

Б. М. ЩИГОЛЕВ

МАТЕМАТИЧЕСКАЯ ОБРАБОТКА НАБЛЮДЕНИЙ

ИЗДАНИЕ ВТОРОЕ, СТЕРЕОТИПНОЕ

*Допущено Министерством
высшего и среднего специального образования РСФСР
в качестве учебника для механико-математических
и физико-математических факультетов университетов*



ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1962

Щеголев Борис Михайлович
Математическая обработка наблюдений

Редактор *И. Т. Резниковский*

Техн. редактор *С. С. Гаврилов*

Корректор *Г. Г. Желтова*

Печать с матриц. Подписано к печати 1/II 1962 г. Бумага 60×90^{1/16}.
Физ. печ. л. 21,50. Условн. печ. л. 21,50. Уч.-изд. л. 22,28. Тираж 22000 экз.
Цена книги 82 коп. Заказ 95.

Государственное издательство физико-математической литературы
Москва, В-71, Ленинский проспект, 15.

1-я типография Трансжелдориздата МПС.
Б. Переяславская ул., 46.

СОДЕРЖАНИЕ

Предисловие	7
Введение	9

ЧАСТЬ I

ДЕЙСТВИЯ С ПРИБЛИЖЕННЫМИ ЧИСЛАМИ

Глава 1. Оценки ошибок приближенных чисел	12
§ 1. Основные задачи теории приближенных вычислений	12
§ 2. Точная ошибка приближенного числа	13
§ 3. Предельная абсолютная погрешность	14
§ 4. Предельная относительная погрешность	17
§ 5. Оценка ошибки по числу верных знаков	19
Глава 2. Погрешности результатов основных арифметических действий	22
§ 6. Сложение	22
§ 7. Статистическая оценка ошибки суммы	24
§ 8. Вычитание близких чисел	26
§ 9. Умножение	28
§ 10. Деление	31
Глава 3. Оценка ошибки функции приближенных аргументов	34
§ 11. Предельные погрешности функции одного независимого переменного	34
§ 12. Погрешности простейших элементарных функций	35
§ 13. Погрешность функции нескольких аргументов	42
§ 14. Понятие об обратной задаче теории приближенных вычислений	48

ЧАСТЬ II

ТОЧЕЧНОЕ ИНТЕРПОЛИРОВАНИЕ

Глава 4. Общие сведения	51
§ 15. Приближение табличных функций; понятие о точечной интерполяции	51
§ 16. Теорема существования интерполяционного полинома	55
§ 17. Интерполяционный полином Лагранжа	57
§ 18. Оценка ошибки точечной интерполяции	60
Глава 5. Интерполирование по таблице с переменным шагом	66
§ 19. Разделенные разности табличной функции	66
§ 20. Способ построения разностных интерполяционных формул	68
§ 21. Интерполяционная формула Ньютона для таблицы с переменным шагом	71

Глава 6. Интерполирование по таблице с постоянным шагом . .	75
§ 22. Обыкновенные и центральные разности табличной функции с постоянным шагом	75
§ 23. Основные свойства обыкновенных разностей	79
§ 24. Способ построения интерполяционных формул для таблиц с постоянным шагом	84
§ 25. Формулы Ньютона для интерполяции вперед и назад	86
§ 26. Формула Стирлинга	92
§ 27. Формула Бесселя (два варианта)	95
§ 28. Общие замечания о применении разностных интерполяционных формул	100

ЧАСТЬ III

СВЕДЕНИЯ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Глава 7. Случайные события; основные понятия и теоремы . .	103
§ 29. Случайные явления	103
§ 30. Классическое определение вероятности	106
§ 31. Примеры вычисления вероятности	107
§ 32. Теорема сложения вероятностей	109
§ 33. Теорема умножения вероятностей	111
§ 34. Полная вероятность; гипотезы	114
§ 35. Априорные и апостериорные вероятности гипотез	116
Глава 8. Задача о повторении испытаний	120
§ 36. Формулировка задачи и вывод основной формулы	120
§ 37. Распределение вероятностей чисел повторений события	122
§ 38. Приближенная формула Лапласа для вычисления вероятности числа повторений события	129
§ 39. Приближенная кривая распределения вероятностей	133
§ 40. Распределение Пуассона (закон редких событий)	134
Глава 9. Дискретные случайные величины	135
§ 41. Случайные величины	135
§ 42. Математическое ожидание дискретной случайной величины	137
§ 43. Теоремы сложения и умножения математических ожиданий	140
§ 44. Дисперсия случайной величины; свойства дисперсии	142
§ 45. Математическое ожидание и дисперсия числа повторений	145
Глава 10. Закон больших чисел	148
§ 46. Лемма Чебышева — Маркова	148
§ 47. Теорема Я. Бернулли	150
§ 48. Предельная теорема Лапласа	154
§ 49. Неравенство и теорема Чебышева	158
§ 50. Замечания о законе больших чисел. Статистические вероятности	162
Глава 11. Непрерывные случайные величины	165
§ 51. Функция распределения непрерывной случайной величины	165
§ 52. Плотность вероятности	167
§ 53. Математическое ожидание, дисперсия и моменты	169
§ 54. Равномерное распределение вероятностей	172
§ 55. Формулировка теоремы Ляпунова. Нормальное распределение вероятностей	174
§ 56. Приближенный вывод нормального закона	175

§ 57. Параметры нормального закона. Кривая Гаусса	178
§ 58. Функция нормального распределения. Вычисление вероятностей	181
§ 59. Моменты нормального распределения	184
§ 60. Понятие о распределениях, отличных от нормального	186
Глава 12. Распределение совокупности двух непрерывных случайных величин	193
§ 61. Плотность вероятности совокупности двух величин	193
§ 62. Условные плотности вероятности	195
§ 63. Нормальное распределение двух случайных величин	198
§ 64. Плотность вероятности нормального распределения	201

ЧАСТЬ IV

ОСНОВЫ ТЕОРИИ СЛУЧАЙНЫХ ОШИБОК ИЗМЕРЕНИЙ (СПОСОБ НАИМЕНЬШИХ КВАДРАТОВ)

Глава 13. Общие сведения об ошибках измерений	209
§ 65. Виды ошибок измерений	209
§ 66. Основная гипотеза теории случайных ошибок. Способы оценки ошибок	212
Глава 14. Обработка равнооточных измерений определенной величины	215
§ 67. Задача обработки измерений определенной величины	215
§ 68. Наиболее вероятное значение измеряемой величины. Способ наименьших квадратов	215
§ 69. Средняя квадратичная ошибка среднего арифметического	218
§ 70. Наиболее вероятное значение средней квадратичной ошибки одного измерения	219
§ 71. Второй вывод приближенного значения измеряемой величины и приближенного значения средней квадратичной ошибки одного измерения	222
§ 72. Пример и схема обработки равнооточных измерений одной величины	224
Глава 15. Обработка неравнооточных измерений определенной величины	227
§ 73. Понятие о неравнооточных измерениях. Веса измерений	227
§ 74. Наиболее вероятное значение измеряемой величины	229
§ 75. Средняя квадратичная ошибка среднего весового	231
§ 76. Наиболее вероятное значение средней квадратичной ошибки измерения с весом единица	232
§ 77. Пример и схема обработки неравнооточных измерений определенной величины	236
Глава 16. Определение нескольких неизвестных из уравнений по способу наименьших квадратов	239
§ 78. Условные и нормальные уравнения. Принцип Лежандра	239
§ 79. Вероятностный смысл принципа Лежандра	242
§ 80. Обобщение принципа Лежандра на неравнооточные условные уравнения. Приведение неравнооточных уравнений к равнооточным	244
§ 81. Приведение нелинейных условных уравнений к линейному виду	246
§ 82. Линейные условные и нормальные уравнения	250

§ 83.	Контроль составления нормальных уравнений	254
§ 84.	Решение системы линейных нормальных уравнений	258
§ 85.	Вычисление весов неизвестных	264
§ 86.	Приближенное значение средней квадратичной ошибки на единицу веса. Средние квадратичные ошибки неизвестных . .	270
§ 87.	Пример и схема решения системы линейных условных уравнений	273
Глава 17.	Эмпирические формулы	277
§ 88.	Постановка задачи	277
§ 89.	Выбор типа формулы	278
§ 90.	Определение значений параметров по принципу Лежандра . .	281
§ 91.	Проверка эмпирической формулы	283
§ 92.	Пример вывода эмпирической формулы	284
ЧАСТЬ V		
ОБРАБОТКА СТАТИСТИЧЕСКОГО МАТЕРИАЛА		
Глава 13.	Обработка одномерной статистической совокупности .	286
§ 93.	Статистические совокупности	286
§ 94.	Дискретное эмпирическое распределение и его числовые характеристики	288
§ 95.	Непрерывное эмпирическое распределение	293
§ 96.	Сравнение эмпирического распределения с теоретическим . .	300
§ 97.	Доверительные вероятности и доверительные границы . . .	305
§ 98.	Графическое представление эмпирической совокупности . .	306
§ 99.	Средние ошибки параметров выборочной совокупности . . .	311
Глава 19.	Элементарная теория корреляции двух величин	313
§ 100.	Эмпирическое распределение двух случайных величин . . .	313
§ 101.	Корреляционная зависимость. Задачи теории корреляции . .	315
§ 102.	Вывод линейной эмпирической формулы	317
§ 103.	Вывод линейных уравнений регрессии	320
§ 104.	Коэффициент корреляции	322
§ 105.	Средние ошибки уравнений регрессии; границы значений коэффициента корреляции	325
§ 106.	Средние ошибки выборочных коэффициентов корреляции и регрессии	327
§ 107.	Вероятностное значение элементарной теории корреляции . .	328
§ 108.	Пример и схема исследования корреляции при большом числе наблюдений	329
§ 109.	Пример исследования корреляции по малому числу наблюдений	337
Литература		339
Приложение		341

ПРЕДИСЛОВИЕ

Эта книга написана по программе курса «Математическая обработка наблюдений» для студентов астрономической специальности механико-математических и физико-математических факультетов университетов. При составлении книги был использован опыт чтения курса в МГУ. В течении семестра студенты обычно успевают прослушать материал, включенный в программу. В книгу включен также дополнительный материал, не входящий в программу; его можно предложить студентам для самостоятельного изучения или включить в курс, если число часов может быть увеличено.

Программа курса несколько шире его названия, так как в него включены не только задачи, связанные с обработкой наблюдений в тесном смысле, но и задачи приближенных вычислений, которые не всегда оказываются задачами обработки наблюдений, хотя их и приходится решать именно в связи с ней. Достаточно указать, например, на точечное интерполирование по таблицам функции, если значения функции вычислены по ее определению (например, с помощью ряда). Такого рода задачи также включены в книгу.

Название «Математическая обработка наблюдений» в подобном расширенном смысле укоренилось, и вряд ли есть надобность менять его.

Задачи, рассмотренные в этой книге применительно к потребностям астрономии, приходится весьма часто решать в самых разнообразных отделах естествознания и техники. Поэтому автор надеется, что книга или по крайней мере некоторые ее части будут полезны не только астрономам.

Книга подразделена на части и главы.

Нумерация глав и параграфов в книге сквозная. Для формул выбрана двойная нумерация: формула (16.10) будет десятой по порядку в главе 16,

В конце книги приведен небольшой список литературы — в основном учебные пособия и руководства. Список составлен применительно к частям книги.

Автор считает своей приятной обязанностью поблагодарить сотрудников кафедры небесной механики МГУ Е. М. Славцеву и А. И. Рыбакова за помощь при подготовке рукописи к печати.

Автор будет очень благодарен читателям за указания на недостатки книги; просьба посылать такие указания по адресу: Москва, В-234, МГУ, Астрономическая обсерватория.

Автор искренно благодарит В. А. Чурикова (г. Тернополь) за сообщение об опечатках, обнаруженных им, которое полностью использовано при подготовке второго издания книги.

Б. М. Щиголев

ВВЕДЕНИЕ

Основой всего естествознания являются *наблюдения и эксперименты*. Особое значение имеют наблюдения и эксперименты, дающие числа — *результаты измерений*. Надлежащая обработка таких чисел приводит к теоретическому осмысливанию результатов наблюдений и к конечной цели естествознания — установлению законов явлений, позволяющих предсказывать ход интересующих нас явлений на будущее.

Все результаты измерений содержат ошибки различного происхождения. Поэтому результаты вычислений с числами, результатами измерений, также содержат ошибки. Очень существенно для практики уметь оценивать как ошибки самих результатов измерений, так и результатов действий над ними, ибо только в этом случае можно с достаточной уверенностью пользоваться выводами из наблюдений. Не менее важна такая организация вычислений и наблюдений, которая обеспечивает по возможности малую ошибку результата. Все сведения о линейных размерах в солнечной системе, в Галактике и т. д. опираются в конечном счете на прямые измерения сравнительно малых величин на поверхности Земли. Эти величины содержат ошибки. Чтобы получить сведения о размерах в солнечной системе, и тем более в Галактике, приходится эти величины умножать на большие числа; ошибки измерений при этом также умножаются и приводят к большим ошибкам результатов. Из этих кратких замечаний ясно, что обработка результатов наблюдений не может выполняться любым способом. Чтобы результаты содержали возможно меньшие ошибки, должны быть разработаны как методы оценок ошибок, так и методы вычислительной работы, обеспечивающие возможно более точные результаты.

