).

Если препарат токсичный, то все или почти все животные гибнут — количество летальных исходов $\xi(\omega)$ большое ($\xi(\omega) \geq l$). В противном случае количество летальных исходов $\xi(\omega)$ малое ($\xi(\omega) < l$), или, что то же, количество выживших животных большое.

Известно, что для токсичного препарата вероятность летального исхода не меньше чем 0,9, а если препарат нетоксичный, то вероятность летального исхода не превышает 0,05. Какое минимальное число n животных необходимо инъецировать, чтобы токсичный препарат обнаруживался (классифицировался как токсичный) с вероятностью не меньшей, чем 0,999, а нетоксичный препарат выдерживал контроль (классифицировался как нетоксичный) с вероятностью 0,98? Найти n, сформулировав поставленную задачу как задачу проверки статистических гипотез (см. задачу 22.1).

Указание. Нулевая гипотеза H_0: препарат токсичный.

22.4 (о контроле чистоты воды). Для нужд некоторого химического производства необходимо, чтобы используемая вода была чистой — содержала мало бактерий. Вода, содержащая в среднем менее одной бактерии на единицу объема, считается пригодной для данного химического производства. Если же среднее количество бактерий на единицу объема воды равно одной или больше, то ее применение недопустимо.

Обычная методика контроля воды на чистоту такая. Берется 10 (в общем случае n) проб воды единичного объема. Потом каждая из этих проб добавляется в колбы с питательной средой, которые содержатся при температуре, благоприятной для роста бактерий. Если проба загрязнена, т. е. содержит по меньшей мере одну бактерию, то колония бактерий будет расти и первоначально прозрачный раствор помутнеет. О чистоте воды судят по количеству ξ загрязненных проб: если это количество не превышает определенного числа l, то вода считается чистой; в противном случае — загрязненной, тогда проводится дополнительная очистка воды, что, естественно, требует труда и затрат, и она снова проходит контроль на чистоту. Однако, если используемая вода содержит значительное количество бактерий, то вызванные этим потери значительно больше.

Каким должно быть l, чтобы вода со средним содержанием бактерий, равным одной на единицу объема, обнаруживалась с вероятностью 0,99? Найти указанное l, сформулировав поставленную задачу как задачу проверки статистических гипотез (см. задачу 22.1).

Замечание. В качестве распределения числа бактерий в единичном объеме воды со средним содержанием λ бактерий на единицу объема естественно рассматривать пуассоновское распределение с параметром λ.

Указание. Нулевая гипотеза H_0: $\lambda = 1$ — среднее содержание бактерий на единицу объема воды равно 1. Альтернативная гипотеза $H = H_\lambda$ $(\lambda < 1)$ — среднее содержание бактерий на единицу объема воды меньше чем 1.

Наблюдаемая случайная величина ξ — количество загрязненных проб — имеет биномиальное распределение с параметрами $(10, \theta)$, где θ — вероятность того, что в пробе будет по меньшей мере одна бактерия. При верной гипотезе H_0: $\lambda = 1$ значение $p = p_0 = 1 - e^{-1}$. Критическое множество для проверки гипотезы H_0 имеет вид $S = \{k : k \leq l\}$.

22.5 (о леди, пробующей чай). Некая леди утверждает, что, попробовав чашку чая с молоком, она может определить, что было сначала налито в чашку — молоко или чай (различает рецепты приготовления чая). При этом леди не претендует на безошибочное определение разницы во вкусе, но утверждает, что, пусть иногда ошибаясь, она чаще определяет правильно, нежели неправильно.

Прежде чем признать наличие у леди способностей различать рецепты приготовления чая, ей предлагают попробовать и классифицировать n пар чашек чая (по паре чашек за завтраком в течение n дней). В каждую пару входит по чашке чая, приготовленного по разным рецептам. Количество ξ правильно классифицированных пар регистрируется. Пусть леди из n пар чашек правильно классифицировала $\xi(\omega)$. Свидетельствует ли это о способностях леди?

Вы — член доброжелательного жюри — не хотите понапрасну отвергнуть способности леди, если они у леди в самом деле имеются. Предложите правило (критерий), которым должно руководствоваться жюри, вынося заключение о способностях леди различать рецепты приготовления чая.

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез (см. задачу 22.1).

Выясните, как меняется функция мощности критерия в зависимости от уровня значимости критерия и количества n пар чашек чая, предлагаемых для классификации.

Замечание. Как член жюри вы, разумеется, не хотите признать за леди упомянутые способности, если она их не имеет; поэтому в качестве нулевой гипотезы предлагаете гипотезу H_0: леди не имеет способностей. Но как член доброжелательного жюри, не хотите понапрасну отвергнуть способности леди, если она их имеет, и поэтому назначаете не слишком малый уровень значимости.

22.6 (о выборочном контроле). Потребитель приобретает крупные партии товаров. Чтобы избежать значительной доли дефектных изделий, осуществляется выборочный контроль. Из каждой партии для контроля выбирается n изделий. Партия принимается, если число дефектных изделий среди n проверенных не превышает l.

Потребитель предъявляет следующие требования к контролю: если доля дефектных изделий в партии меньше чем $8\,\%$, то партия выдерживает контроль и закупается, но партии с $8\,\%$ дефектных изделий должны обнаруживаться с вероятностью не меньшей чем 0,9. У поставщика свои пожелания относительно контроля: он хочет, чтобы партия, доля дефектных изделий в которой составляет 1 %, браковалась с вероятностью не большей чем 0,04.

Какими должны быть объем n выборки из партии (разумеется, минимальный) и значение l?

Сформулировать и решить поставленную задачу в терминах проверки статистических гипотез (см. задачу 22.1 и пример 22.3.1).

Какова вероятность забраковать партию товаров, содержащую 1; 2; 4; 8; 10 % дефектных изделий?

Указание. Нулевая гипотеза H_0: доля дефектных изделий составляет 0,01; альтернативная гипотеза H_p: $p \in (0{,}01; 1)$ (H_p является сложной). Критическое множество $S = \{k : k > l\}$. Требование потребителя формулируется следующим образом:

$$\beta(0{,}08) = \mathsf{P}(S|H_{0{,}08}) \geq 0{,}9,$$

где $\beta(p)$ — функция мощности критерия S.

22.7 (об игральных костях). Имеются две игральные кости: одна — симметричная (вероятность выпадения каждой грани равна 1/6), другая — несимметричная, центр тяжести которой смещен так, что вероятность появления числа очков, большего 3, превышает 1/2.

С целью обнаружить несимметричную кость берем одну (какую именно, неизвестно — кости внешне различить невозможно), подбрасываем ее n раз и регистрируем количество выпавших очков. Какое минимальное количество подбрасываний необходимо выполнить, чтобы симметричная кость классифицировалась как несимметричная с вероятностью не большей чем 0,1, а несимметричная, с вероятностью 0,9 выпадения количества очков большего чем 3, обнаруживалась с вероятностью 0,95?

Решить поставленную задачу, сформулировав ее как задачу проверки статистических гипотез (см. задачу 22.1).

Исследовать, как изменяется функция мощности критерия в зависимости от уровня значимости критерия и числа подбрасываний кости.

Вычислить вероятность того, что несимметричная игральная кость, вероятность выпадения числа очков большего чем 3 которой равна 0,6; 0,7; 0,8; 0,9, будет обнаружена; не будет обнаружена.

Указание. Нулевая гипотеза H_0: игральная кость симметричная, альтернативная гипотеза H_p: $p \in (0{,}05; 1]$. Критическое множество $S = \{k : k > l\}$.

22.8 (о булочках с изюмом). Государственным стандартом установлено, что при выпечке сладких булочек на 1000 изделий должно приходиться 10 000 изюмин (в среднем 10 на одну булочку). У нас, однако, есть сомнение, что весь изюм был израсходован по назначению (его могли, по крайней мере частично, использовать для других целей), и мы хотим проверить, действительно ли это так. Для этого мы покупаем одну булочку и пересчитываем количество изюмин в ней: оказалось, что в булочке $\xi(\omega)$ изюмин. По числу $\xi(\omega)$ изюмин в булочке необходимо сделать заключение, весь ли изюм использован по назначению.

Решить поставленную задачу, сформулировав ее как задачу проверки статистических гипотез (см. задачу 22.1).

Как изменятся выводы, если изюмины пересчитываются в двух булочках?

Указание. Нулевая гипотеза H_0: весь изюм использован по назначению, альтернативная гипотеза H_λ: доля λ изюма разошлась по другим каналам, $\lambda \in (0, 1)$.

22.9 (о качестве инсектицида). Как специалиста с математической подготовкой вас приглашают проанализировать результаты исследовательской работы, связанной с производством инсектицидов. Качество инсектицида определяется “процентом поражения” — процентом насекомых, которые гибнут в популяции, обработанной инсектицидом. Лучший из имеющихся инсектицидов имеет процент поражения 92. В лаборатории создан новый инсектицид INC. Возникает вопрос: является ли этот инсектицид более качественным? Чтобы выяснить, имеются ли различия в качестве инсектицидов, предлагается обработать новым инсектицидом 100 насекомых и подсчитать количество выживших среди них. Если количество выживших насекомых меньше некоторого числа l, то INC качественнее, в противном случае — нет.

Авторы нового инсектицида утверждают, что процент его поражения не менее 98, и хотят, чтобы это свойство нового препарата проявлялось с вероятностью не меньшей чем 0,96. Инсектицид INC дороже имеющихся в продаже препаратов с процентом поражения 92, и покупатель, естественно, отказывается покупать INC, если его процент поражения 92, настаивая, чтобы при тестировании инсектицида INC, процент поражения которого составляет 92, это обнаруживалось с вероятностью не меньшей чем 0,95.

Можно ли, обработав инсектицидом INC 100 насекомых, вынести заключение о его качестве, обеспечив выполнение перечисленных выше условий? Если 100 насекомых недостаточно, то каким минимальным должно быть количество насекомых в обрабатываемой популяции, чтобы можно было ответить на поставленный вопрос?

Дайте ответы на перечисленные вопросы, сформулировав поставленную задачу как задачу проверки статистических гипотез (см. задачу 22.1).

Указание. Нулевая гипотеза H_0: процент поражения инсектицида составляет $\theta = 92$, альтернативная гипотеза H_θ: $\theta \in$ $\in (92; 100]$. Критическое множество $S = \{k : k > l\}$. Требование авторов нового инсектицида:

$$\beta(98) = \mathsf{P}(S|H_{98}) \geq 0{,}96,$$

где $\beta(\theta)$ — функция мощности критерия S.

Глава 23

Проверка гипотез о параметрах нормального распределения

23.1 Проверка гипотезы $a = a_0$ (дисперсия известна)

Чтобы проиллюстрировать идею проверки гипотез о параметрах нормального распределения, не загромождая изложение техническими трудностями, сначала рассмотрим задачу проверки гипотезы H_0: $a = a_0$, когда дисперсия σ^2 известна.

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ из нормального распределения с параметрами $(a; \sigma^2)$. Параметр a неизвестен, дисперсия σ^2 известна, по реализации $\xi(\omega)$ выборки ξ необходимо сделать выводы о его значении. Возможны по меньшей мере два подхода к решению этой задачи: первый — получить оценку неизвестного параметра a (этим мы занимались ранее) и второй — проверить гипотезы о значении параметра a, чем мы и займемся. Относительно значения среднего a нормального распределения выдвигается гипотеза H_0: $a = a_0$ или, что то же, $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ — реализация выборки из $\mathsf{N}_{a_0;\sigma^2}$. Альтернатива H_1 к гипотезе H_0 может быть как односторонняя: если $a \neq a_0$, то $a > a_0$ (она может быть и такой: $a < a_0$), так и двусторонняя: если $a \neq a_0$, то $a > a_0$ или $a < a_0$ (выбор альтернативы определяется рассматриваемой задачей).

Необходимо построить критерий для проверки гипотезы H_0: $a = a_0$.

Выбор статистики для построения критерия. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из $\mathsf{N}_{a;\sigma^2}$. Вне зависимости от того, верна гипотеза H_0: $a = a_0$ или нет, выборочное среднее

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

является состоятельной и несмещенной оценкой параметра a, является "хорошим приближением" параметра a. Последнее обозначает, что *уклонение* $(\overline{\xi} - a)$ *оценки* $\overline{\xi}$ *от параметра* a *малое — лежит в пределах погрешности оценивания*[1] $\sqrt{\mathsf{M}(\overline{\xi} - a)^2} = \sigma/\sqrt{n}$, *т. е. отношение*

$$\varphi_0 = (\overline{\xi} - a)\Big/\frac{\sigma}{\sqrt{n}}$$

является малым. Поэтому, если гипотеза H_0: $a = a_0$ верна, то отношение

$$\varphi = (\overline{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}} = (\overline{\xi} - a)\Big/\frac{\sigma}{\sqrt{n}}$$

малое.

Если же гипотеза H_0: $a = a_0$ неверна (скажем, $a > a_0$), то *уклонение* $\overline{\xi} - a_0 = (\overline{\xi} - a) + (a - a_0)$ *оценки* $\overline{\xi}$ *от гипотетического значения* a_0 *параметра* a *большое — существенно превышает погрешность оценивания* $\sigma/\sqrt{n}$ *параметра* a *оценкой* $\overline{\xi}$, *т. е. отношение*

$$\varphi = (\overline{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}}$$

является большим.

И следовательно, для проверки гипотезы H_0: $a = a_0$ вычисляем отношение

$$\varphi = (\overline{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}},$$

[1] Погрешности при оценивании параметров неизбежны, оценок параметра без погрешностей не бывает. Параметр a величиной $\overline{\xi}$ оценивается с погрешностью $\sqrt{\mathsf{M}(\overline{\xi} - a)^2} = \sigma/\sqrt{n}$. О погрешностях оценивания параметров см. в главе 17, параграф 17.1, пункт "Погрешность оценивания параметра".

— если φ приняло большое значение, то гипотезу H_0 отклоняем, в противном случае — не отклоняем.

Чтобы можно было судить, большое или малое значение приняло отношение
$$\varphi = (\overline{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}},$$
необходимо знать, какие значения принимает отношение φ, когда гипотеза H_0 верна (когда отношение φ является малым) и какие значения принимает отношение φ, когда гипотеза H_0 неверна (когда отношение φ является большим). Другими словами, необходимо знать распределение φ, когда гипотеза H_0 верна и когда она неверна или хотя бы когда гипотеза H_0 верна.

Если гипотеза H_0: $a = a_0$ верна, отношение
$$\varphi = (\overline{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} = (\overline{\xi} - a) \Big/ \frac{\sigma}{\sqrt{n}},$$
имеет нормальное распределение с параметрами $(0; 1)$, поскольку $\overline{\xi}$ имеет нормальное распределение с параметрами $(a, \sigma^2/n)$. Отсюда, в частности, следует, что при верной гипотезе H_0 отношение φ, будучи малым, "почти всегда" — с вероятностью $1 - 2\alpha$ — принадлежат промежутку $(-N_{\alpha;0;1}, N_{\alpha;0;1})$, где $N_{\alpha;0;1}$ — верхний α-предел $\mathsf{N}_{0;1}$-распределения.

Если гипотеза H_0: $a = a_0$ неверна (пусть для определенности $a > a_0$), то отношение
$$\varphi = (\overline{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} = (\overline{\xi} - a) \Big/ \frac{\sigma}{\sqrt{n}} + (a - a_0) \Big/ \frac{\sigma}{\sqrt{n}}$$
имеет нормальное распределение с дисперсией 1 и математическим ожиданием
$$\mathsf{M}\varphi = \mathsf{M}\left((\overline{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} \right) = (a - a_0) \Big/ \frac{\sigma}{\sqrt{n}} = b > 0$$
(см. рис. 23.1.1). Отсюда, в частности, следует, что
$$\mathsf{P}\left\{ b - N_{\alpha;0;1} \le \varphi \le b + N_{\alpha;0;1} \right\} = 1 - 2\alpha$$
и при $n \to \infty$
$$\mathsf{M}\varphi \longrightarrow \infty,$$
$$\varphi = (\overline{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} \xrightarrow{P} \infty.$$

Так что при верной гипотезе H_0: $a = a_0$ отношение

$$\varphi = (\bar{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}},$$

будучи малым, "почти всегда" (с вероятностью $1-2\alpha$) принимает значения из окрестности $(-N_{\alpha;0;1}, N_{\alpha;0;1})$ нуля. Если гипотеза H_0 неверна, то φ с вероятностью $1-2\alpha$ принимает значения из окрестности $(b - N_{\alpha;0;1}, b + N_{\alpha;0;1})$ точки $b = (a - a_0) \Big/ \frac{\sigma}{\sqrt{n}}$ (см. рис. 23.1.1). Поэтому в качестве границ, отделяющих большие значения отношения φ от малых, естественно выбрать числа $-N_{\alpha;0;1}, N_{\alpha;0;1}$. Значения отношения φ, принадлежащие промежутку $(-N_{\alpha;0;1}, N_{\alpha;0;1})$, будем классифицировать как малые, а не принадлежащие — как большие. И для проверки гипотезы H_0: $a = a_0$ вычисляем значение отношения

$$\varphi = (\bar{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}}$$

и выясняем, большое оно или малое, сравнивая φ с границами $-N_{\alpha;0;1}$ и $N_{\alpha;0;1}$, отделяющими малые значения φ от больших. Если φ приняло большое значение, гипотезу H_0 отклоняем, в противном случае — нет.

Критерий для проверки гипотезы $\boldsymbol{H_0}$: $\boldsymbol{a = a_0}$. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из $\mathsf{N}_{a;\sigma^2}$ с известной дисперсией σ^2, $N_{\alpha;0;1}$ — верхний α-предел $\mathsf{N}_{0;1}$-распределения.*

Если гипотезу H_0: $a = a_0$ отклонять при

$$|\varphi| = |\bar{\xi} - a_0| \Big/ \frac{\sigma}{\sqrt{n}} > N_{\alpha;0;1}$$

и не отклонять в противном случае, то с вероятностью 2α гипотеза H_0 будет отклоняться, когда она верна.

Действительно, при верной гипотезе H_0: $a = a_0$ случайная величина

$$\varphi = (\bar{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} = (\bar{\xi} - a) \Big/ \frac{\sigma}{\sqrt{n}}$$

имеет нормальное распределение с параметрами $(0;1)$, поэтому

$$\mathsf{P}\left\{ |\bar{\xi} - a_0| \Big/ \frac{\sigma}{\sqrt{n}} \geq N_{\alpha;0;1} \right\} = 2\alpha.$$

Последнее означает, что при использовании приведенного критерия гипотеза H_0 будет отклоняться, когда она верна, с вероятностью 2α.

Этим критерием мы пользуемся, проверяя гипотезу H_0: $a = a_0$ против двусторонней альтернативы $a < a_0$ или $a > a_0$.

Если альтернатива односторонняя, например, $a > a_0$, то φ сравнивают с $N_{\alpha;0;1}$: при

$$\varphi = (\overline{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}} > N_{\alpha;0;1}$$

гипотезу H_0 отклоняют, в противном случае — нет (уровень значимости этого одностороннего критерия равен α).

Замечание. Критерий для проверки гипотезы H_0: $a = a_0$ против двусторонней альтернативы $a \neq a_0$ как подмножество S пространства $\mathbb{R}^n$ (если $\xi(\omega)$ попадает в S, то H_0 отклоняется, в противном случае — нет) имеет вид:

$$S = \left\{(x_1, x_2, \ldots, x_n) : |\overline{x} - a_0|\Big/\frac{\sigma}{\sqrt{n}} > N_{\alpha;0;1}\right\},$$

где $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, σ^2 — известная дисперсия.

Об ошибках первого и второго рода. Вне зависимости от того, верна гипотеза H_0: $a = a_0$ или нет, отношение

$$\varphi_0 = (\overline{\xi} - a)\Big/\frac{\sigma}{\sqrt{n}}$$

имеет распределение $\mathsf{N}_{0;1}$, и как следствие,

$$\mathsf{P}\left\{|\overline{\xi} - a|\Big/\frac{\sigma}{\sqrt{n}} \leq N_{\alpha;0;1}\right\} = 1 - 2\alpha.$$

Поэтому при верной гипотезе H_0: $a = a_0$ отношение $\varphi = \varphi_0$, “почти всегда” (с вероятностью $1 - 2\alpha$) принимая значения из окрестности $(-N_{\alpha;0;1}; N_{\alpha;0;1})$ точки 0 (см. рис. 23.1.1, слева, соответствующие значения φ обозначены звездочками), может, хотя и изредка (с вероятностью 2α), принять значение вне окрестности $(-N_{\alpha;0;1}; N_{\alpha;0;1})$ (см. рис. 23.1.1, слева, соответствующее значение обозначено точкой). При этом мы гипотезу H_0 отклоним и тем самым совершим ошибку — ошибку первого рода (H_0 отклоняется, хотя она верна).

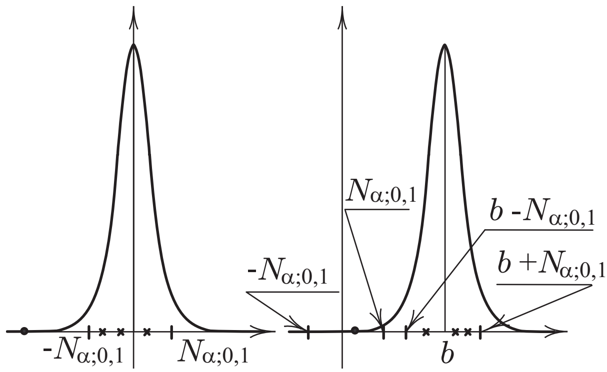

Рис. 23.1.1: Иллюстрация к критерию проверки гипотезы $H_0: a = a_0$, дисперсия известна (единицы по осям различные)

Далее, при неверной гипотезе H_0: $a = a_0$ отношение

$$\varphi = (\overline{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}} = (\overline{\xi} - a)\Big/\frac{\sigma}{\sqrt{n}} + (a - a_0)\Big/\frac{\sigma}{\sqrt{n}} =$$

$$= (\overline{\xi} - a)\Big/\frac{\sigma}{\sqrt{n}} + b,$$

"почти всегда" (с вероятностью $1 - 2\alpha$), принимая значения из промежутка $(b - N_{\alpha;0;1}, b + N_{\alpha;0;1})$ (см. рис. 23.1.1, справа, соответствующие значения обозначены звездочками), может принять значения из промежутка $(-N_{\alpha;0;1}; N_{\alpha;0;1})$ (см. рис. 23.1.1, справа, соответствующее значение обозначено точкой), при этом мы гипотезу H_0 не отклоним и тем самым совершим ошибку — ошибку второго рода (H_0 не отклоняется, хотя она и неверна).

Мощность критерия. Критерий S (критическое множеств) описывается своей функцией мощности

$$\beta(\theta) = \mathsf{P}\,(S\,|H_\theta)\,,\ \theta \in \Theta$$

($\beta(\theta)$ равно вероятности отклонить гипотезу H_0 когда верна гипотеза H_θ, $\theta \in \Theta$).

Критическое множество S для проверки гипотезы H_0: $a = a_0$ о параметре a нормального распределения $\mathsf{N}_{a;\sigma^2}$ (дисперсия σ^2 известна) при двусторонней альтернативе имеет вид:

$$S = \left\{(x_1, x_2, \ldots, x_n) : |\overline{x} - a_0|\Big/\frac{\sigma}{\sqrt{n}} > N_{\alpha;0;1}\right\}.$$

Функция мощности критерия S

$$\beta(a) = \mathsf{P}\,(S|H_a) = \mathsf{P}\,\{(\xi_1, \xi_2, \ldots, \xi_n) \in S|H_a\} =$$

$$= \mathsf{P}\left\{|\overline{\xi} - a_0|\Big/\frac{\sigma}{\sqrt{n}} > N_{\alpha;0;1}|H_a\right\} =$$

$$= \mathsf{P}\left\{|\overline{\xi} - a_0|\Big/\frac{\sigma}{\sqrt{n}} > N_{\alpha;0;1}\right\},\ a \in \mathbb{R}^1,$$

где $\overline{\xi} = \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ вычислена по выборке $\xi_1, \xi_2, \ldots, \xi_n$ из $\mathsf{N}_{a;\sigma^2}$. Подробнее,

$$\beta(a) = 1 - \mathsf{P}\left\{|\overline{\xi} - a_0|\Big/\frac{\sigma}{\sqrt{n}} \le N_{\alpha;0;1}\right\} =$$

$$= 1 - \mathsf{P}\left\{(a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1} \le (\overline{\xi} - a)\Big/\frac{\sigma}{\sqrt{n}} \le\right.$$

$$\left.\le (a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} + N_{\alpha;0;1}\right\} =$$

$$= 1 - \mathsf{N}_{0;1}\left(\left[(a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1};\ (a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} + N_{\alpha;0;1}\right]\right),$$

или

$$\beta(a) = 1 - \left(N_{0;1}\left((a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} + N_{\alpha;0;1}\right) -\right.$$

$$\left.-N_{0;1}\left((a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1}\right)\right).$$

Значение $\mathsf{N}_{0;1}([x, y])$ распределения $\mathsf{N}_{0;1}$ на промежутке $[x, y]$ по функции распределения $N_{0;1}(s)$, $s \in \mathbb{R}^1$, вычисляется так: $\mathsf{N}_{0;1}([x, y]) = N_{0;1}(y) - N_{0;1}(x)$.

Пример (контроль прочности проволоки). *Для стальной проволоки, идущей на изготовление канатов, среднее предельное значение растягивающего усилия должно быть не меньшим 6720 кг/см^2. Серия последовательно проведенных наблюдений показала, что среднее предельное растягивающее усилие, выдерживаемое проволокой, можно считать нормально распределенным, а стандартное уклонение равным 220 кг/см^2.*

С целью контроля величины предельного растягивающего усилия из каждой партии проволоки, поступающей на завод, подвергается испытанию 20 *экземпляров. Результаты одного из таких испытаний приведены ниже:* 6300, 6870, 6720, 6980, 6780, 6780, 6760, 6720, 6630, 6650, 6900, 7130, 6690, 6750, 6560, 6700, 6930, 6720, 6950, 6960.

Можно ли на основании этих данных утверждать, что для данной партии проволоки среднее предельное растягивающее усилие не менее 6720 *кг/см*^{2}*?*

Решить задачу, сформулировав ее в терминах проверки статистических гипотез:

— выбрать нулевую гипотезу и альтернативную;

— выяснить, в чем состоят ошибки первого и второго рода;

— предложить критерий для проверки нулевой гипотезы (*а вместе с ним и уровень значимости*)*;*

— вычислить значения функции мощности критерия в точках $6000 + 50i$, $i = 10, 11, \ldots, 15$*;*

— какова вероятность забраковать партию, для которой среднее предельное растягивающее усилие равно 6500, 6550, 6600*;*

— каким должно быть количество образцов для контроля, чтобы партия со средним предельным растягивающим усилием на разрыв равным 6600 *обнаруживалась с вероятностью* 0,99*;*

— как будет меняться критерий с изменением n; что изменится, если для контроля отбирается 10 *образцов,* 30 *образцов;*

— дать частотную интерпретацию полученным результатам.

Решение. В качестве нулевой (проверяемой) гипотезы рассмотрим гипотезу H_0 — выборка получена из $\mathsf{N}_{a_0;\sigma^2}$, $a_0 = 6720$. Поскольку наша задача — обнаружить проволоку, среднее предельное растягивающее усилие на разрыв которой меньше 6720 кг/см^2, то в качестве альтернативы будем рассматривать одностороннюю альтернативу H_a: $a < a_0$. При таком выборе нулевой и альтернативной гипотез ошибка первого рода будет состоять в том, что проволока со средним предельным растягивающим усилием на разрыв в 6720 кг/см^2 будет браковаться — классифицироваться как проволока со средним предельным растягивающим усилием на разрыв меньшим 6720 кг/см^2. Ошибка второго рода состоит в том, что проволока со средним предельным растягивающим усилием на разрыв, равным a ($a < 6720$), будет классифицироваться как проволока, выдерживающая среднее предельное усилие на разрыв в 6720 кг/см^2 — будет выдерживать контроль.

Для проверки гипотезы H_0: $a = a_0$ ($a_0 = 6720$) против односторонней альтернативы H_1: $a < a_0$ воспользуемся односторонним критерием: если

$$(\bar{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1},$$

то гипотеза H_0 отклоняется, в противном случае — нет (уровень значимости критерия равен α).

В рассматриваемой задаче $n = 20$; $\bar{\xi} = 6775$; $\sigma = 220$. Уровень значимости α выберем равным 0,05, при этом $N_{0,05;0;1} = 1{,}65$ ($N_{0,05;0;1}$ — верхний 0,05-предел $\mathsf{N}_{0;1}$-распределения — найден по табл. 27.1.1). Нормированное уклонение

$$(\bar{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} = (6775 - 6720) \Big/ \frac{220}{\sqrt{20}} = 1{,}13 > -1{,}65 = -N_{0,05;0,1}.$$

Поэтому на 5 % уровне значимости гипотеза H_0 — среднее предельное растягивающее усилие на разрыв равно 6720 кг/см^2, не отклоняется.

Значение функции мощности критерия

$$S = \left\{ (x_1, x_2, \ldots, x_n) : \ (\bar{x} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1} \right\} \qquad (23.1.1)$$

в точке a равно

$$\beta(a) = \mathsf{P}\left\{ (\bar{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1} \,\Big|\, H_a \right\} =$$

$$= \mathsf{P}\left\{ (\bar{\xi} - 6720) \Big/ \frac{220}{\sqrt{20}} < -N_{\alpha;0;1} \right\},$$

где $\bar{\xi} = \frac{1}{n} \sum\limits_{i=1}^{n} \xi_i$ — оценка параметра a, полученная по выборке $\xi_1, \xi_2, \ldots, \xi_n$ из $\mathsf{N}_{a;\sigma^2}$. Поэтому $\beta(a)$ можно переписать так:

$$\beta(a) = \mathsf{P}\left\{ (\bar{\xi} - a_0) \Big/ \frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1} \right\} =$$

$$= \mathsf{P}\left\{ ((\bar{\xi} - a) - (a_0 - a)) \Big/ \frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1} \right\} =$$

$$= \mathsf{P}\left\{(\bar{\xi} - a)\Big/\frac{\sigma}{\sqrt{n}} < (a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1}\right\} =$$

$$= N_{0;1}\left((a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1}\right)$$

или, учитывая, что $a_0 = 6720, \sigma = 220, n = 20$, так

$$\beta(a) = N_{0;1}\left((a_0 - a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1}\right) =$$

$$= N_{0;1}\left((6720 - a)\Big/\frac{220}{\sqrt{20}} - 1{,}65\right)$$

($N_{0;1}(x)$ — значение функции нормального распределения с параметрами $(0; 1)$ в точке x).

Значения $\beta(a)$ функции мощности критерия S в некоторых точках приведены в табл. 23.1.1. График функции мощности критерия (23.1.1) приведен на рис. 23.1.2.

Таблица 23.1.1

a_i	6500	6550	6600	6650	6700	6720	6750
$\beta(a_i)$	0,997	0,965	0,785	0,409	0,108	0,049	0,012

Вероятность забраковать партию (классифицировать как имеющую среднее предельное значение на разрыв меньшее 6720), для которой это значение на самом деле 6500, равна $\beta(a) = \beta(6500) = 0{,}99$, для партии с $a = 6550$ и с $a = 6600$ эти вероятности соответственно равны $\beta(6550) = 0{,}96$ и $\beta(6600) = 0{,}79$.

Количество n образцов, необходимое для обнаружения с вероятностью 0,99 партии со средним предельным значением на разрыв 6600, получаем из соотношения

$$\beta(6600) = N_{0;1}\left((6720 - 6600)\Big/\frac{220}{\sqrt{n}} - 1{,}65\right) = 0{,}99.$$

Из последнего равенства, воспользовавшись таблицей нормального распределения (см. табл. 27.1.1), имеем:

$$120\Big/\frac{220}{\sqrt{n}} - 1{,}65 = 2{,}33.$$

Отсюда получаем, что необходимый объем выборки составляет $n = 54$.

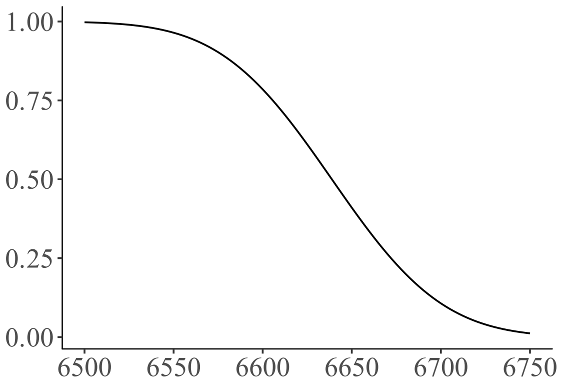

Рис. 23.1.2: Функция мощности критерия (23.1.1)

При изменении n (объема выборки) уровень значимости критерия не меняется, но меняется мощность критерия.

Для $n = 10$ значения функции мощности критерия S в некоторых точках приведены в табл. 23.1.2, а для $n = 30$ — в табл. 23.1.3.

Таблица 23.1.2

a_i	6500	6550	6600	6650	6700	6720	6750
$\beta(a_i)$	0,936	0,788	0,532	0,261	0,087	0,049	0,019

Таблица 23.1.3

a_i	6500	6550	6600	6650	6700	6720	6750
$\beta(a_i)$	0,999	0,996	0,910	0,534	0,125	0,049	0,008

Полученным результатам можно дать следующую частотную интерпретацию (для $n = 20$).

Уровень значимости $\alpha = 0{,}05$ обозначает, что если производить контроль с отбором 20 образцов, то на 100 партий со средним предельным растягивающим усилием на разрыв $a = 6720$ кг/см^2 в среднем пять партий будут браковаться.

Значение $\beta(a) = \beta(6500) = 0{,}99$ функции мощности критерия в точке 6500 интерпретируется так. На 100 партий со средним предельным растягивающим усилием на разрыв $a = 6500$ кг/см^2 в среднем 99 будут браковаться. Аналогично интерпретируются значения $\beta(a)$ в других точках a.

Далее, хотя гипотеза H_0: $a = 6720$ проверяется против альтернативы H_1: $a < 6720$, мы можем вычислять значение функции мощности критерия и в точках $a > 6720$. Например, в точке $a = 6750$ значение функции мощности $\beta(6750) = 0{,}012$ (см. табл. 23.1.1). Последнее обозначает, что при верной гипотезе H_{6750} — среднее предельное усилие на разрыв проволоки равно 6750, гипотеза H_0 с вероятностью 0,012 будет отклоняться, и с вероятностью $1 - \beta(6750) = 0{,}988$ не будет отклоняться, т. е. партия проволоки со средним предельным растягивающим усилием на разрыв равным $a = 6750$ кг/см^2 будет выдерживать контроль с вероятностью 0,988. Так что наш критерий "правильно работает" и при значениях параметра $a > 6720$.

23.2 Проверка гипотезы $a = a_0$ (дисперсия неизвестна)

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ из нормального распределения с параметрами $(a; \sigma^2)$. Параметры a и σ^2 неизвестны. Относительно значения неизвестного нам, но вполне определенного a, выдвигается гипотеза H_0: $a = a_0$, или, что то же, $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ из $\mathsf{N}_{a_0;\sigma^2}$. Альтернатива у гипотезы H_0 может быть как односторонняя: если $a \neq a_0$, то $a > a_0$ (она может быть и такой: $a < a_0$), так и двусторонняя — если $a \neq a_0$, то $a > a_0$ или $a < a_0$. Необходимо построить критерий для проверки этой гипотезы.

Выбор статистики для построения критерия. Из тех же соображений, что и при проверке гипотезы H_0: $a = a_0$, когда σ^2 известна (см. параграф 23.1), уклонение $(\overline{\xi} - a_0)$ оценки $\overline{\xi}$ параметра a от его гипотетического значения a_0 следовало бы сравнивать с погрешностью $\sqrt{\mathsf{M}(\overline{\xi} - a)^2} = \sigma/\sqrt{n}$ оценивания параметра a оценкой $\overline{\xi}$. *Если уклонение* $(\overline{\xi} - a_0)$ *лежит в пределах погрешности оценивания, гипотезу* H_0*:* $a = a_0$ *не от-*

клоняем, если же уклонение $(\overline{\xi} - a_0)$ *существенно превышает погрешность оценивания, гипотезу* H_0 *отклоняем.* Но здесь погрешность оценивания $\sqrt{\mathsf{M}(\overline{\xi} - a)^2} = \sigma/\sqrt{n}$ неизвестна. Поэтому мы будем сравнивать уклонение $(\overline{\xi} - a_0)$ с оценкой $s/\sqrt{n}$ погрешности $\sigma/\sqrt{n}$, используя вместо неизвестной дисперсии σ^2 ее несмещенную состоятельную оценку

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2,$$

— если отношение

$$t = (\overline{\xi} - a_0)\Big/\frac{s}{\sqrt{n}}$$

приняло большое значение, гипотезу H_0 отклоняем, в противном случае — нет.

Чтобы можно было судить, большое или малое значение приняло отношение

$$t = (\overline{\xi} - a_0)\Big/\frac{s}{\sqrt{n}}$$

(и в зависимости от этого отклонять или не отклонять H_0), необходимо знать, какие значения принимает t, когда гипотеза H_0 верна (когда отношение t является малым) и какие значения принимает отношение t, когда гипотеза H_0 неверна (когда отношение t является большим). Другими словами, необходимо знать распределение t, когда гипотеза H_0 верна и когда она неверна.

Если гипотеза H_0 верна, отношение t, которое при этом обозначим через t_{n-1}:

$$t_{n-1} = (\overline{\xi} - a)\Big/\frac{s}{\sqrt{n}},$$

будучи малым, имеет распределение Стьюдента с $(n-1)$ степенями свободы (см. теорему 21.3.1). Отсюда, в частности, следует, что значения отношения t_{n-1} “почти всегда” — с вероятностью $1 - 2\alpha$ — лежат в промежутке $(-t_{\alpha;n-1}, t_{\alpha;n-1})$, где $t_{\alpha;n-1}$ — верхний α-предел распределения Стьюдента с $(n-1)$ степенями свободы, и

$$\mathsf{M}t = \mathsf{M}t_{n-1} = 0.$$

Если же гипотеза H_0: $a = a_0$ неверна (пусть для определенности $a > a_0$), то отношение

$$t = (\bar{\xi} - a_0)\Big/\frac{s}{\sqrt{n}} = (\bar{\xi} - a)\Big/\frac{s}{\sqrt{n}} + (a - a_0)\Big/\frac{s}{\sqrt{n}} =$$

$$= t_{n-1} + (a - a_0)\Big/\frac{s}{\sqrt{n}}$$

имеет так называемое нецентральное распределение Стьюдента. Из равенства

$$t = t_{n-1} + (a - a_0)\Big/\frac{s}{\sqrt{n}}\,,$$

в частности, следует, что

$$\mathsf{P}\left\{(a - a_0)\Big/\frac{s}{\sqrt{n}} - t_{\alpha;n-1} \leq t \leq (a - a_0)\Big/\frac{s}{\sqrt{n}} + t_{\alpha;n-1}\right\} = 1 - 2\alpha$$

и при $n \to \infty$

$$\mathsf{M}t \longrightarrow \infty,$$

$$t = (\bar{\xi} - a_0)\Big/\frac{s}{\sqrt{n}} \xrightarrow{P} \infty.$$

Так что при верной гипотезе H_0: $a = a_0$ отношение

$$t = (\bar{\xi} - a_0)\Big/\frac{s}{\sqrt{n}},$$

"почти всегда" (с вероятностью $1 - 2\alpha$) принимает значения из промежутка $(-t_{\alpha;n-1}, t_{\alpha;n-1})$, если же гипотеза H_0 неверна, то t "почти всегда" (с вероятностью $1 - 2\alpha$) принимает значения из промежутка

$$\left((a - a_0)\Big/\frac{s}{\sqrt{n}} - t_{\alpha;n-1},\ (a - a_0)\Big/\frac{s}{\sqrt{n}} + t_{\alpha;n-1}\right).$$

Поэтому в качестве границ, отделяющих малые значения отношения t от больших значений, естественно выбрать числа $-t_{\alpha;n-1}$, $t_{\alpha;n-1}$. Значения отношения t, принадлежащие промежутку $(-t_{\alpha;n-1}, t_{\alpha;n-1})$, будем классифицировать как малые, а не принадлежащие — как большие. И для проверки гипотезы H_0: $a = a_0$ вычисляем значение отношения

$$t = (\bar{\xi} - a_0)\Big/\frac{s}{\sqrt{n}}$$

и выясняем, большое оно или малое, сравнивая t с границами $-t_{\alpha;n-1}$, $t_{\alpha;n-1}$, отделяющими малые значения отношения t от больших. Если t приняло большое значение, гипотезу H_0: $a = a_0$ отклоняем, в противном случае — нет.

Критерий Стьюдента для проверки гипотезы $\boldsymbol{H_0}$: $\boldsymbol{a = a_0}$. *Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из $\mathsf{N}_{a;\sigma^2}$, $t_{\alpha;n-1}$ — верхний α-предел распределения Стьюдента с $(n-1)$ степенями свободы.*

Если гипотезу $H_0\colon a = a_0$ отклонять при

$$|t| = |\overline{\xi} - a_0| \Big/ \frac{s}{\sqrt{n}} > t_{\alpha;n-1}$$

и не отклонять в противном случае, то с вероятностью 2α гипотеза H_0 будет отклоняться, когда она верна.

Действительно, при верной гипотезе $H_0 : a = a_0$ случайная величина $t = (\overline{\xi} - a_0)/\dfrac{s}{\sqrt{n}}$ имеет распределение Стьюдента с $(n-1)$ степенями свободы, поэтому

$$\mathsf{P}\left\{|\overline{\xi} - a_0| \Big/ \frac{s}{\sqrt{n}} > t_{\alpha;n-1}\right\} = 2\alpha.$$

Последнее равенство означает, что при использовании критерия Стьюдента гипотеза H_0 будет отклоняться, когда она верна, с вероятностью 2α (этим критерием мы пользуемся для проверки гипотезы H_0: $a = a_0$ против двусторонней альтернативы: $a < a_0$ или $a > a_0$).

Если альтернатива односторонняя, например, $a > a_0$, то t сравниваем с $t_{\alpha;n-1}$: при

$$t = (\overline{\xi} - a_0) \Big/ \frac{s}{\sqrt{n}} > t_{\alpha;n-1}$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости этого одностороннего критерия равен α).

З а м е ч а н и е. Критерий для проверки гипотезы H_0: $a = a_0$ как подмножество S выборочного пространства $\mathbb{R}^n$ (если ξ попадает в S, то H_0 отклоняется, в противном случае — нет) имеет вид:

$$S = \left\{(x_1, x_2, \ldots, x_n) : |\overline{x} - a_0| \Big/ \frac{s}{\sqrt{n}} > t_{\alpha;(n-1)}\right\},$$

где

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Об ошибках первого и второго рода. Вне зависимости от того, верна гипотеза $H_0 : a = a_0$ или нет, уклонение $t_{n-1} = (\bar{\xi} - a)\Big/\frac{s}{\sqrt{n}}$ имеет распределение Стьюдента с $n-1$ степенями свободы, и, как следствие,

$$\mathsf{P}\left\{|\overline{\xi} - a|\Big/\frac{s}{\sqrt{n}} \leq t_{\alpha;n-1}\right\} = 1 - 2\alpha,$$

т. е. t_{n-1} “почти всегда” (с вероятностью $1-2\alpha$) принимает значения из окрестности $(-t_{\alpha;n-1}, t_{\alpha;n-1})$ точки 0 (см. рис. 23.2.1). Но изредка (с вероятностью 2α) уклонение $t_{n-1} = (\bar{\xi} - a)\Big/\frac{s}{\sqrt{n}}$ может принимать значения вне окрестности $(-t_{\alpha;n-1}, t_{\alpha;n-1})$. Поэтому при верной гипотезе H_0: $a = a_0$ уклонение

$$t = (\overline{\xi} - a_0)\Big/\frac{s}{\sqrt{n}}$$

может (хотя и с малой вероятностью) принять “большое” значение, т. е. $|t| > t_{\alpha;n-1}$. При этом мы гипотезу H_0 отклоним и тем самым совершим ошибку — ошибку первого рода (H_0 отклоняется, хотя она верна).

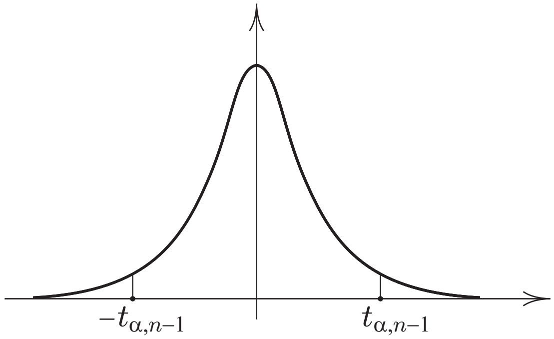

Рис. 23.2.1: Иллюстрация к критерию Стьюдента для проверки гипотезы H_0: $a = a_0$

Далее, при неверной гипотезе H_0: $a = a_0$ уклонение

$$t = (\overline{\xi} - a_0)\Big/\frac{s}{\sqrt{n}} = (\overline{\xi} - a)\Big/\frac{s}{\sqrt{n}} + (a - a_0)\Big/\frac{s}{\sqrt{n}} =$$

$$= t_{n-1} + (a - a_0)\Big/\frac{s}{\sqrt{n}},$$

почти всегда принимая “большие” значения, может принять значение из промежутка $(-t_{\alpha;n-1}, t_{\alpha;n-1})$, при этом мы гипотезу H_0 не отклоним и тем самым совершим ошибку — ошибку второго рода (H_0 не отклонили, хотя она и неверна).

Пример 23.2.1 (эффект использования специальной сеялки). *Для исследования эффекта использования специальной сеялки* 10 *участков земли засеяли при помощи обыкновенной сеялки и* 10 *— специальной, а затем сравнили полученные урожаи зерна.* 20 *участков одинаковой площади были разделены на пары, причем в каждую пару входили смежные участки. Вопрос о том, какой из двух смежных участков должен был обрабатываться специальной машиной, решался подбрасыванием монеты. В таблице приведены разности урожаев с пар смежных участков, засеянных специальной сеялкой и обычной.*

Номер пары	Разность урожаев	Номер пары	Разность урожаев
1	2,4	6	1,6
2	1,0	7	−0,4
3	0,7	8	1,1
4	0,0	9	0,1
5	1,1	10	0,7

Подтверждают ли приведенные данные наличие эффекта использования специальной сеялки, другими словами, дает ли ее использование прибавку урожая?

Решение. В терминах проверки статистических гипотез задачу можно сформулировать следующим образом. Имеем 10 независимых наблюдений случайной величины (см. таблицу) — реализацию $\xi_1(\omega), \xi_2(\omega), \ldots, \xi_{10}(\omega)$ выборки $\xi_1, \xi_2, \ldots, \xi_{10}$ из нормального распределения $\mathsf{N}_{a;\sigma^2}$ (предположение относительно нормального распределения результатов измерений в большинстве случаев оправдывается), параметры a и σ^2 этого распределения неизвестны. Относительно параметра a распределения $\mathsf{N}_{a;\sigma^2}$ выдвигается гипотеза H_0: $a = 0$ об отсутствии эффекта использования специальной сеялки (обидная для разработчиков новой сеялки). Альтернатива односторонняя: $a > 0$ (специальная сеялка конструировалась для повышения урожайности). Отклонение гипотезы H_0: $a = 0$ в пользу альтернативы $a > 0$ будем трактовать как наличие эффекта использования специальной сеялки, неотклонение H_0: $a = 0$ — как отсутствие эффекта.

Необходимо проверить гипотезу H_0. В соответствии с критерием Стьюдента для проверки гипотезы H_0: $a = a_0$ против альтернативы $a > a_0$ вычисляем значение

$$t = (\overline{\xi} - a_0) \Big/ \frac{s}{\sqrt{n}}$$

и сравниваем его с $t_{\alpha;n-1}$ — верхним α-пределом t_{n-1}-распределения. Если

$$t = (\overline{\xi} - a_0) \Big/ \frac{s}{\sqrt{n}} > t_{\alpha;n-1},$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости критерия равен α).

В рассматриваемом примере $n = 10$,

$$\overline{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i = \frac{1}{10}(2{,}4 + 1{,}0 + \ldots + 0{,}7) = 0{,}83;$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\xi_i - \overline{\xi})^2 = \frac{1}{9}((2{,}4 - 0{,}83)^2 +$$

$$+(1{,}0 - 0{,}83)^2 + \ldots + (0{,}7 - 0{,}83)^2) = 0{,}667;$$

$$t = \overline{\xi} \Big/ \sqrt{\frac{s^2}{n}} = 0{,}83 \Big/ \sqrt{\frac{0{,}667}{10}} = 0{,}83/0{,}258 = 3{,}22.$$

Таким образом,

$$t = \overline{\xi} \Big/ \frac{s}{\sqrt{n}} = 3{,}22 > 2{,}26 = t_{0{,}025;9}.$$

Поэтому согласно критерию Стьюдента гипотеза H_0: $a = 0$ отклоняется в пользу альтернативы $a > 0$. Другими словами, гипотеза о том, что выборка получена из нормального распределения со средним 0, противоречит имеющимся данным (см. таблицу).

Полученный результат можно интерпретировать следующим образом. Предположение, что разность урожаев с участков, засеянных при помощи специальной сеялки и обычной, несущественно отклоняется от нуля, противоречит имеющимся данным.

Для сеялки, использование которой не дает эффекта, такие отклонения разности урожаев от нуля, как приведены в таблице, невозможны (точнее, возможны, но крайне редко). Следовательно, эксперимент дает основания утверждать, что эффект использования специальной сеялки существует (к радости её разработчиков).

23.3 Сравнение средних двух выборок

Прежде чем формулировать задачу проверки гипотезы о равенстве двух средних и строить критерий для ее проверки, докажем теорему, которая понадобится при построении критерия.