Выше шла речь об обработке наблюдений в узком смысле, когда производятся действия над числами, непосредственно получаемыми из наблюдений. Однако при построении теории явления и вычисления величин, непосредственно не наблюдаемых, но выводимых путем обработки наблюдений, приходится пользоваться разными математическими приемами, в частности, широко использовать различные функциональные зависимости,

Как известно, функция может быть определена разными способами. В простейшем случае указывается, какие арифметические действия нужно произвести над значением аргумента (или аргументов), чтобы получить значение функции (полиномы, рациональные функции и т. п.). Но функция может быть определена и так, что из определения не видно, как вычислять ее значения (например синус дуги); в подобных случаях определение используется для вывода свойств функции, позволяющих построить бесконечный ряд, который можно считать другим определением функции. Функция может быть определена также интегралом, дифференциальным уравнением и т. п.; каждый из этих способов задания тоже не дает прямых указаний на способ вычисления значения функции по значению аргумента. В подобных случаях приходится либо подыскивать бесконечный ряд, либо прибегать к численным методам решения, которые дают функцию в виде таблицы. В тех случаях, когда функция определена сходящимся бесконечным рядом (обычно степенным), этот ряд используется для составления таблицы значений функции, если эта функция часто встречается в практических задачах. Составление таблицы всегда представляет приближенную операцию, поскольку бесконечный ряд всегда приходится обрывать на том или ином члене. Хотя табличные значения функций не получаются в результате измерений, но и они подобно результатам измерений содержат неизбежные ошибки. Эти ошибки тоже должны оцениваться; необходимо оценивать также и ошибки действий над табличными значениями функций.

Таким образом, есть общие пункты в задачах, связанных с измерениями, и в задачах, в которых используются таблицы функций. Оба типа задач связаны и непосредственно тем, что весьма часто результат измерения является аргументом табличной функции. По этим причинам задачи о действиях над табличными функциями обычно входят в курсы обработки измерений.

Среди ошибок измерений видное место занимают *случайные ошибки*, т. е. такие, величины которых не могут быть указаны до наблюдений. Надо, впрочем, заметить, что они не могут быть указаны и после наблюдений, так как наличие случайных ошибок лишает нас возможности определить точное значение измеряемой величины. Для обработки измерений, содержащих случайные ошибки, должен быть использован аппарат *теории вероятностей*; этот же аппарат необходим и в статистических работах. Чтобы дать возможность студентам пользоваться только одним учебником, с одной стороны, и обеспечить единую систему обозначений, терминологии и т. п., с другой, в курс «Математическая обработка наблюдений» включаются основные сведения по теории вероятностей.

Задачи обработки наблюдений и отчасти связанные с ними задачи приближения функции возникли давно и в первую очередь в связи

с задачами астрономии. Впервые вполне четко они были поставлены в работах французского математика Лежандра (1752—1833) и немецкого математика Гаусса (1777—1855).

В разработке теории ошибок и теории приближения функций видную роль играли русские математики П. Л. Чебышев (1821—1894), А. А. Марков (1856—1922), А. М. Ляпунов (1857—1918). Представители советской школы теории вероятностей и конструктивной теории функций С. Н. Бернштейн, Б. В. Гнеденко, В. Л. Гончаров, А. Н. Колмогоров, В. И. Романовский, А. Я. Хинчин и др. вели и ведут сейчас интенсивную работу в этих областях.

ЧАСТЬ I

ДЕЙСТВИЯ С ПРИБЛИЖЕННЫМИ ЧИСЛАМИ

ГЛАВА I

ОЦЕНКИ ОШИБОК ПРИБЛИЖЕННЫХ ЧИСЕЛ

§ 1. Основные задачи теории приближенных вычислений

В естествознании очень редко приходится иметь дело с *точными* числами. Если астроном исследует движения в тройной звездной системе, то число 3, конечно, является точным. Когда геофизик занимается строением снежинок, то число лучей в каждой отдельной снежинке он тоже определяет точно. Однако число подобных примеров в каждой области естествознания весьма ограничено.

Обычные же результаты измерений всегда являются *приближенными* прежде всего вследствие ограниченной точности измерительных приборов.

Действительно, каждый измерительный инструмент имеет шкалу, на которой промежутки между делениями не могут быть как угодно малыми. Иногда говорят о «пороге чувствительности прибора», подразумевая под этим наименьшую величину, изменение на которую еще регистрируется прибором. Если, например, круг, предназначенный для измерения углов, имеет деления через $10'$ и нониус, дающий $1'$, то можно считать порогом чувствительности этого прибора $1'$, поскольку изменение угла на $1'$ будет отмечено прибором, а изменение на величину, меньшую $1'$, может быть зарегистрировано, но не поддается точному определению. Если измерительный указатель инструмента попадает между двумя делениями, то десятые доли промежутка еще могут быть отсчитаны либо на глаз, либо с помощью нониуса. Таким образом, можно только утверждать, что ошибки измерения меньше половины или одной десятой промежутка между делениями шкалы. Это утверждение будет верным, если нет других источников ошибок, кроме ошибки, зависящей от ограниченной точности измерительного инструмента или прибора.

Есть довольно много измерений, особенно лабораторных, при которых можно указать верхнюю границу абсолютной величины (модуля) ошибки измерения. Ниже мы будем рассматривать именно такие случаи. Впрочем, независимо от фактической возможности

надежно определить эту верхнюю границу, всегда можно допустить ее существование. Действительно, поскольку точное числовое значение измеряемой величины существует как объективная реальность, не зависящая от нас, а измерение дает, вообще говоря, какое-то другое значение, то ошибка всегда ограничена. Мы будем в этом разделе книги предполагать, что имеем дело с приближенными числами, содержащими ошибки любого происхождения, но такими, что можно указать верхнюю границу модуля ошибки.

Первой задачей теории приближенных вычислений будет установление способов оценки ошибок.

Почти всегда приходится производить арифметические (или иные) действия над приближенными числами. Естественно, что результаты таких действий будут тоже приближенными. Отсюда возникает вторая задача — оценить ошибку операции над приближенными числами, если известны оценки ошибок исходных чисел. Эту задачу можно назвать *прямой* задачей теории. Если такая задача решена в буквенном виде, то могут быть поставлены еще две очень важные задачи:

1) *обратная* задача — определение необходимой точности заданных чисел для обеспечения заданной точности результата действий;

2) выяснение условий измерений или вычислений, при которых погрешность результата действий будет по возможности меньшей. При этом под условиями подразумевается либо выбор положений, в которых производятся измерения, либо выбор формул для вычислений.

§ 2. Точная ошибка приближенного числа

Предположим, что некоторая величина (например, угол) имеет определенное числовое значение A , остающееся неизменным во время процесса измерения. Предположим также, что сделанное измерение этой величины дает значение a .

Точной ошибкой Δ_a приближенного числа a назовем разность между точным и приближенным значениями:

$$A - a = \Delta_a. \quad (1.1)$$

Введенное определение удобно тем, что понятие точной ошибки совпадает с понятием *поправки*, ибо

$$A = a + \Delta_a. \quad (1.2)$$

т. е. точная ошибка есть число, которое нужно прибавить к приближенному, чтобы получить точное значение *).

*) Вместо точной ошибки иногда вводят понятие *абсолютной ошибки*. Абсолютной ошибкой $|\Delta|_a$ приближенного числа a называют модуль разности между точным и приближенным значениями величины:

$$|\Delta|_a = |A - a|.$$

Мы не будем пользоваться этим понятием.

Введение понятия «точная ошибка» имеет только теоретическое значение, так как эта ошибка в обычных задачах не может быть фактически определена. Можно отметить только исключительные случаи, когда для исследования точности измерений некоторым методом или с помощью какого-либо прибора производят то же измерение с точностью, значительно большей, чем исследуемое измерение (например, прецизионными приборами). Хотя это последнее измерение дает тоже ошибку, но в этом случае можно принять, что второе измерение дает формально точное значение и определить «точную» ошибку первого измерения, ограничиваясь некоторым количеством знаков.

Пусть, например, измерение угла теодолитом с точностью до $1'$ дало $38^\circ 43'$. Измерим тот же угол универсалом с предельной погрешностью $.5''$; предположим, результат получился $38^\circ 43' 25''$. Считая формально второе значение точным, можно сказать, что точная ошибка равна $0',4$, хотя в действительности это только приближенное значение точной ошибки.

Мы будем в дальнейшем полагать, что величина $|\Delta_a|$ настолько мала по сравнению с $|a|$, что степенями Δ_a выше первой можно пренебрегать. Если будут рассматриваться одновременно несколько приближенных чисел a, b, \dots , то произведениями вида $\Delta_a \Delta_b$ мы также будем пренебрегать.

§ 3. Предельная абсолютная погрешность

Как было сказано в § 1, в исключительных случаях можно определить точную ошибку приближенно в том смысле, что ошибка этого определения гораздо меньше по модулю, чем ошибка основного измерения. Во многих случаях легче определить верхнюю границу модуля точной ошибки.

Назовем *предельной абсолютной погрешностью* ϵ_a приближенного числа a наименьшее положительное число, содержащее одну-две значащие цифры, которое больше или равно модулю точной ошибки, т. е.

$$\epsilon_a \geq |\Delta_a|. \quad (1.3)$$

Это определение требует пояснения. По большей части нетрудно найти положительное число, заведомо большее модуля точной ошибки; его и можно было бы принять за предельную абсолютную погрешность. Если таких чисел можно найти несколько, то по определению, должно быть выбрано наименьшее. Иногда бывает выгодно заменить принятое значение предельной абсолютной погрешности другим, более простым, т. е. имеющим меньше значащих цифр. В таких случаях можно только увеличивать принятое первоначально значение.

Если известна предельная абсолютная погрешность, то для точного значения A можно записать очевидные неравенства

$$a - \varepsilon_a \leq A \leq a + \varepsilon_a *). \quad (1.4)$$

Отметим частный случай, когда предельная абсолютная погрешность числа определяется его записью в десятичной системе. Каждое приближенное число, записанное в десятичной системе счисления, имеет ограниченное число значащих цифр, зависящее от точности измерений или вычислений. Если, например, измеряется длина линейкой с делениями через миллиметр и нониусом, дающим десятые доли, то длина в сантиметрах будет содержать целые сантиметры, десятые и сотые доли сантиметра. По характеру процесса такого измерения можно утверждать, что модуль ошибки измерения меньше 0,01 см, если при пользовании нониусом всегда отмечают меньшее (или большее) деление нониуса, когда конец отрезка не совпадает точно ни с одним делением нониуса (обычный случай). При таком способе измерения можно указать и знак точной ошибки, ибо приближенное значение наверно меньше (или, соответственно, больше) точного значения. Этот способ измерения можно назвать *измерением без округления* по аналогии с соответствующей вычислительной операцией.

Измерение с округлением в нашем примере будет заключаться в том, что ищут деление нониуса, ближайшее к концу измеряемого отрезка, и записывают соответствующую цифру сотых долей. Точная ошибка равняется расстоянию принятого деления нониуса от конца отрезка; при измерении с округлением модуль этого расстояния не превышает половины расстояния между делениями нониуса, которое равняется 0,01 см.

Таким образом, если результат измерения ограничивается сотыми долями сантиметра и измерение производится без округления, то за предельную абсолютную погрешность измерения следует принять 0,01 см, если же при измерении делается округление, то за предельную абсолютную погрешность нужно принять $\frac{1}{2} \cdot 0,01$ см.

Эти выводы справедливы для любых приближенных чисел.

Если приближенное число записано в десятичной системе счисления, то по общепринятому соглашению предельная абсолютная погрешность принимается равной единице последнего знака, если число получено без округления, и половине единицы последнего знака, если число получено с округлением. Последним знаком считается первый справа. Если имеются другие сведения о предельной

*) Заметим, что предлагаемая иногда условная запись

$$A = a \pm \varepsilon_a$$

нецелесообразна, так как подобная запись давно введена в теории случайных ошибок и имеет другой смысл.

абсолютной погрешности, то они должны быть указаны и при операциях надо руководствоваться ими, а не соглашением.

Указанное выше соглашение используется не только в случаях, когда приближенное число получено в результате измерения, но также и в тех случаях, когда приближенное число получается вычислением, и у него отбрасывают несколько знаков справа. Простейший пример такого рода — обращение простой дроби в десятичную с сохранением определенного числа знаков после запятой.

Если знаки просто отбрасываются (в случае десятичной дроби) или заменяются нулями (в целом числе), то по указанному правилу за предельную абсолютную погрешность принимается единица последнего знака. Под «последним» подразумевается знак, за которым направо ничего нет, как в десятичной дроби, либо стоят нули, если число целое *). При простом отбрасывании получаем приближенное значение с точностью до единицы последнего знака с *недостатком*, т. е. меньшее, чем исходное. Можно бы было условиться всегда увеличивать последнюю значащую цифру на единицу, тогда предельная абсолютная погрешность будет тоже равна единице последнего знака, но приближенное значение получалось бы с *избытком*, т. е. больше заданного числа.

Пример 1. Известно, что можно принять $\pi = 3,1416$. Пусть в организуемых вычислениях достаточно иметь два знака после запятой; тогда принимаем $\pi = 3,14$ (или $\pi = 3,15$) и $\epsilon_\pi = 0,01$. Если неизвестно происхождение приближенного значения этого числа, то можно только утверждать, что

$$3,13 < \pi < 3,15;$$

если же известно, что число с недостатком, то левая граница будет 3,14.

Пример 2. Экваториальный радиус Земли R приближенно равен 6384 км. Предположим, что мы хотим заменить это число числом с двумя значащими цифрами. Если только отбрасывать знаки, которые мы считаем лишними, то можно взять $R = 6300$ км с недостатком или $R = 6400$ км с избытком; в обоих случаях $\epsilon_R = 100$ км.

*) Следует иметь в виду возможность некоторой путаницы в способе определения предельной погрешности. Поясним это на простом примере.

Пусть длина, измеренная прецизионным прибором и выраженная в сантиметрах, равна 3,5003. Если достаточно иметь три знака после запятой, то следует принять длину равной 3,500 см, и ни одного нуля нельзя отбросить (в отличие от правил точной арифметики). Действительно, по указанному выше соглашению при нашей записи предельная погрешность равна $\frac{1}{2} \cdot 10^{-3}$ см,

при отбрасывании же одного нуля она была бы равна $\frac{1}{2} \cdot 10^{-2}$ см. Здесь и нули справа — значащие цифры.

Пусть та же длина измерена линейкой, разделенной на мм. тогда получится 3,5 см с предельной погрешностью $\frac{1}{2} \cdot 10^{-1}$ см. По формальным правилам арифметики в этом случае тоже можно написать 3,500 см, но это не изменит оценки ошибки, если только известно, как производилось измерение.

Простое отбрасывание знаков удобно тем, что известен знак ошибки, но при этом не получается приближенного значения, возможно более близкого к точному. С этой точки зрения удобнее применять округление приближенного числа, эквивалентное округлению при измерении. Правило округления общеизвестно и не нуждается в объяснении:

если первая отбрасываемая цифра меньше пяти, то последняя сохраняемая цифра не изменяется; если первая отбрасываемая цифра больше пяти, то последняя сохраняемая цифра увеличивается на единицу;

если отбрасывается одна цифра 5, и следующие цифры неизвестны, то чаще всего применяется соглашение — последнюю сохраняемую цифру не менять, если она четная, и увеличить на единицу, если она нечетная (округление на четную цифру).

Ошибка при округлении оценивается по правилу:

предельная абсолютная погрешность округленного приближения равна половине единицы последнего знака.

Пример 1. Для числа π получаем при округлении приближенное значение: $\pi = 3,14$;

$$\epsilon_{\pi} = \frac{1}{2} 0,01.$$

Пример 2. Для экваториального радиуса Земли имеем:

$$R = 6400 \text{ км}, \quad \epsilon_R = 50 \text{ км}.$$

Некоторое неудобство округления состоит в том, что знак ошибки приближенного числа известен только при получении его. При последующих операциях, когда дано только приближенное число и известно, что оно получено с округлением, можно только утверждать, что

$$a - \epsilon_a \leq A \leq a + \epsilon_a \quad \epsilon_a = \frac{1}{2} \cdot 10^p,$$

где 10^p — единица первой значащей цифры справа.

§ 4. Предельная относительная погрешность

Введенная выше оценка ошибки с помощью предельной абсолютной погрешности позволяет указать границы, в которых находится точное значение числа, но недостаточно характеризует качество измерений, если приближенное число есть результат измерений или вычислений, произведенных по результатам измерений. Чтобы измерение или вычисление считалось хорошим, важна не только малость ϵ_a , но и малость ϵ_a по сравнению с a . Чтобы это пояснить, положим

$$\epsilon_a = \delta \cdot |a|,$$

где δ — положительное число. Тогда можно написать

$$a(1 - \delta) \leq A \leq a(1 + \delta), \text{ если } a > 0,$$

и

$$a(1 + \delta) \leq A \leq a(1 - \delta), \text{ если } a < 0.$$

Из этих неравенств видно, например, что приближенное значение a не определит с уверенностью даже знака числа A , когда $\delta > 1$, так как правая и левая части неравенств в этом случае имеют разные знаки, и действительное значение может иметь как знак «+», так и «-». Отсюда видно, что для характеристики качества приближения важна не только величина предельной абсолютной погрешности, но и ее отношение к величине a . Определение экваториального радиуса Земли при $\epsilon_R = 1$ м можно считать превосходным, измерение размеров даже большой аудитории с такой же предельной абсолютной погрешностью было бы плохим. Этот пример подсказывает еще одно соображение, по которому следует ввести дополнительную оценку погрешности. В обычных задачах a и ϵ_a — числа именованные, и потому числовое значение ϵ_a зависит от выбора единиц измерения, т. е. не является достаточно полной характеристикой неточности. Кроме того, значения предельных абсолютных погрешностей разных измерений не могут быть сравниваемы в тех случаях, когда измеряются величины разных наименований, например, вес и длина и т. п.

В связи с этими соображениями полезно ввести понятие предельной относительной погрешности. *Предельной относительной погрешностью* δ_a приближенного числа a называется отношение его предельной абсолютной погрешности ϵ_a к модулю числа a *):

$$\delta_a = \frac{\epsilon_a}{|a|}. \quad (1.5)$$

Если известно ϵ_a , то можно определить δ_a , и наоборот.

*) Более логично было бы назвать предельной относительной погрешностью отношение ϵ_a к $|A|$. Но эту величину практически определить не удастся, поскольку значение A неизвестно. Сравним две величины:

$$\delta'_a = \frac{\epsilon_a}{A}, \quad \delta_a = \frac{\epsilon_a}{a}.$$

Предполагая для упрощения, что $a > 0$, $A > 0$, и используя определение точной ошибки, получаем

$$\begin{aligned} A &= a + \Delta_a, \\ \delta'_a - \delta_a &= \frac{\epsilon_a}{a} \left(\frac{a}{A} - 1 \right) = \frac{\epsilon_a}{a} \left(\frac{a}{a + \Delta_a} - 1 \right) = \\ &= \frac{\epsilon_a}{a} \left[\frac{1}{\left(1 + \frac{\Delta_a}{a} \right)} - 1 \right] = \frac{\epsilon_a}{a} \left[-\frac{\Delta_a}{a} + \left(\frac{\Delta_a}{a} \right)^2 - \dots \right]. \end{aligned}$$

Величину δ_a довольно часто выражают в процентах или в промилле, т. е. в тысячных долях (знак ‰). При вычислении предельной относительной погрешности стараются получить число с малым количеством значащих цифр. Вычисление делается, как правило, в уме, при этом δ_a упрощается так, чтобы получилось число большее, чем то, которое должно быть по определению. Например, если $\pi = 3,14$ с округлением, то

$$\epsilon_\pi = \frac{1}{2} \cdot 10^{-2}, \quad \delta_\pi = \frac{0,5 \cdot 10^{-2}}{3,14} < \frac{1}{6} \cdot 10^{-2} < \frac{1}{5} \cdot 10^{-2};$$

за предельную относительную погрешность мы примем

$$\delta_\pi = 2 \cdot 10^{-3} = 2\text{‰}.$$

§ 5. Оценка ошибки по числу верных знаков

В практике широко используется связь между числом верных значащих цифр в приближенном числе и предельной относительной погрешностью.

Пусть положительное приближенное число a содержит s верных десятичных знаков. Тогда его десятичное разложение имеет вид

$$a = n_1 \cdot 10^r + n_2 \cdot 10^{r-1} + \dots + n_s 10^p,$$

где n_1, n_2, \dots, n_s — десятичные цифры, причем $n_1 \neq 0$. Целые числа r, p ($r > p$) и целое положительное число s связаны очевидным равенством $p - r = 1 - s$.

Пусть число a получено с округлением. Тогда

$$\epsilon_a = \frac{1}{2} \cdot 10^p.$$

По определению предельной относительной погрешности (1.5)

$$\delta_a = \frac{1}{2} \cdot \frac{10^p}{a}.$$

Подставим теперь в выражение для δ_a вместо a его десятичное разложение, ограничиваясь лишь первым членом; очевидно, от этого значение δ_a может только увеличиться:

$$\delta_a < \frac{1}{2} \cdot \frac{10^p}{n_1 10^r} = \frac{1}{2} \cdot \frac{1}{n_1} \cdot 10^{1-s}.$$

Учитывая, что $\frac{|\Delta_a|}{a} \leq \delta_a$, находим

$$|\delta'_a - \delta_a| \leq \delta_a [\delta_a + \delta_a^2 + \dots].$$

Таким образом, $\delta'_a - \delta_a$ есть величина второго порядка малости, если δ_a — первого порядка. Следовательно, практически несущественно, какой из величин δ_a или δ'_a характеризовать относительную погрешность числа a .

Последнее выражение можно принять за предельную относительную погрешность

$$\delta_a = \frac{10}{2n_1} \cdot 10^{-s}. \quad (1.6)$$

Здесь s — число верных значащих цифр в приближенном числе, n_1 — первая цифра числа. Следует обратить внимание на то, что согласно полученной формуле предельная относительная погрешность зависит только от числа верных знаков и n_1 , но не зависит от места запятой.

Приведем таблицу значений δ_a для разных n_1 и s (δ_a выражено в процентах):

$n_1 \backslash s$	1	2	3	4	5
1	50	5,0	0,50	0,050	0,0050
2	25	2,5	0,25	0,025	0,0025
3	17	1,7	0,17	0,017	0,0017
4	13	1,3	0,13	0,013	0,0013
5	10	1,0	0,10	0,010	0,0010
6	8,4	0,84	0,084	0,0084	0,00084
7	7,2	0,72	0,072	0,0072	0,00072
8	6,7	0,67	0,067	0,0067	0,00067
9	5,6	0,56	0,056	0,0056	0,00056

Для примерной оценки относительной погрешности можно взять среднее значение первой цифры, т. е. положить $n_1 = 5$. Тогда окажется, что предельная относительная погрешность числа, имеющего один верный знак, есть число порядка 10%, в случае двух верных знаков — 1%, в случае трех верных знаков — 0,1% и т. д. Следует отметить, что во многих прикладных задачах оказывается достаточной предельная относительная погрешность порядка десятых долей процента; такие вычисления достаточно вести с тремя значащими цифрами.

Только что мы по числу верных знаков определяли предельную относительную погрешность. Легко решить и обратную задачу — по заданной предельной относительной погрешности определить необходимое число верных знаков. Пусть требуется определить s так, чтобы $\delta_a = 10^{-q}$. Тогда, согласно формуле (1.6)

$$\frac{10}{2n_1} \cdot 10^{-s} \leq 10^{-q}, \quad 10^q \leq 10^s \cdot \frac{2n_1}{10}, \quad 10^s \geq 10^q \cdot \frac{10}{2n_1}.$$

Заменяя n_1 средним числом 5 (n_1 может принимать любые целые значения от 1 до 9), получим $10^s \geq 10^q$, $s \geq q$. Следовательно, в среднем число верных знаков должно быть равно модулю показателя степени при 10 в заданной величине δ_a . Например, если требуется δ_a порядка 1%, то число должно иметь не меньше двух верных знаков.

Получим более надежную (не среднюю) оценку. Для этого используем десятичное разложение числа a :

$$a = n_1 \cdot 10^r + n_2 \cdot 10^{r-1} + \dots + n_s \cdot 10^0.$$

Отсюда

$$(n_1 + 1) \cdot 10^r > a \geq n_1 \cdot 10^r;$$

имеем далее

$$\delta'_a = \frac{1}{2} \cdot \frac{10^p}{a}, \quad \delta'_a \geq \frac{1}{2} \cdot \frac{10^p}{(n_1 + 1)10^r} = \frac{10}{2(n_1 + 1)} \cdot 10^{-s},$$

если a — приближенное число, полученное с округлением. Если задано $\delta_a = 10^{-q}$, то нужно взять s так, чтобы $\delta'_a < \delta_a$, т. е.

$$10^s \geq \frac{10^{q+1}}{2(n_1 + 1)}. \quad (1.7)$$

За необходимое число верных знаков следует принять наименьшее целое число s , удовлетворяющее этому неравенству. Можно получить нечто вроде средней оценки, если принять $n_1 = 4$; тогда $s \geq q$.

По формуле (1.7) составлена нижеследующая таблица, дающая число верных знаков при заданной относительной погрешности:

$\delta\%$ n_1 \ q	100 0	10 1	1 2	0,1 3	0,01 4	0,001 5
1—3 4—9	1 0	2 1	3 2	4 3	5 4	6 5

Эту таблицу можно заменить простым правилом: если первая цифра не превышает трех, то число верных цифр должно быть на единицу больше, чем модуль показателя при 10 в заданной относительной погрешности; в остальных случаях эти числа равны. Значение 0 при $q = 0$, $n_1 = 4 - 9$ означает, что 100-процентная ошибка будет, если в числе нет ни одного верного знака, но все-таки известно, что первая цифра не меньше четырех. Действительно, пусть точное однозначное число равно 5. Если допускается ошибка в 100% числа, значит, модуль ошибки может достигнуть 5, и приближенное число может иметь любое значение от 0 до 10, т. е. и первая цифра не будет верной.

ГЛАВА 2

ПОГРЕШНОСТИ РЕЗУЛЬТАТОВ ОСНОВНЫХ АРИФМЕТИЧЕСКИХ ДЕЙСТВИЙ

В этой главе мы рассмотрим следующую задачу: известны предельные абсолютные или относительные погрешности чисел, над которыми производится сложение, вычитание, умножение или деление; требуется определить погрешность результата.

§ 6. Сложение

Пусть

$$u = a_1 + a_2 + \dots + a_n, \quad (2.1)$$

где a_1, a_2, \dots, a_n — приближенные числа, положительные и отрицательные; их предельные абсолютные погрешности $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$. Обозначим через $\Delta_1, \Delta_2, \dots, \Delta_n$ точные ошибки слагаемых, через Δ_u — точную ошибку u . Очевидно,

$$\Delta_u = \Delta_1 + \Delta_2 + \dots + \Delta_n, \quad (2.2)$$

откуда

$$|\Delta_u| \leq \sum_{k=1}^n |\Delta_k|. \quad (2.3)$$

Так как по определению [см. (1.3)]

$$|\Delta_k| \leq \varepsilon_k, \quad k = 1, 2, \dots, n,$$

то

$$|\Delta_u| \leq \sum_{k=1}^n \varepsilon_k.$$

Поэтому ε_u можно определить равенством

$$\varepsilon_u = \sum_{k=1}^n \varepsilon_k. \quad (2.4)$$

Таким образом, за предельную абсолютную погрешность суммы можно принять сумму предельных абсолютных погрешностей слагаемых.

Пример. Положим $u = 3,14 + 0,843 + 0,0365$; пусть приближенные слагаемые даны с точностью до единицы последнего знака. Тогда

$$\epsilon_u = 0,01 + 0,001 + 0,0001 = 0,0111.$$

Если бы требовалось записать эту оценку проще, то следовало бы взять

$$\epsilon_u = 0,02.$$

Из приведенного примера видно, что нет смысла стремиться получать приближенные слагаемые с разным количеством знаков после запятой; действительно, в нашем примере согласно полученной величине ϵ_u ошибка может превысить одну сотую, и потому в результате не имеет смысла писать тысячные и десятитысячные доли, раз нельзя ручаться за точность даже сотых долей. При сложении приближенных чисел можно поступать двумя способами.