Теорема 23.3.1. *Пусть* $\xi_1, \xi_2, \ldots, \xi_n$ *и* $\eta_1, \eta_2, \ldots, \eta_m$ *— независимые выборки из распределения* $\mathsf{N}_{a;\sigma^2}$*. Тогда отношение*

$$t_{n+m-2} = (\overline{\xi} - \overline{\eta}) \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

имеет распределение Стьюдента с $(n+m-2)$ *степенями свободы, где*

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i, \ \overline{\eta} = \frac{1}{m}\sum_{j=1}^{m} \eta_j,$$

$$s_\xi^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2, \ s_\eta^2 = \frac{1}{m-1}\sum_{j=1}^{m}(\eta_j - \overline{\eta})^2,$$

$$s^2 = \frac{1}{n+m-2}\left((n-1)s_\xi^2 + (m-1)s_\eta^2\right). \tag{23.3.1}$$

Доказательство. Аналогично тому, как это делалось в теореме 21.3.1 о распределении случайного вектора $(\overline{\xi}, s^2)$, выразим $\overline{\xi}, \overline{\eta}, s_\xi^2, s_\eta^2, s^2$ через независимые $\mathsf{N}_{0;1}$-распределенные случайные величины

$$\xi_i' = \frac{\xi_i - a}{\sigma}, \ \eta_j' = \frac{\eta_j - a}{\sigma}, \ i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, m:$$

$$\overline{\xi} = \sigma\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i - a}{\sigma} + a = \sigma\overline{\xi'} + a, \ \overline{\eta} = \sigma\frac{1}{m}\sum_{j=1}^{m}\frac{\eta_j - a}{\sigma} + a = \sigma\overline{\eta'} + a,$$

$$s_\xi^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2 = \sigma^2\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i' - \overline{\xi'})^2 =$$
$$= \sigma^2\frac{1}{n-1}\left(\sum_{i=1}^{n}(\xi_i')^2 - n(\overline{\xi'})^2\right),$$

$$s_\eta^2 = \frac{1}{m-1}\sum_{j=1}^{m}(\eta_j - \overline{\eta})^2 = \sigma^2\frac{1}{m-1}\sum_{j=1}^{m}(\eta_j' - \overline{\eta'})^2 =$$

$$= \sigma^2\frac{1}{m-1}\left(\sum_{j=1}^{m}(\eta_i')^2 - m(\overline{\eta'})^2\right),$$
$$s^2 = \frac{1}{n+m-2}\left((n-1)s_\xi^2 + (m-1)s_\eta^2\right) =$$
$$= \sigma^2\frac{1}{n+m-2}\left(\sum_{i=1}^{n}(\xi_i')^2 + \sum_{j=1}^{m}(\eta_j')^2 - n(\overline{\xi'})^2 - m(\overline{\eta'})^2\right). \tag{23.3.2}$$

Преобразуем стандартный гауссовский вектор

$$\alpha = (\xi_1', \xi_2', \ldots, \xi_n', \eta_1', \eta_2', \ldots, \eta_m')'$$

ортогональным преобразованием из $\mathbb{R}^{n+m}$ в $\mathbb{R}^{n+m}$ с матрицей C специального вида:

$$c_{1j} = 1/\sqrt{n},\ j = 1, 2, \ldots, n;$$

$$c_{ij} = 0,\ i = 1, 2, \ldots, n; j = n+1, n+2, \ldots, n+m;$$

$$c_{n+1,j} = 1/\sqrt{m},\ j = n+1, n+2, \ldots, n+m,$$

$$c_{i,j} = 0,\ i = n+1, n+2, \ldots, n+m; j = 1, 2, \ldots, n;$$

остальные c_{ij} выбраны так, чтобы выполнялись условия ортогональности C ($C'C = CC' = I$). Вектор

$$\zeta = C\alpha \tag{23.3.3}$$

является стандартным гауссовским вектором.

Выпишем представления для случайных величин $\overline{\xi} - \overline{\eta}$ и s^2 через компоненты вектора $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_n, \zeta_{n+1}, \ldots, \zeta_{n+m})'$. Учитывая вид 1-й и $(n+1)$-й строк матрицы C, имеем:

$$\zeta_1 = \sqrt{n}\,\overline{\xi'},\ \zeta_{n+1} = \sqrt{m}\,\overline{\eta'},$$

и, в частности,

$$\overline{\xi} - \overline{\eta} = \sigma\overline{\xi'} + a - (\sigma\overline{\eta'} + a) = \sigma(\overline{\xi'} - \overline{\eta'}) = \sigma\left(\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_{n+1}}{\sqrt{m}}\right). \quad (23.3.4)$$

Далее, поскольку C — ортогональное преобразование, то

$$(\zeta, \zeta) = (C\alpha, C\alpha) = (\alpha, \alpha),$$

подробнее

$$\sum_{i=1}^{n+m} \zeta_i^2 = \sum_{i=1}^{n} (\xi_i')^2 + \sum_{j=1}^{m} (\eta_j')^2,$$

и, учитывая еще, что $\zeta_1 = \sqrt{n}\,\overline{\xi'}$, $\zeta_{n+1} = \sqrt{m}\,\overline{\eta'}$, для s^2 (см. (23.3.2)) получаем:

$$s^2 = \sigma^2 \frac{1}{n+m-2}\left(\sum_{i=1}^{n} (\xi_i')^2 + \sum_{j=1}^{m} (\eta_j')^2 - n(\overline{\xi'})^2 - m(\overline{\eta'})^2\right) =$$

$$= \sigma^2 \frac{1}{n+m-2}\left(\sum_{j=1}^{n+m} \zeta_j^2 - \zeta_1^2 - \zeta_{n+1}^2\right). \quad (23.3.5)$$

Отношение

$$(\overline{\xi} - \overline{\eta}) \bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right),$$

учитывая равенства (23.3.4) и (23.3.5), запишем в виде

$$(\overline{\xi} - \overline{\eta}) \bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right) = \frac{\sigma\left(\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_{n+1}}{\sqrt{m}}\right)\sqrt{\frac{nm}{n+m}}}{\sigma\sqrt{s^2/\sigma^2}} =$$

$$= \frac{\left(\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_{n+1}}{\sqrt{m}}\right)\sqrt{\frac{nm}{n+m}}}{\sqrt{\frac{1}{n+m-2}\left(\sum\limits_{i=1}^{n+m}\zeta_i^2 - \zeta_1^2 - \zeta_{n+1}^2\right)}}. \tag{23.3.6}$$

Числитель и знаменатель последней дроби являются независимыми случайными величинами (как функции от независимых случайных величин). При этом числитель, как сумма независимых нормально распределенных случайных величин, имеет нормальное распределение, его параметры:

$$\mathsf{M}\left(\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_{n+1}}{\sqrt{m}}\right)\sqrt{\frac{nm}{n+m}} = 0,$$

$$\mathsf{D}\left(\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_{n+1}}{\sqrt{m}}\right)\sqrt{\frac{nm}{n+m}} = \frac{nm}{n+m}\mathsf{D}\left(\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_{n+1}}{\sqrt{m}}\right) =$$
$$= \frac{nm}{n+m}\left(\frac{1}{n} + \frac{1}{m}\right) = 1$$

(мы учли, что ζ_1 и ζ_{n+1} как компоненты стандартного гауссовского вектора распределены $\mathsf{N}_{0;1}$). Случайная величина

$$\frac{1}{n+m-2}\left(\sum_{i=1}^{n+m}\zeta_i^2 - \zeta_1^2 - \zeta_{n+1}^2\right)$$

в знаменателе дроби (23.3.6) распределена как $\frac{1}{n+m-2}\,\chi^2_{n+m-2}$, поскольку ζ — стандартный гауссовский вектор. Поэтому отношение

$$(\overline{\xi} - \overline{\eta}) \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right) = \frac{\left(\frac{\zeta_1}{\sqrt{n}} - \frac{\zeta_{n+1}}{\sqrt{m}}\right)\sqrt{\frac{nm}{n+m}}}{\sqrt{\frac{1}{n+m-2}\left(\sum\limits_{i=1}^{n+m}\zeta_i^2 - \zeta_1^2 - \zeta_{n+1}^2\right)}}$$

имеет распределение Стьюдента с $(n+m-2)$ степенями свободы.

Тем самым теорема доказана.

Постановка задачи сравнения средних двух выборок. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ — реализация выборки

$\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ из распределения $\mathsf{N}_{a_\xi;\sigma^2}$ и $\eta(\omega) = (\eta_1(\omega), \eta_2(\omega), \ldots, \eta_m(\omega))$ — реализация выборки $\eta = (\eta_1, \eta_2, \ldots, \eta_m)$ из распределения $\mathsf{N}_{a_\eta;\sigma^2}$, выборки $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ независимы. Средние a_ξ, a_η и дисперсия σ^2 (одна и та же для обеих выборок) неизвестны.

Относительно значений параметров a_ξ и a_η выдвигается гипотеза H_0: $a_\xi - a_\eta = 0$, или, что то же, выборки ξ и η получены из одного и того же нормального распределения. Альтернатива к гипотезе может быть как односторонней, так и двусторонней. Необходимо построить критерий для проверки гипотезы H_0.

Выбор статистики для построения критерия. Вне зависимости от того, верна гипотеза H_0: $a_\xi - a_\eta = 0$ или нет, $\overline{\xi}$ и $\overline{\eta}$ — состоятельные и несмещенные оценки соответственно a_ξ и a_η, и, следовательно, $(\overline{\xi} - \overline{\eta})$ — состоятельная и несмещенная оценка разности $a_\xi - a_\eta$, другими словами, оценка $(\overline{\xi} - \overline{\eta})$ является хорошим приближением разности $(a_\xi - a_\eta)$. Из последнего следует, что при верной гипотезе H_0: $a_\xi - a_\eta = 0$ разность $(\overline{\xi} - \overline{\eta}) - (a_\xi - a_\eta) = \overline{\xi} - \overline{\eta}$ принимает малые значения — лежит в пределах погрешности $\sqrt{\mathsf{M}((\overline{\xi} - \overline{\eta}) - (a_\xi - a_\eta))^2} = \sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}$ оценивания параметра $(a_\xi - a_\eta)$ оценкой $(\overline{\xi} - \overline{\eta})$. Если же гипотеза H_0: $a_\xi - a_\eta = 0$ неверна, разность $(\overline{\xi} - \overline{\eta})$, будучи близкой к $a_\xi - a_\eta$, существенно отличается от нуля — превышает погрешность оценивания. Поэтому для проверки гипотезы H_0: $a_\xi - a_\eta = 0$ следует вычислить разность $(\overline{\xi} - \overline{\eta})$ и сравнить ее с погрешностью оценивания $\sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}$: если $\overline{\xi} - \overline{\eta}$ лежит в пределах погрешности оценивания — отношение $(\overline{\xi} - \overline{\eta}) \Big/ \sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}$ мало — гипотезу H_0 не отклоняем, в противном случае H_0 отклоняем. А поскольку σ^2 неизвестна, то $\overline{\xi} - \overline{\eta}$ будем сравнивать с оценкой $s\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}$ погрешности $\sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}$, используя вместо σ^2 ее несмещенную состоятельную оценку

$$s^2 = \frac{1}{n + m - 2}\left((n-1)s_\xi^2 + (m-1)s_\eta^2\right),$$

— если отношение

$$t = (\overline{\xi} - \overline{\eta}) \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

приняло большое значение, гипотезу H_0 отклоняем, в противном случае — нет.

Чтобы можно было судить, большое или малое значение приняло отношение

$$t = (\overline{\xi} - \overline{\eta}) \Big/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

(и в зависимости от этого отклонять или не отклонять H_0), необходимо знать, какие значения оно принимает, когда гипотеза H_0 верна (когда отношение t принимает малые значения), и когда H_0 неверна. Другими словами, необходимо знать распределение отношения t, когда гипотеза H_0 верна и когда она неверна.

При верной гипотезе H_0 отношение t, которое в этом случае будем обозначать через t_{n+m-2}:

$$t_{n+m-2} = (\overline{\xi} - \overline{\eta}) \Big/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right),$$

будучи малым, имеет распределение Стьюдента с $n+m-2$ степенями свободы (см. теорему 23.3.1). Отсюда, в частности, следует, что t_{n+m-2} “почти всегда” (с вероятностью $1 - 2\alpha$) принадлежит промежутку $(-t_{\alpha;n+m-2}, t_{\alpha;n+m-2})$, где $t_{\alpha;n+m-2}$ — верхний α-предел t-распределения с $n + m - 2$ степенями свободы, и

$$\mathsf{M}t = \mathsf{M}t_{n+m-2} = 0.$$

Если гипотеза H_0: $a_\xi - a_\eta = 0$ неверна (пусть для определенности $a_\xi > a_\eta$), то отношение

$$t = (\overline{\xi} - \overline{\eta}) \Big/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right) =$$

$$= ((\overline{\xi} - a_\xi) - (\overline{\eta} - a_\eta)) \Big/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right) + (a_\xi - a_\eta) \Big/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right) =$$

$$= t_{n+m-2} + (a_\xi - a_\eta) \Big/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right) \qquad (23.3.1)$$

имеет нецентральное распределение Стьюдента. Из представления (23.3.1), в частности, следует, что

$$\mathsf{P}\left\{t(a_\xi, a_\eta) - t_{\alpha;n+m-2} \le t \le t(a_\xi, a_\eta) + t_{\alpha;n+m-2}\right\} = 1 - 2\alpha,$$

где

$$t(a_\xi, a_\eta) = (a_\xi - a_\eta) \bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right),$$

и при $n, m \to \infty$

$$\mathsf{M}t \longrightarrow \infty,$$

$$t = (\overline{\xi} - \overline{\eta}) \bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right) \xrightarrow{P} \infty.$$

Так что при верной гипотезе H_0: $a_\xi = a_\eta$ отношение

$$t = (\overline{\xi} - \overline{\eta}) \bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right),$$

будучи малым, “почти всегда” (с вероятностью $1-2\alpha$) принимает значения из промежутка $(-t_{\alpha;n+m-2}, t_{\alpha;n+m-2})$, если же гипотеза H_0 неверна, то t “почти всегда” (с вероятностью $1 - 2\alpha$) принимает значения из промежутка

$$(t(a_\xi, a_\eta) - t_{\alpha;n+m-2},\ t(a_\xi, a_\eta) + t_{\alpha;n+m-2}).$$

Поэтому в качестве границ, отделяющих большие значения отношения t от малых, естественно выбрать числа $-t_{\alpha;n+m-2}$ и $t_{\alpha;n+m-2}$. Значения уклонения t, принадлежащие промежутку $(-t_{\alpha;n+m-2}, t_{\alpha;n+m-2})$, будем классифицировать как малые, в противном случае — как большие. И для проверки гипотезы H_0: $a_\xi - a_\eta = 0$ вычисляем значение отношения

$$t = (\overline{\xi} - \overline{\eta}) \bigg/ \left(s\sqrt{\frac{n+m}{nm}}\right)$$

и выясняем, большое оно или малое, сравнивая t с границами $-t_{\alpha;(n+m-2)}$ и $t_{\alpha;(n+m-2)}$, отделяющими большие значения отношения t от малых. Если t приняло большое значение, гипотезу отклоняем, в противном случае — нет.

Критерий Стьюдента для проверки гипотезы $\boldsymbol{H_0}$: $\boldsymbol{a_\xi - a_\eta = 0}$. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ — независимые выборки соответственно из нормальных распределений $\mathsf{N}_{a_\xi;\sigma^2}$ и $\mathsf{N}_{a_\eta;\sigma^2}$; $t_{\alpha;n+m-2}$ — верхний α-предел распределения Стьюдента с $(n+m-2)$ степенями свободы. Если гипотезу H_0: $a_\xi - a_\eta = 0$ отклонять при*

$$|t| = |\overline{\xi} - \overline{\eta}| \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right) \geq t_{\alpha;n+m-2}$$

и не отклонять в противном случае, то с вероятностью 2α гипотеза H_0 будет отклоняться, когда она верна.

Действительно, если гипотеза $H_0: a_\xi - a_\eta = 0$ верна, уклонение

$$t = (\overline{\xi} - \overline{\eta}) \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

имеет распределение Стьюдента с $(n+m-2)$ степенями свободы, поэтому

$$\mathsf{P}\left\{ |\overline{\xi} - \overline{\eta}| \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right) \geq t_{\alpha;n+m-2} \right\} = 2\alpha.$$

Последнее равенство означает, что при использовании критерия Стьюдента гипотезу H_0 будем отклонять, когда она верна, с вероятностью 2α.

Этим критерием мы пользуемся для проверки гипотезы H_0: $a_\xi - a_\eta = 0$ против двусторонней альтернативы: $a_\xi - a_\eta < 0$ или $a_\xi - a_\eta > 0$. Если альтернатива односторонняя, например, $a_\xi - a_\eta > 0$, то с $t_{\alpha;n+m-2}$ сравниваем t: при

$$t \geq t_{\alpha;n+m-2}$$

гипотезу H_0 отклоняем, в противном случае не отклоняем (уровень значимости этого критерия равен α).

Пример 23.3.1 (сравнение прочности бетона). *Чтобы проверить, влияет ли на прочность бетона особый способ его приготовления, предположительно повышающий прочность бетона, был проведен эксперимент. Из данной партии сырья были взяты шесть порций, которые были случайным образом разделены на две группы по три порции каждая, и из* *каждой был сделан пробный куб, причем кубы, изготовленные из порций сырья второй группы, подверглись особой обработке. После* 28*-дневной выдержки шести пробных кубов определили их сопротивление на сжатие. Приведены значения нагрузки, при которых начинается разрушение образцов — коротко будем говорить "предел прочности". Получили следующие результаты опыта (бетон* 1 *— стандартного приготовления, бетон* 2 *— особого приготовления):*

Бетон 1	290	311	284
Бетон 2	309	318	318

.

Свидетельствуют ли эти данные о наличии эффекта специальной обработки бетона?

Р е ш е н и е. В терминах проверки статистических гипотез эту задачу можно сформулировать так. Имеются реализации $\xi_1(\omega), \xi_2(\omega), \xi_3(\omega)$ и $\eta_1(\omega), \eta_2(\omega), \eta_3(\omega)$ (см. таблицу) двух независимых выборок соответственно из нормальных распределений $\mathsf{N}_{a_\xi;\sigma^2}$ и $\mathsf{N}_{a_\eta;\sigma^2}$ (предположение о нормальности результатов измерений, как правило, согласуется с опытом); ξ_1, ξ_2, ξ_3 — выборка предела прочности бетона стандартного приготовления (бетона 1), η_1, η_2, η_3 — бетона особого приготовления (бетона 2). Параметры (a_ξ, σ^2) и (a_η, σ^2) нормальных распределений $\mathsf{N}_{a_\xi;\sigma^2}$ и $\mathsf{N}_{a_\eta;\sigma^2}$ неизвестны. Относительно параметров a_ξ и a_η выдвигается гипотеза H_0: $a_\xi = a_\eta$ об отсутствии эффекта обработки (обидная для авторов специальной обработки бетона), другими словами, наши данные (см. таблицу) — реализации выборок из одного и того же нормального распределения. По постановке задачи в качестве гипотезы, альтернативной к гипотезе H_0, следует рассматривать одностороннюю альтернативу $a_\eta > a_\xi$. Отклонение гипотезы H_0 в пользу этой альтернативы интерпретируется как наличие эффекта обработки, неотклонение — как отсутствие.

Согласно критерию Стьюдента для проверки гипотезы H_0: $a_\xi = a_\eta$ при альтернативе $a_\eta > a_\xi$ необходимо значение

$$t = (\bar{\eta} - \bar{\xi}) \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

сравнить с верхним α-пределом $t_{\alpha;(n+m-2)}$ распределения Стьюдента с $(n+m-2)$ степенями свободы. Если при этом

$$t > t_{\alpha;(n+m-2)},$$

то гипотеза H_0 отклоняется, в противном случае — нет (уровень значимости этого критерия равен α).

В рассматриваемой задаче $n = 3$, $m = 3$.

$$\bar{\xi} = \frac{1}{3}(290 + 311 + 284) = 295, \ \bar{\eta} = \frac{1}{3}(309 + 318 + 318) = 315,$$

$$s_\xi^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2, \ s_\eta^2 = \frac{1}{m-1}\sum_{j=1}^{m}(\eta_j - \bar{\eta})^2,$$

$$s^2 = \frac{1}{n+m-2}\left((n-1)s_\xi^2 + (m-1)s_\eta^2\right) = 114,$$

$$t = (\bar{\eta} - \bar{\xi}) \Bigg/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right) = (315 - 295) \Bigg/ \left(\sqrt{\frac{114}{1{,}5}}\right) = 2{,}29.$$

Для $\alpha = 0{,}05$ значение $t_{\alpha;(n+m-2)} = t_{0{,}05;4} = 2{,}13$ (найдено по таблице распределения Стьюдента (см. табл. 27.3.1)). Значение

$$t = 2{,}29 > 2{,}13 = t_{0{,}05;4},$$

поэтому гипотеза H_0: $a_\xi = a_\eta$ отклоняется — противоречит опытным данным.

Полученный результат можно интерпретировать так. Данные по измерению предела прочности бетона на сжатие (см. таблицу) дают основание считать, что особый способ приготовления бетона повышает его прочность.

23.4 Проверка гипотезы $\sigma^2 = \sigma_0^2$

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ из распределения $\mathsf{N}_{a;\sigma^2}$. Параметры a и σ^2 неизвестны. Относительно значения параметра σ^2 выдвигается гипотеза H_0: $\sigma^2 = \sigma_0^2$. Эту гипотезу можно сформулировать так: погрешность σ^2 измерений (наблюдений), представленных выборкой $\xi_1, \xi_2, \ldots, \xi_n$, равна данному значению σ_0^2. Мы будем записывать H_0 в виде H_0: $\sigma^2/\sigma_0^2 = 1$. Альтернатива к гипотезе H_0 может быть как односторонней: если $\sigma^2/\sigma_0^2 \neq 1$, то $\sigma^2/\sigma_0^2 > 1$ (она может быть и такой: $\sigma^2/\sigma_0^2 < 1$), так и двусторонней: если $\sigma^2/\sigma_0^2 \neq 1$, то $\sigma^2/\sigma_0^2 > 1$ или $\sigma^2/\sigma_0^2 < 1$.

Альтернатива определяется решаемой задачей. Отклонение гипотезы $\sigma^2/\sigma_0^2 = 1$ в пользу альтернативы $\sigma^2/\sigma_0^2 > 1$ можно интерпретировать как превышение погрешности измерений, представленных выборкой $\xi_1, \xi_2, \ldots, \xi_n$, данного значения σ_0^2. По реализации $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ выборки $\xi = (\xi_1, \xi_2, \ldots$ $\ldots, \xi_n)$ необходимо вынести заключение о гипотезе H_0: $\sigma^2/\sigma_0^2 = 1$ — отклонить H_0 или не отклонить.

Выбор статистики для построения критерия. Вне зависимости от того, верна гипотеза H_0: $\sigma^2/\sigma_0^2 = 1$ или нет, оценка $s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(\xi_i - \overline{\xi})^2$, полученная по выборке $\xi_1, \xi_2, \ldots, \xi_n$ из $\mathsf{N}_{a;\sigma^2}$, является несмещенной и состоятельной оценкой параметра σ^2, т. е. $\mathsf{M}s^2 = \sigma^2$ и $s^2 \xrightarrow{P} \sigma^2$ при $n \to \infty$. Отсюда

$$\mathsf{M}\frac{s^2}{\sigma^2} = 1 \text{ и } \frac{s^2}{\sigma^2} \xrightarrow{P} 1.$$

Последнее обозначает, что отношение s^2/σ^2 мало уклоняется от 1. Поэтому, если гипотеза H_0 $\sigma^2 = \sigma_0^2$ верна, отношение $s^2/\sigma_0^2 = s^2/\sigma^2$ мало уклоняется от 1. Если же гипотеза H_0: $\sigma^2/\sigma_0^2 = 1$ неверна, например, $\sigma^2/\sigma_0^2 > 1$, то

$$\mathsf{M}\frac{s^2}{\sigma_0^2} = \frac{\sigma^2}{\sigma_0^2} > 1 \text{ и } \frac{s^2}{\sigma_0^2} \xrightarrow{P} \frac{\sigma^2}{\sigma_0^2} > 1,$$

т. е. отношение s^2/σ_0^2 отличается от 1 существенно. И следовательно, для проверки гипотезы H_0: $\sigma^2/\sigma_0^2 = 1$ вычисляем отношение s^2/σ_0^2 и в зависимости от того, приняло оно значение, существенно уклоняющееся от 1, или нет, отклоняем гипотезу H_0 или не отклоняем. Чтобы так можно было поступать, необходимо знать, какие значения принимает отношение s^2/σ_0^2 (необходимо знать распределение отношения s^2/σ_0^2), когда гипотеза H_0 верна и когда H_0 неверна.

Представим отношение s^2/σ_0^2 в виде

$$\frac{s^2}{\sigma_0^2} = \frac{\sigma^2}{\sigma_0^2} \cdot \frac{s^2}{\sigma^2} = \frac{\sigma^2}{\sigma_0^2} \cdot \frac{1}{n-1}\chi^2_{n-1},$$

χ^2_{n-1} — случайная величина, имеющая χ^2-распределение с $(n-1)$ степенями свободы. Отсюда следует, что при верной гипотезе

H_0: $\sigma^2/\sigma_0^2 = 1$ отношение s^2/σ_0^2 распределено как случайная величина $\frac{1}{n-1}\chi^2_{n-1}$ и в, частности, значение s^2/σ_0^2 "почти всегда" (с вероятностью $1-2\alpha$) лежит в окрестности

$$\left(\frac{1}{n-1}\,\chi^2_{1-\alpha;n-1};\,\frac{1}{n-1}\,\chi^2_{\alpha;n-1}\right)$$

точки 1 (мало уклоняются от 1), где $\chi^2_{\alpha;n-1}$ и $\chi^2_{1-\alpha;n-1}$ — соответственно верхний α-предел и верхний $(1-\alpha)$-предел χ^2_{n-1}-распределения.

Если гипотеза H_0: $\sigma^2/\sigma_0^2 = 1$ неверна (пусть для определенности $\sigma^2/\sigma_0^2 > 1$), то из равенства

$$\frac{s^2}{\sigma_0^2} = \frac{\sigma^2}{\sigma_0^2}\frac{s^2}{\sigma^2}$$

имеем, что отношение s^2/σ_0^2 распределено как случайная величина $\frac{\sigma^2}{\sigma_0^2}\frac{1}{n-1}\,\chi^2_{n-1}$ и, в частности, значения s^2/σ_0^2 "почти всегда" (с вероятностью $1-2\alpha$) лежат в окрестности

$$\left(\frac{\sigma^2}{\sigma_0^2}\frac{1}{n-1}\,\chi^2_{1-\alpha;n-1};\,\frac{\sigma^2}{\sigma_0^2}\frac{1}{n-1}\,\chi^2_{\alpha;n-1}\right)$$

точки σ^2/σ_0^2.

Так что при верной гипотезе H_0: $\sigma^2/\sigma_0^2 = 1$ отношение s^2/σ_0^2 "почти всегда" (с вероятностью $1-2\alpha$) принимает значения из окрестности

$$\left(\frac{1}{n-1}\,\chi^2_{1-\alpha;n-1};\,\frac{1}{n-1}\,\chi^2_{\alpha;n-1}\right)$$

точки 1 (мало уклоняется от 1), если же гипотеза H_0 неверна, отношение s^2/σ_0^2 "почти всегда" принимает значения из окрестности

$$\left(\frac{\sigma^2}{\sigma_0^2}\frac{1}{n-1}\,\chi^2_{1-\alpha;n-1};\,\frac{\sigma^2}{\sigma_0^2}\frac{1}{n-1}\,\chi^2_{\alpha;n-1}\right)$$

точки σ^2/σ_0^2 (существенно уклоняется от 1). Следовательно, для проверки гипотезы H_0: $\sigma^2/\sigma_0^2 = 1$ вычисляем значение отношения s^2/σ_0^2 и выясняем, насколько оно отличается от 1. Если s^2/σ_0^2

оказалось лежащим вне окрестности

$$\left(\frac{1}{n-1}\,\chi^2_{1-\alpha;n-1};\frac{1}{n-1}\,\chi^2_{\alpha;n-1}\right),$$

точки 1, мы гипотезу H_0 отклоняем, в противном случае — нет.

Критерий для проверки гипотезы $\boldsymbol{H_0}$: $\boldsymbol{\sigma^2/\sigma_0^2 = 1}$. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из $\mathsf{N}_{a;\sigma^2}$. Если гипотезу H_0 : $\sigma^2/\sigma_0^2 = 1$ отклонять при*

$$\frac{s^2}{\sigma_0^2} \notin \left(\frac{1}{n-1}\chi^2_{1-\alpha;n-1}, \frac{1}{n-1}\chi^2_{\alpha;n-1}\right)$$

и не отклонять в противном случае, то с вероятностью 2α гипотеза H_0 будет отклоняться, когда она верна.

Поскольку при верной гипотезе $H_0 : \sigma^2/\sigma_0^2 = 1$ отношение $s^2/\sigma_0^2 = s^2/\sigma^2$ распределено как $\chi^2_{n-1}/(n-1)$, то

$$\mathsf{P}\left\{\frac{s^2}{\sigma_0^2} \notin \left(\frac{1}{n-1}\,\chi^2_{1-\alpha;n-1}, \frac{1}{n-1}\,\chi^2_{\alpha;n-1}\right)\right\} = 2\alpha.$$

Последнее равенство означает, что уровень значимости приведенного критерия равен 2α. Этим критерием мы пользуемся, если альтернатива двусторонняя: $\sigma^2/\sigma_0^2 < 1$ или $\sigma^2/\sigma_0^2 > 1$.

Если альтернатива односторонняя, например, $\sigma^2/\sigma_0^2 > 1$, то s^2/σ_0^2 необходимо сравнить с $\chi^2_{\alpha;n-1}/(n-1)$: при

$$\frac{s^2}{\sigma_0^2} > \frac{1}{n-1}\chi^2_{\alpha;n-1}$$

гипотезу H_0 отклоняем, в противном случае не отклоняем (уровень значимости этого одностороннего критерия равен α).

Об ошибках первого и второго рода. При верной гипотезе H_0: $\sigma^2/\sigma_0^2 = 1$ отношение $s^2/\sigma_0^2 = s^2/\sigma^2$ распределено как случайная величина $\dfrac{1}{n-1}\,\chi^2_{n-1}$ и, как следствие, значения s^2/σ_0^2 “почти всегда” (с вероятностью $1-2\alpha$) принадлежат окрестности

$$\left(\frac{1}{n-1}\chi^2_{1-\alpha;n-1};\frac{1}{n-1}\chi^2_{\alpha;n-1}\right)$$

точки 1. Но s^2/σ_0^2 может, хотя и изредка (с вероятностью 2α), принимать значения вне окрестности

$$\left(\frac{1}{n-1}\chi^2_{1-\alpha;n-1};\,\frac{1}{n-1}\chi^2_{\alpha;n-1}\right),$$

при этом мы гипотезу H_0 отклоняем и тем самым допускаем ошибку — ошибку первого рода.

Далее, если гипотеза $H_0\colon \sigma^2/\sigma_0^2 = 1$ неверна, отношение s^2/σ_0^2, "почти всегда" (с вероятностью 2α) принимая значения из окрестности

$$\left(\frac{\sigma^2}{\sigma_0^2}\frac{1}{n-1}\,\chi^2_{1-\alpha;n-1};\,\frac{\sigma^2}{\sigma_0^2}\frac{1}{n-1}\,\chi^2_{\alpha;n-1}\right)$$

точки σ^2/σ_0^2, может принять значение из окрестности

$$\left(\frac{1}{n-1}\chi^2_{1-\alpha;n-1};\,\frac{1}{n-1}\chi^2_{\alpha;n-1}\right)$$

точки 1, при этом мы гипотезу H_0 не отклоним и тем самым допустим ошибку — ошибку второго рода (H_0 не отклоняется, хотя она и неверна).

23.5 Проверка гипотезы $\boldsymbol{\sigma_\xi^2 = \sigma_\eta^2}$

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ и $\eta(\omega) = (\eta_1(\omega), \eta_2(\omega), \ldots, \eta_m(\omega))$ — реализации независимых выборок $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \ldots, \eta_m)$ соответственно из распределений $\mathsf{N}_{a_\xi;\sigma_\xi^2}$ и $\mathsf{N}_{a_\eta;\sigma_\eta^2}$. Параметры $(a_\xi; \sigma_\xi^2)$ и $(a_\eta; \sigma_\eta^2)$ неизвестны. Относительно параметров σ_ξ^2 и σ_η^2 выдвигается гипотеза

$$H_0\colon \sigma_\xi^2 = \sigma_\eta^2.$$

Ее удобно записывать в виде $H_0\colon \sigma_\xi^2/\sigma_\eta^2 = 1$. Эту гипотезу можно рассматривать как гипотезу об одинаковой погрешности измерений (наблюдений), представленных выборками $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$. Альтернатива к гипотезе H_0 может быть как двусторонней: $\sigma_\xi^2/\sigma_\eta^2 > 1$ или $\sigma_\xi^2/\sigma_\eta^2 < 1$, так и односторонней: $\sigma_\xi^2/\sigma_\eta^2 < 1$ (односторонняя альтернатива может быть и такой:

$\sigma_\xi^2/\sigma_\eta^2 > 1$). Отклонение гипотезы $\sigma_\xi^2/\sigma_\eta^2 = 1$, скажем, в пользу альтернативы $\sigma_\xi^2/\sigma_\eta^2 > 1$, интерпретируется как бóльшая погрешность в измерениях, представленных выборкой $\xi_1, \xi_2, \ldots, \xi_n$ по сравнению с выборкой $\eta_1, \eta_2, \ldots, \eta_m$. Необходимо построить критерий для проверки гипотезы H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$.

О выборе статистики для построения критерия. Вне зависимости от того, верна гипотеза H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ или нет,

$$s_\xi^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2 \quad \text{и} \quad s_\eta^2 = \frac{1}{m-1}\sum_{j=1}^{m}(\eta_j - \overline{\eta})^2$$

являются несмещенными и состоятельными оценками соответственно параметров σ_ξ^2 и σ_η^2: при $n \to \infty$, $m \to \infty$

$$s_\xi^2 \xrightarrow{P} \sigma_\xi^2, \;\; s_\eta^2 \xrightarrow{P} \sigma_\eta^2.$$

Отсюда следует, что отношение s_ξ^2/s_η^2 мало отклоняется от $\sigma_\xi^2/\sigma_\eta^2$. Поэтому, если $\sigma_\xi^2/\sigma_\eta^2 = 1$ (гипотеза H_0 верна), то отношение s_ξ^2/s_η^2 оценок s_ξ^2 и s_η^2 соответственно дисперсий σ_ξ^2 и σ_η^2 мало уклоняется от 1. Если же гипотеза H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ неверна, например, $\sigma_\xi^2/\sigma_\eta^2 > 1$, то отношение s_ξ^2/s_η^2, будучи близким к $\sigma_\xi^2/\sigma_\eta^2$, отличается от 1 существенно.

Следовательно, для проверки гипотезы H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ вычисляем отношение s_ξ^2/s_η^2 и, в зависимости от того, приняло оно значение, существенно отличающееся от 1, или нет, отклоняем гипотезу H_0 или не отклоняем. Чтобы так можно было поступать, необходимо знать, какие значения принимает отношение s_ξ^2/s_η^2 (необходимо знать распределение отношения s_ξ^2/s_η^2), когда гипотеза H_0 верна и когда она неверна.

Представим отношение s_ξ^2/s_η^2 в виде

$$\frac{s_\xi^2}{s_\eta^2} = \frac{\sigma_\xi^2}{\sigma_\eta^2} \cdot \frac{s_\xi^2/\sigma_\xi^2}{s_\eta^2/\sigma_\eta^2} = \frac{\sigma_\xi^2}{\sigma_\eta^2}\frac{\frac{1}{n-1}\chi_{n-1}^2}{\frac{1}{m-1}\chi_{m-1}^2} = \frac{\sigma_\xi^2}{\sigma_\eta^2}F_{n-1,m-1},$$

т. е.

$$\frac{s_\xi^2}{s_\eta^2} = \frac{\sigma_\xi^2}{\sigma_\eta^2}F_{n-1,m-1}, \tag{23.5.1}$$

где $F_{n-1;m-1}$ — случайная величина, имеющая распределение Фишера с $(n-1, m-1)$ степенями свободы. Из равенства (23.5.1) следует, что при верной гипотезе H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$

$$s_\xi^2/s_\eta^2 = F_{n-1,m-1}.$$

Отсюда, в частности, получаем, что при верной гипотезе H_0 отношение s_ξ^2/s_η^2 "почти всегда" (с вероятностью $1-2\alpha$) принадлежит промежутку

$$\left(F_{1-\alpha;n-1,m-1}; F_{\alpha;n-1,m-1}\right),$$

где $F_{\alpha;n-1,m-1}$ и $F_{1-\alpha;n-1,m-1}$ — соответственно верхний α-предел и верхний $(1-\alpha)$-предел $F_{n-1,m-1}$-распределения. Заметим, что первый момент F-распределения с (n, m) степенями свободы равен $m/(m-2)$.

Если гипотеза H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ неверна (пусть для определенности $\sigma_\xi^2/\sigma_\eta^2 > 1$), то из равенства (23.5.1) следует, что "почти всегда" (с вероятностью $1-2\alpha$) отношение s_ξ^2/s_η^2 принадлежит промежутку

$$\left(\frac{\sigma_\xi^2}{\sigma_\eta^2} F_{1-\alpha;n-1,m-1}; \frac{\sigma_\xi^2}{\sigma_\eta^2} F_{\alpha;n-1,m-1}\right).$$

З а м е ч а н и е. Верхний $(1-\alpha)$-предел $F_{n;m}$-распределения нам будет удобно выражать через верхний α-предел распределения $F_{m,n}$, а именно:

$$F_{1-\alpha;n;m} = \frac{1}{F_{\alpha;m;n}}.$$

Это представление получается так. По определению, $F_{n;m}$-распределение — это распределение случайной величины

$$F_{n;m} = \frac{\chi_n^2/n}{\chi_m^2/m},$$

где χ_n^2 и χ_m^2 — независимые случайные величины, имеющие распределение χ^2 с n и m степенями свободы соответственно.

По определению, $(1-\alpha)$-предел $F_{n;m}$-распределения — это число $F_{1-\alpha;n;m}$, для которого

$$\mathsf{P}\{F_{n;m} \geq F_{1-\alpha;n;m}\} = 1-\alpha.$$

Отсюда имеем:

$$1-\alpha = \mathsf{P}\{F_{n;m} \geq F_{1-\alpha;n;m}\} = \mathsf{P}\left\{\frac{1}{F_{n;m}} \leq \frac{1}{F_{1-\alpha;n;m}}\right\} =$$

$$= \mathsf{P}\left\{F_{m;n} \leq \frac{1}{F_{1-\alpha;n;m}}\right\} = 1 - \mathsf{P}\left\{F_{m;n} > \frac{1}{F_{1-\alpha;n;m}}\right\},$$

или

$$\mathsf{P}\left\{F_{m;n} > \frac{1}{F_{1-\alpha;n;m}}\right\} = \alpha.$$

Из последнего соотношения по определению верхнего α-предела $F_{m;n}$-распределения имеем, что

$$F_{\alpha;m;n} = \frac{1}{F_{1-\alpha;n;m}}$$

или

$$F_{1-\alpha;n;m} = \frac{1}{F_{\alpha;m;n}}.$$

Критерий для проверки гипотезы $\boldsymbol{\sigma_\xi^2/\sigma_\eta^2 = 1}$. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ — независимые выборки соответственно из распределений $\mathsf{N}_{a_\xi;\sigma_\xi^2}$ и $\mathsf{N}_{a_\eta;\sigma_\eta^2}$.*

Если гипотезу $H_0\colon \sigma_\xi^2/\sigma_\eta^2 = 1$ отклонять при

$$\frac{s_\xi^2}{s_\eta^2} \notin \left(\frac{1}{F_{\alpha;m-1;\,n-1}}, F_{\alpha;n-1;m-1}\right)$$

и не отклонять в противном случае, то с вероятностью 2α гипотеза H_0 будет отклоняться, когда она верна.

При верной гипотезе $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$ отношение s_ξ^2/s_η^2 имеет распределение $F_{n-1;m-1}$, поэтому

$$\mathsf{P}\left\{\frac{s_\xi^2}{s_\eta^2} \notin \left(\frac{1}{F_{\alpha;m-1;n-1}}, F_{\alpha;n-1;m-1}\right)\right\} = 2\alpha.$$

Последнее равенство означает, что уровень значимости приведенного критерия равен 2α.

Этим критерием мы пользуемся, если альтернатива двусторонняя: $\sigma_\xi^2/\sigma_\eta^2 < 1$ или $\sigma_\xi^2/\sigma_\eta^2 > 1$. Если альтернатива односторонняя, например $\sigma_\xi^2/\sigma_\eta^2 > 1$, то s_ξ^2/s_η^2 сравниваем с $F_{\alpha;n-1;m-1}$: при

$$\frac{s_\xi^2}{s_\eta^2} > F_{\alpha;n-1;m-1}$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости этого критерия равен α).

З а м е ч а н и е. При проверке гипотезы $H_0\colon \sigma_\xi^2/\sigma_\eta^2 = 1$ не имеет значения, какое из отношений s_ξ^2/s_η^2 или s_η^2/s_ξ^2 рассматривать, что следует из очевидного равенства

$$\left\{\frac{1}{F_{\alpha;m-1;\,n-1}} < \frac{s_\xi^2}{s_\eta^2} < F_{\alpha;n-1;m-1}\right\} =$$

$$= \left\{\frac{1}{F_{\alpha;n-1;m-1}} < \frac{s_\eta^2}{s_\xi^2} < F_{\alpha;m-1;n-1}\right\}.$$

Об ошибках первого и второго рода. При верной гипотезе $H_0\colon \sigma_\xi^2/\sigma_\eta^2 = 1$ отношение s_ξ^2/s_η^2 имеет распределение $F_{n-1;m-1}$ и поэтому “почти всегда” (с вероятностью $1-2\alpha$) значения s_ξ^2/s_η^2 принадлежат окрестности

$$\left(\frac{1}{F_{\alpha;m-1;n-1}}, F_{\alpha;n-1;m-1}\right) \tag{23.5.2}$$

точки 1. Но отношение s_ξ^2/s_η^2 может, хотя и изредка (с вероятностью 2α), принимать значения вне окрестности (23.5.2). При этом мы гипотезу $H_0\colon \sigma_\xi^2/\sigma_\eta^2 = 1$ отклоняем и тем самым совершаем ошибку — ошибку первого рода.

При неверной гипотезе H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ отношение s_ξ^2/s_η^2 может (хотя и изредка) принимать значения из окрестности

$$\left(\frac{1}{F_{\alpha;m-1;n-1}}, F_{\alpha;n-1;m-1}\right),$$

точки 1, при этом мы гипотезу H_0 не отклоняем и тем самым совершаем ошибку — ошибку второго рода.

23.6 Примеры и задачи

Примеры

Пример 23.6.1. *Ниже приведены данные об измерениях неровностей поверхности одной и той же чистоты обработки при помощи двух двойных микроскопов.*

Можно ли считать, что между показаниями приборов нет систематического расхождения?

Микроскоп I: 0,8; 1,9; 3,0; 3,5; 3,8; 2,5; 1,7; 0,9; 1,0; 2,3; 3,3; 3,4.

Микроскоп II: 1,4; 2,1; 3,1; 3,6; 2,7; 1,7; 1,1; 0,2; 1,6; 2,8; 4,0; 4,7.

Решение. Для определенности будем обозначать через $\xi_1, \xi_2, \ldots, \xi_n$ данные измерений, полученные при помощи первого микроскопа, через $\eta_1, \eta_2, \ldots, \eta_m$ — при помощи второго.

В терминах проверки статистических гипотез эту задачу можно сформулировать так. Имеем реализации двух независимых выборок $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ (см. данные) из распределений $\mathsf{N}_{a_\xi;\sigma^2}$ и $\mathsf{N}_{a_\eta;\sigma^2}$ соответственно (предположение о нормальном распределении результатов измерений в большинстве случаев оправдывает себя). Относительно параметров a_ξ и a_η выдвигается гипотеза H_0: $a_\xi = a_\eta$. Это гипотеза об отсутствии систематического расхождения между показаниями приборов. Априори, если $a_\xi \neq a_\eta$, то может быть как $a_\xi < a_\eta$, так и $a_\xi > a_\eta$; поэтому альтернатива двусторонняя. Отклонение гипотезы H_0 в пользу этой альтернативы интерпретируется как наличие систематического расхождения показаний приборов, неотклонение — как отсутствие расхождения.

Согласно критерию Стьюдента для проверки гипотезы H_0: $a_\xi = a_\eta$ против двусторонней альтернативы $a_\xi < a_\eta$ или $a_\xi > a_\eta$ необходимо сравнить значение

$$|t| = |\overline{\xi} - \overline{\eta}| \Bigg/ \left(s\sqrt{\frac{n+m}{nm}} \right)$$

с $t_{\alpha;(n+m-2)}$ — верхним α-пределом $t_{(n+m-2)}$-распределения. Если

$$|t| = |\overline{\xi} - \overline{\eta}| \Bigg/ \left(s\sqrt{\frac{n+m}{nm}} \right) \geq t_{\alpha;(n+m-2)},$$

гипотеза H_0 отклоняется, в противном случае — нет (уровень значимости критерия 2α).

В рассматриваемом примере $n = 12,\ m = 12,$

$$\overline{\xi} = \frac{1}{12}(0{,}8 + 1{,}9 + \ldots + 3{,}4) = 2{,}34,$$

$$\overline{\eta} = \frac{1}{12}(1{,}4 + 2{,}1 + \ldots + 4{,}7) = 2{,}42,$$

$$s^2 = \frac{1}{n+m-2}((n-1){s_\xi}^2 + (m-1){s_\eta}^2) = 1{,}44,$$

$$|t| = |\overline{\xi} - \overline{\eta}| \Big/ \left(s\sqrt{\frac{n+m}{nm}}\right) = |2{,}42 - 2{,}34| \Big/ \sqrt{\frac{1{,}44(12+12)}{(12\cdot 12)}} = 0{,}16,$$

где

$${s_\xi}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2,\ {s_\eta}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\eta_i - \overline{\eta})^2.$$

Так что

$$|t| = 0{,}16 < 2{,}074 = t_{0{,}025;\,22}.$$

Поэтому в соответствии с критерием Стьюдента гипотеза H_0 о равенстве $a_\xi = a_\eta$ на 5 % уровне значимости не отклоняется.

Этот результат можно трактовать так. Предположение об отсутствии систематического расхождения между показаниями микроскопов не противоречит экспериментальным данным. (Такие показания вполне могли быть получены при работе с одним и тем же прибором.) Другими словами, эксперимент не дает оснований говорить о существовании систематического расхождения между показаниями микроскопов.

Пример 23.6.2 (точность измерений). *Определяется предел прочности на разрыв материала на двух разных стендах: A и B. Получены такие выборки значений предела прочности на разрыв:*

Стенд A: 1,32; 1,35; 1,32; 1,35; 1,30; 1,30; 1,37; 1,31; 1,39; 1,39.

Стенд B: 1,35; 1,31; 1,31; 1,41; 1,39; 1,37; 1,32; 1,34.

Выяснить, можно ли считать, что точность измерений предела прочности на разрыв на стендах A и B одинакова.

Решение. В терминах проверки статистических гипотез эту задачу можно сформулировать так. Имеем две реализации (см. данные) независимых выборок из нормальных распределений $\mathsf{N}_{a_\xi;\sigma_\xi^2}$ и $\mathsf{N}_{a_\eta;\sigma_\eta^2}$ (пусть для определенности $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ —

выборка, полученная на стенде A, $\eta = (\eta_1, \eta_2, \ldots, \eta_m)$ — на стенде B). Относительно неизвестных параметров σ_ξ^2 и σ_η^2 выдвигается гипотеза H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ — гипотеза об одинаковой точности (об одинаковой погрешности) измерений на стендах A и B. Альтернатива двусторонняя: $\sigma_\xi^2/\sigma_\eta^2 < 1$ или $\sigma_\xi^2/\sigma_\eta^2 > 1$, поскольку нет никакой априорной информации относительно точности работы стендов A и B. Отклонение гипотезы H_0 в пользу этой альтернативы будем интерпретировать как наличие расхождений в точности измерений на стендах, неотклонение — как отсутствие расхождений.

Согласно критерию для проверки гипотезы H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ против двусторонней альтернативы гипотезу H_0 отклоняем, если

$$\frac{s_\xi^2}{s_\eta^2} \notin \left(\frac{1}{F_{\alpha;(m-1);(n-1)}},\ F_{\alpha;(n-1);(m-1)}\right),$$

и не отклоняем в противном случае.

В рассматриваемом примере $n = 10, m = 8$, $\overline{\xi} = 1{,}34$; $\overline{\eta} = 1{,}35$,

$${s_\xi}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2 = 12{,}2 \cdot 10^{-4};$$

$$s_\eta^2 = \frac{1}{m-1}\sum_{j=1}^{m}(\eta_i - \overline{\eta})^2 = 14{,}0 \cdot 10^{-4};$$

$$\frac{s_\xi^2}{s_\eta^2} = \frac{12{,}2 \cdot 10^{-4}}{14{,}0 \cdot 10^{-4}} = 0{,}87;$$

$$F_{\alpha;(n-1);(m-1)} = F_{0{,}01;9;7} = 6{,}72;$$

$$\frac{1}{F_{\alpha;(m-1);(n-1)}} = \frac{1}{5{,}61} = 0{,}18.$$

Значение s_ξ^2/s_η^2 принадлежит промежутку

$$\left(\frac{1}{F_{\alpha;(m-1);(n-1)}},\ F_{\alpha;(n-1);(m-1)}\right) = (0{,}18;\ 6{,}72),$$

поэтому гипотеза H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$ на 2 % уровне значимости не отклоняется.

Этот результат можно интерпретировать так. Предположение (гипотеза), что стенды A и B имеют одинаковую точность измерений предела прочности на разрыв, не противоречит экспериментальным данным. (Такие данные могли быть получены при работе на одном и том же стенде.) Другими словами, эксперимент не дает оснований утверждать, что точность измерений предела прочности на разрыв на стендах A и B различна.

Задачи

23.1. На протяжении летних каникул 10 учеников находились в спортивном лагере. В начале сезона и после его завершения у них определяли емкость легких (в миллилитрах). По результатам измерений необходимо определить, существенно ли изменился этот показатель под влиянием интенсивных физических упражнений.