Первый способ заключается в том, что все слагаемые «выравниваются» по наименее точному слагаемому. Так как при этом предельные абсолютные погрешности слагаемых будут равны, то предельная абсолютная погрешность суммы равна произведению предельной абсолютной погрешности отдельного слагаемого на число слагаемых. Из этого сейчас же следует, что в сумме теряется в точности один знак после запятой, если число слагаемых однозначное, два знака, если число слагаемых двузначное и т. д.

Следует, однако, заметить, что описанная оценка ошибки пригодна только при небольшом числе слагаемых. Если число слагаемых велико, например 20, 50 и т. п., то оценка ошибки может оказаться сильно преувеличенной, поскольку при таком способе фактически предполагается, что все приближенные слагаемые имеют наибольшие по модулю ошибки и все одного знака. В обычных вычислениях такой случай может, конечно, встретиться, но очень редко. По большей части числа Δ_k имеют различные знаки и в сумме большого числа слагаемых эти ошибки частью взаимно компенсируются. Поэтому действительная ошибка суммы большого числа слагаемых может оказаться значительно меньше того числа, которое получится по описанной «оценке на максимум» (см. следующий параграф).

Второй способ сложения чисел с различными количествами знаков после запятой заключается в том, что отдельно складываются по группам все числа, имеющие одинаковые числа знаков после запятой, и эти суммы округляют до наименьшего числа знаков после запятой.

Пример.

$$\begin{aligned} u &= 3,14 + (0,847 + 0,936) + (0,0746 + 0,0358) = \\ &= 3,14 + 1,783 + 0,1104 \approx 3,14 + 1,783 + 0,110 = \\ &= 3,14 + 1,893 \approx 3,14 + 1,89 = 5,03. \end{aligned}$$

Если все слагаемые даны с точностью до половины единицы последнего знака, то расчет погрешности результата в этом примере имеет вид: предельная погрешность суммы 4-го и 5-го слагаемых равна $2 \cdot 0,00005 = 0,0001$,

после округления этой суммы до трех знаков после запятой прибавляется предельная погрешность 0,0005; всего получается 0,0006; сложение 3-го и 4-го слагаемых с предыдущей суммой дает предельную погрешность $2 \cdot 0,0005 + 0,0006 = 0,0016$, отбрасывание третьего знака после запятой увеличивает погрешность на 0,005, всего получается 0,0066, что можно заметить на 0,007; сложение последней суммы с первым из слагаемых даст предельную погрешность результата — $0,005 + 0,007 = 0,012$.

По первому способу надо было бы все слагаемые округлить до двух знаков после запятой. Тогда мы получили бы

$$u = 3,14 + 0,85 + 0,94 + 0,07 + 0,04 = 5,04$$

с предельной абсолютной погрешностью

$$\frac{1}{2} \cdot 0,01 \cdot 5 = 0,025.$$

Второй способ можно применять в тех случаях, когда число слагаемых не очень мало.

§ 7. Статистическая оценка ошибки суммы *)

Пусть

$$u = \sum_{k=1}^n a_k,$$

где a_k — приближенные числа с точными ошибками Δ_k ; для каждой из ошибок известна предельная абсолютная погрешность ϵ_k , т. е. $\epsilon_k \geq |\Delta_k|$. Точная ошибка суммы

$$\Delta_u = \sum_{k=1}^n \Delta_k.$$

Чтобы построить вероятностные характеристики суммы u , нужно сделать определенные предположения о функции распределения величин Δ_k . В обычных вычислениях с определенным числом знаков после запятой считают, что ошибка случайно выбранного слагаемого есть число случайное, подчиненное равномерному закону распределения в области от $-\epsilon_k$ до $+\epsilon_k$. Согласно предыдущему параграфу точность вычислений мало повышается, если брать слагаемые с разными предельными абсолютными погрешностями. Поэтому мы будем считать все ϵ_k равными одному и тому же числу ϵ .

При этих предположениях плотность вероятности каждой из величин Δ_k будет равна $\frac{1}{2\epsilon}$ и центр распределения равен нулю. Поэтому дисперсия каждой из точных ошибок определится формулой:

$$D(\Delta_k) = \int_{-\epsilon}^{+\epsilon} \frac{1}{2\epsilon} a^2 da = \frac{1}{3} \epsilon^2.$$

Предполагая точные ошибки независимыми, по теореме о дисперсии суммы отсюда находим:

$$D(\Delta_u) = \frac{1}{3} n \epsilon^2.$$

*) Этот параграф следует читать после прохождения третьей части курса.

Для статистической оценки Δ_u требуется еще построить функцию распределения u . Закон распределения суммы конечного числа слагаемых строится довольно громоздко, но мы можем обойтись и без этого. Мы предположим, что сумма подчиняется нормальному закону с центром 0 и дисперсией $\frac{n}{3} \varepsilon^2$. По «правилу сигмы» можем считать

$$P(|\Delta_u| < \varepsilon \sqrt{\frac{n}{3}}) = 0,68,$$

где Δ_u — точная ошибка суммы, по «правилу трех сигм»

$$P(|\Delta_u| < \varepsilon \sqrt{3n}) = 0,9973.$$

Величину $\varepsilon \sqrt{3n}$ и можно принять за предельную абсолютную погрешность суммы n слагаемых. По оценке на максимум мы должны были бы взять na .

Легко видеть, что при $n > 3$ статистическая оценка ошибки суммы дает меньшую предельную погрешность, чем обычная. Если складывать 50 слагаемых, заданных с точностью до 0,005, то точная предельная погрешность будет 0,25, т. е. в сумме ненадежны даже десятые доли, а вероятностная оценка будет $0,005 \cdot \sqrt{150} \approx 0,06$. С вероятностью 0,9973 можно ожидать, что суммарная ошибка не превысит по модулю 0,06*).

Остается, конечно, неясным вопрос о близости распределения суммарной ошибки к нормальному распределению, которым мы пользуемся для определения вероятности. Сделаем расчеты для простейшего случая.

Пусть $n = 2$; по формуле распределения суммы двух слагаемых — распределение Симпсона с базой от -2ε до $+2\varepsilon$, для которого p вероятности $\varphi(\Delta_u)$ определяется формулами:

$$\varphi(\Delta_u) = \frac{1}{2\varepsilon} \left(1 + \frac{\Delta_u}{2\varepsilon}\right), \quad \text{если } -2\varepsilon \leq \Delta_u < 0$$

и

$$\varphi(\Delta_u) = \frac{1}{2\varepsilon} \left(1 - \frac{\Delta_u}{2\varepsilon}\right), \quad \text{когда } 0 \leq \Delta_u \leq 2\varepsilon.$$

(Множитель $\frac{1}{2\varepsilon}$ перед скобкой получается из условия нормирования плотности, вероятности, по которому площадь, ограниченная кривой распределения, равна единице.)

*) Можно, не зная закона распределения суммы ошибок, подсчитать вероятности по одной дисперсии, пользуясь неравенством Чебышева:

$$P(|\Delta_u| < t\sigma_u) > 1 - \frac{1}{t^2}, \quad \sigma_u = \sqrt{D(\Delta_u)}.$$

Если мы хотим получить предельные погрешности с вероятностью $> 0,99$, то $t^{-2} = 0,01$, $t = 10$,

$$P(|\Delta_u| < 10\sigma_u) > 0,99;$$

у нас $\sigma_u = \varepsilon \sqrt{\frac{n}{3}}$, поэтому в нашем примере с вероятностью больше 0,99

«предельная» ошибка равна $10 \cdot 0,005 \sqrt{\frac{50}{3}} \approx 0,2$. Результат этот все-таки несколько лучше, чем оценка по максимуму.

Дисперсия суммы определяется формулой

$$D_2(\Delta_u) = \frac{2}{3} \epsilon^2$$

(значок 2 при D обозначает, что число слагаемых равно двум). Вычислим вероятность того, что модуль ошибки суммы не превышает ее среднего квадратического отклонения:

$$\begin{aligned} P(|\Delta_u| < \epsilon \sqrt{\frac{2}{3}}) &= \int_{-\epsilon \sqrt{\frac{2}{3}}}^{\epsilon \sqrt{\frac{2}{3}}} \varphi(\Delta) d\Delta = \int_{-\epsilon \sqrt{\frac{2}{3}}}^0 \frac{1}{2\epsilon} \left(1 + \frac{\Delta}{2\epsilon}\right) d\Delta + \\ &+ \int_0^{\epsilon \sqrt{\frac{2}{3}}} \frac{1}{2\epsilon} \left(1 - \frac{\Delta}{2\epsilon}\right) d\Delta = \frac{1}{2\epsilon} \left(\Delta - \frac{\Delta^2}{4\epsilon}\right) \Big|_{-\epsilon \sqrt{\frac{2}{3}}}^0 + \\ &+ \frac{1}{2\epsilon} \left(\Delta - \frac{\Delta^2}{4\epsilon}\right) \Big|_0^{\epsilon \sqrt{\frac{2}{3}}} = 2 \left(\frac{1}{2} \sqrt{\frac{2}{3}} - \frac{1}{12}\right) = \sqrt{\frac{2}{3}} - \frac{1}{6}. \\ P(|\Delta_u| < \epsilon \sqrt{\frac{2}{3}}) &\approx 0,65. \end{aligned}$$

Таким образом, даже в случае двух слагаемых вероятность ошибки, не превышающей средней квадратической, оказывается равной 0,65, вместо 0,68 по нормальному закону. Искать вероятность отклонения, не превышающего 3ϵ , здесь нет смысла, так как предельная ошибка по модулю равна 2ϵ , а утроенная средняя квадратичная ошибка будет

$$3\epsilon \sqrt{\frac{2}{3}} = \epsilon \sqrt{6} > 2,4\epsilon.$$

§ 8. Вычитание близких чисел

Вычитание можно считать алгебраическим сложением и при оценке ошибки разности воспользоваться тем результатом, который получен в § 6.

Если

$$u = a - b, \quad (2.5)$$

то

$$\left. \begin{aligned} \epsilon_u &= \epsilon_a + \epsilon_b, \\ \delta_u &= \frac{\epsilon_a + \epsilon_b}{|u|}. \end{aligned} \right\} \quad (2.6)$$

При вычитании встречаются случаи, представляющие затруднения в технике вычисления. Пусть b мало отличается от a . Так как a и b имеют ограниченное число значащих цифр, то в этом случае в разности будет мало значащих цифр, а это означает, что относительная

погрешность разности будет велика. Особенное затруднение будет в том случае, когда результат такого вычитания входит в дальнейшие вычисления. Если в разности будет, например, только два верных знака, то в результате действий над этим числом и другими нельзя получить больше одного, в лучшем случае двух верных знаков. В таких случаях говорят, что потеряна точность, так как никакое повышение точности других чисел (кроме a и b) не может исправить результата.

Из создавшегося положения могут быть два выхода.

Прежде всего можно попытаться заметно увеличить число верных знаков в числах a и b , чтобы в результате сохранилось нужное число верных знаков. Однако уточнение a и b возможно далеко не всегда (например a и b могут быть получены из наблюдений). Тогда стараются перестроить вычислительные формулы так, чтобы устранить разность близких чисел.

Если уменьшаемое мало отличается от вычитаемого, то оба эти числа можно представить в виде

$$a = m + \alpha, \quad b = m + \beta,$$

где α и β — числа, заметно отличающиеся одно от другого. Простейшее преобразование формул заключается в том, что стараются устранить общую часть m преобразованием формул, так, чтобы вычисление свелось к такому:

$$u = \alpha - \beta.$$

Если и этот прием не удастся, то потребуются более сложная перестройка формул.

Пример. Пусть нужно вычислить левую часть формулы

$$(r_1 + r_2 + s)^{\frac{3}{2}} - (r_1 + r_2 - s)^{\frac{3}{2}} = u$$

(известная формула Эйлера, выражающая связь между двумя радиусами-векторами параболы, длиной хорды и временем). В обычных задачах s — малое число по сравнению с $r_1 + r_2$, поэтому прямое вычисление по этой формуле приведет к потере точности. Можно сделать следующее преобразование:

$$u = (r_1 + r_2)^{\frac{3}{2}} \left[\left(1 + \frac{s}{r_1 + r_2} \right)^{\frac{3}{2}} - \left(1 - \frac{s}{r_1 + r_2} \right)^{\frac{3}{2}} \right].$$

По условию дробь $\frac{s}{r_1 + r_2}$ мала по сравнению с единицей; поэтому выражения в круглых скобках можно разложить в биномиальные ряды.

Производя разложение, получим:

$$u = (r_1 + r_2)^{\frac{3}{2}} \left(3\sigma - \frac{1}{8}\sigma^3 + \dots \right), \quad \text{где } \sigma = \frac{s}{r_1 + r_2};$$

это преобразование исключило общую часть в уменьшаемом и вычитаемом (единицы), что и дало возможность повысить точность остатка. Пусть,

например, $r_1 + r_2 = 2,000$, $s = 0,01423$ (числа даны с четырьмя верными знаками); прямое вычисление дает $u = 0,060$ (вычисления с 4-значными таблицами логарифмов), т. е. в результате оказывается только два верных знака, хотя исходные данные имели четыре верных знака. Если же считать по преобразованной формуле, то получим $0,06031$ с сомнительной последней цифрой *).