Ученик	До сезона	После сезона	Ученик	До сезона	После сезона
1	3400	3800	6	3100	3200
2	3600	3700	7	3200	3200
3	3000	3300	8	3400	3300
4	3500	3600	9	3200	3500
5	2900	3100	10	3400	3600

23.2. Один из методов количественного анализа степени износа шины состоит в измерении глубины проникновения щупа[2] в определенном месте шины. Имеется подозрение, что появление значительной части дисперсии измерений связано с действиями контролеров. Чтобы исключить из общей дисперсии измерений указанную ее часть, двум контролерам предложили провести по 12 измерений в одной и той же точке шины. Получили такие результаты:

Контролер X: 121, 121, 126, 130, 127, 131, 127, 124, 125, 119, 126, 123.

Контролер Y: 120, 129, 128, 136, 117, 138, 124, 119, 136, 136, 134, 132.

Существенно ли отличаются дисперсии измерений, проведенных разными контролерами?

23.3. На автоматическом станке обрабатываются втулки. После настройки станка была получена выборка из 10 изделий. Оказалось, что выборочное среднее $\overline{\xi}$ диаметра втулки составляет 2,059 мм, а значение несмещенной оценки дисперсии s^2_ξ равно

[2]Здесь щуп — тонкая продолговатая металлическая пластинка прямоугольной формы.

4,4 мкм^2. Через некоторый промежуток времени с целью контроля настройки станка на данный диаметр втулки снова была получена выборка из 10 изделий, для которой выборочное среднее $\overline{\eta} = 2{,}063$ мм, а значение несмещенной оценки дисперсии s_η^2 равно 8,6 мкм^2. Предположим,что на протяжении указанного промежутка времени изменения в работе станка могут сказаться только на уровне его настройки, но точность работы станка не меняется.

Свидетельствуют ли приведенные данные об изменении уровня настройки станка за промежуток времени, который отделяет моменты получения выборок?

23.4. В задаче 23.3 описывается процесс контроля за работой станка-автомата. При этом предполагается, что точность работы станка не изменяется, хотя значения несмещенных оценок дисперсий после настройки станка и через определенное время составляют соответственно $s_\xi^2 = 4{,}4$ мкм^2 и $s_\eta^2 = 8{,}6$ мкм^2.

Является ли обоснованным предположение о неизменности точности работы станка?

Р е ш е н и е. Точность работы станка характеризуется дисперсией и со временем не возрастает (а дисперсия соответственно не уменьшается). В терминах проверки статистических гипотез предположение о неизменности точности работы станка формулируется как гипотеза H_0: $\sigma_\xi^2/\sigma_\eta^2 = 1$, альтернатива односторонняя: $\sigma_\xi^2/\sigma_\eta^2 < 1$. Значение $s_\xi^2/s_\eta^2 = 0{,}51 > 0{,}31 = 1/F_{0{,}05;9;9} = 1/F_{\alpha;(m-1);(n-1)}$. Поэтому нулевая гипотеза не отклоняется. Таким образом, можно считать, что точность работы станка не уменьшилась (хотя значение оценки дисперсии и возросло с $s_\xi^2 = 4{,}4\,$мкм^2 до $s_\eta^2 = 8{,}6\,$мкм^2, но это допустимое увеличение дисперсии).

23.5. В процессе производства электрических счетчиков с вращающимся диском их работа была синхронизирована с работой стандартного счетчика (постоянная, характеризующая стандартный счетчик, равна 1). Проверка 10 счетчиков, состоящая в определении их постоянной с помощью точных ваттметров и секундомеров, дала такие результаты:

Номер счетчика	Значение постоянной	Номер счетчика	Значение постоянной
1	0,983	6	0,988
2	1,002	7	0,994
3	0,998	8	0,991
4	0,996	9	1,005
5	1,003	10	0,986

Можно ли отклонения от стандарта рассматривать как случайные, или, наоборот, результаты указывают на то, что постоянные отрегулированных счетчиков систематически отклоняются от постоянной стандартного счетчика?

Ответить на этот вопрос, проверив гипотезу о том, что 10 измерений образуют выборку, полученную из нормального распределения со средним 1,000.

23.6. В одном классе из 20 детей наудачу были выбраны 10, которым ежедневно начали выдавать апельсиновый сок. Остальные 10 учеников ежедневно получали молоко. Через некоторое время зафиксировали увеличение массы детей в фунтах (1 фунт = 453,6 г):

Сок	4,0	2,5	3,5	4,0	1,5	1,0	3,5	3,0	2,5	3,5
Молоко	1,5	3,5	2,5	3,0	2,5	2,0	2,0	2,5	1,5	3,0

Среднее увеличение веса одного ученика в группе, где выдавали апельсиновый сок, составило 2,9 фунта, а в группе, где выдавали молоко, — 2,4. Существенно ли отличается увеличение веса детей в группах?

23.7. Исследования, проведенные на протяжении нескольких лет, после того, как в практике лыжного спорта стала использоваться техника коньковых ходов, дают основание считать, что она, в условиях специально подготовленной трассы, обеспечивает на дистанции 15 км у мужчин выигрыш по сравнению с традиционной техникой более чем 1 мин. Ниже приведены результаты в минутах соревнований двух групп лыжников (дистанция 15 км): одни проходили дистанцию традиционным ходом, а другие — коньковым.

Традиционная техника: 37,02; 36,74; 37,82; 38,12; 36,91; 37,98; 38,21; 37,51; 37,56; 38,03.

Техника коньковых ходов: 35,81; 35,61; 35,02; 35,53; 35,84; 35,12; 36,12; 36,49; 35,62; 36,28.

Можно ли по результатам этих соревнований прийти к выводу о наличии более чем одноминутного эффекта техники коньковых ходов?

Решение. Пусть a_ξ и a_η — средние значения результатов при использовании техники традиционных и коньковых ходов соответственно. Нулевая гипотеза H_0: $a_\xi - 1 = a_\eta$, альтернатива односторонняя: $a_\xi - 1 > a_\eta$. Отклонение нулевой гипотезы в пользу альтернативной $a_\xi - 1 > a_\eta$ будет свидетельствовать о наличии более чем одноминутного эффекта коньковых ходов, неотклонение гипотезы H_0 — об отсутствии такого эффекта.

Значение
$$t = ((\bar{\xi} - 1) - \bar{\eta}) \Big/ \left(s\sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$
сравниваем с $t_{\alpha;n+m-2}$ — если $t > t_{\alpha;n+m-2}$, гипотеза H_0 отклоняется, в противном случае — нет. Имеем:
$$t = 3{,}774 > 2{,}10 = t_{0{,}05;18} = t_{\alpha;n+m-2}.$$
Поэтому в соответствии с критерием Стьюдента гипотеза H_0: $a_\xi - 1 = \ a_\eta$ на 5 % уровне значимости отклоняется, что интерпретируется как более чем одноминутный выигрыш при использовании техники коньковых ходов сравнительно с традиционной.

23.8 (длительность обезболивающего действия препарата). Сравнивается действие обезболивающих препаратов A и B. (В некоторых случаях одним из "лекарств" может быть инертное плацебо, которое используется для контроля при исследовании действия другого препарата.)

В группе больных, которые изъявили желание принять участие в эксперименте, насчитывалось восемь человек. Вполне возможно, что у этих больных возраст, пол, общее состояние и т. д. далеко не одинаковые. Поэтому оба препарата дают каждому больному и фиксируют продолжительность обезболивающего действия каждого из них. С целью обеспечения чистоты эксперимента, приняты все разумные меры предосторожности: между приемами обоих препаратов проходит время, исключающее "перекрывание" их действия; четверо больных получают сначала препарат A, а другие четверо — сначала препарат B, при этом ни один из них не знает, какой именно препарат он принимает, и т. д. Результаты эксперимента приведены в таблице (фиксировалась в часах длительность обезболивающего действия препаратов A и B).

Больной	Длительность действия препарата		Разность длительности действия препаратов B и A
	A	B	
1	3,2	3,8	0,6
2	1,6	1,0	−0,6
3	5,7	8,4	2,7
4	2,8	3,6	0,8
5	5,5	5,0	−0,5
6	1,2	3,5	2,3
7	6,1	7,3	1,2
8	2,9	4,8	1,9

Вынести заключение об эффективности действия препаратов в предположении, что:

1° отсутствует априорная информация об эффективности препаратов;

2° эффективность действия препарата B едва ли меньше, чем препарата A.

Ответ дать в терминах проверки статистических гипотез.

23.9. На станке-автомате изготавливается один вид продукции. Критическим размером изделий является внешний диаметр. После настройки станка отобрали 20 изделий. При этом оказалось, что выборочная дисперсия размера внешнего диаметра составляет 0,84 мм^2. Через некоторый промежуток времени с целью контроля точности работы станка (и, при необходимости, его настройки) отобрали 15 изделий. Оказалось, что выборочная дисперсия, вычисленная по ним, равна 1,07 мм^2.

Свидетельствуют ли приведенные данные об изменении точности работы станка?

Глава 24

Критерий χ^2

24.1 Критерии согласия

Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из распределения F, распределение F неизвестно. Относительно распределения F выдвигается гипотеза H_0: $\mathsf{F} = \mathsf{G}$, G — распределение с заданными свойствами, например, G — равномерное на $[0; 1]$ распределение, или, например, распределение G принадлежит классу нормальных распределений. Далее приводятся общие соображения, которыми естественно руководствоваться при построении критериев для проверки гипотез вида H_0: $\mathsf{F} = \mathsf{G}$. (Для наглядности, G — полностью определенное распределение.)

Вне зависимости от того, верна гипотеза H_0: $\mathsf{F} = \mathsf{G}$ или нет, эмпирическое распределение $\hat{\mathsf{F}}_n$, построенное по выборке $\xi_1, \xi_2, \ldots, \xi_n$ из распределения F, "близко" к F — является хорошей аппроксимацией F, и с увеличением n аппроксимация распределения F эмпирическим распределением $\hat{\mathsf{F}}_n$ становится лучше, а именно, при каждом фиксированном x эмпирическая функция распределения $\hat{F}_n(x)$ является состоятельной и несмещенной оценкой функции распределения $F(x)$, а вместе с этим $\hat{\mathsf{F}}_n([a, b))$ является состоятельной и несмещенной оценкой $\mathsf{F}([a, b))$ (для любых $a < b$). Так что при верной гипотезе H_0: $\mathsf{F} = \mathsf{G}$ эмпирическое распределение $\hat{\mathsf{F}}_n$ мало отличается от G, если же $\mathsf{G} \neq \mathsf{F}$, то распределение $\hat{\mathsf{F}}_n$, будучи близким к F, от распределения G отличается существенно.

Поэтому естественно попытаться ввести уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ между эмпирическим распределением $\hat{\mathsf{F}}_n$ и гипотетическим G, причем так, чтобы при верной гипотезе H_0: $\mathsf{F} = \mathsf{G}$ оно принимало малые значения, а если гипотеза $H_0\colon \mathsf{F} = \mathsf{G}$ неверна — большие (заметим, что уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$, как функция выборки, является случайной величиной). Предположим, что нам это удалось. Тогда для проверки гипотезы H_0: $\mathsf{F} = \mathsf{G}$ вычисляем значение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ и, в зависимости от того, большим или малым оно оказалось, отклоняем или не отклоняем гипотезу H_0: $\mathsf{F} = \mathsf{G}$. Чтобы так можно было поступать, необходимо знать, какие значения принимает уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$, когда гипотеза H_0: $\mathsf{F} = \mathsf{G}$ верна (когда уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ малые) и какие значения принимает уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$, когда гипотеза H_0 неверна (когда уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ большие). А поскольку $D(\hat{\mathsf{F}}_n, \mathsf{G})$ является случайной величиной, то последнее фактически обозначает, что необходимо знать распределение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$, когда гипотеза H_0 верна и когда H_0 неверна, или хотя бы когда H_0 верна. Заметим, что если распределение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F})$ известно, оно не может зависеть от распределения F, из которого получена выборка $\xi_1, \xi_2, \ldots, \xi_n$ — распределение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F})$ одно и то же при всех возможных F. И хотя предъявленные к $D(\hat{\mathsf{F}}_n, \mathsf{G})$ требования представляются столь жесткими, что возникает сомнение в существовании таких уклонений, тем не менее они существуют.

Пусть уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ обладает перечисленными выше свойствами, а именно, во-первых, при верной гипотезе H_0: $\mathsf{F} = \mathsf{G}$ уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G}) = D(\hat{\mathsf{F}}_n, \mathsf{F}) = D_n$ принимает малые значения, а если H_0: $\mathsf{F} = \mathsf{G}$ неверна — большие, и, во-вторых, при верной гипотезе H_0 распределение $D(\hat{\mathsf{F}}_n, \mathsf{G}) = D(\hat{\mathsf{F}}_n, \mathsf{F}) = D_n$ известно. Зная распределение D_n, найдем промежуток $[0, D_{\alpha,n}]$, в котором "почти всегда" (с вероятностью не меньшей $1 - \alpha$) лежат значения D_n. Для этого достаточно выбрать $D_{\alpha,n}$ так, чтобы

$$\mathsf{P}\{D_n > D_{\alpha,n}\} = \alpha \tag{24.1.1}$$

(или как наименьшее D, для которого $\mathsf{P}\{D_n > D\} \leq \alpha$). Значения $D(\hat{\mathsf{F}}_n, \mathsf{G})$, принадлежащие промежутку $[0, D_{\alpha,n}]$, будем классифицировать как малые, в противном случае — как большие. Далее для проверки гипотезы H_0: $\mathsf{F} = \mathsf{G}$ вычисляем значение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ и выясняем, большое оно или малое, сравнивая его с $D_{\alpha,n}$. Если при этом $D(\hat{\mathsf{F}}_n, \mathsf{G})$ приняло большое значение:

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) > D_{\alpha,n},$$

гипотезу H_0: $\mathsf{F} = \mathsf{G}$ отклоняем, в противном случае — нет.

Теперь можно сформулировать критерий для проверки гипотезы H_0: $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из распределения G.

Если гипотезу H_0 отклонять при

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) > D_{\alpha,n}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α, гипотеза H_0 будет отклоняться, когда она верна.

Действительно, если гипотеза H_0 верна, т. е. распределение G совпадает с распределением F, то $D(\hat{\mathsf{F}}_n, \mathsf{G}) = D(\hat{\mathsf{F}}_n, \mathsf{F}) = D_n$ и, следовательно,

$$\mathsf{P}\{D(\hat{\mathsf{F}}_n, \mathsf{G}) > D_{\alpha,n}\} = \mathsf{P}\{D_n > D_{\alpha,n}\} \leq \alpha.$$

Так что вероятность отклонения верной гипотезы H_0 не превосходит α.

Вид уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ определяет тот или иной критерий для проверки гипотезы H_0: $\mathsf{F} = \mathsf{G}$.

Мы рассмотрим уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ эмпирического распределения $\hat{\mathsf{F}}_n$ от гипотетического G, предложенные К. Пирсоном и А. Н. Колмогоровым. По ним строятся соответственно критерий χ^2 и критерий А. Н. Колмогорова для проверки гипотезы H_0: $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из распределения G.

24.2 Критерий χ^2 (параметры известны)

Рассматриваемый далее критерий χ^2 для проверки гипотезы H_0: $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из распределения G, строится по уклонению $D(\hat{\mathsf{F}}_n, \mathsf{G})$ между эмпирическим распределением $\hat{\mathsf{F}}_n$ и гипотетическим G, предложенному К. Пирсоном.

Мы рассмотрим два варианта критерия χ^2: 1) гипотетическое распределение G не зависит от неизвестных параметров, например, G — равномерное на $[0; 1]$ распределение; 2) гипотетическое распределение G принадлежит данному классу распределений $\mathsf{G}(\cdot; \theta_1, \theta_2, \ldots, \theta_k)$ (кратко $\mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)$), зависящему от неизвестных параметров $\theta_1, \theta_2, \ldots, \theta_k$, например, G — нормальное распределение (зависит от параметров a и σ^2).

Сначала рассмотрим критерий χ^2 для гипотетического распределения, не зависящего от параметров.

Уклонение Пирсона эмпирического распределения от гипотетического. Два вероятностных распределения F и G, заданные на $X \subset \mathbb{R}^1$, совпадают, если для каждого борелевского множества B из X справедливо равенство $\mathsf{F}(B) = \mathsf{G}(B)$. Если F и G различны, то найдется борелевское множество $B' \subset X$ такое, что $\mathsf{F}(B') \neq \mathsf{G}(B')$. Поэтому в качестве уклонения между распределениями F и G естественно рассматривать величину

$$\sum_{j=1}^{r} c_j(\mathsf{F}(X_j) - \mathsf{G}(X_j))^2,$$

где c_j — неотрицательные константы, а $\{X_j\}$ — разбиение X на непересекающиеся борелевские множества:

$$\bigcup_{j=1}^{r} X_j = X,\ X_i \cap X_j = \emptyset,\ i \neq j,\ 2 \le r < \infty,$$

зачастую $X_j = [a_j, b_j)$, $j = 1, 2, \ldots, r$. Уклонение Пирсона между эмпирическим распределением $\hat{\mathsf{F}}_n$, построенным по выборке $\xi_1, \xi_2, \ldots, \xi_n$, и гипотетическим G из этих соображений и строится. А именно, в качестве уклонения между $\hat{\mathsf{F}}_n$ и G рассматриваем

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{j=1}^{r} c_j(\hat{\mathsf{F}}_n(X_j) - \mathsf{G}(X_j))^2,$$

где $\mathsf{G}(X_j)$ — вероятность попадания выборочного значения в множество X_j, вычисленная по гипотетическому распределению G (далее её будем обозначать $\mathsf{G}(X_j) = p_j$), $\hat{\mathsf{F}}_n(X_j)$ — вероятность попадания выборочного значения в X_j, вычисленная по эмпирическому распределению $\hat{\mathsf{F}}_n$ (численно $\hat{\mathsf{F}}_n(X_j)$ равна частоте ν_j/n попадания выборочного значения в множество X_j, вычисленной по выборке $\xi_1, \xi_2, \ldots, \xi_n$, ν_j — число выборочных значений из $\xi_1, \xi_2, \ldots, \xi_n$, попавших в X_j), c_j — коэффициенты, выбор которых и определяет уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$.

К. Пирсон в качестве c_j предложил рассматривать n/p_j, $j = 1, 2, \ldots, r$. При этом уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ принимает вид

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{j=1}^{r} \frac{n}{p_j}\left(\hat{\mathsf{F}}_n(X_j) - \mathsf{G}(X_j)\right)^2 = \sum_{j=1}^{r} \frac{n}{p_j}\left(\frac{\nu_j}{n} - p_j\right)^2.$$

Вводя уклонение

$$\sum_{j=1}^{r} \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j\right)^2$$

между эмпирическим распределением и гипотетическим, К. Пирсон, по-видимому, исходил из следующих соображений.

Вне зависимости от того, верна гипотеза H_0: $\mathsf{F} = \mathsf{G}$ или нет, частота ν_j/n попадания выборочных значений в X_j, вычисленная по выборке $\xi_1, \xi_2, \ldots, \xi_n$ из F, является несмещенной и состоятельной оценкой $\mathsf{F}(X_j)$, т. е.

$$\mathsf{M}(\nu_j/n) = \mathsf{F}(X_j)$$

и при $n \to \infty$

$$\nu_j/n \xrightarrow{\mathsf{P}} \mathsf{F}(X_j)$$

(для всех $j = 1, 2, \ldots, r$). Поэтому, если гипотеза H_0: $\mathsf{F} = \mathsf{G}$ верна, и, следовательно, $\mathsf{F}(X_j) = \mathsf{G}(X_j) = p_j$, то

$$\nu_j/n - p_j \xrightarrow{\mathsf{P}} 0, \ j = 1, 2, \ldots, r.$$

Причем

$$\nu_j = \sum_{k=1}^{n} I_{X_j}(\xi_k),$$

как сумма независимых одинаково распределенных случайных величин, имеет распределение близкое к нормальному, а именно, распределение случайной величины

$$\frac{\nu_j - \mathsf{M}\nu_j}{\sqrt{\mathsf{D}\nu_j}} = \frac{\nu_j - np_j}{\sqrt{np_j(1-p_j)}} = \sqrt{\frac{n}{p_j(1-p_j)}} \left(\frac{\nu_j}{n} - p_j\right)$$

сходится к нормальному распределению с параметрами $(0, 1)$, $j = 1, 2, \ldots, r$. Поэтому естественно ожидать, что распределение суммы квадратов этих случайных величин имеет распределение близкое к χ^2-распределению. На самом деле при $n \to \infty$ распределение случайной величины

$$\sum_{j=1}^{r} \left(\sqrt{\frac{n}{p_j}} \left(\frac{\nu_j}{n} - p_j\right)\right)^2 = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}}\right)^2 = \sum_{j=1}^{r} \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j\right)^2$$

сходится к χ^2-распределению с $(r-1)$ степенями свободы.

Если же гипотеза H_0: $\mathsf{F} = \mathsf{G}$ неверна, т. е. $\mathsf{F} \neq \mathsf{G}$, то уклонение

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{j=1}^{r} \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j\right)^2$$

принимает большие значения, и при $n \to \infty$ сходится по вероятности к $+\infty$, поскольку

$$\left(\frac{\nu_j}{n} - p_j\right)^2 \xrightarrow{\mathsf{P}} (\mathsf{F}(X_j) - p_j)^2 = (\mathsf{F}(X_j) - \mathsf{G}(X_j))^2\,,$$

а среди чисел

$$(\mathsf{F}(X_j) - \mathsf{G}(X_j))^2\,,\ j = 1, 2, \ldots, r,$$

при достаточно "мелком" разбиении выборочного пространства X найдутся числа строго большие нуля и, следовательно, найдутся X_j, для которых при $n \to +\infty$

$$\frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j\right)^2 \xrightarrow{\mathsf{P}} +\infty,$$

а вместе с ними и $D(\hat{\mathsf{F}}_n, \mathsf{G}) \to +\infty$. Так что уклонение К. Пирсона

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{j=1}^{r} \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j\right)^2 = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}}\right)^2$$

когда гипотеза H_0: $\mathsf{F} = \mathsf{G}$ верна, принимает малые значения (при этом распределение малых уклонений $D(\hat{\mathsf{F}}_n, \mathsf{F})$ близко к распределению χ^2_{r-1}), а когда гипотеза H_0: $\mathsf{F} = \mathsf{G}$ неверна, уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ принимает большие значения и

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) \to +\infty$$

при $n \to \infty$.

Теорема 24.2.1 (о распределении уклонения К. Пирсона). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения* F *на $\mathbb{R}^1$; X_j, $j = 1, 2, \ldots, r$, — разбиение выборочного пространства $X \subset \mathbb{R}^1$ ($\bigcup\limits_{j=1}^{r} X_j = X$, $X_i \cap X_j = \emptyset$, $i \neq j$, $r \geq 2$);*

$p_j = \mathsf{P}\{\xi_k \in X_j\} = \mathsf{F}(X_j)$ *— вероятность попадания выборочного значения в X_j, $j = 1, 2, \ldots, r$, вычисленная по распределению F; ν_j — число выборочных значений $\xi_1, \xi_2, \ldots, \xi_n$, попавших в X_j, $j = 1, 2, \ldots, r$. Тогда при $n \to \infty$ распределение случайной величины*

$$\triangle_n = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}} \right)^2$$

сходится к χ^2-распределению с $r-1$ степенями свободы, т. е.

$$\mathsf{P}\{\triangle_n < x\} \to \mathsf{P}\{\chi^2_{r-1} < x\}$$

для каждого $x \in (-\infty, \infty)$.

Доказательство. Пусть

$$\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_r) = \left(\frac{\nu_1 - np_1}{\sqrt{np_1}}, \frac{\nu_2 - np_2}{\sqrt{np_2}}, \ldots, \frac{\nu_r - np_r}{\sqrt{np_r}} \right). \tag{24.2.1}$$

В силу центральной предельной теоремы (см. теорему 15.1.1) распределение каждой компоненты $\zeta_j = (\nu_j - np_j)/\sqrt{np_j}$ вектора ζ сходится к нормальному распределению. Поэтому естественно ожидать, что

1) распределение F_ζ вектора ζ сходится к распределению F_α некоторого нормального вектора $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r)$;

2) распределение суммы квадратов компонент вектора ζ

$$\triangle_n = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}} \right)^2 = \sum_{j=1}^{r} {\zeta_j}^2 = (\zeta, \zeta) = \triangle_n(\zeta)$$

сходится к χ^2-распределению.

Сначала докажем первое утверждение. Для этого, воспользовавшись теоремой Леви, достаточно установить, что характеристическая функция вектора $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_r)$ сходится к характеристической функции некоторого гауссовского вектора.

Чтобы найти характеристическую функцию вектора ζ (см. (24.2.1)), выпишем сначала характеристическую функцию вектора $\nu = (\nu_1, \nu_2, \ldots, \nu_r)$, описывающего число выборочных значений, попавших в $X_1, X_2, \ldots, X_r$ (ν_j — число выборочных значений из $\xi_1, \xi_2, \ldots, \xi_n$ попавших в X_j, $j = 1, 2, \ldots, r$). Представим вектор $\nu = (\nu_1, \nu_2, \ldots, \nu_r)$ в виде суммы векторов

$$\mu^{(k)} = \left(I_{X_1}(\xi_k), I_{X_2}(\xi_k), \ldots, I_{X_r}(\xi_k)\right), \; k = 1, 2, \ldots, n,$$

а именно:

$$\nu = (\nu_1, \nu_2, \ldots, \nu_r) = \sum_{k=1}^{n} \mu^{(k)}. \tag{24.2.2}$$

Векторы $\mu^{(k)}$, $k = 1, 2, \ldots, n$, независимы и одинаково распределены (как функции от независимых одинаково распределенных случайных величин ξ_k), поэтому характеристическая функция $\varphi_\nu(t)$ вектора (24.2.2) как характеристическая функция суммы независимых случайных величин равна их произведению:

$$\varphi_\nu(t_1, t_2, \ldots, t_r) = \varphi_\nu(t) = \mathsf{M} \exp\{i(t, \nu)\} =$$

$$= \mathsf{M} \exp\left\{ i \left(t, \sum_{k=1}^{n} \mu^{(k)} \right) \right\} = \mathsf{M} \exp\left\{ i \sum_{k=1}^{n} \left(t, \mu^{(k)} \right) \right\} =$$

$$= \mathsf{M} \prod_{k=1}^{n} \exp\left\{ i \left(t, \mu^{(k)} \right) \right\} = \prod_{k=1}^{n} \mathsf{M} \exp\left\{ i \left(t, \mu^{(k)} \right) \right\} =$$

$$= \left(\mathsf{M} \exp\left\{ i \left(t, \mu^{(1)} \right) \right\} \right)^n.$$

Чтобы вычислить $\mathsf{M} \exp\left\{ i \left(t, \mu^{(1)} \right) \right\}$, найдем распределение вектора $\mu^{(1)}$ (такое же распределение имеют все $\mu^{(k)}$, $k = 1, 2, \ldots, n$). Поскольку $X = \bigcup_{j=1}^{r} X_j$, и $X_i \cap X_j = \emptyset$, $i \neq j$, то у вектора

$$\mu^{(1)} = (I_{X_1}(\xi_1), I_{X_2}(\xi_1), \ldots, I_{X_r}(\xi_1))$$

только одна компонента равна 1, остальные равны 0, поэтому

$$\mathsf{P}\{\mu^{(1)} = (\underbrace{0, 0, \ldots, 0}_{j-1}, 1, 0, 0, \ldots, 0)\} =$$

$$= \mathsf{P}\{I_{X_j}(\xi_1) = 1\} = \mathsf{P}\{\xi_1 \in X_j\} = p_j, \tag{24.2.3}$$

$j = 1, 2, \ldots, r$ (p_j — вероятность попадания выборочного значения ξ_1 в X_j). Равенства (24.2.3) и определяют распределение вектора $\mu^{(1)}$.

Случайная величина $\exp\left\{i\left(t,\mu^{(1)}\right)\right\}$ является функцией дискретного вектора $\mu^{(1)}$ с распределением

$$\mathsf{P}\left\{\mu^{(1)}=(1,0,0,\ldots,0)\right\}=p_1,$$

$$\mathsf{P}\left\{\mu^{(1)}=(0,1,0,\ldots,0)\right\}=p_2,$$

$$\vdots$$

$$\mathsf{P}\left\{\mu^{(1)}=(0,0,0,\ldots,1)\right\}=p_r$$

(см. равенство (24.2.3)). Поэтому согласно теореме о вычислении математического ожидания функции от случайной величины (от случайного вектора) по ее распределению (см. теорему 5.3.2) имеем

$$\mathsf{M}\exp\left\{i(t,\mu^{(1)}\right\}=p_1\exp\left\{i(t_1\cdot 1+t_2\cdot 0+\ldots+t_r\cdot 0)\right\}+$$

$$+p_2\exp\left\{i(t_1\cdot 0+t_2\cdot 1+\ldots+t_r\cdot 0)\right\}+\ldots$$

$$\ldots+p_r\exp\ \ i(t_1\cdot 0+t_2\cdot 0+\ldots+t_r\cdot 1)\}=$$

$$=p_1\exp\left\{it_1\right\}+p_2\exp\left\{it_2\right\}+\ldots+p_r\exp\left\{it_r\right\}.$$

Так что характеристическая функция вектора $\nu=(\nu_1,\nu_2,\ldots,\nu_r)$

$$\varphi_\nu(t_1,t_2,\ldots,t_r)=\mathsf{M}\exp\{i(t,\nu)\}=\left(\mathsf{M}\exp\left\{i(t,\mu^{(1)}\right\}\right)^n=$$

$$=(p_1\exp\{it_1\}+p_2\exp\{it_2\}+\ldots+p_r\exp\{it_r\})^n.$$

По характеристической функции вектора ν найдем характеристическую функцию $\varphi_\zeta(t_1,t_2,\ldots,t_r)$ вектора $\zeta=(\zeta_1,\zeta_2,\ldots,\zeta_r)$ (см. (24.2.1)):

$$\varphi_\zeta(t_1,t_2,\ldots,t_r)=\mathsf{M}\exp\left\{i(t,\zeta)\right\}=$$

$$=\mathsf{M}\exp\left\{i\left(t_1\frac{\nu_1-np_1}{\sqrt{np_1}}+t_2\frac{\nu_2-np_2}{\sqrt{np_2}}+\ldots+t_r\frac{\nu_r-np_r}{\sqrt{np_r}}\right)\right\}=$$

$$=\mathsf{M}\exp\left\{i\left(\nu_1\frac{t_1}{\sqrt{np_1}}+\nu_2\frac{t_2}{\sqrt{np_2}}+\ldots+\nu_r\frac{t_r}{\sqrt{np_r}}\right)-\right.$$

$$-i\left(t_1\sqrt{np_1}+t_2\sqrt{np_2}+\ldots+t_r\sqrt{np_r}\right)\Bigg\}=$$

$$=\exp\left\{-i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}\right\}\times$$

$$\times\mathsf{M}\exp\left\{i\left(\nu_1\frac{t_1}{\sqrt{np_1}}+\nu_2\frac{t_2}{\sqrt{np_2}}+\ldots+\nu_r\frac{t_r}{\sqrt{np_r}}\right)\right\}=$$

$$=\exp\left\{-i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}\right\}\times$$

$$\times\left(p_1\exp\left\{i\frac{t_1}{\sqrt{np_1}}\right\}+p_2\exp\left\{i\frac{t_2}{\sqrt{np_2}}\right\}+\ldots+p_r\exp\left\{i\frac{t_r}{\sqrt{np_r}}\right\}\right)^n.$$

При больших n (при $n\to\infty$), учитывая, что

$$e^z=1+z+z^2/2+o(z^2)\ \text{при}\ z\to 0,$$

для $\ln\varphi_\zeta(t_1,t_2,\ldots,t_r)$ имеем:

$$\ln\varphi_\zeta(t_1,t_2,\ldots,t_r)=-i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}+$$

$$+n\ln\left(p_1\exp\left\{i\frac{t_1}{\sqrt{np_1}}\right\}+\ldots+p_r\exp\left\{i\frac{t_r}{\sqrt{np_r}}\right\}\right)=$$

$$=-i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}+$$

$$+n\ln\left(p_1\left(1+i\frac{t_1}{\sqrt{np_1}}-\frac{1}{2}\frac{t_1^2}{np_1}+o(1/n)\right)+\ldots\right.$$

$$\left.\ldots+p_r\left(1+i\frac{t_r}{\sqrt{np_r}}-\frac{1}{2}\frac{t_r^2}{np_r}+o(1/n)\right)\right)=$$

$$=-i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}+$$

$$+n\ln\left(1+\frac{i}{\sqrt{n}}\sum_{k=1}^{r}t_k\sqrt{p_k}-\frac{1}{2n}\sum_{k=1}^{r}t_k^2+o(1/n)\right).$$

При $n\to\infty$ величина

$$z=\frac{i}{\sqrt{n}}\sum_{k=1}^{r}t_k\sqrt{p_k}-\frac{1}{2n}\sum_{k=1}^{r}t_k^2+o(1/n)\to 0,$$

поэтому, воспользовавшись представлением

$$\ln(1+z)=z-z^2/2+o(z^2)\text{ при }z\to 0,$$

для $\ln\varphi_\zeta(t_1,t_2,\ldots,t_r)$ имеем (в явном виде выпишем только члены, стремящиеся к нулю при $n\to\infty$ не быстрее чем $o(1/n)$):

$$\ln\varphi_\zeta(t_1,t_2,\ldots,t_r)=-i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}+$$

$$+n\left(\frac{i}{\sqrt{n}}\sum_{k=1}^{r}t_k\sqrt{p_k}-\frac{1}{2n}\sum_{k=1}^{r}t_k^2-\frac{1}{2}\left(\frac{i}{\sqrt{n}}\sum_{k=1}^{r}t_k\sqrt{p_k}\right)^2+o\left(\frac{1}{n}\right)\right)=$$

$$=-i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}+i\sqrt{n}\sum_{k=1}^{r}t_k\sqrt{p_k}-\frac{1}{2}\sum_{k=1}^{r}t_k^2+$$

$$+\frac{1}{2}\left(\sum_{k=1}^{r}t_k\sqrt{p_k}\right)^2+o(1)=-\frac{1}{2}\sum_{k=1}^{r}t_k^2+\frac{1}{2}\left(\sum_{k=1}^{r}t_k\sqrt{p_k}\right)^2+o(1)=$$

$$=-\frac{1}{2}\left(\sum_{k=1}^{r}t_k^2-\left(\sum_{k=1}^{r}t_k\sqrt{p_k}\right)^2\right)+o(1).$$

Так что характеристическая функция случайного вектора $\zeta = (\zeta_1,\zeta_2,\ldots,\zeta_r)$

$$\varphi_\zeta(t_1,t_2,\ldots,t_r)=$$

$$=\exp\left\{-\frac{1}{2}\left(\sum_{k=1}^{r}t_k^2-\left(\sum_{k=1}^{r}t_k\sqrt{p_k}\right)^2\right)+o(1)\right\}.$$

Последняя при $n \to \infty$ сходится в каждой точке $t = (t_1, t_2, \ldots, t_r)$ к характеристической функции

$$\exp\left\{-\frac{1}{2}\left(\sum_{k=1}^{r} t_k^2 - \left(\sum_{k=1}^{r} t_k\sqrt{p_k}\right)^2\right)\right\} = \varphi_\alpha(t_1, t_2, \ldots, t_r), \tag{24.2.4}$$

являющейся характеристической функцией гауссовского вектора $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ со средним нуль и ковариационной матрицей Γ_α, определяемой квадратичной формой

$$(\Gamma_\alpha t, t) = \sum_{k=1}^{r} t_k^2 - \left(\sum_{k=1}^{r} t_k\sqrt{p_k}\right)^2, \tag{24.2.5}$$

заметим, что эта квадратичная форма неотрицательно определена, поскольку

$$\left(\sum_{k=1}^{r} t_k\sqrt{p_k}\right)^2 \le \left(\sum_{k=1}^{r} t_k^2\right)\left(\sum_{k=1}^{r}(\sqrt{p_k})^2\right) = \sum_{k=1}^{r} t_k^2.$$

Из сходимости

$$\varphi_\zeta(t_1, t_2, \ldots, t_r) \to \varphi_\alpha(t_1, t_2, \ldots, t_r)$$

в силу теоремы Леви (см. теорему 13.5.1), получаем, что при $n \to \infty$ распределение F_ζ вектора $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_r)$ сходится в собственном смысле к распределению F_α гауссовского вектора α (со средним нуль и ковариационной матрицей Γ_α).

Теперь установим, что при $n \to \infty$ распределение случайной величины

$$\triangle_n = \triangle_n(\zeta) = (\zeta, \zeta) = \sum_{j=1}^{r}\left(\frac{\nu_j - np_j}{\sqrt{np_j}}\right)^2$$

сходится к распределению случайной величины

$$\triangle = \triangle(\alpha) = (\alpha, \alpha) = \sum_{k=1}^{r} \alpha_k^2.$$

Воспользуемся теоремой Леви, для этого покажем, что характеристическая функция $\varphi_{\triangle_n}(u)$ случайной величины $\triangle_n = \triangle_n(\zeta)$

сходится к характеристической функции $\varphi_{\triangle}(u)$ случайной величины $\triangle = \triangle(\alpha)$.

Из сходимости распределения F_{ζ} вектора ζ к распределению F_{α} вектора α в силу теоремы о слабой сходимости распределений (см. теорему 12.3.1) следует, что при $n \to \infty$ характеристическая функция

$$\varphi_{\triangle_n}(u) = \mathsf{M}\exp\{iu\triangle_n\} = \mathsf{M}\exp\{iu(\zeta_1^2 + \zeta_2^2 + \ldots + \zeta_r^2)\} =$$

$$= \int\limits_{\mathbb{R}^r} \exp\left\{iu\left(z_1^2 + z_2^2 + \ldots + z_r^2\right)\right\} \mathsf{F}_{\zeta}(d(z_1, z_2, \ldots, z_r))$$

случайной величины $\triangle_n = (\zeta, \zeta)$ сходится к характеристической функции

$$\varphi_{\triangle}(u) = \mathsf{M}\exp\{iu\triangle(\alpha)\} =$$

$$= \int\limits_{\mathbb{R}^r} \exp\left\{iu\left(z_1^2 + z_2^2 + \ldots + z_r^2\right)\right\} \mathsf{F}_{\alpha}(d(z_1, z_2, \ldots, z_r))$$

случайной величины $\triangle = \triangle(\alpha) = (\alpha, \alpha)$ (при каждом фиксированном $u \in \mathbb{R}^1$ функция $\exp\left\{iu(z_1^2 + z_2^2 + \ldots + z_r^2)\right\}$ является непрерывной ограниченной функцией переменных $z_1, z_2, \ldots, z_r$). Поэтому в силу теоремы Леви (см. теорему 13.5.1) распределение случайной величины $\triangle_n(\zeta)$ сходится к распределению случайной величины $\triangle(\alpha)$.

Для завершения доказательства теоремы осталось доказать, что случайная величина $\triangle(\alpha) = \sum\limits_{k=1}^{r} \alpha_k^2 = (\alpha, \alpha)$ — сумма квадратов компонент гауссовского вектора $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r)$ с характеристической функцией (24.2.4), имеет распределение χ^2 с $r-1$ степенями свободы. Для этого рассмотрим гауссовский вектор

$$\beta = C\alpha,$$

полученный из гауссовского вектора α ортогональным преобразованием с матрицей C, первая строка которой имеет вид $(\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_r})$. Из ортогональности преобразования C следует, что

$$(\alpha, \alpha) = (C\alpha, C\alpha) = (\beta, \beta)$$

(здесь и далее β, α, s, t — вектор-столбцы). У гауссовского вектора $\beta = C\alpha$

$$\mathsf{M}\beta = \mathsf{M}C\alpha = C\mathsf{M}\alpha = 0.$$

А ковариационную матрицу $\Gamma_\beta = C\Gamma_\alpha C'$ вектора β по ковариационной матрице Γ_α вектора α получим, сделав в квадратичной форме

$$(\Gamma_\alpha t, t) = \sum_{k=1}^{r} t_k^2 - \left(\sum_{k=1}^{r} t_k\sqrt{p_k}\right)^2$$

замену $t = C's$ $(s = Ct)$. Имеем:

$$(\Gamma_\alpha C's, C's) = \sum_{k=1}^{r} s_k^2 - s_1^2$$

или

$$(C\Gamma_\alpha C's, s) = \sum_{k=2}^{r} s_k^2.$$

Так что

$$(\Gamma_\beta s, s) = \sum_{k=2}^{r} s_k^2$$

и ковариационная матрица вектора β имеет вид

$$\Gamma_\beta = \begin{Vmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{Vmatrix}.$$

Отсюда получаем, что компоненты гауссовского вектора β некоррелированы, а значит и независимы и каждая имеет распределение $\mathsf{N}_{0;1}$, исключая компоненту β_1, вырождающуюся в нуль. Поэтому по определению χ^2-распределения случайная величина

$$(\alpha, \alpha) = (\beta, \beta) = \sum_{k=2}^{r} \beta_k^2$$

имеет распределение χ^2 с $r-1$ степенями свободы.

Тем самым теорема полностью доказана.

Критерий χ^2 (гипотетическое распределение не зависит от параметров). Из теоремы 24.2.1 следует, что при верной гипотезе $H_0 : \mathsf{F} = \mathsf{G}$ для достаточно больших n в качестве распределения уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G}) = D(\hat{\mathsf{F}}_n, \mathsf{F}) = D_n$ можно рассматривать χ^2-распределение с $r-1$ степенями свободы. Поэтому $D_{\alpha;n}$ — минимальное D, для которого

$$\mathsf{P}\left\{D_n > D\right\} \leq \alpha$$

(см. неравенство (24.1.1)), можно считать равным минимальному D, для которого

$$\mathsf{P}\left\{\chi^2_{r-1} > D\right\} \leq \alpha,$$

т. е. можно считать, что $D_{\alpha;n} = \chi^2_{\alpha;r-1}$, где $\chi^2_{\alpha;r-1}$ — верхний α-предел χ^2-распределения с $r-1$ степенями свободы.

Критерий χ^2. *Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из неизвестного распределения F, $\chi^2_{\alpha;r-1}$ — верхний α-предел χ^2-распределения с $r-1$ степенями свободы.*

Если гипотезу H_0: $\mathsf{F} = \mathsf{G}$ (G не зависит от параметров) отклонять при

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}}\right)^2 > \chi^2_{\alpha;r-1}$$

и не отклонять в противном случае, то с вероятностью α гипотеза H_0 будет отклоняться, когда она верна.

З а м е ч а н и е к теореме К. Пирсона. В выражение уклонения К. Пирсона

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}}\right)^2$$

распределения G и $\hat{\mathsf{F}}_n$ входят только через гипотетические вероятности $p_j = \mathsf{G}(X_j)$ и частоты $\nu_j/n = \hat{\mathsf{F}}_n(X_j)$ попадания выборочных значений в подмножества X_j выборочного пространства X. И при доказательстве теоремы о распределении уклонения

$$\triangle_n = \sum_{j=1}^{r}\left(\frac{\nu_j - np_j}{\sqrt{np_j}}\right)^2$$

мы нигде не пользовались тем, что выборочное пространство является подмножеством $\mathbb{R}^1$. Поэтому теоремой К. Пирсона можно пользоваться, рассматривая уклонение между истинным и эмпирическим распределениями и для выборочных пространств, отличных от $\mathbb{R}^1$, см. далее критерий независимости признаков.

Проблема выбора классов. Сначала отметим один момент из практики использования критерия согласия χ^2.

Асимптотическая теория критерия согласия χ^2 справедлива при любом разбиении выборочного пространства X на непересекающиеся классы X_i $(i = 1, 2, \ldots, r)$:

$$X = \bigcup_{i=1}^{r} X_i,\ X_i \cap X_j = \emptyset,\ i \neq j,\ i, j = 1, 2, \ldots, r,$$

по которым группируются наблюдения, если только классы определяются безотносительно к последним (безотносительно к выборке). Однако обычно классы выбираются по выборке, а это обозначает, что границы классов случайны. Можно ли при этом пользоваться критерием χ^2? Оказывается, можно — асимптотическая теория справедлива и в описанной ситуации.

Далее относительно проблемы выбора классов.

1° Если данные в задаче дискретны или мы имеем выборку из дискретного распределения, то задача о выборе классов возникает только в смысле объединения некоторых классов с целью уменьшения погрешности при замене распределения уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F})$ распределением χ^2 как, например, для пуассоновского распределения необходимо объединять классы с малыми гипотетическими вероятностями.

Но по существу проблема выбора классов возникает тогда, когда мы имеем выборку из непрерывного распределения.

2° Метод равных вероятностей выбора классов.

Если из тех или иных соображений число классов r определено, то классы следует выбирать так, чтобы гипотетические вероятности $\mathsf{G}(X_j)$ были равны $1/r$ (и, следовательно, равны между собой) или, по-возможности, как можно ближе к $1/r$. В противном случае погрешности при замене распределения уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F})$ распределением χ^2 будут бо́льшими.

Определившись с границами классов, мы приходим к задаче о выборе числа классов.

3° Проблема выбора числа классов в методе равных вероятностей состоит в следующем. Если число классов r не достаточно велико, то критерий может терять "чувствительность", а именно, для гипотетического распределения G, существенно отличающегося от эмпирического распределения, тем не менее

$$\sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j}} \right)^2 \le \chi^2_{\alpha;r-1}$$

и гипотеза $H_0 : \mathsf{F} = \mathsf{G}$ не будет отклоняться. Если же число классов r слишком большое, то количество выборочных значений, попавших в классы, может быть малым, а это увеличивает погрешность при замене распределения уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F})$ распределением χ^2. На практике руководствуются следующим правилом: r выбираем так, чтобы

$$np_j \ge 10, \;\; j = 1, 2, \ldots, r.$$

Манн и Вальд доказали, что оптимальное (в смысле мощности критерия) число $r = r_n$ классов удовлетворяет соотношению

$$r_n \sim 4 \cdot 2^{1/5} \left(\frac{n}{N_{\alpha;0;1}} \right)^{2/5}$$

при $n \to \infty$, где $N_{\alpha;0;1}$ — верхний α-предел $N_{0;1}$-распределения (α — уровень значимости). Впоследствии было установлено, что без существенных потерь мощности критерия можно считать, что при больших n

$$r_n \approx 2 \cdot 2^{1/5} \left(\frac{n}{N_{\alpha;0;1}} \right)^{2/5}.$$

Рекомендации по применению критерия χ^2.

1° Использовать классы с равными или близкими к равным гипотетическими вероятностями попадания в эти классы.

2° Соотношением

$$r_n \approx 2 \cdot 2^{1/5} \left(\frac{n}{N_{\alpha;0;1}} \right)^{2/5},$$

пользуются для $\alpha = 0{,}05$ начиная с $n \geq 450$, а для $\alpha = 0{,}01$ начиная с $n \geq 300$.

3° Для $\alpha = 0{,}05$ и $n \leq 450$ и для $\alpha = 0{,}01$ и $n \leq 300$ число r классов выбирают по-возможности бóльшим, но при этом должны выполняться условия

$$np_i \geq 10, \; i = 1, 2, \ldots, r.$$

Если при данном разбиении для некоторых классов X_i значения

$$np_i < 10,$$

то рассматривают другое разбиение (чаще всего объединяя некоторые X_j) такое, чтобы для новых $X'_j, \; j = 1, 2, \ldots, r'$ неравенства

$$np'_j \geq 10, \; j = 1, 2, \ldots, r',$$

все таки выполнялись. Если же n настолько мало, что это сделать невозможно, критерием χ^2 не пользуются.

Обычно χ^2-распределение табулировано для числа степеней свободы, меньших 30. Если число степеней свободы больше 30, в качестве χ^2_r-распределения пользуются нормальным распределением с параметрами $(r; 2r)$, см. теорему 21.2.1.

Пример. *Бюффон, подбросив монету* $n = 4040$ *раз, получил* $\nu_1 = 2048$ *выпадений герба и* $\nu_0 = 1992$ *выпадений решетки. Совместима ли гипотеза: вероятность выпадения герба p равна* $1/2$ *(или, что то же, монета симметрична) с этими данными?*

Решение. Эксперимент с подбрасыванием монеты можно описать в терминах независимых наблюдений случайной величины ξ, принимающей два значения: 1, если выпал герб, и 0, если выпала решетка. Гипотеза о симметричности монеты в терминах распределения случайной величины ξ формулируется так: распределением ξ является

$$\mathsf{P}\{\xi = k\} = \mathsf{G}\{k\} = 1/2, \; k = 0, 1. \tag{24.2.6}$$

Проверим эту гипотезу.

Множество выборочных значений $X = \{0; 1\}$ разбивается на два подмножества $X_0 = \{0\}$ и $X_1 = \{1\}$. Вероятности попадания выборочных значений в эти подмножества, вычисленные по гипотетическому распределению (см. (24.2.6)), равны $p_0 = 1/2$, $p_1 = 1/2$. При этом

$$np_0 = 4040 \cdot \frac{1}{2} \geq 10,\ np_1 = 4040 \cdot \frac{1}{2} \geq 10.$$

Поэтому можно воспользоваться критерием χ^2. Значение уклонения между гипотетическим распределением и эмпирическим

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \frac{(\nu_0 - np_0)^2}{np_0} + \frac{(\nu_1 - np_1)^2}{np_1} =$$

$$= \frac{(1992 - 2020)^2}{2020} + \frac{(2048 - 2020)^2}{2020} = 0,776.$$

Значение $\chi^2_{\alpha;(r-1)}$ для $\alpha = 0{,}05$ и $r = 2$, равное $\chi^2_{0{,}05;1} = 3{,}84$ находим в таблице χ^2-распределения (см. табл. 27.2.1). Далее,

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = 0{,}776 < 3{,}84 = \chi^2_{0{,}05;1},$$

и, следовательно, гипотеза: распределением случайной величины ξ является

$$\mathsf{P}\{\xi = k\} = 1/2,\ k = 0, 1,$$

не отклоняется. Другими словами, гипотеза о симметричности монеты не противоречит опытным данным. При подбрасывании 4040 раз заведомо симметричной монеты появление герба 2048 раз вполне естественно и допустимо.

24.3 Критерий χ^2 (параметры неизвестны)

Гипотетическое распределение, не зависящее от параметров, в приложениях встречается редко. Чаще гипотетическое распределение $\mathsf{G}(\cdot; \theta_1, \theta_2, \ldots, \theta_k)$ (далее кратко $\mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)$) зависит от неизвестных параметров $\theta_1, \theta_2, \ldots, \theta_k$, относительно значений которых мы располагаем лишь той информацией, которая может быть извлечена из выборки. Можно ли в этой ситуации пользоваться критерием χ^2 и если да, то как?

Итак, пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из распределения F. Относительно F выдвигается гипотеза

$$H_0 : \mathsf{F} = \mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k),$$

параметры $\theta_1, \theta_2, \ldots, \theta_k$ неизвестны. Другими словами, гипотезу H_0 можно сформулировать так: F принадлежит классу распределений $\mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)$, т. е. совпадает с одним из распределений $\mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)$ при некоторых значениях параметров $\theta_1, \theta_2, \ldots, \theta_k$. Например, гипотезу

$$H_0 : \mathsf{F} = N_{a;\sigma^2},\ a \in \mathbb{R}^1,\ \sigma^2 \in (0, \infty)$$

можно сформулировать так: F является нормальным распределением. Наша задача — проверить гипотезу H_0.

Как и ранее, разобьем выборочное пространство (множество X) на r непересекающихся подмножеств $X_1, X_2, \ldots, X_r$. Обозначим число выборочных значений, попавших в множества $X_1, X_2, \ldots, X_r$ соответственно через $\nu_1, \nu_2, \ldots, \nu_r$, а вероятности попадания выборочных значений в $X_1, X_2, \ldots, X_r$, вычисленные по гипотетическому распределению $\mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)$, соответственно через $p_1(\theta_1, \theta_2, \ldots, \theta_k), \ldots, p_r(\theta_1, \theta_2, \ldots, \theta_k)$. Если бы "истинные значения" параметров θ_j, $j = 1, 2, \ldots, r$, были известны, то оставалось бы вычислить

$$D(\hat{\mathsf{F}}_n, \mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)) = \sum_{i=1}^{r} \frac{(\nu_i - np_i(\theta_1, \theta_2, \ldots, \theta_k))^2}{np_i(\theta_1, \theta_2, \ldots, \theta_k)}$$

и воспользоваться приведенным ранее критерием χ^2. Однако значения параметров θ_j, $j = 1, 2, \ldots, k$, а вместе с ними и вероятности $p_i(\theta_1, \theta_2, \ldots, \theta_k)$ неизвестны. В этой ситуации естественно поступать следующим образом: оценить параметры $\theta_1, \theta_2, \ldots, \theta_k$ по выборке $\xi_1, \xi_2, \ldots, \xi_n$ и в качестве гипотетического распределения рассмотреть $\mathsf{G}(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ ($\hat{\theta}_i$ — оценка θ_i, $i = 1, 2, \ldots, k$), а в качестве вероятности $p_i(\theta_1, \theta_2, \ldots, \theta_k)$ попадания в X_i рассмотреть $p_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$. Но $p_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ как функции выборки являются случайными величинами, и мы не можем утверждать, что при верной гипотезе H_0 уклонение

$$\triangle_n = \sum_{i=1}^{r} \frac{(\nu_i - np_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k))^2}{np_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)}$$

имеет распределение χ^2 с $r-1$ степенями свободы, а следовательно, не можем пользоваться критерием χ^2 в приведенной выше формулировке. В связи с этим, как и ранее, возникает задача: найти распределение уклонения $\triangle_n = \triangle_n(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ или хотя бы его предельное распределение при $n \to \infty$.

Различные методы оценивания параметров дают, вообще говоря, разные оценки и, следовательно, распределение случайной величины $\triangle_n(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ будет в той или иной мере зависеть от метода оценивания. Р. Фишер установил, что если неизвестные параметры $\theta_1, \theta_2, \ldots, \theta_k$ оценивать по выборке методом максимального правдоподобия, то при $n \to \infty$ распределение уклонения $\triangle_n(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ сходится к χ^2-распределению с числом степеней свободы $(r-1-k)$, где k — число параметров, оцененных по выборке. Поэтому для проверки гипотезы $H_0 : \mathsf{F} = \mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)$ можно пользоваться критерием χ^2, сравнивая $D(\hat{\mathsf{F}}_n, \mathsf{G}(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ с $\chi^2_{\alpha;(r-1-k)}$.

Если гипотезу H_0 отклонять при

$$D(\hat{\mathsf{F}}_n, \mathsf{G}(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)) =$$

$$= \sum_{i=1}^{r} \frac{(\nu_i - np_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k))^2}{np_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)} > \chi^2_{\alpha;(r-1-k)}$$

и не отклонять в противном случае, то с вероятностью α гипотеза H_0 будет отклоняться, когда она верна.

Критерием χ^2 в такой формулировке можно пользоваться, если объем выборки n и вероятности $p_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$, $i = 1, 2, \ldots$ $\ldots, r$, попадания выборочных значений в множества X_i, вычисленные по гипотетическому распределению, удовлетворяют неравенствам:

$$np_i(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \geq 10, \; i = 1, 2, \ldots, r.$$

Пример. *В течение Второй мировой войны на Лондон упало 537 самолетов-снарядов. Вся территория Лондона была разделена на 576 участков площадью 0,25 кв.км. В таблице, приведенной ниже, указано число ν_i участков, на которые упало i снарядов.*

i	0	1	2	3	4	5	6	7
ν_i	229	211	93	35	7	0	0	1

С помощью критерия χ^2 проверить гипотезу, что число самолетов-снарядов, упавших на участок, имеет распределение Пуассона.

Решение. Имеется $n = 576$ независимых наблюдений (по числу участков) случайной величины ξ — числа самолетов-снарядов, упавших на участок. Заметим, что наблюдения представлены не выборкой, а таблицей, фактически задающей эмпирическое распределение случайной величины — числа самолетов-снарядов, упавших на участок:

i	0	1	2	3	4	5	6	7
$\frac{\nu_i}{n}$	$\frac{229}{576}$	$\frac{211}{576}$	$\frac{93}{576}$	$\frac{35}{576}$	$\frac{7}{576}$	$\frac{0}{576}$	$\frac{0}{576}$	$\frac{1}{576}$

.

Относительно неизвестного распределения $\mathsf{P}_\xi(k)$, $k = 0, 1, \ldots$, случайной величины ξ выдвигается гипотеза

$$H_0 : \mathsf{P}_\xi(k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \ldots,$$

которую необходимо проверить.

Параметр λ гипотетического распределения неизвестен. В качестве λ рассмотрим значение $\hat{\lambda}$ оценки максимального правдоподобия этого параметра:

$$\hat{\lambda} = \bar{\xi} = \frac{1}{576} \sum_{i=1}^{576} \xi_i =$$

$$= \frac{0 \cdot 229 + 1 \cdot 211 + 2 \cdot 93 + 3 \cdot 35 + 4 \cdot 7 + 7 \cdot 1}{576} \approx 0{,}9$$

(заметим, что хотя выборка $\xi_1, \xi_2, \ldots, \xi_{576}$ явно не дана, оценку $\hat{\lambda} = \bar{\xi}$ получить можно). Так что гипотетическое распределение имеет вид

$$\mathsf{P}_\xi(k) = \frac{0{,}9^k}{k!} e^{-0{,}9},\ k = 0, 1, \ldots,$$

(значения $\mathsf{P}_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, \ldots$, для различных λ табулированы, см., например, табл. 27.6.1).

Далее, пользуясь описанной выше методикой применения критерия χ^2, разбиваем выборочное пространство $X = \{0, 1, 2, \ldots\}$

— множество возможных значений случайной величины ξ на непересекающиеся подмножества X_i, причем так, чтобы для каждого из них выполнялось условие $np_i > 10$. Для подмножеств $X_i = \{i\}$, $i = 0, 1, \ldots, 4$, имеем:

$$X_0 = \{0\}, \ p_0 = \frac{0{,}9^0}{0!}e^{-0{,}9} = 0{,}4066, \ np_0 = 234 > 10;$$

$$X_1 = \{1\}, \ p_1 = \frac{0{,}9^1}{1!}e^{-0{,}9} = 0{,}3659, \ np_1 = 211 > 10;$$

$$X_2 = \{2\}, \ p_2 = 0{,}1646, \ np_2 = 97 > 10;$$

$$X_3 = \{3\}, \ p_3 = 0{,}0493, \ np_3 = 29 > 10;$$

$$X_4 = \{4\}, \ p_4 = 0{,}0111, \ np_4 = 6 < 10.$$

Причем неравенство $np_4 < 10$ имеет место не только для $X_4 = \{4\}$, но и для $X_4' = \{4, 5, 6, \ldots\}$. Поэтому подмножества X_4 в разбиении выборочного пространства $X = \{0, 1, 2, \ldots\}$ не будет, а в качестве X_3 вместо ранее выбранного $X_3 = \{3\}$ рассмотрим $X_3 = \{3, 4, \ldots\}$. Для этого нового X_3 вероятность $\mathsf{P}\{\xi \in X_3\} = \mathsf{P}\{\xi \geq 3\} = 0{,}0629$, а $np_3 = 36$. Так что для каждого X_i, $i = 0, 1, 2, 3$, значение $np_i > 10$. Далее,

$$D(\hat{\mathsf{F}}_{576}, \mathsf{G}) = \sum_{i=0}^{3} \frac{(\nu_i - np_i)^2}{np_i} = \frac{(229 - 234)^2}{234} + \frac{(211 - 211)^2}{211} +$$

$$+ \frac{(93 - 97)^2}{97} + \frac{(43 - 36)^2}{36} = 1{,}53.$$

Значение $D(\hat{\mathsf{F}}_{576}, \mathsf{G})$ сравниваем с $\chi^2_{\alpha;(r-1-k)}$, где k — число параметров, оцененных по выборке, r — количество подмножеств, на которые разбито выборочное пространство. У нас $r = 4$, $k = 1$, $\chi^2_{\alpha;(r-1-k)} = \chi^2_{\alpha;2}$. Значение $\chi^2_{0{,}05;2} = 5{,}99$ найдено по таблице χ^2-распределения (см. табл. 27.2.1). Имеем:

$$D(\hat{\mathsf{F}}_{576}, \mathsf{G}) = 1{,}53 < 5{,}99 = \chi^2_{0{,}05;2},$$

поэтому на 5 % уровне значимости гипотеза

$$H_0 : \mathsf{P}_\xi(k) = \frac{0{,}9^k}{k!}e^{-0{,}9}, \ k = 0, 1, \ldots,$$

о пуассоновском распределении числа самолетов-снарядов, упавших на участок, не отклоняется (не противоречит имеющимся данным). При наблюдении 576 значений случайной величины ξ, заведомо имеющей распределение

$$p_k = \mathsf{P}_\xi(k) = \frac{0{,}9^k}{k!}\, e^{-0{,}9}, k = 0, 1, \ldots,$$

появление значений $0, 1, 2, \ldots$ в количестве, указанном в таблице, вполне естественно и допустимо.

24.4 Критерий χ^2 как критерий независимости

Постановка задачи. Имеется n независимых наблюдений (a_i, b_j), $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$, векторной случайной величины $\zeta = (\xi, \eta)$ со значениями в $\mathbb{R}^2$. При этом значение (a_i, b_j) случайная величина $\zeta = (\xi, \eta)$ приняла ν_{ij} раз, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$ (ясно, что $\sum\limits_{i=1}^{s}\sum\limits_{j=1}^{k}\nu_{ij} = n$). Результаты наблюдений удобно представлять в виде так называемой *таблицы сопряженности признаков* (см. табл. 24.4.1), которая фактически задает эмпирическое распределение случайной величины $\zeta = (\xi, \eta)$ (см. табл. 24.4.2), заметим, что крайний правый столбец табл. 24.4.2 задает эмпирическое распределение ξ, а нижняя строка задает эмпирическое распределение η. Но ни распределение векторной случайной величины $\zeta = (\xi, \eta)$:

$$\mathsf{P}\{\zeta = (a_i, b_j)\} = p_{i,j},\ i = 1, 2, \ldots, s;\ j = 1, 2, \ldots, k \qquad (24.4.1)$$

(см. также табл. 24.4.3), ни распределения ее компонент ξ и η:

$$\mathsf{P}\{\xi = a_i\} = p_{i\cdot},\ i = 1, 2, \ldots, s;\ \mathsf{P}\{\eta = b_j\} = p_{\cdot j},\ j = 1, 2, \ldots, k$$

неизвестны.

Т а б л и ц а 24.4.1. Таблица сопряженности признаков

Значения ξ	Значения η				Сумма
	b_1	b_2	$\cdots$	b_k	
a_1	ν_{11}	ν_{12}	$\cdots$	ν_{1k}	$\nu_{1\cdot}$
a_2	ν_{21}	ν_{22}	$\cdots$	ν_{2k}	$\nu_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_s	ν_{s1}	ν_{s2}	$\cdots$	ν_{sk}	$\nu_{s\cdot}$
Сумма	$\nu_{\cdot 1}$	$\nu_{\cdot 2}$	$\cdots$	$\nu_{\cdot k}$	n

В табл. 24.4.1

$$\nu_{i\cdot} = \sum_{j=1}^{k} \nu_{ij},\ \nu_{\cdot j} = \sum_{i=1}^{s} \nu_{ij},\ i = 1, 2, \ldots, s,\ j = 1, 2, \ldots, k.$$

Т а б л и ц а 24.4.2. Эмпирическое распределение $\zeta = (\xi, \eta)$

Значения ξ	Значения η				Сумма
	b_1	b_2	$\cdots$	b_k	
a_1	ν_{11}/n	ν_{12}/n	$\cdots$	ν_{1k}/n	$\nu_{1\cdot}/n$
a_2	ν_{21}/n	ν_{22}/n	$\cdots$	ν_{2k}/n	$\nu_{2\cdot}/n$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_s	ν_{s1}/n	ν_{s2}/n	$\cdots$	ν_{sk}/n	$\nu_{s\cdot}/n$
Сумма	$\nu_{\cdot 1}/n$	$\nu_{\cdot 2}/n$	$\cdots$	$\nu_{\cdot k}/n$	1

Таблица 24.4.3. Распределение $\zeta = (\xi, \eta)$

Значения ξ	Значения η				Сумма
	b_1	b_2	$\cdots$	b_k	
a_1	p_{11}	p_{12}	$\cdots$	p_{1k}	$p_{1\cdot}$
a_2	p_{21}	p_{22}	$\cdots$	p_{2k}	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_s	p_{s1}	p_{s2}	$\cdots$	p_{sk}	$p_{s\cdot}$
Сумма	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot k}$	1

В табл. 24.4.3

$$p_{ij} = \mathsf{P}\{(\xi, \eta) = (a_i, b_j)\},\ p_{i\cdot} = \sum_{j=1}^{k} p_{ij},\ p_{\cdot j} = \sum_{i=1}^{s} p_{ij}.$$

Относительно совместного распределения случайных величин ξ и η выдвигается гипотеза

$$H_0 : p_{ij} = p_{i\cdot}\, p_{\cdot j},\ i = 1, 2, \ldots, s,\ j = 1, 2, \ldots, k,$$

— совместное распределение $p_{i,j}$, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$ (распределение вектора $\zeta = (\xi, \eta)$) равно произведению распределений компонент (см. табл. 24.4.4)). Другими словами, H_0 — это гипотеза о независимости случайных величин ξ и η (еще говорят, *гипотеза о независимости признаков*). Необходимо проверить H_0.

Для проверки гипотезы

$$H_0 : p_{ij} = p_{i\cdot}\, p_{\cdot j},\ i = 1, 2, \ldots, s,\ j = 1, 2, \ldots, k,$$

воспользуемся критерием χ^2 применительно к случаю гипотетического распределения

$$\mathsf{G} = \mathsf{G}(p_{1\cdot}, p_{2\cdot}, \ldots, p_{s\cdot}, p_{\cdot 1}, p_{\cdot 2}, p_{\cdot k}) = p_{i\cdot} p_{\cdot j},$$

зависящего от неизвестных параметров $p_{i\cdot}$ и $p_{\cdot j}$, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$, причем

$$\sum_{i=1}^{s} p_{i\cdot} = 1,\ \sum_{j=1}^{k} p_{\cdot j} = 1.$$

Таблица 24.4.4. Гипотетическое распределение $\zeta = (\xi, \eta)$

Значения ξ	Значения η				Сумма
	b_1	b_2	$\cdots$	b_k	
a_1	$p_{1\cdot}p_{\cdot 1}$	$p_{1\cdot}p_{\cdot 2}$	$\cdots$	$p_{1\cdot}p_{\cdot k}$	$p_{1\cdot}$
a_2	$p_{2\cdot}p_{\cdot 1}$	$p_{2\cdot}p_{\cdot 2}$	$\cdots$	$p_{2\cdot}p_{\cdot k}$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_s	$p_{s\cdot}p_{\cdot 1}$	$p_{s\cdot}p_{\cdot 2}$	$\cdots$	$p_{s\cdot}p_{\cdot k}$	$p_{s\cdot}$
Сумма	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot k}$	1

Уклонение эмпирического распределения ν_{ij}/n, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$, случайной величины $\zeta = (\xi, \eta)$ от гипотетического из класса распределений $p_{ij} = p_{i\cdot}\, p_{\cdot j}$, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$, равно

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{i=1}^{s} \sum_{j=1}^{k} \frac{(\nu_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}},$$

где $\hat{p}_{i\cdot}$ и $\hat{p}_{\cdot j}$ — оценки максимального правдоподобия соответственно параметров $p_{i\cdot}$ и $p_{\cdot j}$. В качестве разбиения выборочного пространства случайной величины $\zeta = (\xi, \eta)$ рассматриваются подмножества $X_i \times Y_j = (a_i, b_j)$, $i = 1, 2, \ldots, s, j = 1, 2, \ldots, k$.

Найдем оценки максимального правдоподобия неизвестных параметров $p_{i\cdot}$ и $p_{\cdot j}$, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$. Сначала выпишем функцию максимального правдоподобия выборки $\zeta_1, \zeta_2, \ldots, \zeta_n$. Распределением каждой из случайных величин $\zeta_l = (\xi_l, \eta_l)$, $l = 1, 2, \ldots, n$, является

$$\mathsf{P}\{\zeta_l = (a_i, b_j)\} = p_{i\cdot}p_{\cdot j},\ \ i = 1, 2, \ldots, s;\ \ j = 1, 2, \ldots, k.$$

Совместное распределение независимых случайных величин $\zeta_1, \zeta_2, \ldots, \zeta_n$ равно произведению их распределений, так что функция максимального правдоподобия

$$L = L(p_{1\cdot}, p_{2\cdot}, \ldots, p_{s\cdot}, p_{\cdot 1}, p_{\cdot 2}, \ldots, p_{\cdot k}) =$$

$$= \mathsf{P}\{\zeta_1 = (a_{i_1}, b_{j_1}), \zeta_2 = (a_{i_2}, b_{j_2}), \ldots, \zeta_n = (a_{i_n}, b_{j_n})\} =$$

$$= \mathsf{P}\{\zeta_1 = (a_{i_1}, b_{j_1})\}\mathsf{P}\{\zeta_2 = (a_{i_2}, b_{j_2})\}\ldots\mathsf{P}\{\zeta_n = (a_{i_n}, b_{j_n})\}.$$

Число n сомножителей в этом произведении равно $\sum\limits_{i=1}^{s}\sum\limits_{j=1}^{k}\nu_{ij}$, причем среди них каждый сомножитель $\mathsf{P}\{\zeta = (a_i, b_j)\} = p_{i\cdot}p_{\cdot j}$ встречается ν_{ij} раз, $i = 1, 2, \ldots, s$; $j = 1, 2, \ldots, k$. Поэтому функцию правдоподобия можно переписать так:

$$\begin{gathered}
L = L(p_{1\cdot}, p_{2\cdot}, \ldots, p_{s\cdot}, p_{\cdot 1}, p_{\cdot 2}, \ldots, p_{\cdot k}) = \\
= (p_{1\cdot}p_{\cdot 1})^{\nu_{11}}(p_{1\cdot}p_{\cdot 2})^{\nu_{12}}\ldots(p_{1\cdot}p_{\cdot k})^{\nu_{1k}}\times \\
\times(p_{2\cdot}p_{\cdot 1})^{\nu_{21}}(p_{2\cdot}p_{\cdot 2})^{\nu_{22}}\ldots(p_{2\cdot}p_{\cdot k})^{\nu_{2k}}\times\ldots \\
\ldots(p_{s\cdot}p_{\cdot 1})^{\nu_{s1}}(p_{s\cdot}p_{\cdot 2})^{\nu_{s2}}\ldots(p_{s\cdot}p_{\cdot k})^{\nu_{sk}} = \\
= p_{1\cdot}^{\nu_{1\cdot}}p_{2\cdot}^{\nu_{2\cdot}}\ldots p_{s\cdot}^{\nu_{s\cdot}}p_{\cdot 1}^{\nu_{\cdot 1}}p_{\cdot 2}^{\nu_{\cdot 2}}\ldots p_{\cdot k}^{\nu_{\cdot k}} = \prod_{i=1}^{s}p_{i\cdot}^{\nu_{i\cdot}}\prod_{j=1}^{k}p_{\cdot j}^{\nu_{\cdot j}}
\end{gathered}$$

(см. также табл. 24.4.1 и табл. 24.4.4). Логарифмическая функция максимального правдоподобия равна

$$\ln L = \sum_{i=1}^{s}\nu_{i\cdot}\ln p_{i\cdot} + \sum_{j=1}^{k}\nu_{\cdot j}\ln p_{\cdot j}.$$

Оценкой максимального правдоподобия параметров $p_{i\cdot}$, $p_{\cdot j}$, $i = 1, 2, \ldots, s$; $j = 1, 2, \ldots, k$, является точка, в которой функция L (или $\ln L$) достигает наибольшего значения, причем

$$\mathsf{G}_1 = \sum_{i=1}^{s}p_{i\cdot} - 1 = 0,\ \ \mathsf{G}_2 = \sum_{j=1}^{k}p_{\cdot j} - 1 = 0. \qquad (24.4.2)$$

Чтобы при условиях (24.4.2) найти точку максимума функции $\ln L$, сначала найдем ее точки экстремумов. Воспользуемся методом неопределенных множителей Лагранжа: точка экстремума функции $\ln L$, является решением системы уравнений

$$\begin{cases}
\dfrac{\partial}{\partial p_{i\cdot}}(\ln L - \lambda_1\mathsf{G}_1 - \lambda_2\mathsf{G}_2) = 0,\ \ i = 1, 2, \ldots, s; \\
\dfrac{\partial}{\partial p_{\cdot j}}(\ln L - \lambda_1\mathsf{G}_1 - \lambda_2\mathsf{G}_2) = 0,\ \ j = 1, 2, \ldots, k; \\
\mathsf{G}_1 = 0; \\
\mathsf{G}_2 = 0
\end{cases}$$

или, что то же,

$$\begin{cases} \nu_{i\cdot}\dfrac{1}{p_{i\cdot}} - \lambda_1 = 0, \ i = 1, 2, \ldots, s; \\ \nu_{\cdot j}\dfrac{1}{p_{\cdot j}} - \lambda_2 = 0, \ j = 1, 2, \ldots, k; \\ \sum\limits_{i=1}^{s} p_{i\cdot} = 1; \\ \sum\limits_{j=1}^{k} p_{\cdot j} = 1. \end{cases}$$

Из первых s и последующих k уравнений имеем:

$$p_{i\cdot} = \nu_{i\cdot}/\lambda_1, \ i = 1, 2, \ldots, s,$$

$$p_{\cdot j} = \nu_{\cdot j}/\lambda_2, \ j = 1, 2, \ldots, k.$$

Подставляя $p_{i\cdot}$ и $p_{\cdot j}$ в оставшиеся два уравнения, получаем $\lambda_1 = n$, $\lambda_2 = n$ и, следовательно, решением последней системы является

$$p_{i\cdot} = \nu_{i\cdot}/n, \ i = 1, 2, \ldots, s; \ p_{\cdot j} = \nu_{\cdot j}/n, \ j = 1, 2, \ldots, k,$$

(это точки, подозрительные на экстремум). В граничных точках множества

$$0 \le p_{i\cdot} \le 1, \ i = 1, 2, \ldots, s, \ \sum_{i=1}^{s} p_{i\cdot} = 1,$$

$$0 \le p_{\cdot j} \le 1, \ j = 1, 2, \ldots, k, \ \sum_{j=1}^{k} p_{\cdot j} = 1,$$

возможных значений параметров $p_{i\cdot}$ и $p_{\cdot j}$ значение функции правдоподобия

$$L = \prod_{i=1}^{s} p_{i\cdot}^{\nu_{i\cdot}} \prod_{j=1}^{k} p_{\cdot j}^{\nu_{\cdot j}}$$

равно нулю. Поэтому наибольшего значения функция максимального правдоподобия достигает при

$$p_{i\cdot} = \nu_{i\cdot}/n, \ i = 1, 2, \ldots, s; \ p_{\cdot j} = \nu_{\cdot j}/n, \ j = 1, 2, \ldots, k,$$

и следовательно,

$$\hat{p}_{i\cdot} = \nu_{i\cdot}/n,\ i = 1, 2, \ldots, s;\ \hat{p}_{\cdot j} = \nu_{\cdot j}/n,\ j = 1, 2, \ldots, k,$$

являются оценками максимального правдоподобия соответственно параметров $p_{i\cdot}, p_{\cdot j}$, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$. Заметим, что оценкой параметра $p_{i\cdot}$ является частота $\nu_{i\cdot}/n$ попадания ξ в $\{a_i\}$, $i = 1, 2, \ldots, s$, оценкой параметра $p_{\cdot j}$ — частота $\nu_{\cdot j}/n$ попадания η в $\{b_j\}$, $j = 1, 2, \ldots, k$ (см. также табл. 24.4.2).

Получив оценки максимального правдоподобия неизвестных параметров, уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ можно переписать так

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(\nu_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}} = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(\nu_{ij} - (\nu_{i\cdot}\nu_{\cdot j})/n)^2}{(\nu_{i\cdot}\nu_{\cdot j})/n}.$$

Число параметров $p_{i\cdot}, p_{\cdot j}$, оцененных по выборке, равно $(s-1)+$ $+(k-1)$ (поскольку $\sum\limits_{i=1}^{s} p_{i\cdot} = 1$, $\sum\limits_{j=1}^{k} p_{\cdot j} = 1$). Количество подмножеств $X_{i,j}$, $i = 1, 2, \ldots, s$, $j = 1, 2, \ldots, k$, на которые разбито выборочное пространство, равно $r = sk$. Поэтому при $n \to \infty$ распределение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F})$ сходится к χ^2-распределению с $sk - 1 - ((s-1) + (k-1)) = (s-1)(k-1)$ степенями свободы.

Итак, *если гипотезу H_0 о независимости компонент вектора $\zeta = (\xi, \eta)$ отклонять при*

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(\nu_{ij} - (\nu_{i\cdot}\nu_{\cdot j})/n)^2}{(\nu_{i\cdot}\nu_{\cdot j})/n} > \chi^2_{\alpha;(s-1)(k-1)}$$

и не отклонять в противном случае, то с вероятностью α гипотеза H_0 будет отклоняться, когда она верна.

З а м е ч а н и е. Случайные величины ξ и η еще называют признаками и в этом случае говорят о критерии χ^2 для независимости признаков.

При разбиении выборочного пространства $X \times Y$ двумерной случайной величины $\zeta = (\xi, \eta)$ на непересекающиеся подмножества $X_i \times Y_j$ необходимо следить, чтобы для гипотетических вероятностей $\hat{p}_{i\cdot}\hat{p}_{\cdot j}$ попадания значений (ξ, η) в $X_i \times Y_j$ выполнялись неравенства

$$(\hat{p}_{i\cdot}\hat{p}_{\cdot j})n = (\nu_{i\cdot}\nu_{\cdot j})/n \geq 10,\ i = 1, 2, \ldots, s,\ j = 1, 2, \ldots, k,$$

где n — объем выборки.

Пример 24.4.1 (прививка против холеры). *В таблице (Гринвуд, Юл,* 1915) *приведены данные о* 818 *наблюдениях, классифицированных по двум признакам: наличие прививки против холеры и отсутствие заболевания.*

Можно ли на основании этих данных прийти к выводу о зависимости между отсутствием заболевания и наличием прививки?

Наличие прививки	Наличие заболевания		Всего
	Не заболели	Заболели	
Привитые	276	3	279
Не привитые	473	66	539
Всего	749	69	818

Решение. В терминах проверки статистических гипотез поставленная задача формулируется как задача проверки гипотезы о независимости признаков (наличие прививки и отсутствие заболевания).

В рассматриваемом примере $s = 2$, $k = 2$, $n = 818$. Значения ν_{ij}, $\nu_{i\cdot}$, $\nu_{\cdot j}$, $i, j = 1, 2$, приведены в таблице сопряженности признаков. Поскольку

$$(\hat{p}_{i\cdot}\hat{p}_{\cdot j})n = \nu_{i\cdot}\nu_{\cdot j}/n \geq 10, \ \ i, j = 1, 2,$$

то для проверки гипотезы о независимости признаков можно воспользоваться критерием χ^2.

Значение уклонения между эмпирическим распределением $\hat{\mathsf{F}}_n$ и гипотетическим G:

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(\nu_{ij} - (\nu_{i\cdot}\nu_{\cdot j})/n)^2}{(\nu_{i\cdot}\nu_{\cdot j})/n} =$$

$$= \frac{(276 - 279 \cdot 749/818)^2}{279 \cdot 749/818} + \frac{(3 - 279 \cdot 69/818)^2}{279 \cdot 69/818} +$$

$$+ \frac{(473 - 539 \cdot 749/818)^2}{539 \cdot 749/818} + \frac{(66 - 539 \cdot 69/818)^2}{539 \cdot 69/818} = 29{,}61.$$

Таким образом,

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = 29{,}61 > 3{,}84 = \chi^2_{0{,}05;1} = \chi^2_{\alpha;(s-1)(k-1)}.$$

Поэтому согласно критерию χ^2 гипотеза о независимости признаков (отсутствие заболевания и наличие прививки) отклоняется; она противоречит имеющимся данным. Другими словами, приведенные данные наблюдений дают основания утверждать, что эффект прививки против холеры существует.

24.5 Примеры и задачи

Примеры

Пример 24.5.1. *Среди* 2020 *семей с двумя детьми* 527 *имеют двух мальчиков,* 476 *— двух девочек, у остальных* 1017 *семей дети разного пола.*

Можно ли считать, что количество мальчиков в семье, имеющей двух детей, является биномиально распределенной случайной величиной?

Решение. Обозначим через ξ количество мальчиков в семье с двумя детьми, ξ принимает значения: 0, 1, 2. Относительно распределения случайной величины ξ выдвигается гипотеза

$$H_0\colon \mathsf{P}\{\xi = k\} = \mathsf{P}_\xi(k) = C_2^k p^k (1-p)^{2-k},\ k = 0, 1, 2,$$

которую и необходимо проверить.

Гипотетическое распределение зависит от неизвестного параметра p. Оценкой максимального правдоподобия параметра p биномиального распределения

$$\mathsf{P}(k; p) = C_m^k p^k (1-p)^{m-k},\ k = 0, 1, \ldots, m,$$

полученной по выборке $\xi_1, \xi_2, \ldots, \xi_n$, является

$$\hat{p} = \frac{1}{mn} \sum_{k=1}^{n} \xi_k.$$

В рассматриваемом примере значение оценки максимального правдоподобия параметра p равно:

$$\hat{p} = \frac{0 \cdot 476 + 1 \cdot 1017 + 2 \cdot 527}{2 \cdot 2020} = 0{,}513.$$

Таким образом, гипотетическим распределением при значении параметра p, равном $\hat{p} = 0{,}513$, является

$$\mathsf{P}(k) = C_2^k (0{,}513)^k (1 - 0{,}513)^{2-k},\ k = 0, 1, 2.$$

Его можно записать и в таком виде:

$$\mathsf{G} = \begin{pmatrix} 0 & 1 & 2 \\ 0{,}237 & 0{,}500 & 0{,}263 \end{pmatrix}.$$

В качестве разбиения выборочного пространства $X = \{0, 1, 2\}$ рассмотрим $X_0 = \{0\}$, $X_1 = \{1\}$, $X_2 = \{2\}$. Для множеств X_i оценки $\hat{p}_i$ попадания выборочных значений в X_i, $i = 0, 1, 2$, соответственно равны 0,237, 0,500, 0,263, объем выборки $n = 2020$, поэтому

$$n\hat{p}_i \geq 10, \ i = 0, 1, 2,$$

и следовательно, можно воспользоваться критерием χ^2 для проверки гипотезы H_0: $\mathsf{F} = \mathsf{G}(\theta_1, \theta_2, \ldots, \theta_k)$ когда гипотетическое распределение зависит от параметров.

Значение уклонения Пирсона между эмпирическим $\hat{\mathsf{F}}_n$ и гипотетическим G распределениями

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \frac{(\nu_0 - n\hat{p}_0)^2}{n\hat{p}_0} + \frac{(\nu_1 - n\hat{p}_1)^2}{n\hat{p}_1} + \frac{(\nu_2 - n\hat{p}_2)^2}{n\hat{p}_2} =$$

$$= \frac{(476 - 478{,}74)^2}{478{,}74} + \frac{(1017 - 1010)^2}{1010} + \frac{(527 - 531{,}26)^2}{531{,}26} = 0{,}098.$$

Так что

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = 0{,}098 < 3{,}84 = \chi^2_{0{,}05;1} = \chi^2_{\alpha;(r-1-k)}$$

($r - 1 - k = 3 - 1 - 1 = 1$, где $k = 1$ — количество параметров, которые были оценены по выборке, $r = 3$ — число подмножеств, на которые делится выборочное пространство). Поэтому согласно критерию χ^2 гипотеза о биномиальном распределении случайной величины ξ на 5% уровне значимости не отклоняется. Предположение, что количество мальчиков в семьях с двумя детьми имеет биномиальное распределение, не противоречит имеющимся данным.

Задачи

24.1 (распределение $\boldsymbol{\alpha}$-частиц). В эксперименте Резерфорда и Гейгера радиоактивное вещество наблюдалось на протяжении $N = 2612$ промежутков времени (каждый длительностью $1/8$ мин), для которых регистрировалось количество α-частиц, достигающих счетчика.

В таблице приведено количество N_k промежутков времени, на протяжении которых наблюдалось ровно k α-частиц. Общее число частиц $T = \sum\limits_k kN_k = 10\,132$. Среднее их количество, зарегистрированное за один промежуток, составляет $\lambda = T/N = 10\,132/2\,612 = 3{,}879$.

k	N_k	k	N_k
0	57	6	273
1	203	7	139
2	383	8	49
3	525	9	27
4	532	10	10
5	408	> 10	6
		Всего	2612

Воспользовавшись критерием χ^2, проверить гипотезу о пуассоновском распределении числа частиц, достигающих счетчика, за промежуток времени длительностью $1/8$ мин.

24.2 (количество смертей от удара копытом). Далее приведены классические данные фон Борткевича о количестве лиц, убитых ударом копыта лошади в 10 прусских армейских корпусах за 20 лет (1875 – 1895):

i	0	1	2	3	4	5 и более	Всего
n_i	109	65	22	3	1	0	200

В таблице: i — количество смертей в одном корпусе за год; n_i — количество наблюдений, в которых произошло i смертей.

Проверить гипотезу о пуассоновском распределении количества смертей от удара копытом лошади в одном корпусе на протяжении года.

24.3 (смышленость и материальные условия). В таблице приведены результаты обследования 697 школьников.

Мальчики были упорядочены согласно IQ и в соответствии с условиями их жизни дома. При этом использованы обозначения: A — очень способный, B — достаточно умный, C — имеет средние способности, D — недостаточно развит, E — умственно отсталый.

Можно ли считать, что условия жизни (обеспеченность) детей влияют на их смышленость?

Обеспеченность	Смышленость мальчиков					Всего
	A	B	C	D	E	
Хорошая	33	137	125	47	8	350
Плохая	21	127	129	61	9	347
Всего	54	264	254	108	17	697

З а м е ч а н и е. IQ —Intellectual quality (умственные способности) — показатель умственных способностей учеников в баллах, используемый в американской педагогической практике.

24.4 (предвзятость наблюдателя). Перед датчиком, состоящим из круглого диска, разделенного на 10 одинаковых секторов, пронумерованных цифрами от 0 до 9, поставили наблюдателя. Диск вращался с большой скоростью. Время от времени перед ним вспыхивала электрическая лампочка на столь короткое время, что диск казался неподвижным. Наблюдатель должен был смотреть на диск и записывать номер того сектора, который в момент вспышки лампочки находился напротив фиксированного указателя.

Прибор предназначался для получения случайных чисел и в самом деле их доставлял при работе с другим наблюдателем. Но наблюдатель, о котором шла речь выше, снимал показания с заметной предвзятостью.

Частоты появления цифр в 10 000 выполненных им наблюдений приведены в таблице (в таблице частота цифры — количество ее появлений на 10 000 наблюдений).

Проверить, действительно ли приведенные данные указывают на пристрастие наблюдателя.

Цифра	Частота	Цифра	Частота
0	1083	5	1007
1	865	6	1081
2	1053	7	997
3	884	8	1025
4	1057	9	948
		Всего	10 000

З а м е ч а н и е. Если бы наблюдатель был непредвзятым, то цифры появлялись бы приблизительно с одинаковой частотой. Из таблицы, однако, видно, что он имел пристрастие к четным цифрам и предубеждение против нечетных цифр 1, 3 и 9. Причина такого поведения и пристрастия неясна и загадочна, поскольку наблюдатель не должен был производить оценку, а должен был только записывать то, что видел, или думал, что видит. Объяснение, по-видимому, состоит в том, что наблюдатель отдавал сильное предпочтение некоторым цифрам, т. е. фиксировал не те из них, которые в момент вспышки лампочки в самом деле видел напротив указателя. Здесь мы имеем дело с одной из чрезвычайно сильных форм психологической предвзятости.

24.5 (перестройка хромосом в клетке). Рентгеновское облучение вызывает в органических клетках определенные процессы, которые будем называть перестройкой хромосом. Число перестроек хромосом в клетке согласно теории должно подчиняться распределению Пуассона. Ниже приведены результаты эксперимента (подсчитывалось количество перестроек хромосом в клетке под воздействием рентгеновских лучей)

i	0	1	2	3	4 и более	Всего
n_i	434	195	44	9	0	682

.

В таблице: i — количество перестроек хромосом в клетке; n_i — число клеток, в которых наблюдалось i перестроек хромосом.

Согласуется ли с приведенными данными гипотеза о пуассоновском распределении числа перестроек в клетке?

24.6 (случайно и произвольно). Случайно не означает произвольно: случайность подчиняется своим строгим законам. Например, на первый взгляд кажется, что предложить случайную последовательность цифр просто. Но на самом деле это могут сделать лишь очень немногие люди. Последовательность случайных цифр должна удовлетворять ряду требований случайности. В частности, от такой последовательности естественно потребовать, чтобы появление цифр $0, 1, \ldots, 9$ было равновероятным.

Проверьте свои способности: выпишите случайную (с вашей точки зрения) последовательность: 1) из 150 “четырехзначных” чисел (нуль может занимать первое место слева, как и любая другая цифра), 2) из 600 цифр. Проверьте с помощью критерия χ^2, действительно ли в предложенной вами последовательности цифры $0, 1, \ldots, 9$ встречаются с вероятностью $1/10$.

Сравните результаты этих экспериментов.

Примечание. Выписывая последовательность, не просматривайте и не используйте каким-либо другим способом выписанную ее часть: каждое следующее число (каждую следующую цифру) выписывайте так, как будто вы его (ее) пишете впервые, как будто до этого ничего не выписывалось.

Замечание. Встречаются люди, у которых психологические процессы столь хорошо сбалансированы, что они могут извлекать случайные выборки. Обычно такие лица считают себя чрезвычайно одаренными.

24.7 (селекция гороха Г. Менделем). В экспериментах по селекции гороха Г. Мендель наблюдал частоты появления различных видов семян, получаемых в результате скрещивания растений с круглыми желтыми и морщинистыми зелеными семенами (в этой задаче частота — количество семян определенного вида). Эти данные и значения теоретических вероятностей, которые определяются согласно теории наследственности Менделя, приведены в таблице.

Вид семян	Частота	Вероятность
Круглые желтые	315	9/16
Морщинистые желтые	101	3/16
Круглые зеленые	108	3/16
Морщинистые зеленые	32	1/16
Всего	556	1

Согласуются ли вероятности, полученные согласно теории Менделя, с приведенными экспериментальными данными?

24.8 (гипотеза Д'Аламбера). Стохастический эксперимент состоит в подбрасывании двух симметричных монет. Замечательный французский ученый и математик Д'Аламбер считал, что события A — "обе монеты выпали гербом", B — "обе монеты выпали решеткой", C — "монеты выпали разными сторонами", равновероятны и, следовательно, вероятность каждого из них равна $1/3$. А впрочем, и до Д'Аламбера, и после него был известен правильный подход к решению этой задачи: монеты необходимо различать. Для симметричных различимых монет каждому из исходов: ГГ, РР, ГР, РГ необходимо приписать вероятность $1/4$ (буква Г означает появление герба, буква Р — решетки). Тогда

$$\mathsf{P}(A) = \mathsf{P}(\text{ГГ}) = 1/4,\ \mathsf{P}(B) = \mathsf{P}(\text{РР}) = 1/4,$$

$$\mathsf{P}(C) = \mathsf{P}(\{\text{ГР}, \text{РГ}\}) = \mathsf{P}(\text{ГР}) + \mathsf{P}(\text{РГ}) = 1/4 + 1/4 = 1/2.$$

Но, по-видимому, невозможно логически обосновать, почему не прав Д'Аламбер, считая, что события A, B, C равновероятны. В физике элементарных частиц встречаются ситуации, в которых Д'Аламбер скорее прав, чем неправ.

Рассмотрите две модели стохастического эксперимента, состоящего в подбрасывании двух монет:

Модель 1 (гипотеза Д'Аламбера):

$$\Omega^* = \{A, B, C\},\ \mathsf{P}(A) = \frac{1}{3},\ \mathsf{P}(B) = \frac{1}{3},\ \mathsf{P}(C) = \frac{1}{3}$$

(события A, B, C определены выше).

Модель 2:

$$\Omega = \{\text{ГГ}, \text{ГР}, \text{РГ}, \text{РР}\},$$

$$\mathsf{P}(\text{ГГ}) = \frac{1}{4},\ \mathsf{P}(\text{ГР}) = \frac{1}{4},\ \mathsf{P}(\text{РГ}) = \frac{1}{4},\ \mathsf{P}(\text{РР}) = \frac{1}{4}.$$

С целью проверки адекватности описания указанными моделями стохастического эксперимента проведите его достаточное число раз и проверьте:

1° Адекватно ли описывается стохастический эксперимент моделью 1 (согласуется ли гипотеза Д'Аламбера с результатами проведенных экспериментов)?

2° Адекватно ли описывается стохастический эксперимент моделью 2?

Заметим, что, вообще говоря, для описания одного и того же стохастического эксперимента можно предложить не одну модель, которая адекватно его описывает, но из предложенных выше моделей одна заведомо не будет адекватно описывать эксперимент, поскольку вероятности, например, события "монеты легли разными сторонами", вычисленные в разных вероятностных пространствах, различны.

Для проверки адекватности модели 1 будем фиксировать число подбрасываний двух монет, в которых обе монеты легли гербом, обе — решетками, монеты легли разными сторонами.

Для проверки адекватности модели 2 будем регистрировать число подбрасываний двух монет, в которых обе монеты легли гербом, обе — решетками, первая монета легла гербом, а вторая — решеткой, первая монета легла решеткой, а вторая — гербом.

Результаты экспериментов удобно записывать в виде приведенных далее таблиц (первая таблица для проверки адекватности модели 1, вторая — модели 2), где, например, $N_{\text{гр}}$ — количество подбрасываний из N проведенных, в которых на первой монете выпал герб, на второй — решетка.

A	B	C
N_A	N_B	N_C

ГГ	РР	ГР	РГ
$N_{\text{гг}}$	$N_{\text{рр}}$	$N_{\text{гр}}$	$N_{\text{рг}}$

24.9 (о телепатах). Начиная с двадцатых годов двадцатого столетия в Гарвардском университете, университете Дюка и других финансируются экспериментальные исследования по телепатии (чтении мыслей на расстоянии), в которых, в частности, используются карты Зенера с изображением пяти символов: окружность, квадрат, плюс, три волнистые линии, звезда (рис. 24.5.1).

При помощи карт Зенера проверить способности телепата читать мысли на расстоянии. С этой целью провести следующий эксперимент. Из колоды карт Зенера наудачу выбрать одну и зафиксировать результат. Телепат читает выбранную вами карту и также фиксирует результат. Эксперимент провести достаточное количество раз.

Если телепат в самом деле читает мысли, то частота правильно прочитанных символов (окружность читается как окружность, квадрат — как квадрат, ...), т. е. частота успехов будет большой (успех — правильно прочитанный символ, неудача — неправильно прочитанный символ).

По результатам проведенного эксперимента сделать вывод относительно способностей телепата читать мысли на расстоянии.

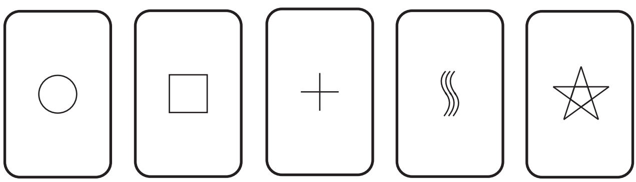

Рис. 24.5.1: Карты Зенера

Указание 1. Сформулировать задачу проверки способностей телепата читать мысли на расстоянии как задачу проверки статистических гипотез. В качестве нулевой рассмотреть гипотезу: телепат мысли не читает. Воспользоваться критерием χ^2.

Указание 2. Проверить телепатические способности своего товарища, знакомого (попросите прочитать ваши мысли).

24.10. Стохастический эксперимент из известной задачи Льюиса Кэррола (см. пример 4.2.1) состоит в последовательном извлечении из урны двух шаров. Можно предложить по меньшей мере две модели (два вероятностных пространства) этого стохастического эксперимента (как и в примере 4.2.1, белый шар обозначим через W, черный — через B, белый шар, вложенный в урну, пометим звездочкой и обозначим через W^*).

Модель 1:

$$\Omega = \{WW, WB, BW\},$$

$$\mathsf{P}(WW) = \frac{1}{3},\ \mathsf{P}(WB) = \frac{1}{3},\ \mathsf{P}(BW) = \frac{1}{3}.$$

Модель 2:

$$\Omega = \{WW^*, W^*W, W^*B, BW^*\},$$

$$\mathsf{P}(WW^*) = \frac{1}{4},\ \mathsf{P}(W^*W) = \frac{1}{4},\ \mathsf{P}(W^*B) = \frac{1}{4},\ \mathsf{P}(BW^*) = \frac{1}{4}.$$

Правдоподобные рассуждения, приведенные в решении к примеру 4.2.1, склоняют нас к мысли, что адекватной моделью стохастического эксперимента является модель 2. Но так ли это на самом деле?

Проведите стохастический эксперимент достаточное число раз и проверьте:

1° Адекватно ли описывается стохастический эксперимент моделью 1?

2° Адекватно ли описывается стохастический эксперимент моделью 2?

Заметим, что хотя для одного и того же эксперимента можно предложить не одну модель, которая адекватно его описывает, для данного стохастического эксперимента по меньшей мере одна из моделей не будет адекватной эксперименту уже хотя бы потому, что вероятности одного и того же события, вычисленные в различных вероятностных пространствах, разные (см. пример 4.2.1).

Стохастический эксперимент будем проводить следующим образом. Возьмем две урны: вспомогательную и основную. Во вспомогательной находятся два шара: белый и черный. Из вспомогательной урны наудачу выбираем один шар и перекладываем в основную, не регистрируя результат. Таким образом, в основной урне находится белый шар (с вероятностью $1/2$) или черный шар. Далее в основную урну кладем белый шар, его при проверке адекватности модели 2 пометим звездочкой. И, наконец, из основной урны последовательно вынимаем оба шара, фиксируя результат эксперимента: WW, WB, BW — в случае проверки адекватности модели 1 и WW^*, W^*W, W^*B, BW^* — в случае проверки адекватности модели 2.

Данные экспериментов удобно записать в виде таблицы (первая — для проверки адекватности модели 1, вторая — модели 2), где, например, через N_{WW} обозначено количество экспериментов из N проведенных, в которых исходом была пара WW.

WW	WB	BW
N_{WW}	N_{WB}	N_{BW}

WW^*	W^*W	W^*B	BW^*
N_{WW^*}	N_{W^*W}	N_{W^*B}	N_{BW^*}

24.11. С 1871 по 1900 г. в Швейцарии родились 1 359 671 мальчик и 1 285 086 девочек.

Согласуется ли с этими данными гипотеза H_0: вероятность рождения мальчика составляет: а) 0,5; б) 0,515?

24.12 (непредвзятость экспериментатора). Каждую из 150 спичек разломайте наудачу на две части. Линейкой с миллиметровыми делениями измерьте с точностью до десятых долей миллиметра длину 300 получившихся частей.

Зафиксируйте десятые доли милиметра. При этом придется оценивать первый знак после запятой в десятичной записи длины части спички в миллиметрах. Зафиксируйте этот знак (последнюю цифру записи).

Результаты измерений удобно представить в виде таблицы:

i	0	1	2	3	4	5	6	7	8	9
n_i										

где i — последняя цифра в записи результата измерения длины части спички $i = 0, 1, \ldots, 9$; n_i — число появлений последней цифры i в измерениях.

Проверьте с помощью критерия χ^2 непредвзятость ваших измерений.

З а м е ч а н и е. Предвзятость при снятии показаний наблюдается достаточно часто. Сомнительно, чтобы была какая-то объективная причина, которая бы обуславливала более частое появление одних цифр сравнительно с другими; поэтому естественно предположить,что неодинаковая частота появления цифр указывает на предвзятость наблюдателя. Даже те исследователи, которые знают о возможности предвзятости и о необходимости осторожности при снятии показаний, все-таки не всегда могут избежать ее.

24.13 (задача о разделе ставки II). Вернемся к задаче о разделе ставки, впервые опубликованной в Венеции в 1494 г. Фра Лука Пачоли (см. пример 4.4.6).

Два игрока играют в справедливую игру, например, подбрасывая симметричную монету — если монета легла гербом, партию выиграл первый игрок, если решеткой — второй. Игру выигрывает тот из игроков, кто первым выиграет 6 партий, при этом он получает весь приз. На самом деле игра остановилась, когда первый игрок выиграл 5 партий, а второй — 3. Как справедливо разделить приз?

Задачу пытались решить (безуспешно) многие знаменитые математики. Решение в 1654 г. независимо друг от друга получили Паскаль и Ферма — приз между игроками следует разделить пропорционально вероятностям $7/8$ и $1/8$ выигрыша шести партий соответственно первым и вторым игроками, если бы игра при счете 5 к 3 не была прервана.