§ 9. Умножение

Пусть

$$u = a \cdot b; \quad (2.7)$$

определим ε_u и δ_u , считая заданными ε_a и ε_b . Для упрощения выкладок предположим, что $a > 0$ и $b > 0$. Точное произведение

$$U = A \cdot B = (a + \Delta_a)(b + \Delta_b) \quad (2.8)$$

или

$$U = u + a\Delta_b + b\Delta_a + \Delta_b\Delta_a.$$

Считая Δ_a и Δ_b малыми по сравнению с a и b , отбросим произведение $\Delta_b\Delta_a$, как малую второго порядка. Тогда

$$U - u = a\Delta_b + b\Delta_a \quad (2.9)$$

я

$$|U - u| \leq a|\Delta_b| + b|\Delta_a|. \quad (2.10)$$

Заменяя модули точных ошибок предельными абсолютными погрешностями, получим верхнюю границу модуля точной ошибки, которую и можно принять за предельную абсолютную погрешность произведения двух сомножителей:

$$\varepsilon_u = a\varepsilon_b + b\varepsilon_a. \quad (2.11)$$

Формула эта легко распространяется на произведение любого числа сомножителей. Если

$$u = a_1 a_2 \dots a_s, \quad (2.12)$$

то

$$\varepsilon_u = a_1 a_2 \dots a_{s-1} \varepsilon_s + a_1 a_2 \dots a_{s-2} a_s \varepsilon_{s-1} + \dots + a_2 a_3 \dots a_s \varepsilon_1. \quad (2.13)$$

*) Заметим, что в рассмотренном примере можно было бы избавиться от разности близких чисел и проще. Умножая и деля левую часть исходной формулы на сумму

$$(r_1 + r_2 + s)^{\frac{s}{2}} + (r_1 + r_2 - s)^{\frac{s}{2}},$$

получим:

$$u = \frac{2(r_1 + r_2)^{\frac{s}{2}} (3\alpha - \sigma^8)}{(1 + \sigma)^{\frac{s}{2}} + (1 - \sigma)^{\frac{s}{2}}}.$$

Отметим, однако, что метод разложения в ряд является более общим, чем рассмотренный частный прием.

При этом предполагается, что все сомножители положительны, если же среди них есть и отрицательные, то их нужно заменить их модулями.

Пример. Определяется длина s окружности радиуса $R = 3,484$ м. Число π может быть взято с любым практически нужным числом верных знаков; возьмем $2\pi = 6,283$. Каждый из сомножителей дан с точностью до $\frac{1}{2} \cdot 10^{-3}$. Поэтому предельная абсолютная погрешность определится формулой:

$$\epsilon_c = (R + 2\pi) \cdot \frac{1}{2} \cdot 10^{-3}.$$

Заметим, что для расчета погрешности нет надобности брать те значения R и π , которые непосредственно даны, так как часто работа по оценке ошибки могла бы стать сложнее основного вычисления.

Подставим в ϵ_c 10 вместо $R + 2\pi$ (что допустимо, поскольку оценка становится грубее). Тогда

$$\epsilon_c = \frac{1}{2} \cdot 10^{-2},$$

т. е. в произведении надо ограничиться сотыми долями; легко видеть, что произведение будет содержать четыре верных знака.

Умножение приближенных чисел может выполняться разными способами с учетом точности результата. При непосредственном умножении нет надобности считать все знаки, которые получаются при вычислениях. В частности, разработаны различные способы сокращенного умножения, которые позволяют упростить работу*). Обычно при сокращенном умножении определяют и знак, следующий за последним сохраняемым, затем его отбрасывают с округлением. В нашем примере нужно считать и тысячные доли, которые после сложения отбрасываются. Таким образом,

$$s = 21,89 \text{ м.}$$

Если вычисления выполняются с таблицами логарифмов, то число считааемых знаков автоматически регулируется числом знаков таблицы.

Но теперь умножения чаще выполняются на арифмометрах. При этом автоматически считаются все знаки, какие формально могут получиться. На результатном счетчике устанавливается запятая по известным правилам арифметики. Производится подсчет точности, определяются те знаки после запятой, которые можно считать верными, и соответственно отмечаются второй запятой те знаки, которые нужно сохранить.

Определим теперь предельную относительную погрешность произведения. По общей формуле для любого числа сомножителей получим

$$\delta_u = \frac{\epsilon_u}{u} = \sum_{k=1}^s \frac{\epsilon_k}{a_k}.$$

*) См. книгу Я. С. Безиковича в списке литературы к ч. I, приведенном в конце книги.

По определению $\delta_k = \frac{\varepsilon_k}{a_k}$, поэтому

$$\delta_u = \sum_{k=1}^8 \delta_k. \quad (2.14)$$

Таким образом, *предельная относительная погрешность произведения равна сумме предельных относительных погрешностей сомножителей.*

Этот результат имеет большое значение. Вспомним, что предельная относительная погрешность тесно связана с числом верных знаков. Поэтому при умножении нет смысла брать сомножители с различными числами верных знаков. Действительно, если, например, в одном сомножителе три верных знака, в другом пять, то произведение будет иметь относительную погрешность, соответствующую числу верных знаков не больше трех. Следовательно, в произведении число верных знаков будет равно или меньше, чем наименьшее из числа верных знаков отдельных сомножителей. Если число сомножителей однозначное, то можно сформулировать правило: произведение однозначного числа сомножителей имеет верных знаков либо столько же, либо на единицу меньше, чем наименьшее из числа верных знаков сомножителей.

Пример. Вычислим площадь круга радиусом 2,37 см.

Площадь P равна произведению трех сомножителей, из них два имеют по три верных знака. Поэтому число π достаточно взять тоже с тремя верными знаками, т. е. 3,14. Относительная погрешность произведения приближенно равна

$$\delta_P = \frac{1}{4} \cdot 10^{-2} + \frac{1}{4} \cdot 10^{-2} + \frac{1}{6} \cdot 10^{-2} = \frac{2}{3} \cdot 10^{-2}$$

при подсчете было взято 2 вместо 2,37, 3 вместо 3,14 (так как эти упрощения увеличивают предельную погрешность, то они допустимы). По грубой оценке, которую можно сделать в уме, площадь будет близка к 20, поэтому предельная абсолютная погрешность равна $1,4 \cdot 10^{-1}$, следовательно, в произведении следует брать только десятые доли, причем последняя цифра определяется неуверенно. Вычисляя площадь, получим $P = 17,6 \text{ см}^2$. Сомножители имели по три верных знака с округлением, произведение имеет два верных знака и один сомнительный с предельной абсолютной погрешностью 0,14.

Применим полученное правило к степени с целым положительным показателем. Если

$$u = x^n, \quad (2.15)$$

то

$$\left. \begin{aligned} \varepsilon_u &= nx^{n-1}\varepsilon_x, \\ \delta_u &= n\delta_x \quad (x > 0). \end{aligned} \right\} \quad (2.16)$$

Число верных знаков находится так же, как и при обыкновенном умножении. При определении ε_u нет, конечно, надобности брать заданное значение x со всеми знаками; обычно берут только первую цифру, увеличенную на единицу.

Пример.

$$u = 3,458^3; \quad \epsilon_u = 3 \cdot 4^3 \cdot \frac{1}{2} \cdot 10^{-3}$$

или (увеличивая) $\epsilon_u = 3 \cdot 10^{-2}$; из этого следует, что в произведении нужно взять только десятые доли, результат будет верен до $\frac{1}{2} \cdot 10^{-1}$, поскольку $\epsilon_u < 5 \cdot 10^{-2}$.

Из этого примера видно, что в произведении можно ручаться только за три знака, когда возвышаемое в степень число имеет четыре верных знака. Расчет получится грубым, но при уточненном расчете за сотые доли в произведении ручаться нельзя. Это можно показать следующим образом. Если мы будем считать с запасными знаками, то $x^2 = 11,958$, $x^3 = 41,351$. Возьмем теперь кубы грани, между которыми заключено точное значение числа. Куб числа заключен между $3,4575^3 = 41,33$ и $3,4585^3 = 41,37$. Этот подсчет наглядно показывает, что сотые доли произведения ненадежны, но ошибка в сотых долях наверно не превышает четырех единиц. Поэтому часто при вычислениях вводят лишний знак (по сравнению с расчетом); в нашем примере мы могли бы принять произведение равным 41,36, но при дальнейшем пользовании этим числом надо было бы учесть, что его предельная абсолютная погрешность не $\frac{1}{2} \cdot 10^{-2}$, а 0,03. В приведенном выше примере порядок вычислений на арифмометре удобно принять таким: возвышаем число в квадрат, снимаем со счетчика четыре знака (сотые доли) — 11,96, затем умножаем на 3,458 и, округляя, получим $u = 41,36$.

Таким образом, при возвышении в квадрат и куб степень может быть взята безусловно с числом верных знаков, на единицу меньшим, чем у основания. Можно взять и столько же знаков, сколько в возвышаемом числе, но последний знак будет сомнителен. Ясно, конечно, что брать в степени больше знаков, чем в основании степени, незаконно.

§ 10. Деление

Пусть

$$u = \frac{a}{b}; \quad (2.17)$$

ϵ_a и ϵ_b известны. Следует определить ϵ_u . Имеем

$$U = \frac{A}{B} = \frac{a + \Delta_a}{b + \Delta_b}, \quad (2.18)$$

$$\Delta_u = U - u = \frac{a + \Delta_a}{b + \Delta_b} - \frac{a}{b} \quad (2.19)$$

или

$$\Delta_u = \frac{b\Delta_a - a\Delta_b}{b^2 + b\Delta_b}. \quad (2.20)$$

Введем обозначение

$$\Delta'_u = \frac{b\Delta_a - a\Delta_b}{b^2}$$

и найдем разность $\Delta'_u - \Delta_u$:

$$\Delta'_u - \Delta_u = \frac{(b\Delta_a - a\Delta_b)\Delta_b}{b(b^2 + b\Delta_b)}.$$

Эта разность есть малая величина второго порядка, если Δ_a и Δ_b считать малыми величинами первого порядка. Поэтому можно принять

$$\Delta_u = \frac{b\Delta_a - a\Delta_b}{b^2},$$

откуда

$$|\Delta_u| \leq \frac{|b| |\Delta_a| + |a| |\Delta_b|}{b^2}. \quad (2.21)$$

Заменяя неизвестные $|\Delta_a|$ и $|\Delta_b|$ их максимумами ϵ_a и ϵ_b , получим

$$\epsilon_u = \frac{|a| \epsilon_b + |b| \epsilon_a}{b^2}. \quad (2.22)$$

Пример. Положим $a = 5,36$, $b = 0,748$, $\epsilon_a = 0,5 \cdot 10^{-2}$, $\epsilon_b = 0,5 \cdot 10^{-3}$; для упрощения расчета ϵ_u возьмем 6 вместо a , 0,8 вместо b в числителе и 0,5 вместо b^2 в знаменателе. Тогда получим

$$\epsilon_u = \frac{3 \cdot 10^{-3} + 4 \cdot 10^{-3}}{0,5} = 1,4 \cdot 10^{-2},$$

примем окончательно $\epsilon_u = 2 \cdot 10^{-2}$ (увеличивая для упрощения оценки).

Строго говоря, из примера следует, что в частном нельзя ругаться за точность сотых долей. В таком случае, как здесь, сотые доли в частном берут, поскольку ошибка в сотых по расчету мало превышает единицу, а расчет делается по максимальной возможной ошибке. Если, однако, над частным предстоит производить еще действия, то при расчете погрешности следующих операций нужно учесть полученную оценку погрешности $2 \cdot 10^{-2}$, а частное следует принять равным 7,17 ($2 \cdot 10^{-2}$), где в скобках указана предельная погрешность *).

Найдем теперь предельную относительную погрешность частного. По определению получаем

$$\delta_u = \frac{|a| \epsilon_b + |b| \epsilon_a}{b^2} : \frac{|a|}{|b|} = \frac{\epsilon_a}{|a|} + \frac{\epsilon_b}{|b|}; \quad \left. \begin{aligned} \delta_u &= \delta_a + \delta_b. \end{aligned} \right\} \quad (2.23)$$

*) Иногда имеет смысл записывать результат так:

$$7,15 < u < 7,19.$$

Этот результат показывает, что вопрос о числе верных знаков в случае деления решается так же, как и при умножении. Поэтому правило, указанное в предыдущем параграфе, можно распространить на совокупности умножений и делений. Сохраняется также положение о бесполезности брать перемножаемые или делимые числа, заметно отличающиеся по числу верных знаков.

Пример 1.

$$a = 2,43; \quad b = 0,216; \quad u = \frac{a}{b};$$

$$\epsilon_a = 0,5 \cdot 10^{-2}; \quad \epsilon_b = 0,5 \cdot 10^{-3}.$$

Для того чтобы иметь возможность оценить ошибку в уме, обычно a и b в числителе формулы для ϵ_u несколько увеличивают, а b в знаменателе уменьшают. В данном случае

$$\epsilon_u = \frac{3 \cdot 0,5 \cdot 10^{-2} + 0,3 \cdot 0,5 \cdot 10^{-2}}{0,2^2} = 75 \cdot 10^{-3}.$$

Можно принять $\epsilon_u = 10^{-1}$. В частном, следовательно, можно ручаться только за десятые доли:

$$u = 11,2.$$

Другой расчет можно сделать по предельной относительной погрешности:

$$\delta_a = \frac{1}{4} \cdot 10^{-2}, \quad \delta_b = \frac{1}{4} \cdot 10^{-2}, \quad \delta_u = \frac{1}{2} \cdot 10^{-2}.$$

Так как первая цифра частного равна единице, то, согласно таблице, приведенной на стр. 21, оно имеет три верных знака.

Оба заключения согласуются с общим правилом: поскольку делимое и делитель имеют по три верных знака, то и частное имеет не больше трех верных знаков.

Пример 2. $a = 3,144$; $b = 0,0536$. Определим $u = \frac{a}{b}$ и оценим ошибку.

По таблице, приведенной на стр. 21, находим:

$$\delta_a = 0,017\% \approx 0,02\%, \quad \delta_b = 0,10\%, \quad \delta_u = 0,12\%.$$

Так как первая цифра частного 5, то приняв (не совсем законно) $\delta_u = 0,1\%$, по таблице найдем, что в частном будет три верных знака. Хотя в частном получилось число верных знаков, равное меньшему числу верных знаков, но при таком малом делителе все-таки имеет значение то, что a дано с четырьмя знаками. Если взять a с тремя верными знаками, то было бы

$$\delta_a = 0,10\%, \quad \delta_b = 0,10\%, \quad \delta_u = 0,2\%;$$

по таблице можно было бы ручаться только за два верных знака. У нас $u = 58,7$; если взять частное до половины сотой доли, то будет $u = 58,66$.