Ну а как на самом деле?

Согласуется ли решение Паскаля и Ферма с опытом? Ведь решения, предлагавшиеся ранее видными математиками, выглядели не менее правдоподобно и тем не менее были ошибочными, в частности, Никколо Тарталья, получивший за одну ночь формулу корней кубического уравнения, считал, что приз необходимо разделить в отношении 2 к 1. Поставленный вопрос это фактически вопрос об адекватности модели эксперименту.

В связи с такой постановкой вопроса о согласии с опытом решения Паскаля и Ферма задачи о разделе ставки проведем эксперимент.

Будем подбрасывать симметричную монету (продолжая прерванную игру). При этом выпадение герба интерпретируем как выигрыш партии первым игроком, решетки — вторым. Монету будем подбрасывать пока не появится герб или три раза (разумеется, кряду) решетка. На этом игра завершается — выигрышем первого игрока, если выпал герб и второго, если выпали три решетки.

Проведем эксперимент достаточно много раз. Результат удобно записать в виде таблицы

G_1	G_2
n_1	n_2

,

где n_1— число экспериментов из n проведенных, закончившихся выигрышем игры первым игроком (игроком G_1), n_2— вторым (игроком G_2).

Далее, воспользовавшись критерием χ^2, проверить, согласуется ли гипотеза H_0 :

G_1	G_2
7/8	1/8

о распределении вероятностей выигрыша игры первым и вторым игроком с результатами опыта:

G_1	G_2
n_1	n_2

.

Глава 25

Непараметрические критерии

Робастность. Статистические выводы основываются на тех или иных предположениях о наблюдениях. Чаще всего явно или неявно делаются предположения о независимости наблюдений и о виде распределений, как правило, о нормальной распределенности наблюдений. Набор таких предположений представляет собой не что иное, как математическую модель наблюдений. Однако нельзя не заметить, что на практике отклонения от модели (бóльшие или меньшие) неизбежны. Остаются ли при этом верными статистические выводы, полученные в предположениях модели (хотя бы тогда, когда отклонения от них не очень большие)? Классические статистические методы, основанные на предположении о нормальности распределений, часто нечувствительны к отклонениям от последних. К сожалению, это имеет место далеко не всегда — отклонения от математической модели, даже незначительные, могут приводить к существенным ошибкам в окончательных результатах. Но, оказывается, существует целый класс методов получения статистических выводов, мало чувствительных к отклонениям от предположений модели. Такие методы называют робастными. К ним, в частности, относятся непараметрические методы. Они не предназначены для определенного параметрического семейства распределений и специально строятся так, чтобы вовсе обходиться без предположений о виде распределения.

25.1 Критерий А. Н. Колмогорова

Первым прорывом в непараметрической статистике была теорема А. Н. Колмогорова о распределении наибольшего уклонения эмпирической функции распределения от теоретической. Это уклонение еще называют статистикой А. Н. Колмогорова. По статистике А. Н. Колмогорова строится критерий, носящий его имя.

Постановка задачи. Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — выборка из непрерывного распределения F (F — неизвестно). Относительно распределения F выдвигается гипотеза H_0: $\mathsf{F} = \mathsf{G}$, где G — полностью определенное (не зависящее от неизвестных параметров) непрерывное распределение. Необходимо построить критерий для проверки этой гипотезы. Общий подход (см. раздел 24.1 в гл. 24) такой. Вводим уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ эмпирического распределения $\hat{\mathsf{F}}_n$, построенного по выборке $\xi_1, \xi_2, \ldots, \xi_n$, от гипотетического G, причем так, чтобы при верной гипотезе H_0: $\mathsf{F} = \mathsf{G}$ уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ принимало малые значения, а при $\mathsf{G} \neq \mathsf{F}$ — большие. Гипотезу H_0: $\mathsf{F} = \mathsf{G}$ отклоняем, если уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ приняло большое значение, и не отклоняем в противном случае. А. Н. Колмогоров в качестве уклонения эмпирического $\hat{\mathsf{F}}_n$, построенного по выборке $\xi_1, \xi_2, \ldots, \xi_n$, от гипотетического распределения G предложил рассматривать

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sup_{x \in \mathbb{R}^1} |G(x) - \hat{F}_n(x)|, \tag{25.1.1}$$

где $\hat{F}_n(x)$ — эмпирическая функция распределения:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x)}(\xi_i), \ \ x \in \mathbb{R}^1. \tag{25.1.2}$$

Свойства статистики А. Н. Колмогорова. Описываются следующими теоремами.

Теорема 25.1.1 (о распределении $\sup\limits_{x} |F(x) - \hat{F}_n(x)|$). *Пусть $\xi_i, i = 1, 2, \ldots, n,$ — выборка из непрерывного распределения F, тогда*

$$\eta_i = F(\xi_i), \ \ i = 1, 2, \ldots, n,$$

— выборка из распределения U, *равномерного на* $[0;1]$, *и*

$$\sup_{x\in(-\infty,+\infty)} |F(x) - \hat{F}_n(x)| = \sup_{t\in(0,1)} |U(t) - \hat{U}_n(t)|, \tag{25.1.3}$$

где $\hat{F}_n(x)$ *и* $\hat{U}_n(t)$ *— эмпирические функции распределения, построенные соответственно по выборкам* $\xi_1, \xi_2, \ldots, \xi_n$ *из* F *и* $\eta_1, \eta_2, \ldots, \eta_n$ *из* U.

Мы докажем теорему в предположении, что функция распределения $F(x)$ строго монотонная, но теорема имеет место и для выборок из любых непрерывных распределений.

Доказательство. Строго монотонная непрерывная функция распределения $F(x)$, $x \in (-\infty, +\infty)$, имеет непрерывную монотонно возрастающую обратную функцию $F^{-1}(t)$, $t \in (0;1)$, причем значения $F^{-1}(t)$ "пробегают" прямую $(-\infty, +\infty)$, когда t "пробегает" интервал $(0;1)$ (см. также рис. 25.1.1).

Случайные величины $\eta_i = F(\xi_i)$, $i = 1, 2, \ldots, n$, независимы, как функции от независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, и каждая η_i, $i = 1, 2, \ldots, n$, равномерно распределена на $[0;1]$. Действительно, по определению

$$F_{\eta_i}(t) = \mathsf{P}\{\eta_i < t\} = \mathsf{P}\{F(\xi_i) < t\}.$$

Значения $F(\xi_i) \in [0,1]$, поэтому, если $t \le 0$, то $F_{\eta_i}(t) = 0$, если $t > 1$, то $F_{\eta_i}(t) = 1$, если $0 < t \le 1$, то, учитывая, что $F(x)$ имеет монотонно возрастающую непрерывную обратную функцию $F^{-1}(t)$, получаем:

$$F_{\eta_i}(t) = \mathsf{P}\{F(\xi_i) < t\} = \mathsf{P}\{\xi_i < F^{-1}(t)\} = F(F^{-1}(t)) = t.$$

Так что

$$F_{\eta_i}(t) = U(t) = \begin{cases} 0 & \text{при } t \le 0, \\ t & \text{при } 0 < t \le 1, \\ 1 & \text{при } t > 1, \end{cases}$$

$i = 1, 2, \ldots, n$.

Покажем теперь, что

$$\sup_{x\in(-\infty,+\infty)} |F(x) - \hat{F}_n(x)| = \sup_{t\in(0,1)} |U(t) - \hat{U}_n(t)|.$$

Поскольку значения $F^{-1}(t)$ "пробегают" $(-\infty, +\infty)$, когда t "пробегает" интервал $(0;1)$, то

$$\sup_{x\in(-\infty,+\infty)} |F(x) - \hat{F}_n(x)| = \sup_{t\in(0;1)} |F(F^{-1}(t)) - \hat{F}_n(F^{-1}(t))| =$$

$$= \sup_{t\in(0,1)} |U(t) - \hat{F}_n(F^{-1}(t))|.$$

Убедимся, что функция

$$\hat{F}_n(F^{-1}(t)) = \hat{U}_n(t), t \in (0,1),$$

является эмпирической функцией распределения, построенной по выборке $\eta_1, \eta_2, \ldots, \eta_n$ из равномерного на промежутке $[0;1]$ распределения.

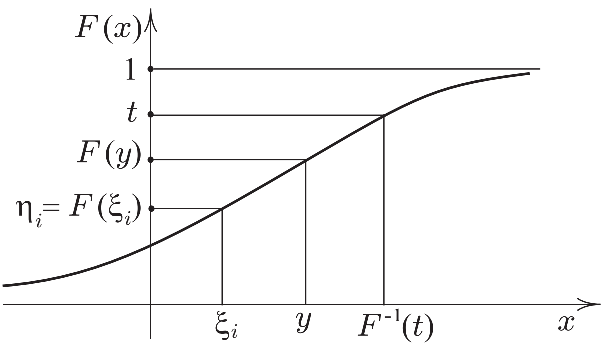

Рис. 25.1.1: Иллюстрация к доказательству теоремы

Значение $\hat{F}_n(x)$ в точке $F^{-1}(t)$ равно

$$\hat{F}_n(F^{-1}(t)) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,F^{-1}(t))}(\xi_i) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} I_{(0,t)}(\eta_i),\ t \in (0;1),$$

последнее следует из того, что значение $I_{(-\infty,F^{-1}(t))}(\xi_i)$ равно 1 тогда и только тогда, когда $-\infty < \xi_i < F^{-1}(t)$ или $0 < F(\xi_i) < t$

или $0 < \eta_i < t$ или $I_{(0,t)}(\eta_i) = 1$. Заметим еще, что поскольку $\eta_i \in (0,1)$, то

$$\frac{1}{n}\sum_{i=1}^{n} I_{(0,t)}(\eta_i) = \hat{U}_n(t),\ t \in (0,1).$$

— эмпирическая функция выборки $\eta_1, \eta_2, \ldots, \eta_n$ из равномерного на $[0,1]$ распределения. Так что $\hat{F}_n(F^{-1}(t)) = \hat{U}_n(t)$ при $t \in (0,1)$.

Следствие. *Для выборки $\xi_1, \xi_2, \ldots, \xi_n$ из любого распределения F с непрерывной монотонно возрастающей функцией распределения $F(x)$ распределение уклонения*

$$\sup_x |F(x) - \hat{F}_n(x)|$$

совпадает с распределением случайной величины

$$\sup_t |U(t) - \hat{U}_n(t)|,$$

и, следовательно, не зависит от распределения F.

При больших n распределение нормированного уклонения

$$\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}}$$

описывается теоремой А. Н. Колмогорова.

Теорема 25.1.2 (А. Н. Колмогорова). *Пусть $\hat{F}_n(x)$ — эмпирическая функция распределения, построенная по выборке $\xi_1, \xi_2, \ldots, \xi_n$ из непрерывного распределения F, тогда при $n \to \infty$ распределение нормированного уклонения*

$$\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}}$$

сходится к распределению Колмогорова: для каждого $\lambda > 0$

$$\mathsf{P}\left\{\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}} < \lambda\right\} \to \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2} = K(\lambda).$$

Построение критерия А. Н. Колмогорова. Из теоремы 25.1.2, в частности, следует, что если $n \to \infty$, то

$$D(\hat{\mathsf{F}}_n, \mathsf{F}) = \sup_x |F(x) - \hat{F}_n(x)| \xrightarrow{\mathsf{P}} 0,$$

т. е. для любого $\varepsilon > 0$ при $n \to \infty$

$$\mathsf{P}\{D(\hat{\mathsf{F}}_n, \mathsf{F}) < \varepsilon\} = \mathsf{P}\{\sup_x |F(x) - \hat{F}_n(x)| < \varepsilon\} =$$

$$= \mathsf{P}\{\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}} < \varepsilon\sqrt{n}\} \to 1.$$

Поэтому, если гипотеза H_0: $\mathsf{F} = \mathsf{G}$ верна, то при $n \to \infty$ уклонение

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = D(\hat{\mathsf{F}}_n, \mathsf{F}) = \sup_x |F(x) - \hat{F}_n(x)|$$

сходится по вероятности к нулю (принимает малые значения). Если гипотеза H_0: $\mathsf{F} = \mathsf{G}$ неверна, т. е. $\mathsf{F} \neq \mathsf{G}$, то

$$\sup_x |G(x) - F(x)| = a > 0,$$

и уклонение

$$\sup_x |G(x) - \hat{F}_n(x)|$$

не стремится к нулю (принимает большие значения), поскольку

$$|G(x) - F(x)| \leq |G(x) - \hat{F}_n(x)| + |F(x) - \hat{F}_n(x)|,$$

$$\sup_x |G(x) - F(x)| - \sup_x |F(x) - \hat{F}_n(x)| \leq \sup_x |G(x) - \hat{F}_n(x)|.$$

Так что, если гипотеза H_0 верна, уклонение

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = D(\hat{\mathsf{F}}_n, \mathsf{F})$$

мало, в противном случае большое. И следовательно, для проверки гипотезы H_0: $\mathsf{F} = \mathsf{G}$ вычисляем значение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ и в зависимости от того, каким оно окажется — большим или малым, отклоняем гипотезу H_0 или не отклоняем. Чтобы так можно было поступать, необходимо знать, какие значения принимает уклонение $D(\hat{\mathsf{F}}_n, \mathsf{G})$ (в идеале найти распределение $D(\hat{\mathsf{F}}_n, \mathsf{G})$), когда гипотеза H_0: $\mathsf{F} = \mathsf{G}$ верна и когда она неверна.

Из теоремы 25.1.1 следует, что при верной гипотезе H_0: $\mathsf{F} = \mathsf{G}$ распределение уклонения

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = D(\hat{\mathsf{F}}_n, \mathsf{F}) = \sup_x |F(x) - \hat{F}_n(x)|$$

не зависит от распределения F, из которого получена выборка, и его можно считать известным — оно совпадает с распределением уклонения $\sup\limits_t |U(t) - \hat{U}_n(t)|$ (см. следствие из теоремы 25.1.1), а при достаточно больших n распределение нормированного уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F}) \Big/ \dfrac{1}{\sqrt{n}}$ близко к распределению А. Н. Колмогорова (см. теорему 25.1.1). Поэтому можно указать промежуток $[0; \varepsilon_{\alpha,n}]$, в котором "почти всегда" (с вероятностью $1-\alpha$) лежат значения $D(\hat{\mathsf{F}}_n, \mathsf{F})$. Для этого достаточно выбрать в качестве $\varepsilon_{\alpha,n}$ наименьшее ε, для которого

$$\mathsf{P}\{\sup_x |F(x) - \hat{F}_n(x)| \geq \varepsilon\} \leq \alpha$$

($\varepsilon_{\alpha,n}$ называют *α-критическим* значением уклонения $D(\hat{\mathsf{F}}_n, \mathsf{F})$).

Критерий А. Н. Колмогорова. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из непрерывного распределения F, $\varepsilon_{\alpha,n}$ — α-критическое значение для $\sup\limits_x |F(x) - \hat{F}_n(x)|$. Если гипотезу H_0: $\mathsf{F} = \mathsf{G}$ отклонять при*

$$\sup_x |G(x) - \hat{F}_n(x)| \geq \varepsilon_{\alpha,n}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α, гипотеза H_0 будет отклоняться, когда она верна.

Действительно, H_0 отклоняется, если

$$\sup_x |G(x) - \hat{F}_n(x)| \geq \varepsilon_{\alpha,n}.$$

При верной гипотезе H_0: $\mathsf{F} = \mathsf{G}$

$$\sup_x |G(x) - \hat{F}_n(x)| = \sup_x |F(x) - \hat{F}_n(x)|,$$

поэтому вероятность отклонения H_0

$$\mathsf{P}\{\sup_x |G(x) - \hat{F}_n(x)| \geq \varepsilon_{\alpha,n}\} = \mathsf{P}\{\sup_x |F(x) - \hat{F}_n(x)| \geq \varepsilon_{\alpha,n}\} \leq \alpha.$$

З а м е ч а н и е 1. При $n \le 100$ по заданным α и n критические значения $\varepsilon_{\alpha,n}$ для $\sup\limits_x |F(x) - \hat{F}_n(x)|$ табулированы (см. табл. 27.7.1).

Для $n \ge 100$ в качестве критического значения $\varepsilon_{\alpha,n}$ можно рассматривать $\lambda_\alpha/\sqrt{n}$, т. е.

$$\varepsilon_{\alpha,n} = \lambda_\alpha/\sqrt{n},$$

где λ_α — верхний α-предел распределения А. Н. Колмогорова, т. е. корень уравнения $K([\lambda_\alpha, +\infty)) = \alpha$. Действительно, в силу теоремы А. Н. Колмогорова функция распределения

$$\mathsf{P}\left\{\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}} < \lambda\right\}$$

сходится к функции распределения $K(\lambda)$. Поэтому при достаточно больших n (начиная уже с $n = 100$) можно считать, что с незначительной погрешностью

$$\mathsf{P}\left\{\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}} < \lambda\right\} = K(\lambda),$$

отсюда

$$\mathsf{P}\left\{\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}} \ge \lambda_\alpha\right\} = K([\lambda_\alpha, +\infty)) = \alpha.$$

Поэтому $\varepsilon_{\alpha,n} = \lambda_\alpha/\sqrt{n}$.

З а м е ч а н и е 2. *Критерием А. Н. Колмогорова можно пользоваться только тогда, когда гипотетическое распределение* G *непрерывно и не зависит от неизвестных параметров.*

Если же гипотетическое распределение зависит от параметров, которые необходимо оценить по выборке, критерием Колмогорова для проверки гипотезы H_0: $\mathsf{F} = \mathsf{G}$ в приведенной выше форме пользоваться нельзя. В этом случае распределение случайной величины $\sup\limits_x |F(x) - \hat{F}_n(x)|$ будет отлично от распределения случайной величины $\sup\limits_{t \in (0,1)} |U(t) - \hat{U}_n(t)|$ (см. теорему 25.1.1), а предельное распределение

$$\sup_x |F(x) - \hat{F}_n(x)| \Big/ \frac{1}{\sqrt{n}}$$

будет отличным от распределения А. Н. Колмогорова.

Пример 25.1.1. *Ниже приведены показания* 14 *механических часов, выставленных в витринах часового магазина:* 9 *ч.* 18 *мин.,* 8 *ч.* 35 *мин.,* 10 *ч.* 45 *мин.,* 11 *ч.* 30 *мин.,* 3 *ч.* 06 *мин.,* 7 *ч.* 50 *мин.,* 4 *ч.* 22 *мин.,* 10 *ч.* 12 *мин.,* 7 *ч.* 47 *мин.,* 4 *ч.* 28 *мин.,* 11 *ч.* 16 *мин.,* 7 *ч.* 08 *мин.,* 5 *ч.* 53 *мин.,* 8 *ч.* 00 *мин.*

Момент остановки часов естественно рассматривать как случайную величину, равномерно распределенную на промежутке $[0; 12]$. *Согласуется ли гипотеза о равномерном на промежутке* $[0; 12]$ *распределении показаний часов с приведенными данными?*

Решение. Запишем показания часов в десятичной форме: 9,30; 8,58; 10,75; 11,5; 3,1; 7,83; 4,36; 10,2; 7,78; 4,47; 11,27; 7,13; 5,88; 8,00. Далее будем иметь дело с этой выборкой.

Показания 14 часов — реализация выборки из некоторого непрерывного распределения F, сосредоточенного на промежутке $[0; 12]$. Относительно неизвестного распределения F выдвигается гипотеза H_0: F — равномерное на $[0; 12]$ распределение. Обозначим равномерное на промежутке $[0; 12]$ распределение через G, его функция распределения

$$G(x) = \begin{cases} 0, & \text{при } x < 0, \\ x/12, & \text{при } 0 \leq x < 12, \\ 1, & \text{при } x \geq 12. \end{cases}$$

Будем проверять гипотезу H_0: $\mathsf{F} = \mathsf{G}$, пользуясь критерием Колмогорова. Для этого вычислим значение уклонения

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sup_x |G(x) - \hat{F}_n(x)|$$

и сравним его с α-критическим значением $\varepsilon_{\alpha;n}$ (здесь $n = 14$, $\hat{F}_n(x)$ — реализация эмпирической функции распределения, построенной по выборке). Если

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) \geq \varepsilon_{\alpha;n},$$

то гипотезу H_0: $\mathsf{F} = \mathsf{G}$ отклоним, в противном случае — нет.

Поскольку $\hat{F}_n(x)$ — кусочно-постоянная функция со скачками в точках $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$, а $G(x)$ — монотонно неубывающая функция, то $\sup\limits_x |G(x) - \hat{F}_n(x)|$ равен наибольшему из чисел

$$\left|G(\xi_k^*) - \hat{F}_n(\xi_k^*)\right|, \ \left|G(\xi_k^*) - \hat{F}_n(\xi_{k+1}^*)\right|,$$

$k = 1, 2, \ldots, n$, где обозначено $\hat{F}_n(\xi^*_{n+1}) = 1$. Вариационный ряд $\xi^*_1, \xi^*_2, \ldots, \xi^*_n$, значения $\hat{F}_n(\xi^*_k)$, $G(\xi^*_k)$ и разности

$$\left|G(\xi^*_k) - \hat{F}_n(\xi^*_k)\right|, \ \left|G(\xi^*_k) - \hat{F}_n(\xi^*_{k+1})\right|,$$

$k = 1, 2, \ldots, n$, сведены в таблицу 25.1.1. Из таблицы имеем:

$$D(\hat{\mathsf{F}}_{14}, \mathsf{G}) = \sup_x |G(x) - \hat{F}_{14}(x)| = 0{,}30.$$

Таблица 25.1.1. К вычислению $\sup_x \left|G(x) - \hat{\mathsf{F}}_{14}(x)\right|$

ξ^*_k	$\hat{F}_{14}(\xi^*_k)$	$G(\xi^*_k) = \xi^*_k/12$	$\left\|G(\xi^*_k) - \hat{F}_{14}(\xi^*_k)\right\|$	$\left\|G(\xi^*_k) - \hat{F}_{14}(\xi^*_{k+1})\right\|$
3,10	0	0,26	0,26	0,19
4,36	0,07	0,36	0,29	0,22
4,47	0,14	0,37	0,23	0,16
5,88	0,21	0,49	0,28	0,20
7,13	0,29	0,59	**0,30**	0,23
7,78	0,36	0,65	0,29	0,22
7,83	0,43	0,65	0,22	0,15
8,00	0,50	0,67	0,17	0,10
8,58	0,57	0,72	0,15	0,08
9,30	0,64	0,78	0,14	0,07
10,20	0,71	0,85	0,14	0,06
10,75	0,79	0,90	0,11	0,04
11,27	0,86	0,94	0,08	0,01
11,50	0,93	0,96	0,03	0,04
	1,00			

Отсюда

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = 0{,}30 < 0{,}35 = \varepsilon_{0{,}05;14}$$

(значение $\varepsilon_{\alpha,n}$ при $\alpha = 0{,}05$ и $n = 14$, равное 0,35, найдено по табл. 27.7.1). Поэтому гипотеза H_0: $\mathsf{F} = \mathsf{G}$ на 5% уровне значимости не отклоняется.

Для выборки объемом $n = 14$ из равномерного на промежутке $[0; 12]$ распределения наибольшее уклонение между эмпирической функцией распределения $\hat{\mathsf{F}}_n$ и функцией $G(x)$ равномерного на промежутке $[0; 12]$ распределения, равное 0,30, естественно и допустимо.

Этот результат можно интерпретировать так. Предположение (гипотеза) о равномерном распределении момента остановки часов не противоречит результатам наблюдений.

25.2 Критерий знаков для повторных наблюдений

Сначала рассмотрим одну задачу. Некий препарат предположительно повышает кровяное давление. Чтобы убедиться, так ли это, измеряют давление у 10 подопытных кроликов, затем кроликам вводят препарат и снова измеряют давление. Результаты регистрируют: давление повысилось, осталось неизменным, понизилось. По этим данным необходимо вынести заключение о воздействии препарата на давление. Для этого естественно подсчитать количество кроликов, давление у которых повысилось (подсчитать количество положительных разностей между давлением после и до введения препарата). Если число положительных разностей оказалось близким к половине имеющихся (в нашем случае к пяти), то заключаем, что препарат не оказывает влияния на давление, если же число положительных разностей будет существенно большим половины, то делаем вывод, что препарат повышает кровяное давление.

По постановке и методам решения эта и многие другие задачи укладываются в приведенную далее общую схему.

Постановка задачи. Имеется n пар случайных величин $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_n, \eta_n)$, относительно которых известно, что разности $\zeta_j = \eta_j - \xi_j$ представимы в виде:

$$\zeta_j = \eta_j - \xi_j = \theta + e_j, \ j = 1, 2, \ldots, n,$$

где θ — константа, а случайные величины $e_1, e_2, \ldots, e_n$ независимые (сами η_j и ξ_j могут быть и зависимы), симметрично распределенные относительно нуля (распределения e_j и $-e_j$ совпадают) и абсолютно непрерывные. Далее предположение об абсолютной непрерывности $e_1, e_2, \ldots, e_n$ будет снято.

Параметр θ неизвестен. Относительно θ выдвигается гипотеза H_0: $\theta = 0$. Альтернатива гипотезе H_0: $\theta = 0$ может быть как односторонней: если $\theta \neq 0$, то $\theta > 0$ (она может быть и такой: $\theta < 0$), так и двусторонней: если $\theta \neq 0$, то $\theta > 0$ или $\theta < 0$.

Наша задача — построить критерий для проверки гипотезы H_0: $\theta = 0$.

Замечание. Пары $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_n, \eta_n)$ часто интерпретируют как $2n$ наблюдений: по два на каждый из n объектов, пациентов, приборов, ... При этом ξ_j называют наблюдением до “обработки”, η_j — после “обработки”, параметр θ называют “эффектом обработки”. Отклонение гипотезы H_0 свидетельствует в пользу наличия эффекта обработки, неотклонение — в пользу отсутствия.

Выбор статистики для построения критерия. Рассмотренный пример подсказывает, что в качестве статистики для построения критерия для проверки гипотезы $H_0\colon \theta = 0$ естественно рассматривать количество μ положительных разностей

$$\zeta_j = \eta_j - \xi_j = \theta + e_j,\ j = 1, 2, \ldots, n.$$

Случайную величину μ представим в виде суммы

$$\mu = \sum_{j=1}^{n} I_{(0,+\infty)}(\zeta_j)$$

независимых случайных величин

$$I_{(0,+\infty)}(\zeta_j) = I_{(0,+\infty)}(\theta + e_j),\ j = 1, 2, \ldots, n.$$

При верной гипотезе $H_0\colon \theta = 0$

$$\mathsf{P}\{I_{(0,+\infty)}(\zeta_j) = 1\} = \mathsf{P}\{\zeta_j > 0\} = \mathsf{P}\{e_j > 0\} = 1/2$$

— поскольку e_j, $j = 1, 2, \ldots, n$, симметрично распределены относительно нуля и абсолютно непрерывны, то

$$1 = \mathsf{P}\{e_j > 0\} + \mathsf{P}\{e_j < 0\} + \mathsf{P}\{e_j = 0\} =$$

$$= \mathsf{P}\{e_j > 0\} + \mathsf{P}\{-e_j < 0\} = \mathsf{P}\{e_j > 0\} + \mathsf{P}\{e_j > 0\} = 2\mathsf{P}\{e_j > 0\}.$$

Поэтому случайная величина μ, как сумма независимых одинаково распределенных случайных величин, каждая с распределением

$$\mathsf{P}\{I_{(0,+\infty)}(\zeta_j) = s\} = 1/2,\ s = 0, 1,$$

имеет биномиальное распределение с параметрами $(n; 1/2)$:

$$\mathsf{P}\{\mu = k\} = C_n^k\,(1/2)^n\,,\ k = 0, 1, \ldots, n,$$

и, как следствие,

$$\mathsf{M}\mu = n/2.$$

Если гипотеза $H_0\colon \theta = 0$ неверна, например $\theta > 0$, то (при достаточно больших θ)

$$\mathsf{P}\{I_{(0,+\infty)}(\zeta_j) = 1\} = \mathsf{P}\{\zeta_j > 0\} =$$
$$= \mathsf{P}\{\theta + e_j > 0\} = \mathsf{P}\{e_j > -\theta\} = p_j(\theta) \geq 1/2, \ j = 0, 1, \ldots, n,$$

и, как следствие,

$$\mathsf{M}\mu \geq n/2.$$

Так что при верной гипотезе H_0: $\theta = 0$ случайная величина μ мало уклоняется от $n/2$, а при неверной гипотезе H_0 уклонение от $n/2$ существенное. И, следовательно, для проверки гипотезы H_0: $\theta = 0$ подсчитываем количество μ положительных разностей $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \ldots, n$, и, в зависимости от того, существенно μ уклоняется от $n/2$ или несущественно, отклоняем гипотезу H_0: $\theta = 0$ или не отклоняем. Границы, отделяющие значения μ, мало уклоняющиеся от $n/2$, от значений μ, уклоняющихся от $n/2$ существенно, находим по распределению случайной величины μ_0, имеющей биномиальное распределение с параметрами $(n; 1/2)$, другими словами, по распределению случайной величины μ, когда гипотеза H_0: $\theta = 0$ верна.

Определение. Пусть случайная величина μ_0 имеет биномиальное распределение с параметрами $(n; 1/2)$, $0 < \alpha < 1$. Минимальное целое m, удовлетворяющее условию

$$\mathsf{P}\{\mu_0 > m\} \leq \alpha,$$

или, что то же, условию

$$\sum_{k=m+1}^{n} C_n^k \, (1/2)^n \leq \alpha,$$

будем называть *верхним α-пределом биномиального распределения* с параметрами $(n; 1/2)$ и будем обозначать $m_{\alpha;n}$.

Случайная величина μ_0 "почти всегда" (с вероятностью не меньшей $1 - 2\alpha$) принимает значения из промежутка с концами $n - m_{\alpha;n}$ и $m_{\alpha;n}$.

Значения числа μ положительных разностей, принадлежащих промежутку $\mu \in [n - m_{\alpha;n}, m_{\alpha;n}]$, будем классифицировать как мало уклоняющиеся от $n/2$, в противном случае — как уклоняющиеся от $n/2$ существенно.

И, следовательно, для проверки гипотезы H_0: $\theta = 0$ подсчитываем количество μ положительных разностей $\zeta_j = \eta_j - \xi_j$ и выясняем, существенно оно уклоняется от $n/2$ или нет, сравнивая его с границами $n - m_{\alpha;n}$ и $m_{\alpha;n}$, отделяющими большие уклонения μ от малых. Если $\mu \notin [n - m_{\alpha;n}, m_{\alpha;n}]$, гипотезу H_0: $\theta = 0$ отклоняем, в противном случае — не отклоняем.

Критерий знаков. *Пусть для n пар (ξ_j, η_j), $j = 1, 2, \ldots, n$, случайных величин, разности $\zeta_j = \eta_j - \xi_j$ представимы в виде:*

$$\zeta_j = \eta_j - \xi_j = \theta + e_j,\ j = 1, 2, \ldots, n,$$

где случайные величины $e_1, e_2, \ldots, e_n$ независимые, симметрично распределенные относительно нуля и абсолютно непрерывные, μ — число положительных разностей среди $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \ldots, n$.

Если гипотезу H_0: $\theta = 0$ отклонять при

$$\mu > m_{\alpha;n}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α, гипотеза H_0 будет отклоняться, когда она верна (альтернатива односторонняя: $\theta > 0$).

Если гипотезу H_0 отклонять при

$$\mu \notin [n - m_{\alpha;n}, m_{\alpha;n}]$$

и не отклонять в противном случае, то с вероятностью, не превосходящей 2α, гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя: $\theta < 0$ или $\theta > 0$).

Доказательство. Пусть гипотеза H_0 верна, тогда μ имеет биномиальное распределение с параметрами $(n; 1/2)$, поэтому

$$\mathsf{P}\{\mu > m_{\alpha;n}\} \leq \alpha,$$

$$\mathsf{P}\{\mu \notin [n - m_{\alpha;n}, m_{\alpha;n}]\} =$$

$$= \mathsf{P}\{\mu \in [0, n - m_{\alpha;n})\} + \mathsf{P}\{\mu \in (m_{\alpha;n}, n]\} \leq 2\alpha.$$

Последнее означает, что, пользуясь приведенным критерием, верную гипотезу H_0 будем отклонять с вероятностью, не превосходящей α, если альтернатива односторонняя, и с вероятностью, не превосходящей 2α, если альтернатива двусторонняя.

Об ошибках первого и второго рода. При верной гипотезе H_0: $\theta = 0$ случайная величина μ — число положительных разностей $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \ldots, n$, — имеет биномиальное распределение с параметрами $(n; 1/2)$. Поэтому μ, принимая значения $0, 1, 2, \ldots, n$, и почти всегда мало уклоняясь от $n/2$, изредка все же может принять значение, значительно уклоняющееся от $n/2$ (т. е. значение из $[0, n - m_{\alpha;n}) \cup (m_{\alpha;n}, n]$). При этом мы гипотезу

H_0: $\theta = 0$ отклоним и тем самым совершим ошибку — ошибку первого рода (ее вероятность не превосходит выбранного уровня значимости).

Если гипотеза H_0 неверна, например, $\theta > 0$, то μ, почти всегда принимая большие значения (значения из $(m_{\alpha;n}, n]$), может, хотя и изредка, принять значение, близкое к $n/2$, при этом мы гипотезу H_0 не отклоним и тем самым совершим ошибку — ошибку второго рода.

25.3 Критерий знаков при наличии связей

Снимем теперь требование абсолютной непрерывности распределений разностей $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \ldots, n$. Тогда разности ζ_j могут обращаться в нуль с ненулевой вероятностью (говорят, что имеются связи). Можно ли пользоваться критерием знаков в этой ситуации, и если да, то как? Оказывается, что да — можно, при этом необходимо разности равные нулю отбросить (как будто их и не было) и применять критерий знаков к оставшимся отличным от нуля разностям. Чтобы убедиться в этом, нам понадобится следующая теорема.

Теорема. *Пусть ζ_j, $j = 1, 2, \ldots, n$, — независимые случайные величины, такие, что* $\mathsf{P}\{\zeta_j > 0\} = p > 0$, $\mathsf{P}\{\zeta_j < 0\} = q > 0$, $\mathsf{P}\{\zeta_j = 0\} = f > 0$, $p + q + f = 1$, s *— число отличных от нуля среди ζ_j, $j = 1, 2, \ldots, n$, а μ — количество положительных среди них.*

Условным распределением случайной величины μ относительно события $\{s = i\}$ $(i = 1, 2, \ldots, n)$ является биномиальное распределение с параметрами $(i; p/(p+q))$:

$$\mathsf{P}\{\mu = k | s = i\} = C_i^k \left(\frac{p}{p+q}\right)^k \left(\frac{q}{p+q}\right)^{i-k}, \ k = 0, 1, \ldots, i. \tag{25.3.1}$$

Д о к а з а т е л ь с т в о. Вычислим $\mathsf{P}\{\mu = k, s = i\}$ и $\mathsf{P}\{s = i\}$.

Случайные величины μ и s являются функциями от случайного вектора $\zeta^* = (\operatorname{sign}\zeta_1, \operatorname{sign}\zeta_2, \ldots, \operatorname{sign}\zeta_n)$. Найдем его распределение. Пусть $x = (x_1, x_2, \ldots, x_n)$ — последовательность, составленная из $+1$, -1 и 0. Учитывая, что ζ_j, $j = 1, 2, \ldots, n$, — независимые случайные величины, имеем:

$$\mathsf{P}_{\zeta^*}(x) = \mathsf{P}\{\zeta^* = x\} =$$

$$= \mathsf{P}\{(\operatorname{sign}\zeta_1, \operatorname{sign}\zeta_2, \ldots, \operatorname{sign}\zeta_n) = (x_1, x_2, \ldots, x_n)\} =$$
$$= \mathsf{P}\{\operatorname{sign}\zeta_1 = x_1, \operatorname{sign}\zeta_2 = x_2, \ldots, \operatorname{sign}\zeta_n = x_n\} =$$
$$= \prod_{j=1}^{n} \mathsf{P}\{\operatorname{sign}\zeta_j = x_j\} = p^{\mu(x)} q^{s(x)-\mu(x)} f^{n-s(x)},$$

где $\mu(x)$ равно количеству $+1$ в последовательности $x = (x_1, x_2, \ldots, x_n)$, $(n - s(x))$ — количеству нулей, а $s(x)$ равно количеству элементов последовательности $x = (x_1, x_2, \ldots, x_n)$, отличных от нуля. Зная распределение ζ^*, получаем:

$$\mathsf{P}\{\mu = k, s = i\} = \mathsf{P}\{\mu(\zeta^*) = k, s(\zeta^*) = i\} = \sum_{x:\mu(x)=k; s(x)=i} \mathsf{P}_{\zeta^*}(x) =$$

$$= \sum_{x:\mu(x)=k; s(x)=i} p^{\mu(x)} q^{s(x)-\mu(x)} f^{n-s(x)} = \sum_{x:\mu(x)=k; s(x)=i} p^k q^{i-k} f^{n-i} =$$
$$= C_n^i C_i^k p^k q^{i-k} f^{n-i},\ i = 0, 1, 2, \ldots, n;\ k = 0, 1, \ldots, i.$$

Далее, случайная величина s — число $\zeta_j, j = 1, 2, \ldots, n$, отличных от нуля, имеет биномиальное распределение с параметрами $(n, p + q)$:

$$\mathsf{P}\{s = i\} = C_n^i (p + q)^i f^{n-i},\ i = 0, 1, 2, \ldots, n,$$

как число успехов в n испытаниях Бернулли, “успех” — случайная величина ζ_j отлична от нуля, “неудача” — ζ_j равна нулю, $j = 1, 2, \ldots, n$; вероятность успеха равна $p+q$, вероятность неудачи равна $1 - (p + q) = f$. Отсюда имеем, что

$$\mathsf{P}\{\mu = k | s = i\} = \frac{\mathsf{P}\{\mu = k, s = i\}}{\mathsf{P}\{s = i\}} = \frac{C_n^i C_i^k p^k q^{i-k} f^{n-i}}{C_n^i (p + q)^i f^{n-i}} =$$

$$= \frac{C_i^k p^k q^{i-k}}{(p + q)^i} = C_i^k \left(\frac{p}{p + q}\right)^k \left(\frac{q}{p + q}\right)^{i-k},\ k = 0, 1, 2, \ldots, i. \tag{25.3.2}$$

Тем самым теорема доказана.

Следствие. *Пусть случайные величины ζ_j, $j = 1, 2, \ldots, n$, независимы и симметрично распределены относительно нуля, тогда для каждого i $(i = 1, 2, \ldots, n)$:*

$$\mathsf{P}\{\mu = k | s = i\} = C_i^k \left(1/2\right)^i,\ k = 0, 1, \ldots, i. \tag{25.3.3}$$

Так как ζ_j, $j = 1, 2, \ldots, n$, симметрично распределены относительно нуля, то

$$p = \mathsf{P}\{\zeta_j > 0\} = \mathsf{P}\{-\zeta_j > 0\} = \mathsf{P}\{\zeta_j < 0\} = q.$$

А при $p = q$ условное распределение (25.3.1) принимает вид (25.3.3).

Далее, пусть случайные величины $\zeta_1, \zeta_2, \ldots, \zeta_n$ независимы и симметрично распределены относительно нуля, s — количество отличных от нуля среди $\zeta_1, \zeta_2, \ldots, \zeta_n$. Определим функцию $m_{\alpha;s}$ от случайной величины s следующим образом: для каждого значения i, $i = 1, 2, \ldots, n$, принятого случайной величиной s, определим $m_{\alpha;i}$ как верхний α-предел биномиального распределения с параметрами $(i; 1/2)$, т. е. $m_{\alpha;i}$ — минимальное m_i, для которого

$$\sum_{k=m_i+1}^{i} C_i^k \left(\frac{1}{2}\right)^i \le \alpha.$$

Тогда

$$\mathsf{P}\{\mu > m_{\alpha;i} | s = i\} = \sum_{k=m_{\alpha;i}+1}^{i} \mathsf{P}\{\mu = k | s = i\} =$$

$$= \sum_{k=m_{\alpha;i}+1}^{i} C_i^k \, (1/2)^i \le \alpha,$$

а вместе с этим и $\mathsf{P}\{\mu > m_{\alpha;s} | s = i\} \le \alpha$, поскольку

$$\mathsf{P}\{\mu > m_{\alpha;s} | s = i\} = \frac{\mathsf{P}\{\mu > m_{\alpha;s}, s = i\}}{\mathsf{P}\{s = i\}} =$$

$$= \frac{\mathsf{P}\{\mu > m_{\alpha;i}, s = i\}}{\mathsf{P}\{s = i\}} = \mathsf{P}\{\mu > m_{\alpha;i} | s = i\} \le \alpha.$$

Для $i = 0$ естественно положить $m_{\alpha;0} = 0$ и

$$\mathsf{P}\{\mu > m_{\alpha;s} | s = 0\} = 0.$$

Критерий знаков при наличии связей. *Пусть* (ξ_i, η_i), $i = 1, 2, \ldots, n$, *— n пар случайных величин, для которых разности* $\zeta_j = \eta_j - \xi_j$ *представимы в виде*

$$\zeta_j = \eta_j - \xi_j = \theta + e_j,\ j = 1, 2, \ldots, n,$$

где случайные величины $e_1, e_2, \ldots, e_n$ *независимы и распределены симметрично относительно нуля;* s *— число разностей* ζ_j *отличных от нуля,* μ *— число положительных среди них.*

Если гипотезу H_0*:* $\theta = 0$ *отклонять при*

$$\mu > m_{\alpha;s}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α*, гипотеза* H_0 *будет отклоняться, когда она верна (альтернатива односторонняя:* $\theta > 0$*).*

Если гипотезу H_0 *отклонять при*

$$\mu \notin [s - m_{\alpha;s}, m_{\alpha;s}]$$

и не отклонять в противном случае, то с вероятностью, не превосходящей 2α*, гипотеза* H_0 *будет отклоняться, когда она верна (альтернатива двусторонняя:* $\theta < 0$ *или* $\theta > 0$*).*

Доказательство. При верной гипотезе H_0 для случайных величин μ и $m_{\alpha;s}$ имеют место соотношения

$$\mathsf{P}\{\mu > m_{\alpha;s} | s = i\} \leq \alpha,\ i = 0, 1, 2, \ldots, n.$$

Отсюда, воспользовавшись формулой полной вероятности, получаем оценку

$$\mathsf{P}\{\mu > m_{\alpha;s}\} = \sum_{i=0}^{n} \mathsf{P}\{\mu > m_{\alpha;s} | s = i\} \mathsf{P}\{s = i\} \leq$$

$$\leq \sum_{i=0}^{n} \alpha \mathsf{P}\{s = i\} \leq \alpha,$$

т. е.

$$\mathsf{P}\{\mu > m_{\alpha;s}\} \leq \alpha.$$

Так что пользуясь критерием знаков для проверки гипотезы H_0: $\theta = 0$ против односторонней альтернативы $\theta > 0$, верную гипотезу H_0 будем отклонять с вероятностью, не превосходящей α.

Далее, поскольку при верной гипотезе H_0 имеют место неравенства

$$\mathsf{P}\{\mu \in (m_{\alpha;s}, s]\} \le \alpha,\ \mathsf{P}\{\mu \in [0, s - m_{\alpha;s})\} \le \alpha$$

(последнее неравенство следует из первого и равенств $C_n^k = C_n^{n-k}$), то

$$\mathsf{P}\{\mu \notin [s - m_{\alpha;s}; m_{\alpha;s}]\} =$$

$$= \mathsf{P}\{\mu \in [0, s - m_{\alpha;s})\} + \mathsf{P}\{\mu \in (m_{\alpha;s}, s]\} \le 2\alpha.$$

Поэтому, пользуясь критерием знаков для проверки гипотезы H_0: $\theta = 0$ против альтернативы: $\theta > 0$ или $\theta < 0$, верную гипотезу H_0 будем отклонять с вероятностью, не превосходящей 2α.

Пример 25.3.1. *В экспериментах по искусственному стимулированию дождя были замерены дождевые осадки в течение 16 пар дней, причем в каждой паре один день облака засеивали стимулятором, в другой — нет. Для каждой пары день засеивания выбирался случайным образом. В таблице приведены замеры осадков, полученные специальными приборами за эти 16 пар дней.*

Свидетельствуют ли приведенные данные о наличии эффекта засеивания?

Номер пары	Уровень осадков		Номер пары	Уровень осадков	
	с зас.	без зас.		с зас.	без зас.
1	0	1,37	9	0	0
2	2,09	0	10	1,87	0,62
3	0,07	0	11	2,50	0
4	0,30	0,10	12	3,15	5,54
5	0	0,44	13	0,15	0,01
6	2,55	0	14	2,96	0
7	1,62	1,01	15	0	0
8	0	0,54	16	0	0,75

Решение. Сформулируем задачу в терминах проверки статистических гипотез.

Имеется 16 пар наблюдений $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_{16}, \eta_{16})$ (η_j — замер уровня осадков после засеивания, ξ_j — без засеивания). Независимость пар наблюдений обеспечивается случайным выбором дней засеивания.

Естественно предположить, что случайные величины η_j и ξ_j связаны соотношениями

$$\zeta_j = \eta_j - \xi_j = \theta + e_j,$$

$j = 1, 2, \ldots, n$, причем e_j, $j = 1, 2, \ldots, n$, независимы и распределены симметрично относительно нуля. Относительно параметра θ выдвигается гипотеза H_0: $\theta = 0$ (гипотеза об отсутствии эффекта засеивания облаков стимулятором). Поскольку облака засеиваются с целью вызвать осадки, в качестве альтернативы естественно рассматривать одностороннюю альтернативу: $\theta > 0$. Отклонение гипотезы H_0: $\theta = 0$ в пользу альтернативы $\theta > 0$ будет свидетельствовать о наличии эффекта стимулятора, неотклонение — об отсутствии.

Для проверки гипотезы H_0 воспользуемся критерием знаков. Сначала определим знаки разностей $\eta_j - \xi_j$, $j = 1, 2, \ldots, 16$:

$$-,\ +,\ +,\ +,\ -,\ +,\ +,\ -, 0,\ +,\ +,\ -,\ +,\ +,\ 0,\ -.$$

Некоторые из них обращаются в нуль, поэтому будем пользоваться критерием знаков при наличии связей. Всего разностей $n = 16$, отличных от нуля $s = 14$. Разности, обращающиеся в нуль, учитывать не будем (как будто их и не было). Итак, имеется 14 отличных от нуля разностей, из них $\mu = 9$ положительных. Значение $m_{\alpha;s}$ для уровня значимости $\alpha = 0{,}025$ и $s = 14$, равное $m_{0{,}025;14} = 11$, определяем по табл. 27.9.1. Так что

$$\mu = 9 \leq 11 = m_{0{,}025;14}.$$

Поэтому согласно критерию знаков для проверки гипотезы H_0: $\theta = 0$ против альтернативы $\theta > 0$ гипотеза H_0 не отклоняется.

Полученный результат можно интерпретировать так. Предположение об отсутствии эффекта засеивания облаков стимулятором не противоречит опытным данным (9 положительных разностей на 14 ненулевых наблюдений естественны и допустимы). Такие результаты (см. таблицу) можно рассматривать как типичные для замеров осадков, если бы замеры производились в 16 парах случайно выбранных дней без какого-либо засеивания облаков.

25.4 Двухвыборочный критерий знаков

Постановка задачи. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_n$ — независимые выборки соответственно из распределений F и G. Распределения F и G связаны соотношением $G(x) = F(x - \theta)$. Параметр θ неизвестен. Необходимо выяснить, отлично значение θ от нуля или нет. Будем решать поставленную задачу, проверяя гипотезу H_0: $\theta = 0$ (нулевую гипотезу можно сформулировать и так H_0: $\mathsf{F} = \mathsf{G}$). Альтернатива к гипотезе H_0: $\theta = 0$ может быть как односторонней, так и двусторонней.

Гипотезу H_0: $\theta = 0$ можно проверять, пользуясь критерием знаков в той же формулировке, что и для парных наблюдений: подсчитать число μ положительных разностей $\zeta_j = \eta_j - \xi_j$ среди $\zeta_1, \zeta_2, \ldots, \zeta_n$ и сравнивать его с $m_{\alpha;s}$ (см. критерий знаков при наличии связей). Чтобы убедиться в справедливости этого утверждения, покажем, что разности $\zeta_j = \eta_j - \xi_j$ можно представить в виде

$$\zeta_j = \eta_j - \xi_j = \theta + e_j,$$

где e_j, $j = 1, 2, \ldots, n$, — независимые, симметрично распределенные относительно нуля случайные величины.

Запишем ζ_j в виде

$$\zeta_j = \eta_j - \xi_j = \theta + ((\eta_j - \theta) - \xi_j) = \theta + e_j,$$

где $e_j = (\eta_j - \theta) - \xi_j$, $j = 1, 2, \ldots, n$. Случайные величины $e_1, e_2, \ldots, e_n$ независимы как функции от независимых случайных величин. Убедимся еще, что $e_1, e_2, \ldots, e_n$ распределены симметрично относительно нуля (распределения e_j и $-e_j$ совпадают). Для этого вычислим характеристические функции случайных величин e_j и $-e_j$. Характеристическая функция e_j :

$$\begin{aligned}\mathsf{M}\exp\{ite_j\} &= \mathsf{M}\exp\{it((\eta_j - \theta) - \xi_j)\} = \\ &= \mathsf{M}\exp\{it(\eta_j - \theta)\}\exp\{i(-t)\xi_j\} = \\ &= \mathsf{M}\exp\{it(\eta_j - \theta)\}\mathsf{M}\exp\{i(-t)\xi_j\}.\end{aligned} \tag{25.4.1}$$

Обозначим через $\varphi(t)$ характеристическую функцию случайной величины ξ_j. Распределение случайной величины $\eta_j - \theta$ совпадает с распределением F случайной величины ξ_j :

$$\mathsf{P}\{\eta_j - \theta < x\} = \mathsf{P}\{\eta_j < x + \theta\} = G(x + \theta) = F((x + \theta) - \theta) = F(x),$$

поэтому характеристическая функция случайной величины $\eta_j - \theta$ совпадает с характеристической функций $\varphi(t)$ случайной величины ξ_j и для характеристических функций случайных величин e_j и $-e_j$ получаем представления:

$$\mathsf{M}\exp\{ite_j\} = \varphi(t)\varphi(-t),$$

$$\mathsf{M}\exp\{it(-e_j)\} = \varphi(-t)\varphi(t)$$

(см. (25.4.1)). Характеристические функции случайных величин e_j и $-e_j$ совпадают, поэтому в силу теоремы единственности совпадают и их распределения, т. е. случайные величины e_j симметрично распределены относительно нуля.