Контроль по максимуму и минимуму частного дает $58,59 < u < 58,72$, что и показывает наглядно неточность сотых долей и некоторую ненадежность десятых, ибо может быть 6 или 7, согласно указанным границам. Следует отметить, что среднее из граничных значений совпадает с тем, которое получится при вычислении сотых. Это до некоторой степени оправдывает употребление в подобных случаях запасного знака; если делимое имеет четыре знака, делитель три знака, то нередко в частном берут четыре знака, помня, конечно, что четвертый знак ненадежен.

ГЛАВА 3

ОЦЕНКА ОШИБКИ ФУНКЦИИ ПРИБЛИЖЕННЫХ АРГУМЕНТОВ

§ 11. Предельные погрешности функции одного независимого переменного

Пусть

$$U = f(X), \quad (3.1)$$

где $f(X)$ — непрерывная дифференцируемая функция. Если вместо точного значения аргумента A подставить его приближенное значение a , то полученное значение функции $f(a)$ будет также *приближенным*. Найдем выражение предельной абсолютной погрешности функции ϵ_u через предельную абсолютную погрешность приближенного значения аргумента ϵ_a . Точное значение $f(A)$ можно представить в виде

$$U = f(A) = f(a + \Delta_a) = f(a) + f'(a) \Delta_a + \dots \quad (3.2)$$

где Δ_a — точная ошибка a . Считая Δ_a малой величиной и полагая, что членами второго и высших порядков этой величины можно пренебречь, получим

$$\Delta_u \approx f'(a) \Delta_a; \quad (3.3)$$

отсюда

$$\epsilon_u = |f'(a)| \epsilon_a. \quad (3.4)$$

Таким образом, *предельная абсолютная погрешность функции одного аргумента равна произведению модуля производной на предельную абсолютную погрешность аргумента.*

Для предельной относительной погрешности получим формулу

$$\delta_u = \left| \frac{f'(a)}{f(a)} \right| \cdot |a| \cdot \delta_a. \quad (3.5)$$

Из этой формулы видно, что предельная относительная погрешность пропорциональна логарифмической производной функции и значению аргумента. Заметим, однако, что построенная таким способом оценка погрешности функции в некоторых случаях может оказаться чересчур грубой.

Пример. Пусть определяется $\operatorname{tg} 85^\circ$, причем угол дан с точностью до $0,5 = 0,0087$ рад. Тогда

$$\epsilon_{\operatorname{tg}} = \frac{1}{\cos^2 85^\circ} \cdot 0,0087 = 0,0087 \cdot 132 = 1,15.$$

По таблицам $\operatorname{tg} 85^\circ = 11,4$. Поэтому неизвестное действительное значение заключено между $11,4 - 1,15 = 10,25$ и $11,4 + 1,15 = 12,55$. Проверим это прямым определением тангенсов крайних значений

$$\operatorname{tg} 84,5^\circ = 10,4; \quad \operatorname{tg} 85,5^\circ = 12,7.$$

Эти границы следует считать точными; те же границы, которые мы установили с помощью предельной абсолютной погрешности, должны совпадать с точными в вычисляемых трех знаках. Заметное отличие границ, превышающее допустимую погрешность вычислений, означает, что предельная абсолютная погрешность неточна. Легко показать, что в нашем случае пренебрежение в ряде членами 2-го и более высокого порядка малости при трехзначных вычислениях было недопустимо. Действительно, первый отброшенный член ряда превышает 0,1; следовательно, в погрешности ненадежны и десятые доли.

Можно указать общее правило проверки пригодности способа определения предельной абсолютной погрешности функции. Предполагая существование и непрерывность второй производной рассматриваемой функции, можем написать по формуле Тэйлора:

$$f(A) - f(a) = f'(a) \Delta_a + \frac{1}{2} f''(\xi) \Delta_a^2, \quad (3.6)$$

где ξ — число, заключенное между a и $a + \Delta_a$. При «линеаризации» правой части мы делаем ошибку:

$$R = \frac{1}{2} f''(\xi) \Delta_a^2.$$

Отсюда получается оценка:

$$|R| \leq \frac{1}{2} M_2 \epsilon_a^2, \quad (3.7)$$

где $M_2 \geq |f''(x)|$, при $a - \epsilon_a \leq x \leq a + \epsilon_a$. Если $|R|$ — число, превышающее единицу первого отбрасываемого знака, то «линеаризация» груба; надо вести счет с меньшей точностью, или ввести поправку $\frac{1}{2} M_2 \epsilon_a^2$ в оценку ошибки.

§ 12. Погрешности простейших элементарных функций

В этом параграфе мы рассмотрим погрешности основных элементарных функций.

1. Функции $\sin x$ и $\cos x$

Если $u = \sin a$, то согласно общей формуле (3.4)

$$\epsilon_u = |\cos a| \cdot \epsilon_a; \quad (3.8)$$

отсюда видно, что

$$\varepsilon_u \leq \varepsilon_a. \quad (3.9)$$

Аналогично получим

$$\varepsilon_{\cos a} = |\sin a| \varepsilon_a. \quad (3.10)$$

Из полученных выражений можно вывести следующее указание об организации вычислений. Если значения углов в формулах приближенные, то для уменьшения погрешности следует брать синус угла, приведенного к 1-й четверти, если угол больше 45° , и косинус угла, когда угол меньше 45° . Правило это нужно применять в тех случаях, когда возможно выбирать вычисляемую тригонометрическую функцию.

Пример. В ряде астрономических задач приходится производить такие вычисления: известны $\alpha = r \sin \delta$ и $\beta = r \cos \delta$; найти δ и r . Определяем $\operatorname{tg} \delta = \frac{\alpha}{\beta}$; по тангенсу находим приближенно угол δ ; для определения r можно по своему выбору воспользоваться одной из формул $r = \frac{\alpha}{\sin \delta}$ или $r = \frac{\beta}{\cos \delta}$.

Выбор той или другой из этих формул следует делать по указанному выше правилу.

Пусть $\alpha = 2,364$, $\beta = 1,575$, $\operatorname{tg} \delta = 1,501$, $\delta = 56^\circ 20'$. Так как $\delta > 45^\circ$, то для определения r лучше взять $\sin \delta$, поскольку погрешность синуса здесь меньше погрешности косинуса. Вычисления располагаются по схеме:

$$\begin{array}{l|l} \alpha^* = r \sin \delta & 2,364^* \\ \sin \delta & 0,832 \end{array} \quad (3)$$

$$\begin{array}{l|l} \beta^* = r \cos \delta & 1,575^* \\ \operatorname{tg} \delta & 1,501 \end{array} \quad (1)$$

$$\begin{array}{l|l} \delta & 56^\circ 20' \end{array} \quad (2)$$

$$\begin{array}{l|l} r & 2,841 \end{array} \quad (5)$$

$$\begin{array}{l|l} \cos \delta & 0,554 \end{array} \quad (4)$$

$$\begin{array}{l|l} (r) & 2,843. \end{array} \quad (6)$$

Звездочки отмечают заданные значения α и β , а номера в скобках указывают последовательность операций.

Последние две строчки дают контроль вычислений; скобка, в которую заключено r , означает контрольное значение.

2. Функции $\operatorname{tg} x$ и $\operatorname{ctg} x$

Если $u = \operatorname{tg} a$, $v = \operatorname{ctg} a$, то

$$\left. \begin{array}{l} \varepsilon_u = \frac{\varepsilon_a}{\cos^2 a}, \\ \varepsilon_v = \frac{\varepsilon_a}{\sin^2 a}. \end{array} \right\} \quad (3.11)$$

Из этих формул видно, что

$$\varepsilon_{\operatorname{tg} a} \geq \varepsilon_a; \quad \varepsilon_{\operatorname{ctg} a} \geq \varepsilon_a. \quad (3.12)$$

Если угол a близок к нулю, то погрешность тангенса близка к минимальной, а погрешность котангенса велика, следовательно, пригодность формулы для ϵ подлежит проверке. К противоположным выводам мы придем, когда угол близок к 90° . Соответственно сказанному и следует организовать вычисления, если можно выбрать формулы, содержащие либо тангенс, либо котангенс.

3. Определение угла по тригонометрической функции

Пусть $u = \arcsin a$. Тогда

$$\epsilon_u = \frac{\epsilon_a}{\sqrt{1-a^2}}. \quad (3.13)$$

Из этой формулы можно сделать следующие заключения:

- а) $\epsilon_u \geq \epsilon_a$; равенство будет при $a=0$, т. е. $u=0$;
- б) если угол нужно определять по синусу, то ошибка будет мала только при условии, что угол (приведенный) мал;
- в) невыгодно определять по синусу угол, близкий к 90° , поскольку ошибка угла будет гораздо больше, чем ошибка синуса. Если, например, $u = \arcsin 0,984$, то

$$\epsilon_u = \frac{5 \cdot 10^{-4}}{0,17} = 3 \cdot 10^{-3},$$

т. е. ошибка в угле может достигать $10'$. То же можно усмотреть непосредственно из таблиц, в которых при сохранении трех верных знаков $\sin 79^\circ 40' = 0,984$ и $\sin 79^\circ 50' = 0,984$.

Такие же заключения можно сделать и в случае определения угла по косинусу — с той разницей, что по косинусу выгодно определять угол, близкий к 90° и невыгодно вычислять угол, близкий к нулю.

Пусть $u = \operatorname{arctg} a$, $v = \operatorname{arctg} a$. Тогда

$$\epsilon_u = \frac{\epsilon_a}{1+a^2}, \quad \epsilon_v = \frac{\epsilon_a}{1+a^2}. \quad (3.14)$$

Из этих формул видно, что выгоднее всего определять угол по тангенсу или котангенсу, ибо ошибка угла меньше ошибки тангенса. Наиболее неблагоприятен случай, когда малый угол определяется по тангенсу или близкий к 90° по котангенсу, но и в этом случае ошибка угла близка к ошибке функции. Поэтому стараются строить вычислительные схемы так, чтобы углы вычислять по тангенсу или котангенсу.

Пример. Рассмотрим одну из часто встречающихся в астрономических вычислениях задач — определение экваториальных сферических координат по прямоугольным координатам x , y , z . При помощи формул,

связывающих эти координаты, составляется следующая схема вычислений:

$$\begin{array}{rcl}
 x^* = r \cos \delta \cos \alpha & \left| \begin{array}{l} 3,1448^* \\ \cos \alpha \\ 0,82509 \end{array} \right. & (3) \\
 y^* = r \cos \delta \sin \alpha & \left| \begin{array}{l} -2,1534^* \\ -0,68475 \end{array} \right. & (1) \\
 y : x = \operatorname{tg} \alpha & & \\
 x : \cos \alpha = r \cos \delta & \left| \begin{array}{l} 3,8115 \\ \cos \delta \\ 0,98906 \end{array} \right. & (4) \\
 z^* = r \sin \delta & \left| \begin{array}{l} 0,56843^* \\ 0,14914 \end{array} \right. & (7) \\
 z : r \cos \delta = \operatorname{tg} \delta & & (5) \\
 \alpha & \left| \begin{array}{l} 325^\circ 35' 54'' \\ \delta \\ 8^\circ 28' 58'' \end{array} \right. & (2) \\
 \delta & & (6) \\
 r \cos \delta : \cos \delta = r & \left| \begin{array}{l} 3,8537 \end{array} \right. & (8)
 \end{array}$$

Вычисления пятизначные, поэтому углы даются с точностью до $1''$. Значения, отмеченные звездочками, заданы. Номера в скобках указывают последовательность действий.

4. Степень и корень

Если $u = a^n$, то

$$\left. \begin{array}{l} \varepsilon_u = |n| a^{n-1} \varepsilon_a, \\ \delta_u = |n| \delta_a. \end{array} \right\} \quad (3.15)$$

Из этих простых формул следует, что в случае $|n| > 1$ относительная ошибка степени больше относительной ошибки основания; если же $|n| < 1$ (в частности, извлечение корня целой степени), то возведение в степень n уменьшает ошибку. Если число n порядка $2-3$ или $\frac{1}{2}$, $\frac{1}{3}$, то из формулы для относительной погрешности следует, что число значащих цифр результата можно брать таким же, как число цифр возвышаемого числа.

Пример.

$$\pi^2 = 9,86, \quad \sqrt{\pi} = 1,77, \quad \text{если принять } \pi = 3,14,$$

$$\pi^2 = 9,8696, \quad \sqrt{\pi} = 1,7725, \quad \text{если взять } \pi = 3,1416.$$

Значение π^2 или $\sqrt{\pi}$, полученное с пятью знаками, можно условно считать точным, а число, полученное с тремя знаками, приближенным. Сравнение трех значащих цифр в «точном» и приближенном значении в нашем случае показывает, что в результате верны либо три знака (при вычислении корня), либо, если быть формально строгим, два знака (при вычислении степени), ибо при округлении пятизначного значения π^2 до трех знаков надо брать 9,87 (по правилу округления), а получено 9,86.

5. Логарифм и показательная функция

Пусть $u = \lg a$. Так как $a = 10^u$, то $u = \lg e \ln a$, где $\ln a$ — натуральный логарифм, $\lg e = 0,4343 < 0,5$. Поэтому можно принять

$$\epsilon_u = \frac{1}{2} \frac{\epsilon_a}{a} = \frac{1}{2} \delta_a. \quad (3.16)$$

Предельная относительная погрешность δ_a приближенного числа a не зависит, как было указано в § 5, гл. I, от места запятой, но в основном зависит от числа верных знаков в этом числе. Если a получено с округлением, то согласно § 5 можно принять

$$\delta_a = \frac{1}{2C \cdot 10^{s-1}}, \quad (3.17)$$

где C — первая значащая цифра, s — число верных знаков. Отсюда следует, что

$$\epsilon_u = \frac{1}{4C} \cdot \frac{1}{10^{s-1}}, \quad (3.18)$$

причем, если $C > 2\frac{1}{2}$, то

$$\epsilon_u < \frac{1}{10^s}, \quad (3.19a)$$

если же $C < 2\frac{1}{2}$, то

$$\epsilon_u > \frac{1}{10^s}. \quad (3.19b)$$

На основании этого можно сформулировать следующее правило: если первая цифра приближенного числа больше двух, то логарифм его имеет столько верных знаков после запятой, сколько верных значащих цифр в числе; если первая цифра меньше трех, то последняя цифра логарифма, взятого с пятью знаками, может быть не совсем надежна — возможна ошибка в одну, две единицы последнего знака (так как при $C = 1$ имеем $\epsilon_u = 2,5 \cdot 10^{-8}$). Обычно пользуются несколько упрощенным правилом: число знаков после запятой в логарифмах должно равняться числу верных знаков приближенного числа. Во многих прикладных задачах технического характера бывает достаточна относительная погрешность, соответствующая трем или четырем верным знакам, так как исходные данные содержат столько же знаков. Если наблюдения дают три знака, то вычисления можно выполнять на логарифмической линейке; в случае 4-значных результатов наблюдения можно пользоваться либо арифмометром, либо 4-значными таблицами логарифмов. При количестве знаков больше чем 3 или 4 в настоящее время чаще всего пользуются арифмометрами.

Легко исследовать погрешность обратной операции, т. е. деления числа по заданному его логарифму (показательная функция или антилогарифм).

Из формулы для предельной абсолютной погрешности логарифма следует

$$\delta_a = \frac{1}{0,4343} \epsilon_u. \quad (3.20)$$

По определению предельной погрешности правую часть можно увеличить, поэтому примем

$$\delta_a = 2,4 \epsilon_u \quad (\text{если } u = \lg a). \quad (3.21)$$

Если логарифм (u) имеет s знаков после запятой, причем u получено с округлением, то

$$\left. \begin{aligned} \epsilon_u &= \frac{1}{2} \cdot 10^{-s}, \\ \delta_a &= 1,2 \cdot 10^{-s}, \end{aligned} \right\} \quad (3.22)$$

а это означает, что почти всегда число имеет s верных знаков. Если логарифм получен без округления, то относительная погрешность антилогарифма в 2,5 раза больше предельной погрешности логарифма. Поэтому число имеет либо s верных знаков, либо $s - 1$.

6. Логарифмы тригонометрических функций

а) $u = \lg \sin a$;

$$\epsilon_u = 0,4343 \operatorname{ctg} a \cdot \epsilon_a. \quad (3.23)$$

Будем считать, что угол приведен к I четверти. Чтобы предельная погрешность была меньше предельной погрешности угла, должно выполняться неравенство $\operatorname{ctg} a \cdot 0,4343 \leq 1$ или $\operatorname{tg} a \geq 0,4343$, т. е. $a > 23^\circ 20'$.

Отсюда видно, что при вычислениях невыгодно брать логарифм синуса угла, меньшего, чем $23,5^\circ$. Читателю рекомендуется самостоятельно показать, что невыгодно брать логарифм косинуса угла, если угол больше $66,5^\circ$ *). Если угол меньше $66,5^\circ$, но больше $23,5^\circ$, то предельные погрешности и логарифма синуса и логарифма косинуса меньше, чем предельная погрешность угла (отметим, что именно это и вкладывается в понятие «выгодности» вычисления).

б) $u = \lg \operatorname{tg} a$;

$$\epsilon_u = 0,4343 \operatorname{ctg} a \sec^2 a \cdot \epsilon_a = \frac{0,8686}{\sin 2a} \epsilon_a. \quad (3.24)$$

Чтобы погрешность функции была меньше погрешности угла, должно выполняться неравенство $\sin 2a \geq 0,8686$, т. е. $59^\circ 30' < 2a < 120^\circ 30'$ или (грубее) $30^\circ < a < 60^\circ$.

*) Следует помнить, что в расчетах погрешности предельную погрешность угла нужно выражать в радианах.

Так как $\lg \operatorname{ctg} a = -\lg \operatorname{tg} a$, то оценка ошибки для логарифма котангенса оказывается такой же, как для логарифма тангенса.

На основании сказанного можно сформулировать следующие правила рациональной организации вычислений при пользовании логарифмами четырех тригонометрических функций.

Если угол меньше $23^{\circ},5$, надо брать логарифм косинуса (перестроить, если можно, формулы); когда угол заключен между $23^{\circ},5$ и 30° , то выгоднее брать логарифмы синуса и косинуса; если дан угол между 30° и 60° , то можно пользоваться логарифмами всех четырех функций; при значениях угла от 60 до $66^{\circ},5$ следует брать логарифмы синуса и косинуса; наконец, если заданный угол больше $66^{\circ},5$, выгоден только логарифм синуса.

7. Определение углов по логарифмам тригонометрических функций

а) Если $\lg \sin u = a$, то $\sin u = 10^a$ и $u = \arcsin 10^a$, поэтому

$$\epsilon_u = \frac{10^a}{\sqrt{1-10^{2a}}} \cdot \frac{1}{0,4343} \epsilon_a \quad (3.25a)$$

или

$$\epsilon_u = \frac{\operatorname{tg} u}{0,4343} \epsilon_a. \quad (3.256)$$

Как и в предыдущих параграфах, критерием выгоды при приближенном вычисления мы будем считать выполнение условия $\epsilon_u \leq \epsilon_a$. Это условие приводит в рассматриваемом случае к неравенству $u \leq 23^{\circ},5$. Таким образом, вычислять угол по логарифму его синуса выгоднее при углах меньше $23^{\circ},5$. Аналогично можно показать, что вычислять угол по десятичному логарифму косинуса выгодно, если угол больше $66^{\circ},5$.

Если необходимо вычислять угол именно по логарифму синуса или косинуса, то придерживаются правила: углы меньше 45° вычисляют (если это возможно) по логарифму синуса, а углы больше 45° — по логарифму косинуса. При этом углы от $23^{\circ},5$ до $66^{\circ},5$ получают с большей предельной погрешностью, чем погрешность использованного логарифма. В частности, если угол равен 45° , то погрешность угла в 2,3 раза больше погрешности логарифма синуса (или косинуса).

Пример. Дано $\lg \sin u = 9,3847 - 10$. По 4-значным таблицам логарифмов тригонометрических функций найдем $u = 14^{\circ} 02'$. По таблицам натуральных значений тригонометрических функций $\operatorname{tg} u = 0,2425$. Поэтому

$$\epsilon_u = \frac{0,2425}{0,4343} \cdot \frac{1}{2} \cdot 10^{-4} = 3 \cdot 10^{-5}.$$

Оценка ошибки угла получена в радианах; переход к градусной мере дает $\epsilon_u = 6'' = 0',1$. Так как угол достаточно мал, то он может быть вычислен с точностью до $0',1$. Интерполяцией получаем $u = 14^{\circ} 02',0$,

б) Если $\lg \operatorname{tg} u = a$, то $\operatorname{tg} u = 10^a$, $u = \operatorname{arctg} 10^a$.

$$\epsilon_u = \frac{10^a}{1 + 10^{2a}} \cdot \frac{\epsilon_a}{0,4343} \quad (3.26a)$$

или

$$\epsilon_u = \frac{\sin 2u}{0,8686} \epsilon_a. \quad (3.26b)$$

Условие выгодности вычисления угла по логарифму тангенса будет $\sin 2u \leq 0,8686$,

т. е. $2u < 60^\circ$ или $2u > 120^\circ$. Отсюда следует, что по логарифму тангенса выгодно определять углы от 0 до 30° и от 60 до 90° . В области от 30 до 60° погрешность угла больше погрешности использованного логарифма тангенса, но это превышение невелико: если угол близок к 45° , то его предельная погрешность в 1,15 раза больше погрешности логарифма тангенса, по которому определяется угол.

Так как $\lg \operatorname{ctg} u = -\lg \operatorname{tg} u$, то предельная погрешность угла по логарифму котангенса вычисляется так же, как и в предыдущем случае.

Сравнение погрешностей при определении угла по логарифмам четырех тригонометрических функций показывает, что выгоднее всего определять угол по логарифму тангенса или котангенса, так как во всей области от 0 до 90° предельная погрешность угла не превышает произведения предельной погрешности заданного логарифма на число 1,15, а в двух третях этой области (от 0 до 30° и от 60 до 90°) предельная погрешность угла меньше предельной погрешности логарифма тангенса. Поэтому и при логарифмических вычислениях формулы и схемы подбираются по возможности так, чтобы углы вычислялись по логарифмам тангенсов (или котангенсов).

§ 13. Погрешность функции нескольких аргументов

Для сокращения записи мы ограничимся отысканием погрешности функции двух аргументов. Формула, которая будет выведена, легко обобщается на случай произвольного количества аргументов.

Пусть $U = f(x, y)$ есть непрерывная дифференцируемая функция в рассматриваемой области значений аргументов x и y . Предположим, что в U вместо точных значений аргументов подставлены их приближенные значения a и b ; тогда мы получим приближенное значение функции $u = f(a, b)$. Требуется вычислить предельную абсолютную погрешность приближенного значения функции, считая известными предельные абсолютные погрешности аргументов ϵ_a и ϵ_b .

Обозначим точные ошибки аргументов через Δ_a, Δ_b . Из определения точной ошибки (1.1) следуют равенства:

$$A = a + \Delta_a, \quad B = b + \Delta_b,$$

где A и B — точные значения аргументов. Точное значение функции

$$U = f(a + \Delta_a, b + \Delta_b).$$

Полагая, что точные ошибки суть малые числа, квадратами и более высокими степенями которых можно пренебречь, разложим правую часть последнего равенства по степеням точных ошибок и оборвем разложение на первых степенях ошибок. Тогда получим

$$\Delta_u = U - u = \left(\frac{\partial U}{\partial x} \right)_{\substack{x=a \\ y=b}} \cdot \Delta_a + \left(\frac{\partial U}{\partial y} \right)_{\substack{x=a \\ y=b}} \cdot \Delta_b. \quad (3.27)$$

В этом равенстве Δ_u представляет приближенную величину точной ошибки, так как в разложении сохранены только первые два члена. Из написанного равенства следует, что

$$|\Delta_u| \leq \left| \frac{\partial U}{\partial x} \right|_{\substack{x=a \\ y=b}} \cdot |\Delta_a| + \left| \frac{\partial U}{\partial y} \right|_{\substack{x=a \\ y=b}} \cdot |\Delta_b|.$$

Заменяя модули точных ошибок их предельными абсолютными погрешностями, получим верхнюю границу модуля ошибки функции — величину, которую можно принять за предельную абсолютную погрешность функции:

$$\varepsilon_u = \left| \frac{\partial U}{\partial x} \right|_{\substack{x=a \\ y=b}} \cdot \varepsilon_a + \left| \frac{\partial U}{\partial y} \right|_{\substack{x=a \\ y=b}} \cdot \varepsilon_b. \quad (3.28)$$

Эту формулу иногда называют *дифференциальной формулой оценки ошибки*, так как правая часть аналогична выражению полного дифференциала. Действительно, если в таком выражении дифференциалы аргументов заменить на предельные абсолютные погрешности, и частные производные функции заменить их модулями, то получится выведенная формула. Она используется как для расчета оценки в конкретных числовых задачах, так и для анализа точности и выяснения условий, при которых точность результата будет выше, т. е. предельная абсолютная погрешность функции меньше.

Заметим, что иногда для рассматриваемой задачи пишут формулу, с которой мы начали, т. е. первые два члена разложения; это имеет смысл только в тех случаях, когда считаются заданными точные ошибки аргументов; тогда формула дает приближенное значение ошибки функции. Если же точные ошибки аргументов неизвестны, как это бывает по большей части, а известны только предельные погрешности, то следует применять выведенную в этом параграфе формулу.

Рассмотрим некоторые примеры применения полученной формулы.

Пример 1. Простейшая формула для определения времени захода и восхода звезд имеет вид

$$t = \arccos(-\operatorname{tg} \varphi \operatorname{tg} \delta),$$

где φ — широта места, δ — склонение звезды, t — часовой угол, подлежащий определению; при пользовании этой формулой мы предполагаем $|\operatorname{tg} \varphi \operatorname{tg} \delta| < 1$, т. е. восход и заход существуют.

По выведенной выше формуле получаем следующее выражение для предельной абсолютной погрешности часового угла:

$$\varepsilon_t = \frac{|\sin \delta \sec \varphi| \varepsilon_\varphi + |\sin \varphi \sec \delta| \varepsilon_\delta}{\sqrt{\cos(\varphi + \delta) \cos(\varphi - \delta)}}.$$

Симметрия формулы относительно φ и δ объясняется тем, что этим свойством обладает и та формула, по которой производится вычисление.

Сделаем расчет предельной погрешности для вычисления часового угла восхода Марса в Симферополе 3 февраля 1948 г.

По таблице широт крупных городов Советского Союза, помещенной в «Курсе общей астрономии» С. Н. Блажко, находим $\varphi = 44^\circ 57'$; по Астрономическому календарю ВАГО на 1948 г. получим, что $\delta = 14^\circ 20'$. Так как эти числа берутся из более точных таблиц с округлением, то можно считать, что $\varepsilon_\varphi = \varepsilon_\delta = 0',5$ или $\varepsilon_\varphi = \varepsilon_\delta = 0,00015 \text{ рад}$. (Напомним, что в этих расчетах предельная погрешность угла должна быть дана в радианах, поэтому величину $0',5 = 30''$ надо было разделить на число секунд в радиане, т. е. на число 206 265; вместо него было взято 200 000, что законно, так как такая замена не уменьшает, а увеличивает предельную погрешность.) Все множители для упрощения расчета следует заменять возможно более простыми, причем так, чтобы погрешность не уменьшалась. Легко видеть, что в нашей задаче все углы нужно увеличить, поэтому принимаем: $\varphi = 45^\circ$, $\delta = 15^\circ$, $\varphi + \delta = 60^\circ$, $\varphi - \delta = 31^\circ$. С этими значениями углов, пользуясь двухзначными или трехзначными таблицами тригонометрических функций или логарифмической линейкой, получим

$$\varepsilon_t = 0,37 \varepsilon_\varphi + 0,74 \varepsilon_\delta = 0,00018 \text{ рад}.$$

Отсюда следует, что часовой угол вычисляется по нашей формуле с предельной абсолютной погрешностью, равной примерно $36''$.