Таким образом, для независимых выборок $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_n$ соответственно из распределений F и G, связанных соотношением $G(x) = \mathsf{F}(x - \theta)$, разности $\zeta_j = \eta_j - \xi_j$ представимы в виде:

$$\zeta_j = \eta_j - \xi_j = \theta + e_j,$$

$j = 1, 2, \ldots, n$, причем случайные величины $e_1, e_2, \ldots, e_n$ являются независимыми и симметрично распределенными относительно нуля. Поэтому для проверки гипотезы H_0: $\theta = 0$ можно пользоваться критерием знаков в приведенной ранее формулировке.

25.5 Критерий Вилкоксона

Первый прорыв в непараметрической статистике сделал А. Н. Колмогоров (статистика А. Н. Колмогорова), второй — Вилкоксон (ранговые критерии).

Мы рассмотрим критерий Вилкоксона для проверки гипотезы об одинаковой распределенности случайных величин.

Постановка задачи. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ — независимые выборки соответственно из абсолютно непрерывных распределений F и G с плотностями f и g (далее через $\xi_1, \xi_2, \ldots, \xi_n$ будем обозначать выборку меньшего объема, через $\eta_1, \eta_2, \ldots, \eta_m$ — большего). Распределения F и G связаны соотношением $G(x) = F(x - \theta)$, но ни распределения F и G, ни значение параметра θ неизвестны. Относительно параметра θ выдвигается гипотеза $H_0 : \theta = 0$ (ее можно также сформулировать в виде H_0: $\mathsf{F} = \mathsf{G}$). Альтернатива гипотезе H_0 может быть как односторонней: если $\theta \neq 0$, то $\theta > 0$ (она может быть и такой: $\theta < 0$), так и двусторонней: если $\theta \neq 0$, то $\theta > 0$ или $\theta < 0$.

Ниже строится критерий Вилкоксона для проверки гипотезы H_0: $\theta = 0$. Построение критерия основано на следующих соображениях.

По плотности $f(x)$ случайной величины ξ вероятность попадания ξ в данный промежуток $[a, b]$ вычисляется так:

$$\mathsf{P}\{\xi \in [a, b]\} = \int_a^b f(x)dx$$

(численно вероятность $\mathsf{P}\{\xi \in [a,b]\}$ равна площади криволинейной трапеции с основанием $[a,b]$, ограниченной сверху плотностью $f(x)$ (см. рис. 25.5.1)). Поэтому вне зависимости от того, верна гипотеза H_0 или нет, выборочные значения ξ_i и η_j по отношению к своим плотностям $f(x)$ и $g(x)$ располагаются так, как это изображено на рис. 25.5.1 и 25.5.2, а именно "в большинстве" сосредоточены на тех промежутках, где плотности $f(x)$ и $g(x)$ принимают большие значения (на рис. 25.5.1 и 25.5.2 выборочные значения $\xi_1, \xi_2, \ldots, \xi_n$ изображены точками, а выборочные значения $\eta_1, \eta_2, \ldots, \eta_m$ — звездочками). Отсюда имеем, что если $\mathsf{G} = \mathsf{F}$, т. е. $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ — независимые выборки из одного и того же распределения, то выборочные значения $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ (и соответствующие им точки и звездочки) "перемешаны хорошо" (см. рис. 25.5.1). Если же $\theta \neq 0$, например $\theta > 0$, то выборочные значения $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ (и соответствующие им точки и звездочки) "перемешаны плохо" (см. рис. 25.5.2) — большинство звездочек лежит правее точек, поскольку $g(x) = f(x - \theta)$, $\theta > 0$.

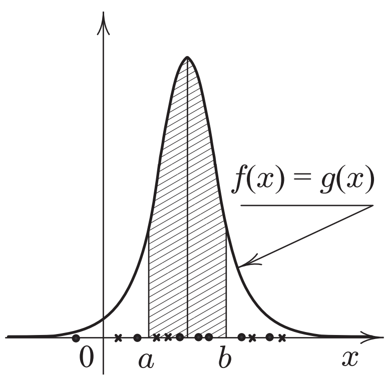

Рис. 25.5.1: Нулевая гипотеза $H_0 : \theta = 0$ верна, точки и звездочки перемешаны хорошо

Так что для проверки гипотезы H_0: $\mathsf{G} = \mathsf{F}$ естественно расположить выборочные значения $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ в общий вариационный ряд:

$$\eta\xi\eta\xi\eta\eta\eta \ldots \xi\eta\eta$$

(индексы опущены — в них нет необходимости). Если при этом ξ и η в общем вариационном ряду окажутся "перемешанными хорошо", гипотезу H_0 не отклоняем, "плохо" — отклоняем.

Для того, чтобы так можно было поступать, необходимо иметь количественную характеристику интуитивно ясного понятия “перемешанности” букв ξ и η в общем вариационном ряду, такой характеристикой является статистика Вилкоксона.

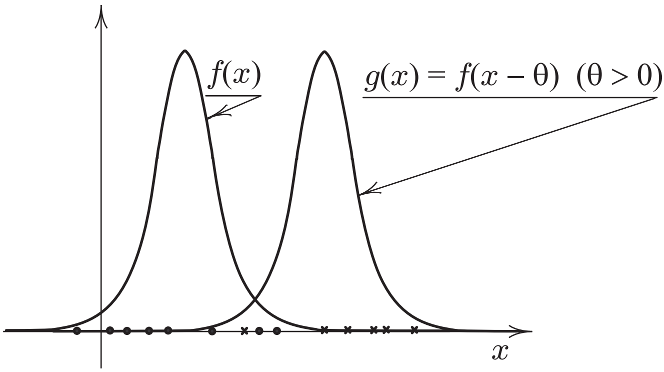

Рис. 25.5.2: Нулевая гипотеза H_0: $\theta = 0$ неверна (альтернатива $\theta > 0$), точки и звездочки перемешаны плохо

Статистика Вилкоксона. *Рангом* числа ζ_i в последовательности чисел $\zeta_1, \zeta_2, \ldots, \zeta_k$ будем называть номер места, который оно получает в вариационном ряду $\zeta_1^*, \zeta_2^*, \ldots, \zeta_k^*$.

Ранг 1 получает наименьшее из чисел $\zeta_1, \zeta_2, \ldots, \zeta_k$, ранг 2 — второе по величине и т. д., ранг k получает наибольшее из чисел $\zeta_1, \zeta_2, \ldots, \zeta_k$. Если среди чисел $\zeta_1, \zeta_2, \ldots, \zeta_k$ есть совпадающие, то каждому из них мы приписываем средний ранг. Например, если $\zeta_i = 8{,}6$, $\zeta_j = 8{,}6$, $\zeta_l = 8{,}6$ и перед 8,6 в вариационном ряду расположены 4 выборочных значения, то каждому из $\zeta_i, \zeta_j, \zeta_l$ приписывается один и тот же ранг $(5 + 6 + 7)/3 = 6$.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ — независимые выборки соответственно из распределений F и G. Расположив выборочные значения в общий вариационный ряд, мы получим перестановку с повторениями из n букв ξ и m букв η $(n \le m)$. Множество всех таких перестановок обозначим через Ω, сами перестановки будем обозначать через ω. Определим на Ω функцию $W = W(\omega)$ как сумму рангов выборочных значений выборки меньшего объема. Так что если в перестановке $\omega = (\xi\eta\xi\eta\eta\ldots\xi)$ буквы ξ располагаются на местах с номерами $r_1, r_2, \ldots, r_n$, то

$$W(\omega) = r_1 + r_2 + \cdots + r_n.$$

Функция $W = W(\omega)$ называется *статистикой Вилкоксона.*

О степени перемешанности букв ξ и η в перестановке $\omega = (\xi\eta\xi\eta\eta\ldots\xi)$ можно судить по значениям, принимаемым статистикой Вилкоксона $W(\omega)$: большие значения $W(\omega)$, когда ξ в основном идут после η, или малые, когда большинство ξ располагаются до η, свидетельствуют о плохой перемешанности ξ и η в перестановке ω. Средние значения $W(\omega)$ говорят о хорошей перемешанности ξ и η в перестановке ω.

Замечание. Поскольку $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ — выборки из непрерывных распределений F и G, среди $\xi_1, \xi_2, \ldots, \xi_n, \eta_1, \eta_2, \ldots, \eta_m$ с вероятностью 1 нет совпадающих. Но у реализаций выборок $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$, полученных в эксперименте, могут встречаться совпадающие значения, это происходит за счёт регистрации выборочных значений с конечным числом знаков.

Распределение статистики Вилкоксона. Независимые выборки $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ соответственно из распределений F и G ($G(x) = F(x - \theta)$) задают на множестве Ω перестановок с повторениями из n букв ξ и m букв η вероятность P. Вероятность P зависит от параметра θ. При $\theta = 0$ (или, что то же, $\mathsf{G} = \mathsf{F}$) все перестановки ω равновероятны, если же $\theta \neq 0$, например, $\theta > 0$, то чаще встречаются перестановки, у которых в основном сначала идут ξ, а затем η (поскольку $G(x) = F(x-\theta)$), такие перестановки имеют бóльшую вероятность по сравнению с другими.

Распределение статистики Вилкоксона — случайной величины $W = W(\omega)$, определенной на дискретном пространстве Ω, вычисляется так:

$$\mathsf{P}\{W = l\} = \sum_{\omega: W(\omega) = l} \mathsf{P}(\omega).$$

При верной гипотезе H_0: $\theta = 0$ (или, что то же, $\mathsf{G} = \mathsf{F}$) все перестановки ω равновероятны, поэтому каждой из них необходимо приписать одну и ту же вероятность

$$\mathsf{P}(\omega) = 1/C_{n+m}^n$$

(число элементов в Ω равно C_{n+m}^n), и, следовательно,

$$\mathsf{P}\{W = l\} = \sum_{\omega: W(\omega) = l} \frac{1}{C_{n+m}^n}. \tag{25.5.1}$$

О симметричности распределения статистики W при $\mathsf{G}=\mathsf{F}$. Заметим, что $n(n+1)/2$ — минимально возможное значение случайной величины W, $nm+n(n+1)/2$ — максимально возможное, $n(n+m+1)/2$ — среднее значение W.

Далее статистику W при верной гипотезе H_0: $\mathsf{G}=\mathsf{F}$ будем обозначать через W_0.

Теорема. *Если* $\mathsf{G}=\mathsf{F}$, *то случайная величина* W_0 *распределена симметрично относительно середины* $n(n+m+1)/2$ *промежутка* $[n(n+1)/2,\ nm+n(n+1)/2]$ *— значения* W_0, *равноудаленные от точки* $n(n+m+1)/2$, *принимаются с равными вероятностями.*

Доказательство. Сначала покажем, что значения случайной величины $W_0=W_0(\omega)$, равноудаленные от концов промежутка $[n(n+1)/2,\ nm+n(n+1)/2]$, принимаются с равными вероятностями.

Рассмотрим две перестановки с повторениями — перестановку $\omega=(\xi\xi\ldots\xi\eta\eta\ldots\eta)$, у которой сначала располагаются ξ, а затем η, и перестановку $\omega^*=(\eta\ldots\eta\eta\xi\ldots\xi\xi)$, полученную из перестановки ω записью букв в обратном порядке (полученную "зеркальным" отображением). Для этих перестановок

$$W_0(\omega)=\frac{n(n+1)}{2},\ W_0(\omega^*)=nm+\frac{n(n+1)}{2}.$$

Будем в перестановке ω последовательно менять местами соседние буквы ξ и η, а в перестановке ω^* соответствующие соседние буквы η и ξ, при этом в перестановке ω буквы ξ "перемещаются" вправо, а в перестановке ω^* — влево; за перестановками оставим те же обозначения. В результате каждой такой перемены мест значение $W_0(\omega)$ увеличивается на единицу, а значение $W_0(\omega^*)$ уменьшается на единицу. После первой перемены мест соседних букв ξ и η

$$W_0(\omega)=\frac{n(n+1)}{2}+1,\ W_0(\omega^*)=nm+\frac{n(n+1)}{2}-1,$$

после второй перемены мест —

$$W_0(\omega)=\frac{n(n+1)}{2}+2,\ W_0(\omega^*)=nm+\frac{n(n+1)}{2}-2$$

и т. д. Так что для каждого k, $k=0,1,\ldots,u$ ($u=[n(n+m+1)/2]-n(n+1)/2$, $[x]$ — целая часть x) число перестановок, для которых

$$W_0(\omega)=\frac{n(n+1)}{2}+k,$$

и число перестановок, для которых

$$W_0(\omega^*) = nm + \frac{n(n+1)}{2} - k,$$

одинаково. Поэтому значения $n(n+1)/2+k$ и $nm+n(n+1)/2 - k$ случайной величины W_0, равноудаленные от концов промежутка $[n(n+1)/2, nm + n(n+1)/2]$, принимаются с равными вероятностями. А, следовательно, с равными вероятностями принимаются и значения W_0, равноудаленные от середины

$$n(n+m+1)/2$$

промежутка $[n(n+1)/2, nm + n(n+1)/2]$, поскольку значения

$$W_0(\omega) = n(n+1)/2 + k \quad \text{и} \quad W_0^*(\omega) = nm + n(n+1)/2 - k,$$

$k = 0, 1, \ldots, u$, равноудаленные от концов промежутка

$$\left[n(n+1)/2,\ nm + n(n+1)/2\right],$$

равноудалены и от его середины $n(n+m+1)/2$:

$$\frac{n(n+1)}{2} + k = \frac{n(n+m+1)}{2} - \frac{n(n+m+1)}{2} + \frac{n(n+1)}{2} + k =$$

$$= \frac{n(n+m+1)}{2} - \left(\frac{nm}{2} - k\right) = \frac{n(n+m+1)}{2} - s,$$

$$nm + \frac{n(n+1)}{2} - k = \frac{n(n+m+1)}{2} -$$

$$- \frac{n(n+m+1)}{2} + nm + \frac{n(n+1)}{2} - k =$$

$$= \frac{n(n+m+1)}{2} + \left(\frac{nm}{2} - k\right) = \frac{n(n+m+1)}{2} + s.$$

Тем самым утверждение доказано.

Следствие. *Если* $\mathsf{G} = \mathsf{F}$*, то*

$$\mathsf{M}W_0 = \frac{n(n+m+1)}{2}.$$

Случайная величина $W_0 - n(n+m+1)/2$ значения s и $-s$ принимает с одинаковыми вероятностями. Поэтому

$$\mathsf{M}\,(W_0 - n(n+m+1)/2) = 0,$$

и, следовательно,

$$\mathsf{M}W_0 = n(n+m+1)/2.$$

Определение. *Нижним α-пределом* $W_{\alpha;n;m}$ статистики Вилкоксона называется наибольшее целое t, для которого

$$\mathsf{P}\{W_0 \le t\} \le \alpha.$$

По определению

$$\mathsf{P}\{W_0 \le W_{\alpha;n;m}\} \le \alpha.$$

Из симметричности распределения случайной величины W_0 на промежутке

$$[n(n+1)/2, nm + n(n+1)/2]$$

и определения $W_{\alpha;n;m}$ получаем:

$$\mathsf{P}\{W_0 \ge n(n+m+1) - W_{\alpha;n;m}\} \le \alpha.$$

По заданным α, n, m значения $W_{\alpha;n;m}$ табулированы (см., например, табл. 27.8.1).

Выбор статистики для построения критерия. При верной гипотезе H_0: $\theta = 0$ случайная величина $W = W_0$ мало уклоняется от $n(n+m+1)/2$ по сравнению с уклонением W от $n(n+m+1)/2$, когда гипотеза H_0 неверна, а именно: если $\theta = 0$, то

$$\mathsf{M}\,(W_0 - n(n+m+1)/2) = 0,$$

и W_0 почти всегда (с вероятностью, не меньшей $1 - 2\alpha$) принимает значения из промежутка

$$(W_{\alpha;n;m};\ n(n+m+1) - W_{\alpha;n;m}).$$

Если гипотеза H_0: $\theta = 0$ неверна, например $\theta > 0$, то перестановкам с повторениями из n букв ξ и m букв η, у которых сначала в большинстве идут ξ, а затем η, приписываются бо́льшие вероятности, поэтому W с большой вероятностью принимает малые значения (значения из промежутка $[n(n+1)/2;\ W_{\alpha;n;m}]$) и, следовательно, существенно уклоняется от $n(n+m+1)/2$.

Так что для проверки гипотезы H_0: $\theta = 0$ (альтернатива $\theta \neq 0$) естественно вычислить значение W и выяснить, насколько оно отклоняется от $n(n+m+1)/2$. Если W отклоняется от $n(n+m+1)/2$, значительно, т. е.

$$W \notin (W_{\alpha;n;m}; n(n+m+1) - W_{\alpha;n;m}),$$

гипотезу H_0 отклоняем, в противном случае — не отклоняем.

Критерий Вилкоксона. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_m$ — независимые выборки из абсолютно непрерывных распределений F и G соответственно, $G(x) = F(x-\theta)$. Если гипотезу H_0: $\theta = 0$ отклонять при*

$$W \leq W_{\alpha;n;m}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α, гипотеза H_0 будет отклоняться, когда она верна (альтернатива односторонняя: $\theta > 0$).

Если гипотезу H_0 отклонять при

$$W \geq n(n+m+1) - W_{\alpha;n;m}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α, гипотеза H_0 будет отклоняться, когда она верна (альтернатива односторонняя: $\theta < 0$).

Если гипотезу H_0 отклонять при

$$W \notin (W_{\alpha;n;m}; n(n+m+1) - W_{\alpha;n;m})$$

и не отклонять в противном случае, то с вероятностью, не превосходящей 2α, гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя: $\theta < 0$ или $\theta > 0$).

Доказательство. При верной гипотезе H_0

$$\mathsf{P}\{W \leq W_{\alpha;n;m}\} \leq \alpha, \ \mathsf{P}\{W \geq n(n+m+1) - W_{\alpha;n;m}\} \leq \alpha,$$

$$\mathsf{P}\{W \notin (W_{\alpha;n;m}; n(n+m+1) - W_{\alpha;n;m})\} \leq 2\alpha.$$

Поэтому, пользуясь критерием Вилкоксона, верную гипотезу H_0: $\theta = 0$ мы будем отклонять с вероятностью, не превосходящей α, когда она проверяется против односторонней альтернативы $\theta > 0$ (или $\theta < 0$) и с вероятностью, не превосходящей 2α, когда она проверяется против двусторонней альтернативы $\theta < 0$ или $\theta > 0$.

Критерий Вилкоксона при больших $\boldsymbol{n, m}$. Непосредственное вычисление $W_{\alpha;n;m}$ довольно затруднительно и с ростом n и m трудности возрастают. Но для больших n и m значение $W_{\alpha;n;m}$ можно получить, пользуясь тем, что при $\min\{n, m\} \to \infty$ статистика W_0 асимптотически нормальна.

Теорема. *Пусть* $\xi_1, \xi_2, \ldots, \xi_n$ *и* $\eta_1, \eta_2, \ldots, \eta_m$ — *независимые выборки из непрерывного распределения. При* $\min\{n, m\} \to \infty$ *случайная величина* W_0 *асимптотически нормальна со средним* $n(n+m+1)/2$ *и дисперсией* $nm(n+m+1)/12$, *т. е.*

$$\mathsf{P}\left\{\frac{W_0 - \frac{1}{2}n(n+m+1)}{\sqrt{\frac{1}{12}nm(n+m+1)}} < x\right\} \to \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} \exp\left\{-s^2/2\right\} ds$$

для каждого $x \in \mathbb{R}^1$.

Нормальное распределение хорошо аппроксимирует распределение W_0, а именно, погрешность при замене

$$\mathsf{P}\left\{\frac{W_0 - \frac{1}{2}n(n+m+1)}{\sqrt{\frac{1}{12}nm(n+m+1)}} < x\right\} \tag{25.5.2}$$

значением

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} \exp\left\{-s^2/2\right\} ds$$

функции распределения нормального распределения с параметрами $(0;1)$ мала, уже начиная с n, m, для которых

$$\min\{n, m\} \geq 6 \text{ и } n+m \geq 20.$$

Поэтому при $\min\{n, m\} \geq 6$ и $n+m \geq 20$ можно считать, что вероятность (25.5.2) равна

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} \exp\left\{-s^2/2\right\} ds,$$

в частности, для $x = z_\alpha$ — α-квантили нормального распределения с параметрами $(0;1)$:

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z_\alpha} \exp\left\{-s^2/2\right\} ds = \alpha,$$

можно считать, что

$$\mathsf{P}\left\{\frac{W_0-\frac{1}{2}n(n+m+1)}{\sqrt{\frac{1}{12}nm(n+m+1)}}\leq z_\alpha\right\}=\alpha.$$

Отсюда

$$\mathsf{P}\left\{W_0\leq\frac{1}{2}n(n+m+1)+z_\alpha\sqrt{\frac{1}{12}nm(n+m+1)}\right\}=\alpha.$$

Так что при $\min\{n,m\}\geq 6$, $n+m\geq 20$ значения $W_{\alpha;n,m}$ и $n(n+m+1)-W_{\alpha;n,m}$ можно получить так:

$$W_{\alpha;n,m}=\frac{1}{2}n(n+m+1)+z_\alpha\sqrt{\frac{1}{12}nm(n+m+1)},$$

$$n(n+m+1)-W_{\alpha;n,m}=$$

$$=\frac{1}{2}n(n+m+1)-z_\alpha\sqrt{\frac{1}{12}nm(n+m+1)},$$

где z_α — α-квантиль нормального распределения с параметрами $(0;1)$.

Об ошибках первого и второго рода. При верной гипотезе H_0: $\theta=0$ (пусть для определенности альтернатива двусторонняя) случайная величина $W=W_0$ — сумма рангов выборочных значений выборки меньшего объема, мало уклоняясь от середины $n(n+m+1)/2$ промежутка $[n(n+1)/2, nm+n(n+1)/2]$ и почти всегда (с вероятностью не меньшей $1-2\alpha$) принимая значения из промежутка $(W_{\alpha;n,m}; n(n+m+1)-W_{\alpha;n,m})$, может, хотя и изредка, принимать малые значения (т. е. $W_0\leq W_{\alpha;n,m}$) или большие (т. е. $W_0\geq n(n+m+1)-W_{\alpha;n,m}$). При этом мы гипотезу H_0 отклоняем и тем самым совершаем ошибку — ошибку первого рода.

Если гипотеза H_0 неверна, то W, почти всегда принимая большие или малые значения, может, хотя и изредка, принять значение, близкое к $n(n+m+1)/2$, т. е.

$$W\in(W_{\alpha;n,m}; n(n+m+1)-W_{\alpha;n,m}),$$

при этом мы гипотезу H_0 не отклоним и тем самым совершим ошибку — ошибку второго рода.

Пример 25.5.1. *Четное число мышей было рассажено по одной в клетки, объединенные случайным образом в две группы с одинаковым количеством клеток: группа A — контрольная, группа B — испытуемая. Мышам из группы B был введен препарат, который, предположительно, угнетает палочку Коха (возбудитель туберкулеза). После этого всех животных в случайном порядке заражали туберкулезом. (Отметим, что обычно экспериментаторы сначала вносят инфекцию животным контрольной группы, а затем испытуемой, что совершенно неправильно.) В таблице приведены дни смерти мышей после инфицирования, при этом данные об одной из мышей были утеряны.*

Группа A	5	6	7	7	8	8	8	9	12	
Группа B	7	8	8	8	9	9	12	13	14	17

Из предварительных экспериментов известно, что применяемый препарат не является токсичным, поэтому можно считать, что продолжительность жизни в испытуемой группе не меньше, чем в контрольной.

Можно ли на основании приведенных данных говорить об эффекте препарата?

Решить поставленную задачу, сформулировав ее как задачу проверки статистических гипотез.

Р е ш е н и е. В терминах проверки статистических гипотез задачу об эффекте препарата можно сформулировать так. Имеются реализации независимых выборок $\xi_1, \xi_2, \ldots, \xi_9$ и $\eta_1, \eta_2, \ldots \ldots, \eta_{10}$ соответственно из распределений F и G, связанных соотношением $G(x) = F(x - \theta)$ (F — распределение продолжительности жизни мышей в контрольной группе, G — в испытуемой). Относительно параметра θ выдвигается гипотеза H_0: $\theta = 0$ (гипотеза об отсутствии эффекта препарата). Поскольку мы интересуемся наличием противотуберкулезного эффекта препарат и к тому же известно, что препарат заведомо не токсичен (продолжительность жизни в испытуемой группе не меньше, чем в контрольной), то в качестве альтернативы к гипотезе H_0: $\theta = 0$ следует рассматривать одностороннюю альтернативу $\theta > 0$. Необходимо проверить гипотезу H_0 против альтернативы $\theta > 0$, т. е. по реализациям выборок сделать заключение — отклонять H_0 или не отклонять. Отклонение гипотезы H_0: $\theta = 0$ в пользу альтернативы $\theta > 0$ будем интерпретировать как наличие противотуберкулезного эффекта препарата, неотклонение — как отсутствие эффекта.

Продолжительность жизни мышей естественно рассматривать как непрерывно распределенную случайную величину, поэтому в качестве критерия для проверки гипотезы H_0 можно воспользоваться критерием Вилкоксона, что мы и сделаем.

Расположим выборочные значения $\xi_1, \xi_2, \ldots, \xi_9$ и $\eta_1, \eta_2, \ldots$ $\ldots, \eta_{10}$ в общий вариационный ряд, подчеркнув выборочные значения выборки меньшего объема:

$$\underline{5}\ \underline{6}\ \underline{7}\ \underline{7}\ 7\ \underline{8}\ \underline{8}\ \underline{8}\ 8\ 8\ 8\ \underline{9}\ 9\ 9\ \underline{12}\ 12\ 13\ 14\ 17.$$

Подсчитаем значение W — суммы рангов выборочных значений выборки меньшего объема (сумму рангов подчеркнутых выборочных значений):

$$W = 1 + 2 + 2 \cdot \frac{3+4+5}{3} + 3 \cdot \frac{6+7+8+9+10+11}{6} + \\ + \frac{12+13+14}{3} + \frac{15+16}{2} = 65.$$

Далее, для $\alpha = 0{,}05$, $n = 9$, $m = 10$

$$W_{\alpha;n;m} = W_{0{,}05;9;10} = 69$$

(см. табл. 27.8.1). Отсюда

$$W = 65 < 69 = W_{\alpha;n;m},$$

и, следовательно, гипотеза H_0: $\theta = 0$ об отсутствии эффекта отклоняется в пользу односторонней альтернативы $\theta > 0$ — она противоречит опытным данным. Для препарата, не оказывающего противотуберкулезного эффекта, такое превышение продолжительности жизни в испытуемой группе по сравнению с контрольной практически невозможно. Так что указанное превышение продолжительности жизни мышей в испытуемой группе свидетельствует в пользу наличия эффекта препарата.

25.6 Примеры и задачи

Примеры

Пример 25.6.1. *Воспользовавшись таблицей случайных чисел* (*табл.* 27.10.1), *получить выборку объемом* 10 *из стандартного нормального распределения.*

С помощью критерия Колмогорова проверить, действительно ли эта выборка получена из указанного распределения.

Решение. Если случайная величина η равномерно распределена на промежутке [0;1] и $F(x)$ — возрастающая непрерывная функция распределения, то случайная величина ξ, определенная равенством

$$\xi = F^{-1}(\eta),$$

имеет своей функцией распределения $F(x)$. В самом деле,

$$F_\xi(x) = \mathsf{P}\{\xi < x\} = \mathsf{P}\{F^{-1}(\eta) < x\} =$$

$$= \mathsf{P}\{\eta < F(x)\} = F_\eta\left(F(x)\right) = F(x).$$

Это утверждение дает возможность по выборке $\eta = (\eta_1, \eta_2, \ldots$ $\ldots, \eta_n)$ из равномерного на промежутке [0;1] распределения строить выборку $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ из данного распределения F, а именно:

$$\xi_i = F^{-1}(\eta_i),\ i = 1, 2, \ldots, n,$$

является выборкой из распределения F. В частности, если в качестве $F(x)$ рассмотреть функцию стандартного нормального распределения

$$N_{0;1}(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left\{-\frac{s^2}{2}\right\} ds,$$

то

$$\xi_i = N_{0;1}^{-1}(\eta_i),\ i = 1, 2, \ldots, 10,$$

является выборкой из $\mathsf{N}_{0;1}$.

Выборку из равномерного на отрезке [0;1] распределения можно получить, воспользовавшись таблицей случайных чисел (табл. 27.10.1).

Выберем из таблицы 27.10.1 последовательно 10 чисел (для определенности — четырехзначных). Выбор можно начинать с любого места таблицы (скажем, с верхнего правого угла) и продолжать любым предварительно оговоренным способом. Например, двигаясь по диагонали, получим:

1009	5420	2689	2529	7080
3407	5718	1656	7048	7835

Числа

0,1009	0,5420	0,2689	0,2529	0,7080
0,3407	0,5718	0,1656	0,7048	0,7835

из отрезка [0;1] можно рассматривать как реализацию выборки $\eta_1, \eta_2, \ldots, \eta_{10}$ из равномерного на промежутке $[0;1]$ распределения. Реализацию выборки $\xi_1, \xi_2, \ldots, \xi_{10}$ из стандартного нормального распределения по выборке $\eta_1, \eta_2, \ldots, \eta_{10}$ получим так:

$$\xi_i = N_{0;1}^{-1}(\eta_i) = \Phi^{-1}(\eta_i),$$

$i = 1, 2, \ldots, 10$ (функция $N_{0;1}(x) = \Phi(x)$ табулирована, см. табл. 27.1.1). Имеем:

−1,27	0,11	−0,62	−0,67	0,55
−0,41	0,18	−0,97	0,54	0,78

(последний знак получен методом линейной интерполяции).

Далее, пользуясь критерием Колмогорова, проверим нулевую гипотезу о том, что $\xi_1, \xi_2, \ldots, \xi_{10}$ — выборка из распределения $\mathsf{N}_{0;1}$. Для этого вычислим

$$D(\hat{\mathsf{F}}_n, \mathsf{G}) = \sup_x \left| G(x) - \hat{F}_n(x) \right|,$$

где $n = 10$, $G(x) = N_{0;1}(x)$ — функция нормального распределения с параметрами (0;1), $\hat{F}_n(x)$ — реализация эмпирической функции распределения, построенной по выборке $\xi_1, \xi_2, \ldots, \xi_{10}$:

$$D(\hat{\mathsf{F}}_n, \mathsf{N}_{0;1}) = \sup_x \left| N_{0;1}(x) - \hat{F}_n(x) \right| = 0{,}21.$$

И сравним значение уклонения $D(\hat{\mathsf{F}}_n, \mathsf{G})$ с критическим значением $\varepsilon_{\alpha;n}$:

$$D(\hat{\mathsf{F}}_n, \mathsf{N}_{0;1}) = 0{,}21 < 0{,}4087 = \varepsilon_{0{,}05;10} = \varepsilon_{\alpha;n}$$

(критическое значение $\varepsilon_{0{,}05;10}$ получено из табл. 27.7.1). Поэтому, в соответствии с критерием Колмогорова, гипотеза H_0: $\xi_1, \xi_2, \ldots, \xi_{10}$ — выборка из нормального распределения с параметрами (0;1), на 5% уровне значимости не отклоняется.

Для выборки объемом 10 из нормального распределения с параметрами (0;1) максимальное отклонение между эмпирической функцией распределения $\hat{F}_{10}(x)$ и функцией распределения $N_{0;1}(x)$, равное 0,21, не является большим, такое отклонение естественно и допустимо.

Пример 25.6.2 (смещение результатов измерений у контролеров). *Один из методов количественного анализа степени износа шины состоит в измерении глубины проникновения щупа*[1] *в канавку протектора в определенном месте шины.*

В рамках дорожно-эксплуатационных исследований два контролера измеряют глубину канавок на шинах после каждого эксперимента. Основная трудность проведения измерений состоит в том, что каждый контролер имеет свою, присущую только ему смещенность результатов измерений, связанную с силой давления на щуп (от силы давления зависит глубина проникновения щупа в канавку протектора).

Результаты 10 *измерений, выполненных контролерами в* 10 *фиксированных точках шины, приведены в таблице.*

Есть предположение, что контролер I получает более высокие результаты, чем контролер II.

Согласуется ли это предположение с результатами эксперимента?

Точка	Контролер I	Контролер II	Точка	Контролер I	Контролер II
1	126	125	6	159	152
2	128	120	7	152	150
3	157	163	8	138	136
4	131	118	9	138	140
5	142	129	10	142	136

Решение. Результаты измерений, выполненных первым контролером, обозначим через η_j, вторым — через ξ_j, $j = 1, 2, \ldots, 10$. Разности $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \ldots, 10$, естественно считать представимыми в виде

$$\zeta_j = \theta + e_j,\ \ j = 1, 2, \ldots, 10,$$

где e_j — независимые симметрично распределенные случайные величины. Параметр θ характеризует отличия, если они имеются, в силе давления контролеров I и II на щуп: если $\theta = 0$, то отличий нет, если $\theta \neq 0$, то они имеются.

По данным измерений необходимо прийти к заключению, действительно ли контролер I получает более высокие результаты по сравнению с контролером II.

В терминах проверки статистических гипотез эту задачу можно сформулировать так. Относительно параметра θ выдвигаем

[1] Здесь щуп — тонкая продолговатая металлическая пластинка прямоугольной формы.

гипотезу H_0: $\theta = 0$ (контролеры получают одинаковые результаты). В качестве альтернативы к гипотезе H_0 выбираем одностороннюю альтернативу: $\theta > 0$ (поскольку имеется подозрение, что контролер I получает более высокие результаты). Отклонение гипотезы H_0: $\theta = 0$ будем трактовать в пользу более высоких результатов контролера I по сравнению с результатами контролера II, неотклонение — как отсутствие различий в измерениях контролеров.

Для проверки гипотезы H_0 воспользуемся критерием знаков. Имеем такие знаки разностей:

$$+ + - + + + + + - + .$$

Всего разностей 10 (все они отличны от нуля), количество μ положительных среди них равно 8. При этом

$$\mu = 8 \leq 8 = m_{0,025;10} = m_{\alpha;n}.$$

Поэтому, в соответствии с критерием знаков, для проверки гипотезы H_0: $\theta = 0$ против альтернативы $\theta > 0$ гипотеза H_0 на уровне значимости 0,025 не отклоняется (появление восьми положительных разностей из 10 возможных допустимо).

Этот результат можно трактовать так. Гипотеза о том, что результаты измерений (см. таблицу), полученные контролерами, одинаковы (сила давления на щуп одинакова), не противоречит данным эксперимента. Такие результаты могли быть получены одним и тем же контролером. Эксперимент не дает оснований утверждать, что результаты измерений, полученные контролером I, выше по сравнению с результатами, полученными контролером II.

Задачи

25.1 (анкетирование "Преподаватель глазами студентов" — эффект экзаменационной оценки). Вопрос о непредвзятости оценивания тех или иных качеств, величин, параметров весьма интересен. При этом предвзятость в оценивании встречается заметно чаще, чем непредвзятость.

В некоторых ситуациях причины предвзятости можно объяснить, в других они абсолютно загадочны и непонятны (см., например, задачу 24.4 из гл. 24, где приведен пример сильного предубеждения при считывании цифр с вращающегося с большой скоростью круга).

При обработке результатов анкетирования "Преподаватель глазами студентов", как и при любом другом оценивании, возникают вопросы: 1) о непредвзятости оценивания (которую едва ли следует ожидать) и 2) о причинах предвзятости (в этом оценивании вполне понятных).

Мы, исследуя возможные причины предвзятого оценивания, рассмотрим:

1° эффект экзаменационной оценки (задача 25.1);

2° эффект практического занятия — зависимость оценки преподавателя студентами от того, проводит преподаватель в группе (параллельно с чтением лекций) практические занятия или нет (задача 25.2).

Необходимо выяснить, имеют ли место перечисленные эффекты.

Можно исследовать и другие факторы как возможные причины предвзятого оценивания.

Эффект экзаменационной оценки. Еще до начала анкетирования высказывались соображения, что оценка характеристик преподавателя студентами не может быть непредвзятой. Как на причину предвзятого оценивания указывалось на то, что студенту на экзамене преподаватель выставляет оценку, и поэтому, оценивая преподавателя, студент сознательно или подсознательно учитывает результат экзамена — полученную оценку, особенно если она неудовлетворительная.

В связи с выдвинутой гипотезой о наличии эффекта экзаменационной оценки возникает вопрос: согласуется ли она с экспериментом (результатами анкетирования)?

В табл. 25.6.1 (столбцы 2 и 3) приведены результаты анкетирования в студенческих группах: столбец 2 — в группе ПМ-84-4, столбец 3 — в группе ПМ-84-1. При этом обе студенческие группы находились в одинаковых условиях относительно преподавателя: в обеих группах на протяжении года он читал лекции, вел практические и лабораторные занятия, принимал зачеты, экзамены. Однако результаты экзамена в этих группах заметно отличаются количеством полученных студентами двоек (см. табл. 25.6.2 и примечание 1).

Сформулировать задачу об эффекте экзаменационной оценки как задачу проверки статистических гипотез и решить ее:

выбрать основную гипотезу и альтернативные (что означает отклонение основной гипотезы? ее неотклонение?);

предложить уровень значимости и критерий для проверки основной гипотезы;

проверить основную гипотезу;

дать частотную интерпретацию полученным результатам.

Таблица 25.6.1. Оценки студентами характеристик преподавателя

Характеристика преподавателя	Оценка					
1	2	3	4	5	6	7
1. Преподает ясно и доступно	8,4	7,3	7,9	7,9	7,1	8,1
2. Разъясняет сложные места	8,3	7,3	8,5	8,0	7,5	8,3
3. Выделяет главные моменты	8,5	7,5	8,5	8,4	7,8	8,5
4. Умеет вызвать и поддержать интерес аудитории к предмету	8,6	6,4	7,9	8,7	5,8	7,2
5. Следит за реакцией аудитории	8,4	8,3	7,9	7,3	6,1	7,6
6. Задает вопросы, побуждает к дискуссии	8,2	5,5	7,2	6,0	4,7	6,7
7. Соблюдает логическую последовательность изложения	8,5	7,7	8,9	8,6	8,4	8,8
8. Демонстрирует культуру речи, четкость дикции, нормальный темп изложения	7,0	3,9	6,5	6,4	5,5	6,5
9. Умеет снять напряжение	8,1	5,6	7,1	5,8	5,1	6,5
10. Ориентирует на использование материала в будущей профессиональной деятельности	6,9	6,1	7,9	5,1	3,9	6,5
11. Творческий подход и интерес к своему делу	8,5	7,2	7,9	7,9	7,1	8,1
12. Доброжелательность и такт по отношению к студентам	8,0	6,6	7,9	8,1	7,7	7,9
13. Терпение	8,2	5,8	8,5	8,0	7,1	8,3
14. Требовательность	8,8	8,3	8,9	8,7	8,8	8,8
15. Заинтересованность в успехах студентов	8,4	6,4	8,0	7,7	7,2	7,8
16. Объективность в оценивании знаний студентов	8,7	6,2	7,4	7,7	7,3	7,5
17. Уважительное отношение к студентам	8,6	6,5	8,5	8,3	7,9	8,4
18. Располагает к себе высокой эрудицией, манерой поведения	8,8	6,5	7,9	8,4	8,1	8,2

Примечание. 1. В табл. 25.6.1 приведены результаты оценивания автора как преподавателя студентами групп ПМ-84-4, ПМ-84-1, ПМ-85-1, ПМ-85-2, ПМ-85-3, ПМ-85-4, а именно, оценки в баллах характеристик преподавателя — среднее в группе или в нескольких группах (максимально возможная оценка характеристики — 9 баллов):

столбец 2 — в группе ПМ-84-4; на экзамене шесть студентов получили неудовлетворительные оценки (четверо из них отказались отвечать после ознакомления с содержанием экзаменационного билета, т. е. фактически сами оценили свои знания как неудовлетворительные (см. табл. 25.6.2));

столбец 3 — в группе ПМ-84-1; на экзамене девять студентов получили неудовлетворительные оценки (см. табл. 25.6.2);

столбец 4 — в группах ПМ-85-1, ПМ-85-2, где лектор вел также практические занятия;

столбец 5 — в группах ПМ-85-3, ПМ-85-4, где лектор не вел практических занятий (их проводил другой преподаватель);

столбец 6 — в группах ПМ-85-1, ПМ-85-2, ПМ-85-3, ПМ-85-4 (анкетирование проведено до экзамена);

столбец 7 — в группах ПМ-85-1, ПМ-85-2, ПМ-85-3, ПМ-85-4 (анкетирование проведено после экзамена).

Анкета приведена в оригинальном виде.

Таблица 25.6.2. Результаты экзамена

Оценка на экзамене	Количество оценок в группах			
	ПМ-84-4	ПМ-84-1	ПМ-85-1 ПМ-85-2	ПМ-85-3 ПМ-85-5
Отлично	1	4	3	4
Хорошо	7	2	10	6
Удовлетворительно	4	6	7	6
Неудовлетворительно	6	9	11	15

Примечание. 2. В табл. 25.6.2 приведены результаты первой сдачи экзамена: речь идет об экзамене по курсу “Теория вероятностей и математическая статистика” на факультете прикладной математики Днепропетровского государственного университета во время зимних сессий 1986/87 и 1987/88 учебных годов.

25.2 (влияние работы метронома на плавность речи). В таблице приведены результаты эксперимента, в котором изучалось влияние работы метронома на речь людей, страдающих заиканием.

Номер участника	Число заиканий при условии			Номер участника	Число заиканий при условии		
	N	R	A		N	R	A
1	15	3	5	7	10	0	2
2	11	3	3	8	8	0	3
3	18	1	3	9	13	0	2
4	21	5	4	10	4	1	0
5	6	2	2	11	11	2	4
6	17	0	2	12	17	2	1

Обследовалось 12 человек с тяжелой формой заболевания. Каждый из них импровизировал трехминутную речь при условиях N, R, A:

N — говорить без метронома;

R — говорить при регулярной (ритмичной) работе метронома (120 ударов за минуту), причем человек был предварительно проинструктирован о необходимости произносить один слог слова на каждый удар метронома;

A — говорить при неритмичной работе метронома, работающего со случайными интервалами между ударами (от 0,3 до 0,7 с), совершая при этом в среднем те же 120 ударов в минуту (при условии A, как и при условии R, человек должен произносить один слог на каждый удар метронома).

Приведенные данные, безусловно, свидетельствуют о том, что работа метронома уменьшает количество заиканий (почему?). А существуют ли отличия во влиянии на заикание ритмично и неритмично работающих метрономов?

25.3 (анкетирование “Преподаватель глазами студентов” — эффект практического занятия). При оценивании студентами преподавателя выдвигается предположение о наличии эффекта практического занятия: если в группе параллельно с чтением лекций преподаватель ведет практические (лабораторные) занятия, то оценка его студентами выше.

В табл. 25.6.1 приведены оценки характеристик преподавателя студентами: столбец 4 — в группах ПМ-85-1 и ПМ-85-2, где преподаватель читал лекции и вел практические и лабораторные занятия, а столбец 5 — в группах ПМ-85-3 и ПМ-85-4, где практические и лабораторные занятия вел другой преподаватель (анкетирование проведено после экзамена).

Проверить, существует ли эффект практического (лабораторного) занятия.

Сформулировать задачу о наличии эффекта практического (лабораторного) занятия при оценивании преподавателя как задачу проверки статистических гипотез и решить ее (подробнее см. задачу 25.1).

25.4 (момент последнего уравнивания). Эксперимент состоит в последовательном подбрасывании монеты $2n = 40$ раз и регистрации момента (номера испытания), когда в последний раз количество выпавших “гербов” и “решеток” было одинаково. Будем говорить: наблюдается момент последнего уравнивания количества “гербов” и “решеток”. В эксперименте этими моментами являются $0, 2, 4, \ldots, 40$.

Относительно распределения момента последнего уравнивания выдвигается гипотеза: момент $2k$ последнего уравнивания имеет распределение арксинуса; подробнее — случайная величина $2k/2n = k/n$ имеет распределение арксинуса, т. е. ее функцией распределения является

$$A(x) = \begin{cases} 0, & \text{если } x \le 0; \\ \frac{2}{\pi} \arcsin \sqrt{x}, & \text{если } 0 < x \le 1; \\ 1, & \text{если } x > 1. \end{cases}$$

Провести эксперимент пять раз, фиксируя момент последнего уравнивания количества "гербов" и "решеток".

Можно ли на основании полученных данных считать, что момент последнего уравнивания имеет распределение арксинуса?

Ответ дать в терминах проверки статистических гипотез. Воспользоваться критерием Колмогорова.

З а м е ч а н и е 1. Эксперимент удобно проводить так. Подбрасываем монету и выпадение "герба" обозначаем через $+1$, а "решетки" через -1. После 40 подбрасываний получаем последовательность, образованную из чисел $+1$ и -1. Последовательно складывая члены последовательности, начиная с первого, зафиксируем момент, когда сумма будет равна нулю — это будет момент второго уравнивания (момент первого уравнивания равен нулю); и так продолжаем, пока не найдем значение момента последнего уравнивания (см. также задачу 25.5).

З а м е ч а н и е 2. Ниже приведены значения функции распределения арксинуса $A(x)$ в точках $0{,}05 \cdot i$, $i = 0, 1, 2, \ldots, 10$:

x	$A(x)$	x	$A(x)$
0,00	0,000	0,30	0,369
0,05	0,144	0,35	0,403
0,10	0,205	0,40	0,436
0,15	0,253	0,45	0,468
0,20	0,295	0,50	0,500
0,25	0,333		

Для вычисления значений $A(x)$, если $0{,}5 < x < 1{,}0$, можно воспользоваться соотношением

$$A(x) = 1 - A(1 - x).$$

25.5. Провести серию из 10 экспериментов, каждый из которых состоит в подбрасывании 16 монет и регистрации количества s_i выпавших "гербов" (i — номер эксперимента, $i = 1, 2, \ldots, 10$). Рассмотреть последовательность чисел

$$\xi_i = \frac{s_i - 8}{2}, \quad i = 1, 2, \ldots, 10.$$

Можно ли считать, что $\xi_1, \xi_2, \ldots, \xi_{10}$ является выборкой из стандартного нормального распределения?

Из каких соображений выдвинута сформулированная гипотеза?

Примечание. Эксперимент, состоящий в подбрасывании 16 монет, удобно провести, например, так: поместить в коробку все 16 монет, а потом, хорошо встряхнув ее, подсчитать количество выпавших "гербов". Эксперимент повторить необходимое количество раз.

25.6 (время пребывания на "положительной" стороне). Эксперимент состоит в последовательном подбрасывании монеты $2n = 40$ раз (одно подбрасывание на единицу времени). Появление "герба" будем обозначать через $+1$, а появление "решетки" — через -1. Таким образом, на k-м шаге, $k = 1, 2, \ldots$ $\ldots, 2n$, имеем символ ε_k, равный $+1$ или -1 в зависимости от того, какой стороной легла монета. Пусть

$$s_k = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_k,\ s_0 = 0,\ k = 1, 2, \ldots, 2n,$$

где s_k — разность между количеством "плюсов" и "минусов" (между количеством "гербов" и "решеток"). Если воспользоваться геометрической терминологией и системой координат (t, x), то результат эксперимента можно представить в виде ломаной с вершинами в точках (k, s_k), $k = 1, 2, \ldots, 2n$ (рис. 25.6.1).

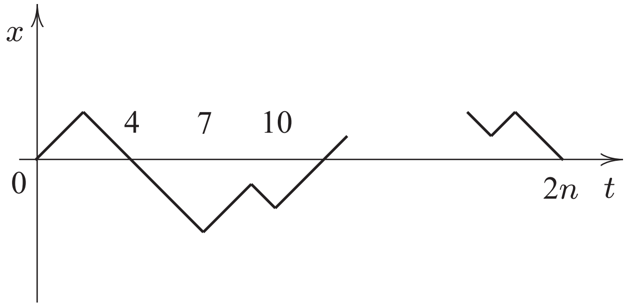

Рис. 25.6.1: Результат эксперимента — ломаная с вершинами в точках (k, s_k), $k = 0, 1, \ldots, 2n$

Вычислим время, когда монета находилась на "положительной" стороне, т. е. когда разность между количеством "гербов" и "решеток" была положительной. Например, если в результате первых 10 подбрасываний монеты получена последовательность $+1, +1, -1, -1, -1, -1, -1, +1, +1, -1$ (см. также рис. 25.6.1), то время пребывания монеты на “положительной” стороне составляет 4 единицы.

Относительно времени пребывания на “положительной” стороне выдвигается гипотеза: время пребывания на “положительной” стороне имеет распределение арксинуса. Подробнее, распределение арксинуса (см. задачу 25.4) имеет случайная величина $2k/2n = k/n$, где через $2k$ обозначено количество единиц времени (из $2n$ единиц), когда монета пребывала на “положительной” стороне.

Провести эксперимент пять раз, фиксируя время пребывания монеты на “положительной” стороне.

Можно ли на основании полученных данных заключить, что время пребывания монеты на “положительной” стороне имеет распределение арксинуса?

Ответ дать в терминах проверки статистических гипотез, воспользовавшись критерием Колмогорова и замечанием 1 к задаче 25.4.

25.7. Шейки рабочей части сверл обрабатываются на шлифовальном станке. Номинальный диаметр шейки составляет 9,8 мм с техническим допуском 0,04 мм.

Были измерены диаметры рабочей части шейки 16 сверл. Получены такие их значения (в миллиметрах): 9,76; 9,78; 9,81; 9,77; 9,75; 9,78; 9,75; 9,77; 9,74; 9,78; 9,77; 9,83; 9,78; 9,81; 9,79; 9,80.

Можно ли заключить, что номинальный диаметр шейки равен 9,8 мм с техническим допуском 0,04 мм при норме отхода 5 %, т. е. что размер диаметра шейки сверла находится в пределах от 9,76 до 9,84 мм с вероятностью 0,95?

У к а з а н и е 1. Сформулировать поставленную задачу как задачу проверки статистических гипотез и воспользоваться критерием Колмогорова.

У к а з а н и е 2. Задачу решить в предположении, что результаты измерений являются выборкой из нормального распределения.

Глава 26

Линейная регрессия

26.1 Нормальная линейная регрессия

Задача восстановления зависимости. В математической статистике часто встречается следующая задача.

Известно, что величины y и x связаны функциональной зависимостью, сама же зависимость $y = f(x)$ неизвестна. Мы имеем возможность в данных, вполне определенных точках $x_1, x_2, \dots, x_n$ наблюдать соответствующие значения $y_1, y_2, \dots, y_n$ величины y, но не точно, а с некоторыми погрешностями e_j, $j = 1, 2, \dots, n$ (наблюдений, измерений без погрешностей не бывает), т. е. наблюдая y_j, мы фактически в качестве результата наблюдения получаем

$$\xi_j = y_j + e_j,\ j = 1, 2, \dots, n.$$

Погрешности $e_j, j = 1, 2 \dots n$, неизвестны, однако их естественно считать независимыми (наблюдения производятся независимо) нормально распределенными случайными величинами со средними $\mathsf{M}e_j$ равными нулю (при этом $\mathsf{M}\xi_j = y_j$, $j = 1, 2, \dots$ $\dots, n$) и одной и той же неизвестной дисперсией σ^2 (последняя как раз и определяет погрешность наблюдений).

По результатам наблюдений $\xi_1, \xi_2, \dots, \xi_n$ величины y в точках $x_1, x_2, \dots, x_n$ необходимо восстановить (найти) зависимость величины y от x.

Далее мы будем рассматривать поставленную задачу восстановления зависимости $y = f(x)$ в часто встречающейся ситуации, когда из тех или иных соображений известно (можно считать), что зависимость $y = f(x)$ величины y от x линейная, т. е.

$$y = a + bx.$$

В этом случае восстановление зависимости сводится к определению (оцениванию) параметров a, b, σ^2 по наблюдениям $\xi_1, \xi_2, \ldots$ $\ldots, \xi_n$ величин $y_1, y_2, \ldots, y_n$ в точках $x_1, x_2, \ldots, x_n$. (Разумеется, коль скоро значения y_j, $j = 1, 2 \ldots n$, наблюдаются с погрешностью, то и параметры a, b, σ^2 неизбежно будут определяться с погрешностью.)

В строгой математической постановке задача восстановления (определения) зависимости $y = a + bx$ формулируется следующим образом.

Дана последовательность независимых нормально распределенных случайных величин ξ_j, $j = 1, 2, \ldots, n$, соответственно со средними $y_j = a + bx_j$ и одной и той же дисперсией σ^2. Параметры a, b, σ^2 неизвестны, и их необходимо оценить по $\xi_1, \xi_2, \ldots, \xi_n$.

Функцию

$$y = a + bx,$$

задающую зависимость y от x, называют простой линейной регрессией, кратко — линейной регрессией (“простая” обозначает зависимость от одной переменной). Когда переменная y является функцией от x, то говорим “регрессия y на x”, когда x является функцией от y, то говорим “регрессия x на y”.

Простая линейная регрессия. Далее простую линейную регрессию $y = a + bx$ мы будем записывать в виде

$$y = \alpha + \beta(x - \bar{x}),$$

где $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, $x_1, x_2, \ldots, x_n$ — значения переменной x; α, β, σ^2 — неизвестные параметры.

Оценки неизвестных параметров α, β, σ^2 мы получим методом максимального правдоподобия. При этом будем пользоваться обозначениями:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2,\ S_1 = \sqrt{S_1^2},$$

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i,\ S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2,\ S_2 = \sqrt{S_2^2},$$

$$R_{12} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(\xi_i - \overline{\xi}).$$

Теорема 26.1.1 (об оценках параметров $\boldsymbol{\alpha, \beta, \sigma^2}$). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые нормально распределенные случайные величины со средними*

$$y_i = \alpha + \beta(x_i - \overline{x}),\ i = 1, 2, \ldots, n,$$

и дисперсией σ^2.

Оценками максимального правдоподобия параметров α, β, σ^2 являются

$$\hat{\alpha} = \overline{\xi}, \quad \hat{\beta} = \frac{R_{12}}{S_1^2}, \quad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - (\hat{\alpha} + \hat{\beta}(x_i - \overline{x})))^2.$$

Доказательство. Поскольку $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины нормально распределенные с параметрами $(\alpha + \beta(x_i - \overline{x}); \sigma^2)$, $i = 1, 2, \ldots, n$, то их совместная плотность

$$f(t_1, t_2, ..., t_n; \alpha, \beta, \sigma^2) =$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(t_i - (\alpha + \beta(x_i - \overline{x})))^2\right\} =$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(t_i - (\alpha + \beta(x_i - \overline{x})))^2\right\}.$$

Функция максимального правдоподобия выборки $\xi_1, \xi_2, \ldots, \xi_n$ имеет вид

$$L(\alpha, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\xi_i - (\alpha + \beta(x_i - \overline{x})))^2\right\} =$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} Q(\alpha, \beta)\right\},$$

где

$$Q(\alpha, \beta) = \sum_{i=1}^{n}(\xi_i - (\alpha + \beta(x_i - \overline{x})))^2. \tag{26.1.1}$$

Точка $(\hat{\alpha}, \hat{\beta}, \widehat{\sigma^2})$, в которой функция максимального правдоподобия достигает наибольшего значения, является оценкой максимального правдоподобия параметров α, β, σ^2. Чтобы найти эту точку, отметим, что если в точке $(\hat{\alpha}, \hat{\beta})$ функция $Q(\alpha, \beta)$ (см. (26.1.1)) достигает наименьшего значения:

$$Q(\hat{\alpha}, \hat{\beta}) = Q^* = \min_{\alpha, \beta} Q(\alpha, \beta),$$

а в точке $\hat{\theta} = \widehat{\sigma^2}$ функция

$$\varphi(\theta) = \frac{1}{(2\pi\theta)^{n/2}} \exp\left\{-\frac{1}{2\theta} Q^*\right\} \tag{26.1.2}$$

достигает наибольшего значения, то функция $L(\alpha, \beta, \sigma^2)$ в точке $(\hat{\alpha}, \hat{\beta}, \widehat{\sigma^2})$ достигает наибольшего значения. В самом деле, если $(\alpha, \beta, \sigma^2)$ — произвольная точка, то

$$L(\alpha, \beta, \sigma^2) =$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} Q(\alpha, \beta)\right\} \le \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} Q^*\right\} \le$$

$$\le \frac{1}{(2\pi\widehat{\sigma^2})^{n/2}} \exp\left\{-\frac{1}{2\widehat{\sigma^2}} Q^*\right\} = \frac{1}{(2\pi\widehat{\sigma^2})^{n/2}} \exp\left\{-\frac{1}{2\widehat{\sigma^2}} Q(\hat{\alpha}, \hat{\beta})\right\}.$$

Так что для того, чтобы найти точку, в которой $L(\alpha, \beta, \sigma^2)$ достигает наибольшего значения, достаточно найти точку $(\hat{\alpha}, \hat{\beta})$, в которой функция $Q(\alpha, \beta)$ (см. (26.1.1)) достигает наименьшего значения, а потом точку $\hat{\theta} = \widehat{\sigma^2}$, в которой функция $\varphi(\theta)$ (см. (26.1.2)) достигает наибольшего значения.

Сначала найдем точку $(\hat{\alpha}, \hat{\beta})$, в которой функция $Q(\alpha, \beta)$ достигает наименьшего значения, для этого представим $Q(\alpha, \beta)$ в виде

$$Q(\alpha, \beta) = \sum_{i=1}^{n} (\xi_i - (\alpha + \beta(x_i - \overline{x})))^2 =$$

$$= \sum_{i=1}^{n} ((\xi_i - \overline{\xi}) - (\alpha - \overline{\xi}) - \beta(x_i - \overline{x}))^2 =$$

$$= \sum_{i=1}^{n} (\xi_i - \overline{\xi})^2 + n(\alpha - \overline{\xi})^2 + \beta^2 \sum_{i=1}^{n} (x_i - \overline{x})^2 -$$

$$-2(\alpha-\overline{\xi})\sum_{i=1}^{n}(\xi_i-\overline{\xi})-2\beta\sum_{i=1}^{n}(\xi_i-\overline{\xi})(x_i-\overline{x})+2\beta(\alpha-\overline{\xi})\sum_{i=1}^{n}(x_i-\overline{x}).$$

Заметив, что

$$\sum_{i=1}^{n}(x_i-\overline{x})=0,\ \sum_{i=1}^{n}(\xi_i-\overline{\xi})=0, \tag{26.1.3}$$

получим для $Q(\alpha,\beta)$ представление

$$Q(\alpha,\beta)=n\left((\alpha-\overline{\xi})^2+\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^2+\beta^2\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2-\right.$$

$$\left.-2\beta\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\overline{\xi})(x_i-\overline{x})\right)=n((\overline{\xi}-\alpha)^2+S_2^2+\beta^2S_1^2-2\beta R_{12})=$$

$$=n((\overline{\xi}-\alpha)^2+S_2^2+S_1^2(\beta^2-2\beta(R_{12}/S_1^2))).$$

Дополнив квадратный относительно β трехчлен до полного квадрата, получим

$$Q(\alpha,\beta)=n\left((\overline{\xi}-\alpha)^2+S_1^2\left(\beta-\frac{R_{12}}{S_1^2}\right)^2+S_2^2-\frac{R_{12}^2}{S_1^2}\right).$$

Из этого представления функции $Q(\alpha,\beta)$ двух переменных следует, что $Q(\alpha,\beta)$ достигает наименьшего значения в точке

$$\alpha=\hat{\alpha}=\overline{\xi},\ \beta=\hat{\beta}=\frac{R_{12}}{S_1^2}.$$

Заметим, что наименьшее значение функции $Q(\alpha,\beta)$ равно

$$Q^*=Q(\hat{\alpha},\hat{\beta})=\sum_{i=1}^{n}(\xi_i-(\hat{\alpha}+\hat{\beta}(x_i-\overline{x})))^2=n\left(S_2^2-\frac{R_{12}^2}{S_1^2}\right). \tag{26.1.4}$$

Точка, в которой функция

$$\varphi(\theta)=\frac{1}{(2\pi\theta)^{n/2}}\exp\left\{-\frac{1}{2\theta}Q^*\right\}$$

достигает наибольшего значения, совпадает с точкой, в которой наибольшего значения достигает функция

$$\ln\varphi(\theta) = -\frac{n}{2}\ln(2\pi\theta) - \frac{Q^*}{2\theta}$$

(в силу монотонности $y = \ln x$). Легко убедиться, что этой точкой является

$$\hat{\theta} = \frac{1}{n}Q^*.$$

Так что

$$\hat{\theta} = \widehat{\sigma^2} = \frac{1}{n}Q^* = \frac{1}{n}Q(\hat{\alpha}, \hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - (\hat{\alpha} + \hat{\beta}(x_i - \overline{x})))^2$$

— оценка максимального правдоподобия параметра $\theta = \sigma^2$. Ее часто удобно записывать в виде

$$\widehat{\sigma^2} = \frac{1}{n}Q(\hat{\alpha}, \hat{\beta}) = S_2^2 - \frac{R_{12}^2}{S_1^2} = S_2^2\left(1 - \frac{R_{12}^2}{S_1^2 S_2^2}\right).$$

Теорема доказана.

Замечание 1. Далее мы будем использовать обозначение $\hat{\sigma} = \sqrt{\widehat{\sigma^2}}$.

Замечание 2. Знак $\sim$ обозначает “распределено как”, “имеет распределение”.

Теорема 26.1.2 (о распределении оценок $\hat{\alpha}, \hat{\beta}, \widehat{\sigma^2}$). *Оценки максимального правдоподобия*

$$\hat{\alpha} = \overline{\xi};\;\; \hat{\beta} = \frac{R_{12}}{S_1^2};\;\; \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - (\hat{\alpha} + \hat{\beta}(x_i - \overline{x})))^2$$

параметров α, β, σ^2 нормальной регрессии

$$y = \alpha + \beta(x - \overline{x})$$

обладают следующими свойствами:

1) *$\hat{\alpha}, \hat{\beta}, \widehat{\sigma^2}$ являются независимыми в совокупности случайными величинами;*

2) $\hat{\alpha}, \hat{\beta}$ *— несмещенные и состоятельные оценки соответственно* α *и* β*;* $\hat{\sigma^2}$ *— асимптотически несмещенная и состоятельная оценка* σ^2*;*

3)

$$(\hat{\alpha} - \alpha) \Big/ \frac{\hat{\sigma}}{\sqrt{n-2}} \sim t_{n-2}, \tag{26.1.5}$$

$$(\hat{\beta} - \beta) \Big/ \frac{\hat{\sigma}}{S_1\sqrt{n-2}} \sim t_{n-2}, \tag{26.1.6}$$

$$\frac{n\hat{\sigma^2}}{\sigma^2} \sim \chi^2_{n-2}, \tag{26.1.7}$$

$$\frac{\Big((\hat{\alpha} - \alpha)^2 + S_1^2(\hat{\beta} - \beta)^2\Big)\Big/ 2}{\hat{\sigma^2}/(n-2)} \sim F_{2;n-2}. \tag{26.1.8}$$

Проверка гипотез о параметрах простой нормальной регрессии. Из соотношений (26.1.5), (26.1.6), (26.1.7), (26.1.8) можно получить критерии для проверки гипотез о коэффициентах α, β и дисперсии σ^2 и можно построить доверительные интервалы для них.

Далее, как обычно, обозначаем через $t_{\gamma;k}$, $\chi^2_{\gamma;k}$ верхние γ-пределы соответственно распределений t_k, χ^2_k, а через $F_{\gamma;n,m}$ — верхний γ-предел распределения $\mathsf{F}_{n,m}$.

Критерий для проверки гипотезы $\boldsymbol{H_0 : \alpha = \alpha_0}$. *Если гипотезу* $H_0 : \alpha = \alpha_0$ *отклонять при*

$$\left|(\hat{\alpha} - \alpha_0) \Big/ \frac{\hat{\sigma}}{\sqrt{n-2}}\right| > t_{\gamma;n-2} \tag{26.1.9}$$

и не отклонять в противном случае, то с вероятностью 2γ *гипотеза* H_0 *будет отклоняться, когда она верна (альтернатива двусторонняя).*

Критерий для проверки гипотезы $\boldsymbol{H_0 : \beta = \beta_0}$. *Если гипотезу* $H_0 : \beta = \beta_0$ *отклонять при*

$$\left|(\hat{\beta} - \beta_0) \Big/ \frac{\hat{\sigma}}{S_1\sqrt{n-2}}\right| > t_{\gamma;n-2} \tag{26.1.10}$$

и не отклонять в противном случае, то с вероятностью 2γ *гипотеза* H_0 *будет отклоняться, когда она верна (альтернатива двусторонняя).*

В частности, критерий для проверки гипотезы $H_0 : \beta = 0$ (гипотеза о значимости регрессии) формулируется так: *Если гипотезу $H_0 : \beta = 0$ отклонять при*

$$\left| \hat{\beta} \Big/ \frac{\hat{\sigma}}{S_1\sqrt{n-2}} \right| > t_{\gamma;n-2} \tag{26.1.11}$$

и не отклонять в противном случае, то с вероятностью 2γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя).

Критерий для проверки гипотезы $H_0 : \sigma^2 = \sigma_0^2$. *Если гипотезу $H_0 : \sigma^2 = \sigma_0^2$ отклонять при*

$$\frac{n\hat{\sigma^2}}{\sigma_0^2} > \chi^2_{\gamma;n-2} \tag{26.1.12}$$

и не отклонять в противном случае, то с вероятностью γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива односторонняя: $\sigma^2 > \sigma_0^2$).

Критерий для проверки гипотезы $H_0 : \alpha = \alpha_0, \beta = \beta_0$. *Если гипотезу $H_0 : \alpha = \alpha_0, \beta = \beta_0$ отклонять при*

$$\frac{((\hat{\alpha} - \alpha_0)^2 + S_1^2(\hat{\beta} - \beta_0)^2)/2}{\hat{\sigma^2}/(n-2)} > F_{\gamma;2;n-2} \tag{26.1.13}$$

и не отклонять в противном случае, то с вероятностью γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива: $\alpha \neq \alpha_0$ или $\beta \neq \beta_0$).

Доверительные интервалы для параметров простой линейной регрессии. Доверительные интервалы для параметров α, β, σ^2 простой линейной регрессии очевидным образом получаются из соотношений (26.1.5), (26.1.6) (26.1.7):

$$\hat{\alpha} - t_{\gamma;n-2}\frac{\hat{\sigma}}{\sqrt{n-2}} \leq \alpha \leq \hat{\alpha} + t_{\gamma;n-2}\frac{\hat{\sigma}}{\sqrt{n-2}}, \tag{26.1.14}$$

$$\hat{\beta} - t_{\gamma;n-2}\frac{\hat{\sigma}}{S_1\sqrt{n-2}} \leq \beta \leq \hat{\beta} + t_{\gamma;n-2}\frac{\hat{\sigma}}{S_1\sqrt{n-2}}, \tag{26.1.15}$$

$$\frac{n\hat{\sigma^2}}{\chi^2_{\gamma;n-2}} < \sigma^2 < \frac{n\hat{\sigma^2}}{\chi^2_{1-\gamma;n-2}}. \tag{26.1.16}$$

Коэффициент доверия каждого из этих доверительных интервалов равен $1-2\gamma$.

Совместная доверительная область для $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$. Из соотношения (26.1.8) получаем

$$\mathsf{P}\left\{\frac{((\alpha-\hat{\alpha})^2+S_1^2(\beta-\hat{\beta})^2)/2}{\hat{\sigma^2}/(n-2)}\le F_{\gamma;2;n-2}\right\}=1-\gamma,$$

или

$$\mathsf{P}\left\{\frac{(\alpha-\hat{\alpha})^2}{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(n-2)}+\frac{(\beta-\hat{\beta})^2}{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(S_1^2(n-2))}\le 1\right\}=1-\gamma, \tag{26.1.17}$$

где $F_{\gamma;2;n-2}$ — верхний γ-предел F-распределения с $(2;n-2)$ степенями свободы.

Равенство (26.1.17) обозначает, что множество точек (x,y), задаваемое неравенством

$$\frac{(x-\hat{\alpha})^2}{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(n-2)}+\frac{(y-\hat{\beta})^2}{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(S_1^2(n-2))}\le 1, \tag{26.1.18}$$

является доверительной областью для параметров (α,β) с коэффициентом доверия $1-\gamma$ (доверительная область (26.1.18) “накрывает” неизвестные параметры (α,β) линейной регрессии $y=\alpha+\beta(x-\overline{x})$ с вероятностью $1-\gamma$). Границей этой области является эллипс

$$\frac{(x-\hat{\alpha})^2}{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(n-2)}+\frac{(y-\hat{\beta})^2}{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(S_1^2(n-2))}=1 \tag{26.1.19}$$

с центром в точке $(\hat{\alpha},\hat{\beta})$ и полуосями

$$\sqrt{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(n-2)},\ \sqrt{2F_{\gamma;2;n-2}\,\hat{\sigma^2}/(S_1^2(n-2))}.$$

Проверка адекватности регрессии. Известно, что величина y связана функциональной зависимостью с величиной x, но сама эта зависимость неизвестна. В точках $x_1,x_2,\ldots,x_n$ с некоторыми погрешностями e_j, $j=1,2,\ldots,n$, мы наблюдаем соответствующие значения $y_1,y_2,\ldots,y_n$, т. е. наблюдая y_j, фактически получаем

$$\xi_j=y_j+e_j,\ j=1,2,\ldots,n. \tag{26.1.20}$$

Погрешности наблюдений e_j, $j = 1, 2, \ldots, n$, — независимые нормально распределенные случайные величины со средними 0 и одной и той же неизвестной нам дисперсией σ^2 (которая фактически и определяет погрешность наблюдений). При этом

$$\mathsf{M}\xi_j = y_j,\ j = 1, 2, \ldots, n.$$

Относительно функциональной зависимости y от x (из тех или иных соображений) выдвигается гипотеза H_0: зависимость y от x имеет вид

$$y = f(x),$$

где $f(x)$ — вполне определенная функция (или $f(x)$ принадлежит данному классу функций). Гипотеза

$$H_0 : y = f(x)$$

— это гипотеза об адекватности описания зависимости y от x функцией $y = f(x)$ (регрессией $y = f(x)$). По результатам наблюдений $\xi_1, \xi_2, \ldots, \xi_n$ значений $y_1, y_2, \ldots, y_n$ соответственно в точках $x_1, x_2, \ldots, x_n$ необходимо сделать выводы о гипотезе H_0 — отклонять H_0 или не отклонять. Если гипотеза $H_0 : y = f(x)$ не отклоняется, то мы говорим, что регрессия $y = f(x)$ адекватно (хорошо) описывает зависимость y от x, в противном случае — нет.

При наличии повторных наблюдений можно построить критерий для проверки гипотезы $H_0 : y = f(x)$, по меньшей мере для проверки гипотезы $H_0 : y = \alpha + \beta(x - \overline{x})$.

Под *повторными наблюдениями* значения y_j (в точке x_j) будем понимать независимые нормально распределенные случайные величины $\xi_{j,1}, \xi_{j,2}, \ldots, \xi_{j,n_j}$ со средним y_j:

$$\mathsf{M}\xi_{j,\nu} = y_j,\ \nu = 1, 2, \ldots, n_j, \tag{26.1.21}$$

и дисперсией σ^2, одной и той же для всех наблюдений $\xi_{j,\nu}$:

$$\mathsf{D}\xi_{j,\nu} = \sigma^2,\ \nu = 1, 2, \ldots, n_j,$$

n_j — количество повторных наблюдений в точке x_j. Всего имеется k точек x_j, $j = 1, 2, \ldots, k$, в каждой из которых проводилось соответственно по n_j наблюдений. Количество всех наблюдений, как и ранее, будем обозначать через n, ясно что

$$\sum_{j=1}^{k} n_j = n.$$

Наличие повторных наблюдений $\xi_{j,\nu}$, $\nu = 1, 2, \ldots, n_j$, значений y_j в точках x_j, $j = 1, 2, \ldots, k$, дает возможность оценить погрешность наблюдений σ^2 — в качестве оценки s_1^2 для σ^2 естественно рассмотреть

$$s_1^2 = \frac{1}{n-k}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(\xi_{j,\nu} - \overline{\xi_j})^2,$$

где

$$\overline{\xi_j} = \frac{1}{n_j}\sum_{\nu=1}^{n_j}\xi_{j,\nu},\ j = 1, 2, \ldots, k,\ n = \sum_{j=1}^{k} n_j$$

(s_1^2 — несмещенная оценка σ^2).

В терминах наблюдений $\xi_{j,\nu}$, $\nu = 1, 2, \ldots, n_j, j = 1, 2, \ldots, k$, гипотеза H_0 — зависимость y от x имеет вид $y = f(x)$, формулируется так: у независимых нормально распределенных случайных величин $\xi_{j,\nu}$ со средними $\mathsf{M}\xi_{j,\nu} = y_j$, $\nu = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, k$, и одной и той же дисперсией σ^2

$$y_j = f(x_j),\ j = 1, 2, \ldots, k.$$

Идея построения критерия для проверки гипотезы H_0: зависимость y от x описывается функцией $y = f(x)$, состоит в следующем. Тем или иным способом (чаще всего методом максимального правдоподобия или методом наименьших квадратов) по результатам наблюдений $\xi_{j,\nu}$, $\nu = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, k$, оцениваем гипотетическую зависимость $y = f(x)$ — получаем эмпирическую регрессию $y = \hat{f}(x)$, а вместе с ней и оценки $\hat{f}(x_j)$ соответственно величин $y_j, j = 1, 2, \ldots, k$, вычисленные по эмпирической регрессии $y = \hat{f}(x)$ (если проверяется гипотеза H_0: $y = \alpha + \beta(x - \overline{x})$, эмпирическая регрессия имеет вид $y = \hat{\alpha} + \hat{\beta}(x - \overline{x})$). С другой стороны, вне зависимости от выдвинутой гипотезы H_0: $y = f(x)$, по результатам наблюдений $\xi_{j,\nu}$, $\nu = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, k$, всегда можно оценить значения y_j в точках x_j, $j = 1, 2, \ldots, k$, — в качестве оценок для y_j естественно рассмотреть их несмещенные оценки

$$\overline{\xi_j} = \frac{1}{n_j}\sum_{\nu=1}^{n_j}\xi_{j,\nu},\ j = 1, 2, \ldots, k$$

(так как $\mathsf{M}\xi_{j,\nu} = y_j$, то $\overline{\xi_j}$ — несмещенная оценка y_j, $j = 1, 2, \ldots, k$).

И если функция $y = f(x)$ адекватно (хорошо) описывает зависимость y от x (гипотеза H_0 верна), то уклонения оценок $\hat{f}(x_j)$ для y_j, вычисленных по эмпирической регрессии $y = \hat{f}(x)$ в точках x_j, от несмещенных оценок $\overline{\xi_j}$ тех же значений y_j, $j = 1, 2, \ldots, k$, должны быть малыми. Если же функция $y = f(x)$ неадекватно описывает зависимость y от x (гипотеза H_0 неверна), уклонения между оценками $\hat{f}(x_j)$ и $\overline{\xi_j}$ для y_j будут большими. Количественно уклонение между оценками $\overline{\xi_j}$ и $\hat{f}(x_j)$ для y_j, $j = 1, 2, \ldots, k$, естественно описывать величиной

$$s_2^2 = c\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(\overline{\xi}_j - \hat{f}(x_j))^2 \tag{26.1.22}$$

(c — некоторый нормирующий множитель).

Поэтому для проверки гипотезы H_0: $y = f(x)$ вычисляем уклонение s_2^2. Если s_2^2 приняло малое значение — лежит в пределах погрешности наблюдений σ^2, т. е. отношение s_2^2/σ^2 малое, мы гипотезу H_0 не отклоняем. Если значение уклонения s_2^2 превышает погрешность наблюдений σ^2, т. е. отношение s_2^2/σ^2 большое, мы гипотезу H_0 отклоняем. А поскольку погрешность наблюдений σ^2 неизвестна, мы будем сравнивать уклонение s_2^2 не с σ^2, а с ее несмещенной оценкой

$$s_1^2 = \frac{1}{n-k}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(\xi_{j,\nu} - \overline{\xi_j})^2$$

— если отношение s_2^2/s_1^2 приняло большое значение — гипотезу H_0 отклоняем, в противном случае — нет.

Чтобы можно было судить, большое или малое значение приняло отношение s_2^2/s_1^2 (и в зависимости от этого отклонять или не отклонять H_0), необходимо знать, какие значения принимает отношение s_2^2/s_1^2, когда гипотеза H_0 верна (когда отношение s_2^2/s_1^2 является малым) и какие значения принимает отношение s_2^2/s_1^2, когда гипотеза H_0 неверна (когда отношение s_2^2/s_1^2 является большим). Другими словами, необходимо знать распределение отношения s_2^2/s_1^2, когда гипотеза H_0 верна и когда она неверна, или хотя бы когда гипотеза H_0 верна.

В случае проверки гипотезы H_0: зависимость y от x описывается функцией

$$y = \alpha + \beta(x - \overline{x}),$$

уклонение s_2^2 (см. (26.1.22)) между несмещенными оценками $\overline{\xi_j}$ для y_j и оценками $\hat{f}(x_j) = \hat{\alpha} + \hat{\beta}(x_j - \overline{x})$ для y_j, полученными по эмпирической линии регрессии

$$y = \hat{\alpha} + \hat{\beta}(x - \overline{x}),$$

имеет вид

$$s_2^2 = c\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(\overline{\xi}_j - \hat{f}(x_j))^2 = \frac{1}{k-2}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(\overline{\xi_j} - (\hat{\alpha} + \hat{\beta}(x_j - \overline{x})))^2 =$$

$$= \frac{1}{k-2}\sum_{j=1}^{k} n_j(\overline{\xi_j} - (\hat{\alpha} + \hat{\beta}(x_j - \overline{x})))^2 \qquad (26.1.23)$$

(в качестве c выбрано $1/(k-2)$). И, что важно, *при верной гипотезе H_0 отношение s_2^2/s_1^2 имеет $\mathsf{F}_{k-2;n-k}$-распределение.* Последнее дает возможность получить следующий критерий для проверки гипотезы

$$H_0 : \mathsf{M}\xi_{j,\nu} = \alpha + \beta(x_j - \overline{x}),\ \nu = 1, 2, \ldots, n_j,\ j = 1, 2, \ldots, k,$$

— гипотезы об адекватности описания зависимости y от x линейной регрессией $y = \alpha + \beta(x - \overline{x})$.

Если гипотезу H_0 об адекватности линейной регрессии отклонять при

$$\frac{s_2^2}{s_1^2} > F_{\gamma;k-2;n-k}$$

и не отклонять в противном случае, то с вероятностью γ гипотезу будем отклонять, когда она верна.

З а м е ч а н и е 1. В случае повторных наблюдений, выписывая значение $\overline{x}$ и значения оценок $\hat{\alpha}$, $\hat{\beta}, \hat{\sigma^2}$ соответственно параметров α, β, σ^2, удобно пользоваться двойными индексами не только для наблюдений:

$$\xi_{j,\nu},\ \nu = 1, 2, \ldots, n_j;\ j = 1, 2, \ldots, k,$$

но и для точек x_j, в которых проводились наблюдения, а именно

$$x_{j,\nu} = x_j,\ \nu = 1, 2, \ldots, n_j,\ j = 1, 2, \ldots, k.$$

При этом для $\overline{x}$, S_1^2, $\overline{\xi}$, S_2^2, R_{12} (см. теорему 26.1.1 об оценках параметров) мы получим следующие выражения:

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j} x_{j,\nu} = \frac{1}{n}\sum_{j=1}^{k} n_j x_j,$$

$$S_1^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(x_{j,\nu} - \overline{x})^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(x_j - \overline{x})^2 = \frac{1}{n}\sum_{j=1}^{k} n_j (x_j - \overline{x})^2,$$

$$\overline{\xi} = \frac{1}{n}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}\xi_{j,\nu},$$

$$S_2^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(\xi_{j,\nu} - \overline{\xi})^2,$$

$$R_{12} = \frac{1}{n}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(x_{j,\nu} - \overline{x})(\xi_{j,\nu} - \overline{\xi}) = \frac{1}{n}\sum_{j=1}^{k}\sum_{\nu=1}^{n_j}(x_j - \overline{x})(\xi_{j,\nu} - \overline{\xi}).$$

Проверка адекватности теории эксперименту. Для величины y известно теоретически предсказанное (вычисленное) значение x. Величина y наблюдается в эксперименте n раз и для каждого наблюдения y_j мы имеем его теоретически предсказанное значение x_j, $j = 1, 2, \ldots, n$.

В эксперименте мы наблюдаем значения y_j с погрешностями, т. е. наблюдая y_j, мы фактически получаем

$$\xi_j = y_j + e_j,\ j = 1, 2, \ldots, n,$$

где e_j — погрешность j-го наблюдения $j = 1, 2, \ldots, n$ (избежать погрешностей в наблюдениях невозможно — наблюдений без погрешностей не бывает). Поэтому, даже если теоретические значения $x_1, x_2, \ldots, x_n$ точно предсказывают $y_1, y_2, \ldots, y_n$, т. е. $y_1 = x_1, y_2 = x_2, \ldots, y_n = x_n$ — теория согласуется с экспериментом, наблюдаемые значения $\xi_1, \xi_2, \ldots, \xi_n$ соответственно величин $y_1, y_2, \ldots, y_n$ и их теоретические значения $x_1, x_2, \ldots, x_n$ совпадать не будут — погрешности

$$\xi_j - x_j = e_j,\ j = 1, 2, \ldots, n,$$

неизбежно отличны от нуля.

В связи с этим возникает вопрос: "Можно ли по теоретически предсказанным значениям $x_1, x_2, \ldots, x_n$ величин $y_1, y_2, \ldots, y_n$ и их наблюдаемым значениям $\xi_1, \xi_2, \ldots, \xi_n$ сделать выводы о согласии (или несогласии) теории с экспериментом?" Другими словами, можно ли считать, что расхождения $\xi_1 - x_1 = e_1$, $\xi_2 - x_2 = e_2, \ldots, \xi_n - x_n = e_n$ между наблюдаемыми значениями $\xi_1, \xi_2, \ldots, \xi_n$ величин $y_1, y_2, \ldots, y_n$ и их теоретическими значениями $x_1, x_2, \ldots, x_n$ находятся в пределах погрешностей наблюдений или же эти расхождения свидетельствуют о неадекватности теории наблюдениям?

Один из возможных подходов состоит в следующем. Пусть из тех или иных соображений известно (можно считать), что величина y и её предсказанное теорией значение x связаны линейной зависимостью

$$y = \alpha + \beta(x - \overline{x}).$$

По наблюдениям $\xi_1, \xi_2, \ldots, \xi_n$ можно получить оценки $\hat{\alpha}, \hat{\beta}$, для коэффициентов линейной регрессии $y = \alpha + \beta(x - \overline{x})$ и проверить те или иные гипотезы о параметрах α, β. В рассматриваемой задаче об адекватности теории эксперименту представляет интерес результаты проверки гипотез

$$1)\ H_{0;1} : \beta = 0;\ \ 2)\ H_{0;2} : \beta = 1;$$

$$3)\ H_{0;3} : \beta = \beta_0 = 1,\ \alpha = \alpha_0 = \overline{x},$$

которые дают возможность говорить о согласии или несогласии теории с экспериментом.

1. Гипотеза $H_{0;1}$: $\beta = 0$. Если гипотеза $H_{0;1}$: $\beta = 0$ не отклоняется, то зависимость y от x имеет вид

$$y = \alpha + \beta(x - \overline{x}) = \alpha.$$

Последнее обозначает, что величина y не зависит от x, т. е. значение x_j, "претендующее" на роль предсказанного значения y_j, на самом деле таковым не является.

2. Гипотеза $H_{0;2}$: $\beta = 1$. Неотклонение гипотезы свидетельствует в пользу согласия теории с экспериментом, если при этом $\alpha - \overline{x} \neq 0$, то присутствует систематическая ошибка.

3. Гипотеза $H_{0;3}$: $\beta = \beta_0 = 1$, $\alpha = \alpha_0 = \overline{x}$. Если гипотеза $H_{0;3}$: $\beta = \beta_0 = 1$, $\alpha = \alpha_0 = \overline{x}$ не отклоняется, то можно считать, что зависимость $y = \alpha + \beta(x - \overline{x})$ величины y от x имеет вид

$$y = x,$$

что естественно интерпретировать как согласие теории с экспериментом.

Важное замечание. В справедливости (или ошибочности) предположения о линейной зависимости $y = \alpha + \beta(x - \overline{x})$ между наблюдаемыми в опыте значениями y и предсказанными теорией значениями x можно убедиться, проверив гипотезу об адекватности линейной регрессии $y = \alpha + \beta(x - \overline{x})$ (при наличии повторных наблюдений это всегда можно сделать). Если при этом гипотеза об адекватности линейной регрессии не отклоняется, то можно воспользоваться изложенным подходом проверки согласия теории с экспериментом (с опытными данными), в противном случае — нет.

Пример 26.1.1 (эксперимент Эддингтона). *Рассмотрим пример проверки адекватности теории эксперименту. Речь идет о проверке общей теории относительности по отклонению луча света в поле тяготения. Данные для статистической обработки взяты из сборника статей Альберта Эйнштейна “Физика и реальность” (М.,“Наука”, 1965).*

Схема опыта, проведенного под руководством Эддингтона, состоит в следующем (см. рис. 26.1.1). Пусть звезда лежит примерно в плоскости земной орбиты. Тогда в момент, когда Земля находится в положении 1, звезда видна в направлении 1.

Через полгода Земля окажется в положении 2, и если бы луч света в поле тяготения Солнца не отклонялся, звезда из положения 2 Земли в направлении 2 не наблюдалась бы. Тем не менее звезда из положения 2 наблюдается (см. рис. 26.1.1), но она как бы смещена и это смещение можно вычислить. (Разумеется, наблюдать звезду из положения 2 Земли в направлении 2 можно лишь в момент полного солнечного затмения — когда Солнце “не мешает” наблюдать звезду.)

Для наблюдения было выбрано 7 звезд. Их видимые перемещения (векторы на небесной сфере, которые из-за малости можно считать векторами на плоскости) разлагались по двум осям координат. Полученные результаты (в угловых секундах) приведены в таблицах 26.1.1, 26.1.2, где x_i — вычисленная координата вектора перемещения, ξ_i — наблюденная координата вектора перемещения.

Таблица 26.1.1. Первая координата

x_i	−0,22	+0,31	+0,10	+0,12	+0,04	+0,09	+0,85
ξ_i	−0,19	+0,29	+0,11	+0,20	+0,10	−0,08	+0,95

Таблица 26.1.2. Вторая координата

x_i	+0,02	−0,43	+0,74	+0,87	+0,40	+0,32	−0,09
ξ_i	+0,16	−0,46	+0,83	+1,00	+0,57	+0,35	−0,27

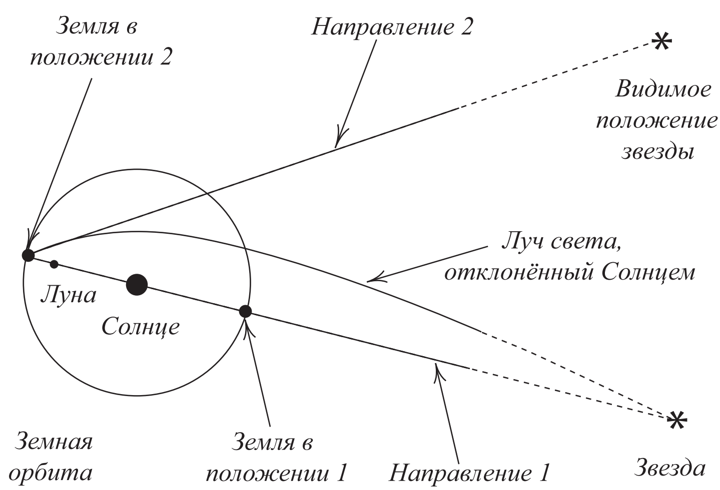

Рис. 26.1.1: Отклонение луча света звезды в поле тяготения Солнца

Согласуются ли вычисленные значения координат векторов перемещений с наблюденными?

Решение. Рассмотрим вопрос о согласии вычисленных значений координат с наблюденными для первой координаты (см. табл. 26.1.1).

Проведенные ранее наблюдения дают основания считать, что вычисленные (предсказанные теорией) значения x координат и наблюденные y связаны линейной зависимостью

$$y = \alpha + \beta(x - \overline{x}).$$

Здесь

$$\overline{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 0{,}184.$$

По результатам наблюдений (см. табл. 26.1.1) получим оценки неизвестных параметров регрессии:

$$\hat{\alpha} = \overline{\xi} = 0{,}197;\ \hat{\beta} = 1{,}081,\ \hat{\sigma^2} = 0{,}006,\ S_1^2 = 0{,}0948$$

(см. теорему 26.1.1).

Сначала проверим гипотезу $H_{0;1} : \beta = 0$, неотклонение которой свидетельствует об отсутствии какой-либо связи между теорией и экспериментом.

Согласно (26.1.11),

$$\left|\frac{\hat{\beta}}{\hat{\sigma}/(S_1\sqrt{n-2})}\right| = \left|\frac{1{,}081}{\sqrt{0{,}006/(0{,}0948\cdot 5)}}\right| = 9{,}61 > t_{0{,}01;5} = 3{,}365.$$

Так что обидная для Эйнштейна гипотеза об отсутствии связи теории с опытом отклоняется.

Далее проверим гипотезу $H_{0;2} : \beta = \beta_0 = 1$.

Согласно (26.1.11),

$$\left|\frac{\hat{\beta}-\beta_0}{\hat{\sigma}/(S_1\sqrt{n-2})}\right| = \left|\frac{1{,}081-1}{\sqrt{0{,}006/(0{,}0948\cdot 5)}}\right| = 0{,}72 < t_{0{,}01;5} = 3{,}365.$$

И, следовательно, гипотеза $H_{0;2} : \beta = \beta_0 = 1$ не отклоняется. Последнее можно трактовать как согласие опыта с экспериментом, но при этом возможна систематическая ошибка (если $\alpha \neq \overline{x}$).

В согласии (или несогласии) теории с опытом можно убедиться, проверяя гипотезу

$$H_{0;3} : \beta = 1,\ \alpha = 0{,}184.$$

Согласно (26.1.13)

$$\frac{((\hat{\alpha}-\alpha)^2 + S_1^2(\hat{\beta}-\beta)^2)/2}{\hat{\sigma^2}/(n-2)} =$$

$$= \frac{((0{,}197-0{,}184)^2 + 0{,}0948(1{,}081-1)^2)/2}{0{,}006/5} =$$

$$= 0{,}33 < F_{0,01;2;5} = 13{,}3.$$

Поэтому можно считать, что $\beta = 1$, $\alpha = 0{,}184$, а, следовательно, зависимость y от x имеет вид

$$y = x,$$

подтверждающий согласие теории с опытом по первой координате.

Аналогично можно убедиться в согласии теории с опытом по второй координате.

26.2 Примеры и задачи

Примеры

Пример 26.2.1. *В “Основах химии” Д. И. Менделеева приводятся данные о растворимости азотно-кислого натрия $NaNO_3$ в зависимости от температуры воды.*

Число условных частей $NaNO_3$, растворившихся в 100 частях воды, при соответствующих температурах приведено в таблице (x — температура в градусах, ξ — растворимость в условных частях на 100 частей воды).

x	ξ	x	ξ
0	66,7	29	92,9
4	71,0	36	99,4
10	76,3	51	113,6
15	80,6	68	125,1
21	85,7		

Теоретические соображения позволяют считать, что количественная сторона этого явления достаточно хорошо описывается линейной зависимостью:

$$y = \alpha + \beta(x - \overline{x}).$$

Найти оценки максимального правдоподобия параметров α, β, дисперсии σ^2. Построить доверительные интервалы для параметров α, β, σ^2. Проверить гипотезу $H_0 : \beta = 0$ о значимости регрессии.

Решение.

$$\overline{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = \frac{234}{9} = 26,\ \overline{\xi} = \frac{1}{9}\sum_{i=1}^{9} \xi_i = \frac{811{,}3}{9} = 90{,}14,$$

$$S_1^2 = \frac{1}{9}\sum_{i=1}^{9}(x_i - \overline{x})^2 = \frac{4060}{9} = 451{,}11,$$

$$S_2^2 = \frac{1}{9}\sum_{i=1}^{9}(\xi_i - \overline{\xi})^2 = \frac{3083{,}98}{9} = 342{,}66,$$

$$R_{12} = \frac{1}{9}\sum_{i=1}^{9}(x_i - \overline{x})(\xi_i - \overline{\xi}) = \frac{3534{,}8}{9} = 392{,}75.$$

Оценки максимального правдоподобия коэффициентов α, β простой линейной регрессии:

$$\hat{\alpha} = \overline{\xi} = 90{,}14,$$

$$\hat{\beta} = \frac{R_{12}}{S_1^2} = \frac{392{,}75}{451{,}11} = 0{,}87.$$

Оценка максимального правдоподобия дисперсии σ^2:

$$\hat{\sigma^2} = S_2^2\left(1 - \frac{R_{12}^2}{S_1^2 S_2^2}\right) = 342{,}66\left(1 - \frac{392{,}75^2}{451{,}11 \cdot 342{,}66}\right) = 0{,}7158.$$

Уравнение регрессии имеет вид

$$y = 90{,}14 + 0{,}87(x - 26),$$

или

$$y = 0{,}87x + 67{,}51.$$

Проверим гипотезу $H_0 : \beta = 0$ о значимости регрессии. Поскольку

$$\left|\frac{\hat{\beta}}{\hat{\sigma}/\sqrt{S_1^2(n-2)}}\right| = \left|\frac{0{,}87}{\sqrt{0{,}7158/451{,}11(9-2)}}\right| =$$

$$= 57{,}83 > t_{0{,}05;7} = 1{,}895,$$

то гипотеза $H_0 : \beta = 0$ отклоняется, линейная регрессия значима. Это можно интерпретировать так: с повышением температуры воды количество растворенного вещества (количество растворенного азотно-кислого натрия) возрастает.

Согласно (26.1.14), (26.1.15), (26.1.16), доверительными интервалами с коэффициентом надежности 0,9 для параметров регрессии являются: (89,538; 90,750) — для параметра α; (0,842; 0,899) — для параметра β; (0,457; 2,968) — для параметра σ^2.

Задачи.

26.1. Проверить согласие теории с опытом по второй координате (см. пример 26.1.1 и данные табл. 26.1.2).

Примечание. Оценки параметров регрессии:

$$\hat{\alpha} = 0{,}311,\ \hat{\beta} = 1{,}154,\ \hat{\sigma^2} = 0{,}009,\ S_1^2 = 0{,}1831.$$

26.2 (тормозной путь и скорость). При изучении движения уличного транспорта фиксировалось расстояние s, пройденное автомобилем по инерции после сигнала "остановиться" (тормозной путь) в зависимости от скорости v. Наблюдения проводились для различных автомобилей, с разными водителями, различным поверхностным покрытием дороги и т. д. Результаты наблюдений приведены в таблице, где s — тормозной путь автомобиля в метрах, v — скорость автомобиля в км/час.

v	s	v	s	v	s	v	s
6	0,6	19	7,3	26	9,8	32	14,6
	3,0		8,5		12,2		15,6
11	1,2	21	7,9	27	9,8		17,1
	6,7		10,4		12,2		19,5
13	4,9		10,4		15,2	35	20,1
14	3,0		14,0	29	12,8	37	16,5
16	5,5	23	7,9		17,1	39	21,3
	7,9		11,0		23,2		28,0
	10,4		18,3		25,6		28,4
18	5,2		24,4	31	11,0		36,6
	8,5	24	6,1		14,0	40	25,9
19	4,3		7,9		20,8		
	6,1		16,5	32	9,8		

Предполагая, что зависимость между тормозным путем автомобиля и скоростью линейная:

$$s = \alpha + \beta(v - \overline{v}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2; построить доверительные интервалы для параметров α, β, σ^2; проверить гипотезу $H_0 : \beta = 0$ о значимости регрессии. Проверить гипотезу об адекватности линейной регрессии.

На плоскости (v, s) изобразить исходные данные и построить эмпирическую линию регрессии.

26.3 (остеопения и количество выводимого из организма кальция). Потеря кальция в костной ткани человека — остеопения 1, 2, 3, 4 степени в своих тяжелых формах (4-я степень остеопении — остеопороз) ведет к разрушению костной ткани. В настоящее время остеопороз приобретает формы широко распространенного заболевания. Степень потери кальция в костной ткани человека определяют при помощи ультразвуковой денситометрии, которая мало доступна как по причине ее дороговизны, так и недостаточной обеспеченности соответствующей аппаратурой. Предлагается метод диагностирования степени остеопении — по содержанию кальция, выводимого с мочой. Обозначим через T результат денситометрического исследования, характеризующего плотность костной ткани и содержание кальция в ней. Если значение T принадлежит промежутку $[-i-1; -i)$, то состояние костной ткани определяется как i-я степень остеопении, $i = 1, 2, 3, 4$. При значении T из промежутка $[-2; -1)$ степень остеопении классифицируется как средняя, при $T \leq -2$ — как тяжелая. Через Ca обозначим количество кальция, выводимого с мочой. Исходя из данных Института гастроэнтерологии Академии медицинских наук Украины (см. таблицу), убедиться, что между величинами T и Ca существует линейная зависимость — постройте линейную регрессию T на Ca. Проверьте гипотезы о значимости и адекватности линейной регрессии. По известной зависимости T от Ca дифференцируйте тяжелую форму остеопении (остеопороз) по количеству выводимого Ca.

Ca	T	Ca	T	Ca	T	Ca	T
2,2	0,23	3,12	−1,63	3,84	−2,1	4,4	−1,77
	0,6	3,2	−0,21	4,1	−1,51	4,63	−1,61
2,24	−1,21		2,06	4,12	−1,86	4,8	−1,04
2,32	−0,04		−1,2		−0,65	5,8	−2,13
2,5	0,14		−1,49	4,2	−0,8	5,84	−2,94
2,6	0,49	3,35	−0,9		−0,6	5,9	−2,34
	0,07	3,5	−1,4		0,54	6,1	−2,56
2,62	−0,02		0,23		−2,23	6,7	−1,62
2,64	−0,27	3,6	−0,15	4,24	0,15	6,8	−2,67
2,67	−1,62		−0,2		0,0	6,82	−2,96
2,76	−0,05		−0,2	4,3	−1,43	7,82	−4,08
2,8	−1,84		1,0	4,32	−1,8		−3,61
	−2,04	3,64	−1,46	4,4	−1,61		−2,91

На плоскости (Ca, T) изобразить исходные данные и построить эмпирическую линию регрессии.

26.4 (потребление вина и смерть от сердечного приступа). Полезно ли вино для здоровья? Имеются данные, свидетельствующие о том, что потребление вина в умеренных количествах способствует предотвращению сердечных приступов.

В таблице приведены данные о годовом потреблении вина (в литрах l алкоголя выпитого с вином) на человека и количестве смертей в год от сердечных заболеваний (количество n смертей на 100 000 человек) в 19 развитых странах.

Предполагая, что зависимость между количеством n смертей от сердечных приступов и потреблением l вина линейная:

$$n = \alpha + \beta(l - \bar{l}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2, построить доверительные интервалы для α, β, σ^2, проверить гипотезу о значимости регрессии.

Страна	l	n	Страна	l	n
Австралия	2,5	211	Нидерланды	1,8	167
Австрия	3,9	167	Новая Зеландия	1,9	266
Бельгия	2,9	131	Норвегия	0,8	227
Канада	2,4	191	Испания	6,5	86
Дания	2,9	220	Швеция	1,6	207
Финляндия	0,8	297	Швейцария	5,8	115
Франция	9,1	71	Великобритания	1,3	285
Исландия	0,8	211	США	1,2	199
Ирландия	0,7	300	Германия	2,7	172
Италия	7,9	107			

На плоскости (n, l) изобразить исходные данные и построить эмпирическую линию регрессии.

26.5 (стоимость эксплуатации самолета и его “возраст”). Дирекция авиакомпании с целью планирования расходов хочет установить количественную зависимость стоимости эксплуатации самолета от продолжительности его эксплуатации (от “возраста” самолета). Со временем из-за старения деталей и узлов самолета компания несет большие затраты на поддержание его в рабочем состоянии, в частности, необходимо чаще проводить ремонтно-профилактические работы, заменять отдельные узлы.

В таблице приведены данные о стоимости эксплуатации самолетов и их “возрасте” (x — “возраст самолета” в годах, y — стоимость эксплуатации самолета в течение полугода в долларах).

Предполагая, что зависимость стоимости эксплуатации самолета от его “возраста” линейная, найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2; построить доверительные интервалы для параметров α, β, σ^2. Проверить гипотезу о значимости линейной регрессии, гипотезу об адекватности линейной регрессии.

x	y	x	y
4,5	619	5,5	987
	1049	0,5	163
	1033		182
4,0	495	6,0	764
	723		1373
	681	1,0	978
5,0	890		466
	1522		549
	1194		

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

26.6. В таблице приведены результаты эксперимента по изучению нового метода измерения скорости крови. Можно ли считать, что оба метода дают одинаковые результаты измерения скорости крови?

x	y	x	y	x	y
1190	1115	1900	1830	2720	2630
1455	1425	1920	1920	2710	2740
1550	1515	1960	1970	2530	2390
1730	1795	2295	2300	2900	2800
1745	1715	2335	2280	2760	2630
1770	1710	2490	2520	3010	2970

x — оценка скорости крови, полученная стандартным методом, y — новым методом.

Указание. Предположим, что скорость крови, полученная с помощью нового метода, линейно зависит от скорости крови, полученной стандартным методом:

$$y = \alpha + \beta(x - \overline{x}).$$

Проверьте гипотезу

$$H_0 : \beta = 1,\ \alpha = \overline{x}.$$

Глава 27

Таблицы математической статистики

В эту главу включены все необходимые для работы с учебником таблицы. В частности, приведены квантили, верхние пределы (критические значения) основных распределений математической статистики: стандартного нормального, Стьюдента, Фишера, χ^2-распределения.

Пусть F — абсолютно непрерывное распределение. Для каждого $\beta \in (0;1)$ число x_β, являющееся решением уравнения

$$F(x_\beta) = \beta,$$

или, что то же, $\mathsf{F}((-\infty, x_\beta)) = \beta$, называется *$\beta$-квантилью* распределения F.

Для каждого $\alpha \in (0;1)$ число z_α, являющееся решением уравнения

$$1 - F(z_\alpha) = \alpha,$$

или, что то же, $\mathsf{F}([z_\alpha, +\infty)) = \alpha$, будем называть верхним *α-пределом* (верхним *100α-процентной пределом*, *100α-процентной точкой*, *100α-критическим значением*) распределения F.

Очевидно, верхний α-предел z_α распределения F и его $(1-\alpha)$-квантиль совпадают: $x_{1-\alpha} = z_\alpha$.

27.1 Нормальное распределение

В табл. 27.1.1 приведены значения функции $\Phi(t)$ нормального распределения с параметрами (0;1) (квантили нормального распределения): для заданных t табулированы значения функции

$$N_{0;1}(t) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{t} \exp\left\{-\frac{s^2}{2}\right\} ds.$$

Для каждого t значение $N_{0;1}(t)$ численно равно площади заштрихованной на рис. 27.1.1 фигуры.

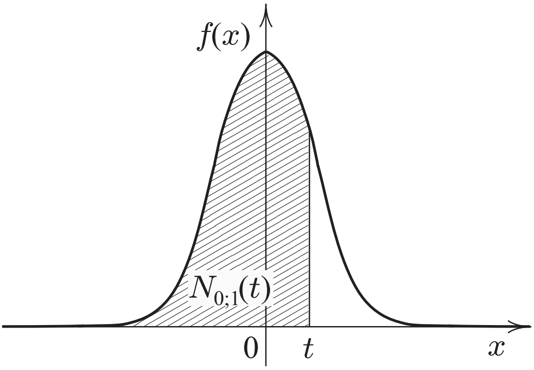

Рис. 27.1.1: К определению квантили нормального распределения; $f(x)$ — плотность распределения $N_{0;1}$

Значение $N_{a;\sigma^2}(x)$ — функции нормального распределения с параметрами $(a;\sigma^2)$ — вычисляется по значениям табулированной функции $N_{0;1}(x) = \Phi(x)$ нормального распределения с параметрами $(0;1)$ следующим образом:

$$N_{a;\sigma^2}(x) = \Phi\left(\frac{x-a}{\sigma}\right).$$

Таблица 27.1.1 допускает линейную интерполяцию.

Таблица 27.1.1. Значения функции $\Phi(t)$

t	0	1	2	3	4	5	6	7	8	9
−0,0	,5000	,4960	,4920	,4880	,4840	,4801	,4761	,4721	,4681	,4641
−0,1	,4602	,4562	,4522	,4483	,4443	,4404	,4364	,4325	,4286	,4247
−0,2	,4207	,4168	,4129	,4090	,4052	,4013	,3974	,3936	,3897	,3859
−0,3	,3821	,3783	,3745	,3707	,3669	,3632	,3594	,3557	,3520	,3483
−0,4	,3446	,3409	,3372	,3336	,3300	,3264	,3228	,3192	,3156	,3121
−0,5	,3085	,3050	,3015	,2981	,2946	,2912	,2877	,2843	,2810	,2776
−0,6	,2743	,2709	,2676	,2643	,2611	,2578	,2546	,2514	,2483	,2451
−0,7	,2420	,2389	,2358	,2327	,2297	,2266	,2236	,2206	,2177	,2148
−0,8	,2119	,2090	,2061	,2033	,2005	,1977	,1949	,1922	,1894	,1867
−0,9	,1841	,1814	,1788	,1762	,1736	,1711	,1685	,1660	,1635	,1611
−1,0	,1587	,1562	,1539	,1515	,1492	,1469	,1446	,1423	,1401	,1379
−1,1	,1357	,1335	,1314	,1292	,1271	,1251	,1230	,1210	,1190	,1170
−1,2	,1151	,1131	,1112	,1093	,1075	,1056	,1038	,1020	,1003	,0985
−1,3	,0968	,0951	,0934	,0918	,0901	,0885	,0869	,0853	,0838	,0823
−1,4	,0808	,0793	,0778	,0764	,0749	,0735	,0721	,0708	,0694	,0681
−1,5	,0668	,0655	,0643	,0630	,0618	,0606	,0594	,0582	,0571	,0559
−1,6	,0548	,0537	,0526	,0516	,0505	,0495	,0485	,0475	,0465	,0455
−1,7	,0446	,0436	,0427	,0418	,0409	,0401	,0392	,0384	,0375	,0367
−1,8	,0359	,0351	,0344	,0336	,0339	,0322	,0314	,0307	,0301	,0294
−1,9	,0288	,0281	,0274	,0268	,0262	,0256	,0250	,0244	,0239	,0233
−2,0	,0228	,0222	,0217	,0212	,0207	,0202	,0197	,0192	,0188	,0183
−2,1	,0179	,0174	,0170	,0166	,0162	,0158	,0154	,0150	,0146	,0143
−2,2	,0139	,0136	,0132	,0129	,0125	,0122	,0119	,0116	,0113	,0110
−2,3	,0107	,0104	,0102	,0099	,0096	,0094	,0091	,0089	,0087	,0084
−2,4	,0082	,0080	,0078	,0075	,0073	,0071	,0069	,0068	,0066	,0064
−2,5	,0062	,0060	,0059	,0057	,0055	,0054	,0052	,0051	,0049	,0048
−2,6	,0047	,0045	,0044	,0043	,0041	,0040	,0039	,0038	,0037	,0036
−2,7	,0035	,0034	,0033	,0032	,0031	,0030	,0029	,0028	,0027	,0026
−2,8	,0026	,0025	,0024	,0023	,0023	,0022	,0021	,0021	,0020	,0019
−2,9	,0019	,0018	,0018	,0017	,0016	,0016	,0015	,0015	,0014	,0014

t	−3,0	−3,1	−3,2	−3,3	−3,4	−3,5	−3,6	−3,7	−3,8	−3,9
$\Phi(t)$	,0013	,0010	,0007	,0005	,0003	,0002	,0002	,0001	,0001	,0000

Таблица 27.1.1 (окончание)

t	0	1	2	3	4	5	6	7	8	9
0,0	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7257	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549
0,7	,7580	,7611	,7642	,7673	,7703	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9842	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9900	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9923	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986

t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t)$	,9987	,9990	,9993	,9995	,9997	,9998	,9998	,9999	,9999	,1000

27.2 Распределение χ^2

χ^2-распределением с n степенями свободы (коротко *χ^2_n-распределением*) будем называть распределение случайной величины

$$\chi^2_n = \sum_{i=1}^{n} \xi_i^2, \tag{27.2.1}$$

где $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, каждая с распределением $\mathsf{N}_{0;1}$.

В табл. 27.2.1 приведены значения функции $\chi^2_{\alpha;n}$, или, что то же, верхние α-пределы (100α-критические значения) распределения Пирсона.

Для данных α и n значение $\chi^2_{\alpha;n}$ определяется как решение уравнения

$$\int\limits_{\chi^2_{\alpha;n}}^{+\infty} f(x)dx = \alpha,$$

где $f(x)$ — плотность χ^2_n-распределения (см. рис. 27.2.1).

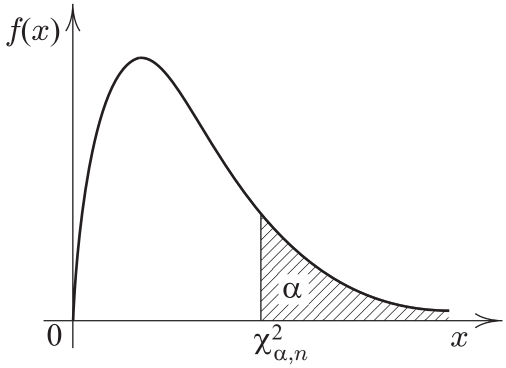

Рис. 27.2.1: К определению $\chi^2_{\alpha;n}$ — верхнего α-предела χ^2_n-распределения, $f(x)$ — плотность χ^2_n-распределения

Таблица 27.2.1. Значения функции $\chi^2_{\alpha;n}$

n	Значения α							
	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010
1	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,64
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21
3	0,12	0,22	0,35	0,58	6,23	7,82	9,35	11,34
4	0,30	0,48	0,71	1,06	7,78	9,48	11,14	13,28
5	0,55	0,83	1,14	1,61	9,24	11,07	12,83	15,09
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09
9	2,90	2,70	3,32	4,17	14,68	16,92	19,02	21,67
10	2,56	3,25	3,94	4,86	15,99	18,31	20,48	23,21
11	3,05	3,82	4,58	5,58	17,28	19,68	21,92	24,72
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58
16	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00
17	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41
18	7,02	8,23	9,39	10,86	25,99	28,87	31,53	34,81
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57
21	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64
24	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,92
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31
26	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,26
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89

Т а б л и ц а 27.2.1 (окончание)

n	Значения α							
	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010
40	22,16	24,43	26,51	29,05	51,80	55,76	59,34	63,69
50	29,71	32,36	34,76	37,69	63,17	67,50	71,42	76,15
60	37,48	40,48	43,19	46,46	74,40	79,08	83,30	88,38
70	45,44	48,76	51,74	55,33	85,53	90,53	95,02	100,4
80	53,54	57,15	60,39	64,28	96,58	101,9	106,6	112,3
90	61,75	65,65	69,13	73,29	107,6	113,1	118,1	124,1
100	70,06	74,22	77,93	82,36	118,5	124,3	129,6	135,8

27.3 Распределение Стьюдента

Распределением Стьюдента или *t-распределением с n степенями свободы* (коротко *t_n-распределением*) будем называть распределение случайной величины

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}}, \tag{27.3.1}$$

где ξ и χ_n^2 — независимые случайные величины, ξ распределена $N_{0;1}$, а χ_n^2 имеет χ^2-распределение с n степенями свободы.

В табл. 27.3.1 приведены значения функции $t_{\alpha;n}$, или, что то же, верхние α-пределы (100α-критические значения) распределения Стьюдента с n степенями свободы (t_n-распределения).

Для данных α и n значение $t_{\alpha;n}$ определяется как решение уравнения

$$\int\limits_{t_{\alpha;n}}^{+\infty} f(x)dx = \alpha,$$

где $f(x)$ — плотность t_n-распределения, $t_{\alpha;n}$ — число, отсекающее правый “хвост” t_n-распределения, на который приходится “масса” α (см. рис. 27.3.1).

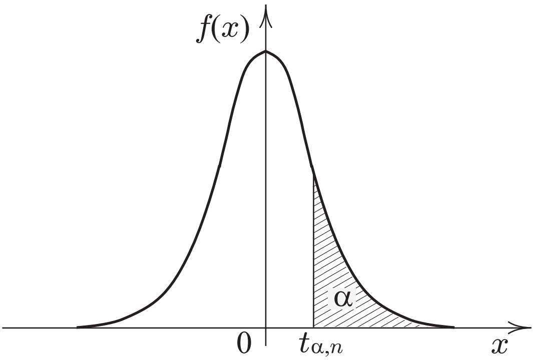

Рис. 27.3.1: К определению $t_{\alpha;n}$ — верхнего α-предела t_n-распределения, $f(x)$ — плотность t_n-распределения

Таблица 27.3.1. Значения функции $t_{\alpha;n}$

	Значения α					Значения α			
n	0,050	0,025	0,010	0,005	n	0,050	0,025	0,010	0,005
1	6,314	12,706	31,821	63,657	18	1,734	2,101	2,552	2,878
2	2,920	4,303	6,965	9,925	19	1,729	2,093	2,539	2,861
3	2,353	3,182	4,541	5,841	20	1,725	2,086	2,528	2,845
4	2,132	2,776	3,747	4,604	21	1,721	2,080	2,518	2,831
5	2,015	2,571	3,365	4,032	22	1,717	2,074	2,508	2,819
6	1,943	2,447	3,143	3,707	23	1,714	2,069	2,500	2,807
7	1,895	2,365	2,998	3,499	24	1,711	2,064	2,492	2,797
8	1,860	2,306	2,896	3,355	25	1,708	2,060	2,485	2,787
9	1,833	2,262	2,821	3,250	26	1,706	2,056	2,479	2,779
10	1,812	2,228	2,764	3,169	27	1,703	2,052	2,473	2,771
11	1,796	2,201	2,718	3,106	28	1,701	2,048	2,467	2,763
12	1,782	2,179	2,681	3,055	29	1,699	2,045	2,462	2,756
13	1,771	2,160	2,650	3,012	30	1,697	2,042	2,457	2,750
14	1,761	2,145	2,624	2,977	40	1,684	2,021	2,423	2,704
15	1,753	2,131	2,602	2,947	60	1,671	2,000	2,390	2,660
16	1,746	2,120	2,583	2,921	120	1,658	1,980	2,358	2,617
17	1,740	2,110	2,567	2,898	∞	1,645	1,960	2,326	2,576

27.4 Распределение Фишера

Распределением Фишера или F-*распределением с* n, m *степенями свободы* (коротко $\mathsf{F}_{n,m}$-*распределением*) будем называть распределение случайной величины

$$F_{n;m} = \frac{\frac{1}{n}\chi_n^2}{\frac{1}{m}\chi_m^2}, \tag{27.4.1}$$

где χ_n^2 и χ_m^2 — независимые случайные величины, χ_n^2 имеет распределение Пирсона с n степенями свободы, χ_m^2 — с m степенями свободы.

В табл. 27.4.1 и 27.4.2 приведены значения функции $F_{\alpha;n;m}$, или, что то же, верхние α-границы (100α-критические значения) $\mathsf{F}_{n;m}$-распределения.

Для данных α, n, m значение $F_{\alpha;n;m}$ определяется как решение уравнения

$$\int\limits_{F_{\alpha;n;m}}^{+\infty} f(x)dx = \alpha,$$

где $f(x)$ — плотность $\mathsf{F}_{n;m}$-распределения (см. рис. 27.4.1).

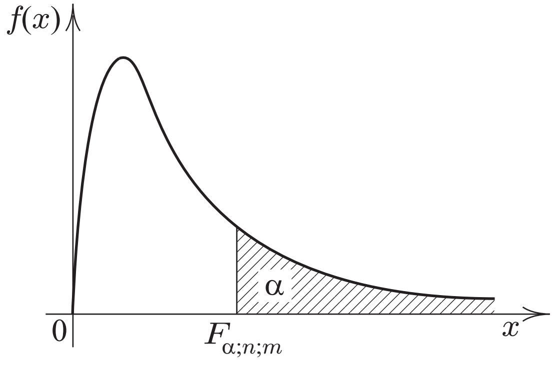

Рис. 27.4.1: К определению $F_{\alpha;n;m}$ — верхнего α-предела $\mathsf{F}_{n;m}$-распределения, $f(x)$ — плотность $\mathsf{F}_{n;m}$-распределения

Таблица 27.4.1. Значения функции $F_{\alpha;n;m}$
(уровень значимости 0,05)

m	n (число степеней свободы числителя)									
	1	2	3	4	5	6	7	8	9	10
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93

Таблица 27.4.1 (окончание)

m	n (число степеней свободы числителя)									
	12	14	16	18	20	30	40	50	60	100
3	8,74	8,71	8,69	8,67	8,66	8,62	8,59	8,58	8,57	8,55
4	5,91	5,87	5,84	5,82	5,80	5,75	5,72	5,70	5,69	5,66
5	4,68	4,64	4,60	4,58	4,56	4,50	4,46	4,44	4,43	4,41
6	4,00	3,96	3,92	3,90	3,87	3,81	3,77	3,75	3,74	3,71
7	3,57	3,53	3,49	3,47	3,44	3,38	3,34	3,32	3,30	3,27
8	3,28	3,24	3,20	3,17	3,15	3,08	3,04	3,02	3,01	2,97
9	3,07	3,03	2,99	2,96	2,94	2,86	2,83	2,80	2,79	2,76
10	2,91	2,86	2,83	2,80	2,77	2,70	2,66	2,64	2,62	2,59
11	2,79	2,74	2,70	2,67	2,65	2,57	2,53	2,51	2,49	2,46
12	2,69	2,64	2,60	2,57	2,54	2,47	2,43	2,40	2,38	2,35
13	2,60	2,55	2,51	2,48	2,46	2,38	2,34	2,31	2,30	2,26
14	2,53	2,48	2,44	2,41	2,39	2,31	2,27	2,24	2,22	2,19
15	2,48	2,42	2,38	2,35	2,33	2,25	2,20	2,18	2,16	2,12
16	2,42	2,37	2,33	2,30	2,28	2,19	2,15	2,12	2,11	2,07
17	2,38	2,33	2,29	2,26	2,23	2,15	2,10	2,08	2,06	2,02
18	2,34	2,29	2,25	2,22	2,19	2,11	2,06	2,04	2,02	1,98
19	2,31	2,26	2,21	2,18	2,16	2,07	2,03	2,00	1,98	1,94
20	2,28	2,22	2,18	2,15	2,12	2,04	1,99	1,97	1,95	1,91
22	2,23	2,17	2,13	2,10	2,07	1,98	1,94	1,91	1,89	1,85
24	2,18	2,13	2,09	2,05	2,03	1,94	1,89	1,86	1,84	1,80
26	2,15	2,09	2,05	2,02	1,99	1,90	1,85	1,82	1,80	1,76
28	2,12	2,06	2,02	1,99	1,96	1,87	1,82	1,79	1,77	1,73
30	2,09	2,04	1,99	1,96	1,93	1,84	1,79	1,76	1,74	1,70
40	2,00	1,95	1,90	1,87	1,84	1,74	1,69	1,66	1,64	1,59
50	1,95	1,89	1,85	1,81	1,78	1,69	1,63	1,60	1,58	1,52
60	1,92	1,86	1,82	1,78	1,75	1,65	1,59	1,56	1,53	1,48
100	1,85	1,79	1,75	1,71	1,68	1,57	1,52	1,48	1,45	1,39

Таблица 27.4.2. Значения функции $F_{\alpha;n;m}$
(уровень значимости 0,01)

m	n (число степеней свободы числителя)									
	1	2	3	4	5	6	7	8	9	10
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,79	2,70
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50

Таблица 27.4.2 (окончание)

m	n (число степеней свободы числителя)									
	12	14	16	18	20	30	40	50	60	100
3	27,1	26,9	26,8	26,8	26,7	26,5	26,4	26,4	26,3	26,2
4	14,4	14,2	14,2	14,1	14,0	13,8	13,7	13,7	13,7	13,6
5	9,89	9,77	9,68	9,61	9,55	9,38	9,29	9,24	9,20	9,13
6	7,72	7,60	7,52	7,45	7,40	7,23	7,14	7,09	7,06	6,99
7	6,47	6,36	6,27	6,21	6,16	5,99	5,91	5,86	5,82	5,75
8	5,67	5,56	5,84	5,41	5,36	5,20	5,12	5,07	5,03	4,96
9	5,11	5,00	4,92	4,86	4,81	4,65	4,57	4,52	4,48	4,42
10	4,71	4,60	4,52	4,46	4,41	4,25	4,17	4,12	4,08	4,01
11	4,40	4,29	4,21	4,15	4,10	3,94	3,86	3,81	3,78	3,71
12	4,16	4,05	3,97	3,91	3,86	3,70	3,62	3,57	3,54	3,47
13	3,96	3,86	3,78	3,72	3,66	3,51	3,43	3,38	3,34	3,27
14	3,80	3,70	3,62	3,56	3,51	3,35	3,27	3,22	3,18	3,11
15	3,67	3,56	3,49	3,42	3,37	3,21	3,13	3,08	3,05	2,98
16	3,55	3,45	3,37	3,31	3,26	3,10	3,02	2,97	2,93	2,86
17	3,46	3,35	3,27	3,21	3,16	3,00	2,92	2,87	2,83	2,76
18	3,37	3,27	3,19	3,13	3,08	2,92	2,84	2,78	2,75	2,68
19	3,30	3,19	3,12	3,05	3,00	2,84	2,76	2,71	2,67	2,60
20	3,23	3,13	3,05	2,99	2,94	2,78	2,69	2,64	2,61	2,54
22	3,12	3,02	2,94	2,88	2,83	2,67	2,58	2,53	2,50	2,42
24	3,03	2,93	2,85	2,79	2,74	2,58	2,49	2,44	2,40	2,33
26	2,96	2,86	2,78	2,72	2,66	2,50	2,42	2,36	2,33	2,25
28	2,90	2,79	2,72	2,65	2,60	2,44	2,35	2,30	2,26	2,19
30	2,84	2,74	2,66	2,60	2,55	2,39	2,30	2,25	2,21	2,13
40	2,66	2,56	2,48	2,42	2,37	2,20	2,11	2,06	2,02	1,94
50	2,56	2,46	2,38	2,32	2,27	2,10	2,01	1,95	1,91	1,82
60	2,50	2,39	2,31	2,25	2,20	2,03	1,94	1,88	1,84	1,75
100	2,37	2,26	2,19	2,12	2,07	1,89	1,80	1,73	1,69	1,60

27.5 Биномиальное распределение

Таблица 27.5.1. Значения функции $P(i;n,p) = C_n^i p^i (1-p)^{n-i}$

n	i	p = 0,1	0,2	0,3	0,4	0,5		
10	0	0,34868	0,10737	0,02825	0,00605	0,00098	10	10
	1	,38742	,26844	,12106	,04031	,00977	9	
	2	,19371	,30199	,23347	,12093	,04395	8	
	3	,05740	,20133	,26683	,21499	,11719	7	
	4	,01116	,08808	,20012	,25082	,20508	6	
	5	0,00149	0,02642	0,10292	0,20066	0,24609	5	
	6	,00014	,00551	,03676	,11148	,20508	4	
	7	,00001	,00079	,00900	,04247	,11719	3	
	8		,00007	,00145	,01062	,04395	2	
	9			,00014	,00157	,00977	1	
	10			,00001	,00010	,00098	0	
15	0	0,20589	0,03518	0,00475	0,00047	0,00003	15	15
	1	,34315	,13194	,03052	,00470	,00046	14	
	2	,26690	,23090	,09156	,02194	,00320	13	
	3	,12851	,25014	,17004	,06339	,01389	12	
	4	,04284	,18760	,21862	,12678	,04166	11	
	5	0,01047	0,10318	0,20613	0,18594	0,09164	10	
	6	,00194	,04299	,14724	,20660	,15274	9	
	7	,00028	,01382	,08113	,17708	,19638	8	
	8	,00003	,00345	,03477	,11806	,19638	7	
	9		,00067	,01159	,06121	,15274	6	
		0,9	0,8	0,7	0,6	0,5 = p	i	n

Таблица 27.5.1 (продолжение)

n	i	p = 0,1	0,2	0,3	0,4	0,5		
15	10		0,00010	0,00298	0,02449	0,09164	5	15
	11		,00001	,00058	,00742	,04166	4	
	12			,00008	,00165	,01389	3	
	13			,00001	,00025	,00320	2	
	14				,00002	,00046	1	
	15					,00003	0	
20	0	0,12158	0,01153	0,00080	0,00004		20	20
	1	,27017	,05765	,00684	,00049	,00002	19	
	2	,28518	,13691	,02785	,00309	,00018	18	
	3	,19012	,20536	,07160	,01235	,00109	17	
	4	,08978	,21820	,13042	,03499	,00462	16	
	5	0,03192	0,17456	0,17886	0,07465	0,01479	15	
	6	,00887	,10910	,19164	,12441	,03696	14	
	7	,00197	,05455	,16426	,16588	,07393	13	
	8	,00036	,02216	,11440	,17971	,12013	12	
	9	,00005	,00739	,06537	,15974	,16018	11	
	10	0,00001	0,00203	0,03082	0,11714	0,17620	10	
	11		,00046	,01201	,07099	,16018	9	
	12		,00009	,00386	,03550	,12013	8	
	13		,00001	,00102	,01456	,07393	7	
	14			,00022	,00485	,03696	6	
	15			0,00004	0,00129	0,01479	5	
	16			,00001	,00027	,00462	4	
	17				,00004	,00109	3	
	18					,00018	2	
	19					,00002	1	
	20						0	
		0,9	0,8	0,7	0,6	0,5 = p	i	n

Таблица 27.5.1 (окончание)

n	i	p = 0,1	0,2	0,3	0,4	0,5		
25	0	0,07179	0,00378	0,00013			25	25
	1	,19942	,02361	,00144	,00005		24	
	2	,26589	,07084	,00739	,00038	,00001	23	
	3	,22650	,13577	,02428	,00194	,00007	22	
	4	,13842	,18668	,05723	,00710	,00038	21	
	5	0,06459	0,19602	0,10302	0,01989	0,00158	20	
	6	,02392	,16335	,14717	,04420	,00528	19	
	7	,00722	,11084	,17119	,07999	,01433	18	
	8	,00180	,06235	,16508	,11998	,03223	17	
	9	,00038	,02944	,13364	,15109	,06089	16	
	10	0,00007	0,01178	0,09164	0,16116	0,09742	15	
	11	,00001	,00401	,05355	,14651	,13284	14	
	12		,00117	,02678	,11395	,15498	13	
	13		,00029	,01148	,07597	,15498	12	
	14		,00006	,00422	,04341	,13284	11	
	15		0,00001	0,00132	0,02122	0,09742	10	
	16			,00035	,00884	,06089	9	
	17			,00008	,00312	,03223	8	
	18			,00002	,00092	,01433	7	
	19				,00023	,00528	6	
	20				0,00005	0,00158	5	
	21				,00001	,00038	4	
	22					,00007	3	
	23					,00001	2	
		p = 0,9	0,8	0,7	0,6	0,5	i	n

27.6 Распределение Пуассона

Таблица 27.6.1. Значения функции

$$P_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

k	Значения λ								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	,9048	,8187	,7408	,6703	,6065	,5488	,4966	,4493	,4066
1	,0905	,1637	,2222	,2681	,3033	,3293	,3476	,3595	,3659
2	,0045	,0164	,0333	,0536	,0758	,0988	,1217	,1438	,1647
3	,0002	,0011	,0033	,0072	,0126	,0198	,0284	,0383	,0494
4		,0001	,0003	,0007	,0016	,0030	,0050	,0077	,0111
5				,0001	,0002	,0004	,0007	,0012	,0020
6							,0001	,0002	,0003

k	Значения λ								
	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	,3679	,1353	,0498	,0183	, 0067	,0025	,0009	,0003	,0001
1	,3679	,2707	,1494	,0733	, 0337	,0149	,0064	,0027	,0011
2	,1839	,2707	,2240	,1465	, 0842	,0446	,0223	,0107	,0050
3	,0613	,1804	,2240	,1954	, 1404	,0892	,0521	,0286	,0150
4	,0153	,0902	,1680	,1954	, 1755	,1339	,0912	,0573	,0337
5	,0031	,0361	,1008	,1563	, 1755	,1606	,1277	,0916	,0607
6	,0005	,0120	,0504	,1042	, 1462	,1606	,1490	,1221	,0911
7	,0001	,0034	,0216	,0595	, 1044	,1377	,1490	,1396	,1171
8		,0009	,0081	,0298	, 0653	,1033	,1304	,1396	,1318
9		,0002	,0027	,0132	, 0363	,0688	,1014	,1241	,1318
10			,0008	,0053	, 0181	,0413	,0710	,0993	,1186
11			,0002	,0019	, 0082	,0225	,0452	,0722	,0970
12			,0001	,0006	, 0034	,0113	,0264	,0481	,0728
13				,0002	, 0013	,0052	,0142	,0296	,0504
14				,0001	, 0005	,0022	,0071	,0169	,0324
15					, 0002	,0009	,0033	,0090	,0194
16						,0003	,0014	,0045	,0109
17						,0001	,0006	,0021	,0058

27.7 Критерий А. Н. Колмогорова. Критические значения

В табл. 27.7.1 приведены критические значения $\varepsilon_{\alpha;n}$ супремума модуля разности истинной и эмпирической функций распределений.

Значение $\varepsilon_{\alpha;n}$ для данных α и n определяется как минимальное ε, для которого

$$\mathsf{P}\{\sup_x |F(x) - \hat{F}_n(x)| \geq \varepsilon\} \leq \alpha.$$

Таблица 27.7.1. Критические значения $\varepsilon_{\alpha;n}$ для супремума модуля разности истинной и эмпирической функций распределения

n	Значения α			n	Значения α		
	0,05	0,02	0,01		0,05	0,02	0,01
1	0,9750	0,9900	0,9950	25	0,2640	0,2952	0,3166
2	0,8419	0,9000	0,9293	30	0,2417	0,2702	0,2899
3	0,7076	0,7846	0,8290	35	0,2243	0,2507	0,2690
4	0,6239	0,6889	0,7342	40	0,2101	0,2349	0,2520
5	0,5633	0,6272	0,6685	45	0,1984	0,2218	0,2380
6	0,5193	0,5774	0,6166	50	0,1884	0,2107	0,2260
7	0,4834	0,5384	0,5758	55	0,1798	0,2011	0,2157
8	0,4543	0,5065	0,5418	60	0,1723	0,1927	0,2067
9	0,4300	0,4796	0,5133	65	0,1657	0,1853	0,1988
10	0,4093	0,4566	0,4889	70	0,1598	0,1786	0,1917
11	0,3912	0,4367	0,4677	75	0,1544	0,1727	0,1853
12	0,3754	0,4192	0,4491	80	0,1496	0,1673	0,1795
13	0,3614	0,4036	0,4325	85	0,1452	0,1624	0,1742
14	0,3489	0,3897	0,4176	90	0,1412	0,1579	0,1694
15	0,3376	0,3771	0,4042	95	0,1375	0,1537	0,1649
20	0,2941	0,3287	0,3524	100	0,1340	0,1499	0,1608

При $n > 100$ можно пользоваться асимптотическими границами

$$\varepsilon_{0,05;n} = \frac{1{,}36}{\sqrt{n}}; \quad \varepsilon_{0,01;n} = \frac{1{,}63}{\sqrt{n}}.$$

27.8 Критерий Вилкоксона. Нижние критические значения

В табл. 27.8.1 приведены нижние критические значения $W_{\alpha;n;m}$ распределения W — суммы рангов выборки меньшего объема.

Значение $W_{\alpha;n;m}$ для данных α (уровня значимости), n и m — объемов выборок (n — объем меньшей выборки, m — бо́льшей) определяется как наибольшее целое t, для которого

$$\mathsf{P}\{W \le t\} \le \alpha.$$

При значениях n и m больших, чем приведенные в таблице (а фактически при n и m, которые удовлетворяют неравенствам $\min\{n, m\} \ge 6, m + n \ge 20$), можно считать, что $W_{\alpha;n;m}$ равно

$$\frac{1}{2}n(n+m+1) + z_\alpha\sqrt{\frac{1}{12}nm(n+m+1)},$$

где z_α — это α-квантиль нормального распределения с параметрами (0;1) (см. табл. 27.1.1).

Таблица 27.8.1. Нижние критические значения $W_{\alpha;n;m}$ распределения W

Объемы		Значения α				Объемы		Значения α			
n	m	0,005	0,01	0,025	0,05	n	m	0,005	0,01	0,025	0,05
6	6	23	24	26	28	6	18	37	40	45	49
	7	24	25	27	29		19	38	41	46	51
	8	25	27	29	31	7	7	32	34	36	39
	9	26	28	31	33		8	34	35	38	41
	10	27	29	32	35		9	35	37	40	43
	11	28	30	34	37		10	37	39	42	45
	12	30	32	35	38		11	38	40	44	47
	13	31	33	37	40		12	40	42	46	49
	14	32	34	38	42		13	41	44	48	52
	15	33	36	40	44		14	43	45	50	54
	16	34	37	42	46		15	44	47	52	56
	17	36	39	43	47		16	46	49	54	58

Таблица 27.8.1 (окончание)

Объемы n	m	0,005	0,01	0,025	0,05	Объемы n	m	0,005	0,01	0,025	0,05
7	17	47	51	56	61	9	13	65	68	73	78
	18	49	52	58	63		14	67	71	76	81
8	8	43	45	49	51		15	69	73	79	84
	9	45	47	51	54		16	72	76	82	87
	10	47	49	53	56	10	10	71	74	78	82
	11	49	51	55	59		11	73	77	81	86
	12	51	53	58	62		12	76	79	84	89
	13	53	56	60	64		13	79	82	88	92
	14	54	58	62	67		14	81	85	91	96
	15	56	60	65	69		15	84	88	94	99
	16	58	62	67	72	11	11	87	91	96	100
	17	60	64	70	75		12	90	94	99	104
9	9	56	59	62	66		13	93	97	103	108
	10	58	61	65	69		14	96	100	106	112
	11	61	63	68	72	12	12	105	109	115	120
	12	63	66	71	75		13	109	113	119	125

(Значения α — столбцы 0,005; 0,01; 0,025; 0,05.)

27.9 Критерий знаков. Границы критической области

В табл. 27.9.1 приведены левая $(n - m_{\alpha;n})$ (левый столбец) и правая $m_{\alpha;n}$ (правый столбец) границы критической области критерия знаков (пределы области отклонения гипотезы H_0: $\theta = 0$).

Значение $m_{\alpha;n}$ для данных α (уровня значимости) и n (числа отличных от нуля разностей) определяется как минимальное целое m, для которого $\mathsf{P}\{\mu > m\} \leq \alpha$, где μ — биномиально распределенная случайная величина с параметрами n и $1/2$.

Критические области критерия знаков:

$(m_{\alpha;n}; n]$ для односторонней альтернативы $\theta > 0$;

$[0; n - m_{\alpha;n})$ для односторонней альтернативы $\theta < 0$;

$[0; n - m_{\alpha;n}) \cup (m_{\alpha;n}; n]$ для двусторонней альтернативы $\theta < 0$ или $\theta > 0$;

Уровень значимости одностороннего критерия не превышает α, двустороннего — 2α.

Таблица 27.9.1. Границы критических областей критерия знаков

n	Значения α						n	Значения α					
	0,025		0,010		0,005			0,025		0,010		0,005	
5	0	5	0	5	0	5	25	8	17	7	18	6	19
6	1	5	0	6	0	6	26	8	18	7	19	7	19
7	1	6	1	6	0	7	27	8	19	8	19	7	20
8	1	7	1	7	1	7	28	9	19	8	20	7	21
9	2	7	1	8	1	8	29	9	20	8	21	8	21
10	2	8	1	9	1	9	30	10	20	9	21	8	22
11	2	9	2	9	1	10	31	10	21	9	22	8	23
12	3	9	2	10	2	10	32	10	22	9	23	9	23
13	3	10	2	11	2	11	33	11	22	10	23	9	24
14	3	11	3	11	2	12	34	11	23	10	24	10	24
15	4	11	3	12	3	12	35	12	23	11	24	10	25
16	4	12	3	13	3	13	36	12	24	11	25	10	26
17	5	12	4	13	3	14	37	13	24	11	26	11	26
18	5	13	4	14	4	14	38	13	25	12	26	11	27
19	5	14	5	14	4	15	39	13	26	12	27	12	27
20	6	14	5	15	4	16	40	14	26	13	27	12	28
21	6	15	5	16	5	16	41	14	27	13	28	12	29
22	6	16	6	16	5	17	42	15	27	14	28	13	29
23	7	16	6	17	5	18	43	15	28	14	29	13	30
24	7	17	6	18	6	18	44	16	28	14	30	14	30

Таблица 27.9.1 (окончание)

n	Значения α						n	Значения α					
	0,025		0,010		0,005			0,025		0,010		0,005	
45	16	29	15	30	14	31	73	28	45	27	46	26	47
46	16	30	15	31	14	32	74	29	45	27	47	26	48
47	17	30	16	32	15	32	75	29	46	27	48	26	49
48	17	31	16	32	15	33	76	29	47	28	48	27	49
49	18	31	16	33	16	33	77	30	47	28	49	27	50
50	18	32	17	33	16	34	78	30	48	29	49	28	50
51	19	32	17	34	16	35	79	31	48	29	50	28	51
52	19	33	18	34	17	35	80	31	49	30	50	29	51
53	19	34	18	35	17	36	81	32	49	30	51	29	52
54	20	34	19	35	18	36	82	32	50	31	51	29	53
55	20	35	19	36	18	37	83	33	50	31	52	30	53
56	21	35	19	37	18	38	84	33	51	31	53	30	54
57	21	36	20	37	19	38	85	33	52	32	53	31	54
58	22	36	20	38	19	39	86	34	52	32	54	31	55
59	22	37	21	38	20	39	87	34	53	33	54	32	55
60	22	38	21	39	20	40	88	35	53	33	55	32	56
61	23	38	21	40	21	40	89	35	54	34	55	32	57
62	23	39	22	40	21	41	90	36	54	34	56	33	57
63	24	39	22	41	21	42	91	36	55	34	57	33	58
64	24	40	23	41	22	42	92	37	55	35	57	34	58
65	25	40	23	42	22	43	93	37	56	35	58	34	59
66	25	41	24	42	23	43	94	38	56	36	58	35	59
67	26	41	24	43	23	44	95	38	57	36	59	35	60
68	26	42	24	44	23	45	96	38	58	37	59	35	61
69	26	43	25	44	24	45	97	39	58	37	60	36	61
70	27	43	25	45	24	46	98	39	59	38	60	36	62
71	27	44	26	45	25	46	99	40	59	38	61	37	62
72	28	44	26	46	25	47	100	40	60	38	62	37	63

27.10 Равномерно распределенные случайные числа

Приведенные в табл. 27.10.1 цифры можно рассматривать как реализации независимых и одинаково распределенных случайных величин, принимающих значения 0,1,2,3,4,5,6,7,8,9 с одной и той же вероятностью 0,1.

Табулированные цифры сгруппированы по две. Пары можно рассматривать как реализации независимых и одинаково распределенных случайных величин, принимающих значения от 00 до 99 с одной и той же вероятностью 0,01.

Аналогичные утверждения можно сформулировать, если цифры группировать по три, четыре и т. д.

Рассмотрим группы из k цифр как целые числа. Умножим каждое из них на 10^{-k}. Полученные числа можно считать реализациями независимых случайных величин, равномерно распределенных на отрезке $[0;1]$.

Таблица 27.10.1. Равномерно распределенные случайные числа

10 09 73 25 33	76 52 01 35 86	34 67 35 48 76	80 95 90 91 17
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 47 64
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77
66 06 57 47 17	34 07 27 68 50	36 69 73 61 70	65 81 33 98 85
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 54 06 10
88 68 54 02 00	86 50 75 84 01	36 76 66 79 51	90 36 47 64 93
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68

Таблица 27.10.1 (продолжение)

65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	53 14 03 33 40
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59
23 52 37 83 17	73 20 88 98 37	68 93 59 14 16	26 25 22 96 63
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91
35 96 31 53 07	26 89 80 93 54	33 35 13 54 62	77 97 45 00 24
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92
46 05 88 52 36	01 39 09 22 86	77 28 14 40 77	93 91 08 36 47
32 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 04 79 00 33
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56
98 08 62 48 26	45 24 02 84 04	44 99 90 88 96	39 09 47 34 07
33 18 51 62 32	41 94 15 09 49	89 43 54 85 81	88 69 54 19 94
80 95 10 04 06	96 38 27 07 74	20 15 12 33 87	25 01 62 52 98
79 75 24 91 40	71 96 12 82 96	69 86 10 25 91	74 85 22 05 39
18 63 33 25 37	98 14 50 65 71	31 01 02 46 74	05 45 56 14 27

Т а б л и ц а 27.10.1 (продолжение)

74 02 94 39 02	77 55 73 22 70	97 79 01 71 19	52 52 75 80 21
54 17 84 56 11	80 99 33 71 43	05 33 51 29 69	56 12 71 92 55
11 66 44 98 83	52 07 98 48 27	59 38 17 15 39	09 97 33 34 40
48 32 47 79 28	31 24 96 47 10	02 29 53 68 70	32 30 75 75 46
69 07 49 41 38	87 63 79 19 76	35 58 40 44 01	10 51 82 16 15
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50
73 18 95 02 07	47 67 72 62 69	62 29 06 44 64	27 12 46 70 18
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 04 75
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06
77 51 30 38 20	86 83 42 99 01	68 41 48 27 74	51 90 81 39 80
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42
33 71 34 80 07	93 58 47 28 69	51 92 66 47 21	58 30 32 98 22
85 27 48 68 93	11 30 32 92 70	28 83 43 41 37	73 51 59 04 00
84 13 38 96 40	44 03 55 21 66	73 85 27 00 91	61 22 26 05 61
56 73 21 62 34	17 39 59 61 31	10 12 39 16 22	85 49 65 75 60
65 13 85 68 06	87 64 88 52 61	34 31 36 58 61	45 87 52 10 69
38 00 10 21 76	81 71 91 17 11	71 60 29 29 37	74 21 96 40 49
37 40 29 63 97	01 30 47 75 86	56 27 11 00 86	47 32 46 26 05
97 12 54 03 48	87 08 33 14 17	21 81 53 92 50	75 23 76 20 47
21 82 64 11 34	47 14 33 40 72	64 63 88 59 02	49 13 90 64 41
73 13 54 27 42	95 71 90 90 35	85 79 47 42 96	08 78 98 81 56

Таблица 27.10.1 (окончание)

07 63 87 79 29	03 06 11 80 72	96 20 74 41 56	23 82 19 95 38
60 52 88 34 41	07 95 41 98 14	59 17 52 06 95	05 53 35 21 39
83 59 63 56 55	06 95 89 29 83	05 12 80 97 19	77 43 35 37 83
10 85 06 27 46	99 59 91 05 07	13 49 90 63 19	53 07 57 18 39
39 82 09 89 52	43 62 26 31 47	64 42 18 08 14	43 80 00 93 51
59 58 00 64 78	75 56 97 88 00	88 83 55 44 86	23 76 80 61 56
38 50 80 73 41	23 79 34 87 63	90 82 29 70 22	17 71 90 42 07
30 69 27 06 68	94 68 81 61 27	56 19 68 00 91	82 06 76 34 00
65 44 39 56 59	18 28 82 74 37	49 63 22 40 41	08 33 76 56 76
27 26 75 02 64	13 19 27 22 94	07 47 74 46 06	17 98 54 89 11
91 30 70 69 91	19 07 22 42 10	36 69 95 37 28	28 82 53 57 93
68 43 49 46 88	84 47 31 36 22	62 12 69 84 08	12 84 38 25 90
48 90 81 58 77	54 74 52 45 91	35 70 00 47 54	83 82 45 26 92
06 91 34 51 97	42 67 27 86 01	11 88 30 95 28	63 01 19 89 01
10 45 51 60 19	14 21 03 37 12	91 34 23 78 21	88 32 58 08 51
12 88 39 73 43	65 02 76 11 84	04 28 50 13 92	17 97 41 50 77
21 77 83 09 76	38 80 73 69 61	31 64 94 20 96	63 28 10 20 23
19 52 35 95 15	65 12 25 96 59	86 28 36 82 58	69 57 21 37 98
67 24 55 26 70	35 58 31 65 63	79 24 68 66 86	76 46 33 42 22
60 58 44 73 77	07 50 03 79 92	45 13 42 65 29	26 76 08 36 37
53 85 34 13 77	36 06 69 48 50	58 83 87 38 59	49 36 47 33 31
24 63 73 87 36	74 38 48 93 42	52 62 30 79 92	12 36 91 86 01
83 08 01 24 51	38 99 22 28 15	07 75 95 17 77	97 37 72 75 85
15 44 42 43 34	36 15 19 90 73	27 49 37 09 39	85 13 03 25 52
60 79 01 81 57	57 17 86 57 62	11 16 17 85 76	45 81 95 29 79
03 99 11 04 61	93 71 61 68 94	66 08 32 46 53	84 60 95 82 32
38 55 59 55 54	32 88 65 97 80	08 35 56 08 60	29 73 54 77 62
17 54 67 37 04	92 05 24 62 15	55 12 12 92 81	59 07 60 79 36
32 64 35 28 61	95 81 90 68 31	00 91 19 89 36	76 35 59 37 79
69 57 26 87 77	39 51 03 59 05	14 06 04 06 19	29 54 96 96 16

Литература

[1] *Боровков А. А.* Курс теории вероятностей. — М.: Наука, 1972. — 288 с.

[2] *Браунли К. А.* Статистическая теория и методология в науке и технике. — М.: Наука, 1977. — 408 с.

[3] *Ван дер Варден Б. Л.* Математическая статистика. — М.: Изд-во иностр. лит., 1960. — 434 с.

[4] *Гихман И. И.* Теория вероятностей и математическая статистика/ И. И. Гихман, А. В. Скороход, М. И. Ядренко. — К.: Вища шк. Головное изд-во, 1979. — 320 с.

[5] *Гнеденко Б. В.* Курс теории вероятностей. — М.: Наука, 1987. — 400 с.

[6] *Гихман И. И., Скороход А. В.* Введение в теорию случайных процессов. — М.: Наука, 1965. — 656 с.

[7] *Дороговцев А. Я.* Теория вероятностей. Сб. задач/ А. Я. Дороговцев, Д. С. Сильвестров, А. В. Скороход, М. И. Ядренко — К.: Вища шк., 1980. — 432 с.

[8] *Ежов И. И.* Элементы комбинаторики/ И. И. Ежов, А. В. Скороход, М. И. Ядренко — М.: Наука, 1977. — 80 с.

[9] *Крамер Г.* Математические методы статистики. — 2-е изд., перераб. — М.: Мир, 1975. — 648 с.

[10] *Коваленко И. Н., Гнеденко Б. В.* Теория вероятностей. — К.: Вища школа, 1990. — 328 с.

[11] *Колмогоров А. Н.* Основные понятия теории вероятностей. — М.: Наука, 1974. — 120 с.

[12] *Мешалкин Л. Д.* Сборник задач по теории вероятностей. — М.: Изд-во Моск. ун-та, 1963. — 155 с.

[13] *Розанов Ю. А.* Теория вероятностей, случайные процессы и математическая статистика — М.: Наука, 1989. — 320 с.

[14] *Скороход А. В.* Элементы теории вероятностей и случайных процессов. — К.: Вища школа, 1980. — 344 с.

[15] *Скороход А. В.* Вероятность вокруг нас. — К.: Наукова думка, 1980. — 196 с.

[16] *Тутубалин В. Н.* Теория вероятностей. — М.: Изд-во Моск. ун-та, 1972. — 230 с.

[17] *Турчин В. Н.* Теория вероятностей и математическая статистика: В 2 т. — Днепропетровск: Изд-во Днепропетр. ун-та, 1995 — 1996. — Т.1, ч.1. — 228 с.; Т.1., ч. 2. — 83 с.; Т.2. — 248 с.

[18] *Турчин В. М.* Математична статистика. — К.: Видавничий центр “Академія”, 1999. — 240 с.

[19] *Турчин В. М.* Теорія ймовірностей. Основні поняття, приклади, задачі. — К.: А.С.К., 2003. — 208 с.

[20] *Турчин В. М.* Теорія ймовірностей і математична статистика: основні поняття, приклади, задачі. — Д.: Вид-во Дніпропетр. ун-ту, 2005. — 470 с.

[21] *Турчин В. Н.* Теория вероятностей и математическая статистика. — Д.: Изд-во Днепропетр. ун-та, 2008. — 656 с.

[22] *Феллер В.* Введение в теорию вероятностей и ее приложения: В 2 т.— 3-е изд. — М.: Мир, 1984. — Т.1. — 527 с.; Т.2. — 751 с.

[23] *Ширяев А. Н.* Вероятность. — М.: Наука, 1980. — 576 с.

[24] *Ядренко М. Й.* Дискретна математика. — К.: Вид.-поліграф. центр “Експрес”, 2003. — 244 с.

[25] *David F. N., Pearson E. S.* Elementary Statistical Exercises. — Cambridge: University Press, 1961 — 108 p.

Оглавление

Навчальне видання

ВАЛЕРІЙ МИКОЛАЙОВИЧ ТУРЧИН

**ТЕОРІЯ ЙМОВІРНОСТЕЙ
І МАТЕМАТИЧНА СТАТИСТИКА**

Підручник для студентів вищих навчальних закладів

(Російською мовою)

Редактор Козаченко Ю.В.
Художник Ткаченко К.Д.
Оригінал-макет Турчин В.М.

Підписано до друку 15.06.2018. Формат $60{\times}90\frac{1}{16}$. Папір офсетний.
Друк офсетний. Гарнітура Computer modern. Ум. др. арк. 47.
Наклад 500 прим. 1-й запуск —1—35 прим. Зам. № 142.

Видавництво і друкарня ПП “Ліра ЛТД”.

49107, м. Дніпро, вул. Наукова, 5.
Свідоцтво про внесення до Держреєстру
ДК № 6042 від 26.02.2018.

ТУРЧИН
Валерий Николаевич

Автор 122 научных и научно-методических работ, 16 учебников и учебных пособий по теории вероятностей и математической статистике. Заведующий кафедрой статистики и теории вероятностей Днепропетровского национального университета имени Олеся Гончара.