Рассмотрим теперь вопрос о наиболее благоприятных условиях наблюдений, т. е. таких, при которых предельная погрешность была бы возможно меньше. Так как ни выбор места наблюдения, ни выбор небесного тела не может быть произвольным в рассматриваемой задаче, то решение вопроса может быть только, так сказать, пассивным; это значит, что можно только установить, для каких небесных тел и для каких широт можно вычислить часовой угол восхода и захода светила с меньшей точностью и для каких — с большей. Поскольку оба аргумента φ и δ входят симметрично, то достаточно исследовать влияние на результат только одного аргумента. Легко видеть, что при возрастании φ от 0 до 90° и при постоянном δ коэффициенты при ε_φ и при ε_δ монотонно растут. Аналогичное заключение можно сделать при условии, что δ монотонно растет, а φ остается постоянным. Из этого следует, что с наименьшей погрешностью часовые углы восхода и захода светила определяются вблизи земного экватора и для светил, близких к небесному экватору.

Пример 2. Определение астрономической широты места наблюдения опирается на основную формулу:

$$\cos z = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos t,$$

где z — геоцентрическое зенитное расстояние используемого светила, φ — известная широта, δ — склонение светила,

$$t = \tau - \alpha, \quad \tau = T + u;$$

α — прямое восхождение светила, T — отсчет хронометра, u — поправка хронометра. Склонение и прямое восхождение берутся из Ежегодника на момент наблюдения, причем должны быть взяты геоцентрические значения. Предположим, что α и δ не содержит ошибок, точнее, что ошибки координат гораздо меньше ошибок тех величин, которые получаются из наблюдений.

Поэтому будем считать, что ошибки содержатся только в измеренном зенитном расстоянии и в величине τ , так как невозможно отделить ошибку отсчета хронометра от ошибки в поправке хронометра. Вместо того, чтобы выражать определяемую широту явно через другие величины, мы поступим следующим образом: продифференцируем обе части формулы; после дифференцирования и замены дифференциалов точными ошибками получим

$$-\sin z \Delta_z = \cos \varphi \sin \delta \Delta_\varphi - \sin \varphi \cos \delta \cos t \Delta_\tau - \cos \varphi \cos \delta \sin t \Delta_\tau.$$

Из сферического треугольника с вершинами в зените, полюсе и звезде получаются формулы

$$\cos \varphi \sin \delta - \sin \varphi \cos \delta \cos t = -\sin z \cos A,$$

$$\cos \delta \sin t = \sin z \sin A,$$

где A — азимут.

Используя эти формулы, запишем выражение для точной ошибки широты в достаточно простом виде

$$\Delta_\varphi = \sec A \Delta_z - \cos \varphi \operatorname{tg} A \Delta_\tau.$$

Замена ошибок предельными погрешностями и введение модулей всех множителей дает формулу для вычисления предельной абсолютной погрешности широты

$$\epsilon_\varphi = |\sec A| \epsilon_z + \cos \varphi |\operatorname{tg} A| \epsilon_\tau.$$

Способ применения этой формулы в частных случаях очевиден, поэтому числового примера мы рассматривать не будем. Из формулы легко вывести условие минимума предельной погрешности широты: предельная абсолютная погрешность широты будет иметь наименьшее значение, если звезда находится в меридиане ($A = 0^\circ$ или 180°). В этом случае коэффициент при ϵ_z имеет минимальное значение — единицу, а коэффициент при ϵ_τ обращается в нуль.

Из сказанного можно извлечь два практических указания: 1) наблюдения для определения широты надо производить, когда звезда находится вблизи меридиана (не так легко обеспечить наблюдения точно в меридиане); 2) весьма важно, чтобы зенитное расстояние наблюдалось возможно точнее, так как при всех обстоятельствах имеем очевидное неравенство $\epsilon_\varphi \gg \epsilon_z$, причем знак равенства будет только при наблюдениях в меридиане.

Пример 3. Для определения поправки часов используется та же формула, что и для определения широты, но при этом считается, что широта известна.

Формулу для вычисления поправки часов следует написать в виде

$$\cos z = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos (T - \alpha + u)$$

(обозначения те же, что и в предыдущем примере).

Как и в предыдущем примере, не будем учитывать ошибок сферических экваториальных координат. В формуле предыдущего параграфа для расчета ошибки сделаем замену

$$\Delta_\tau = \Delta_T + \Delta_u$$

и перепишем ее так, чтобы выделить Δ_u . Тогда

$$\Delta_u = -\Delta_T - \sec \varphi \operatorname{ctg} A \Delta_\varphi.$$

Переход к предельным абсолютным погрешностям дает формулу.

$$\epsilon_u = \epsilon_T + \sec \varphi | \operatorname{ctg} A | \epsilon_\varphi + \sec \varphi | \operatorname{cosec} A | \epsilon_\pi.$$

Вычисление предельной погрешности по этой формуле не представляет никаких затруднений, причем, как обычно в таких вычислениях, можно брать грубо приближенные значения углов A и φ , но так, чтобы предельная погрешность не уменьшалась, а увеличивалась.

Из полученной формулы ясно видно, что определение поправки часов будет сделано с минимальной погрешностью, если светило находится в первом вертикале (азимут равен 90 или 270°). Минимальное значение ϵ_u равно $\epsilon_T + \epsilon_\pi \sec \varphi$.

Пример 4. Ускорение силы тяжести определяется с помощью оборотного маятника по формуле

$$g = \pi^2 \frac{l}{P^2};$$

l — приведенная длина маятника, P — период колебания. Наблюдения дали значения величин и их предельные погрешности:

$$l = 50,02 \text{ см}, \quad \epsilon_l = 0,01 \text{ см},$$

$$P = 0,7098 \text{ сек}, \quad \epsilon_P = 0,0001 \text{ сек}.$$

Вычислим ускорение силы тяжести g и его предельную погрешность; значение π надо принять равным $3,1416$, т. е. положить $\epsilon_\pi = 0,00005$.

Применяя основную формулу к нашему случаю, получим

$$\epsilon_g = \frac{2\pi l}{P^3} \epsilon_\pi + \frac{\pi^2}{P^3} \epsilon_l + \frac{2\pi^3 l}{P^5} \epsilon_P,$$

которую удобнее переписать в виде

$$\epsilon_g = \pi P^{-3} (2Pl\epsilon_\pi + \pi P\epsilon_l + 2\pi l\epsilon_P).$$

Для упрощения примем в скобке $P=1$, $l=50$, $\pi=4$, $2\pi=7$, а множитель перед скобкой положим равным 10 по следующей прикидке: $\pi=3,14$, $0,7^3=0,343$, поэтому произведение имеет значение около девяти, но мы несколько уменьшили величину, поэтому правильнее будет принять значение 10 ; тогда получим $\epsilon_g = 10 \cdot (100 \cdot 5 \cdot 10^{-5} + 4 \cdot 10^{-2} + 7 \cdot 50 \cdot 10^{-4})$ или $\epsilon_g = 0,8 \text{ см/сек}^2$. Погрешность заметно преувеличена, но зато результат получается весьма просто, весь расчет можно сделать в уме. Более аккуратное вычисление, приведенное в книге К. П. Яковлева «Математическая обработка результатов измерений», откуда заимствован этот пример, дает оценку ошибки $0,5 \text{ см/сек}^2$. После выполнения вычислений получим $g = 978,0$. Если учесть полученную здесь оценку ошибки, то следует принять

$$g = 978 \text{ см/сек}^2; \quad \epsilon_g = 1 \text{ см/сек}^2.$$

Пример 5. При определении орбиты малой планеты по трем наблюдениям приходится вычислять среднюю аномалию M , зная эксцентриситет e и эксцентрическую аномалию E . Это вычисление делается при помощи уравнения Кеплера:

$$M = E - e \sin E.$$

Примем $E = 43^\circ 35' 16''$, $e = 0,14136$. Требуется вычислить среднюю аномалию M и оценить ее предельную абсолютную погрешность.

Пусть в значениях E и e верны все написанные значащие цифры и известно, что эти числа получены с округлением. Предельная абсолютная погрешность эксцентриситета равна при этих условиях 0,000005; погрешность эксцентрической аномалии равна $0'',5$; обращая ее в радианы (для чего можно разделить на 200 000, вместо деления на 206 265; см. выше, стр. 44). Мы получим $\epsilon_E = 0,0000025$. Предельную погрешность средней аномалии вычисляем по формуле

$$\epsilon_M = (1 - e \cos E) \epsilon_E + |\sin E| \epsilon_e.$$

В этой формуле нет надобности писать модуль производной по эксцентрической аномалии, так как величина, стоящая в скобке, наверно положительна, ибо $e < 1$, а модуль косинуса не превышает единицы.

Для расчета предельной погрешности примем $\cos E = 0,72$ (с недостатком), $e = 0,14$ (с недостатком), $\sin E = 0,70$ (с избытком); тогда $e \cos E = 0,10$, $1 - e \cos E = 0,9$,

$$\epsilon_M = 0,9 \cdot 2,5 \cdot 10^{-6} + 0,7 \cdot 5 \cdot 10^{-6}.$$

Незначительно увеличивая предельную погрешность, получим $\epsilon_M = 0,000006$ рад. Перевод оценки в секунды дуги умножением на 206 265 или на большее число дает $\epsilon_M = 1'',3$.

Сделаем одно замечание. Для вычисления M либо нужно выразить в радианах угол E , либо выразить в секундах дуги (или градусах) эксцентриситет. Это объясняется тем, что уравнение Кеплера $M = E - e \sin E$ есть формула, выведенная средствами анализа, и потому углы в ней естественно выражены в радианах. Чтобы иметь возможность пользоваться обыкновенными таблицами тригонометрических функций, в которых аргумент выражен в градусной мере, нужно все члены уравнения умножить на число секунд (или градусов) в радиане. После этого углы E и M выражаются в секундах, а в члене $e \sin E$ переводной множитель можно отнести либо к e , либо к $\sin E$; принято формально считать, что эксцентриситет есть угол в радианах, поэтому его обращают в секунды дуги. Выражая эксцентриситет в секундах умножением на 206 265, получим в нашей задаче $e = 29\,158''$. Умножая затем это число на пятизначную величину синуса 0,68947, получим $(E - M)'' = 20\,104''$; поэтому $E - M = 5^\circ 35' 04''$. Окончательно $M = 38^\circ 00' 12''$, $\epsilon_M = 1'',3$. Следовательно, в величине средней аномалии нельзя ручаться за целые секунды: в них возможна ошибка в одну секунду с плюсом или минусом (может быть $11''$, $12''$ или $13''$).

Пример 6. Вычислим полную поверхность P прямого кругового конуса, радиус основания которого $r = 3,4$ см, образующая $l = 7,6$ см, и определим ее предельную относительную погрешность результата P .

Полная поверхность определяется по формуле

$$P = \pi r^2 + 2\pi r l.$$

Формула для вычисления предельной абсолютной погрешности имеет вид

$$\epsilon_P = (r^2 + 2rl) \epsilon_r + \pi (2r + 2l) \epsilon_r + 2\pi r \epsilon_l.$$

Примем $\pi = 3,14$, предельные абсолютные погрешности аргументов будут

$$\epsilon_r = 0,5 \cdot 10^{-2}, \quad \epsilon_r = \epsilon_l = 0,5 \cdot 10^{-1}.$$

Если коэффициенты при предельных погрешностях аргументов вычислять с заданными значениями их, то

$$\varepsilon_P = 64\varepsilon_x + 70\varepsilon_r + 22\varepsilon_l \quad \text{или} \quad \varepsilon_P = 5 \text{ см}^3.$$

Таким образом, в значении полной поверхности можно ручаться только за десятки см^3 . Вычисления дают $P = 200 \text{ см}^3$, $\varepsilon_P = 5 \text{ см}^3$. В P верны знаки 2 и 0; $\delta_P = 2,5\%$.

§ 14. Понятие об обратной задаче теории приближенных вычислений

Пусть U — функция n независимых переменных x, z, \dots, w , приближенные значения которых можно несколько варьировать по их точности.

Предположим, что нам заранее задана нужная точность функции, т. е. задана предельная абсолютная погрешность величины u . Требуется определить предельные абсолютные погрешности аргументов так, чтобы обеспечить заданную погрешность функции. Такую задачу мы условимся называть *обратной задачей приближенных вычислений*.

Если число аргументов функции больше одного, то задача, очевидно, не будет определенной, так как задано только одно число (погрешность ε_u), а определить нужно столько неизвестных предельных погрешностей, сколько имеется аргументов. В практических задачах употребляется нередко соглашение, которое называют *методом равных влияний*. Смысл этого соглашения заключается в следующем. Напишем (согласно 3.28) выражение предельной погрешности функции через предельные ошибки аргументов:

$$\varepsilon_u = \left| \frac{\partial U}{\partial x} \right|_{\substack{x=a \\ y=b \\ \dots \\ w=t}} \varepsilon_a + \left| \frac{\partial U}{\partial y} \right|_{\substack{x=a \\ y=b \\ \dots \\ w=t}} \varepsilon_b + \dots + \left| \frac{\partial U}{\partial w} \right|_{\substack{x=a \\ y=b \\ \dots \\ w=t}} \varepsilon_w.$$

Метод равных влияний состоит в том, что вводится соглашение так подбирать предельные погрешности аргументов, чтобы все члены в правой части имели одинаковые значения:

$$\left| \frac{\partial U}{\partial x} \right| \varepsilon_a = \left| \frac{\partial U}{\partial y} \right| \varepsilon_b = \dots = \frac{\varepsilon_u}{n}, \quad (3.29)$$

где n — число аргументов. Эти равенства и дадут выражения для предельных погрешностей аргументов:

$$\varepsilon_a = \frac{\varepsilon_u}{n \left| \frac{\partial U}{\partial x} \right|}, \quad \varepsilon_b = \frac{\varepsilon_u}{n \left| \frac{\partial U}{\partial y} \right|}, \dots \quad (3.30)$$

Полученные по этому способу предельные погрешности могут оказаться не все достижимыми по условиям наблюдений. В таких

случаях, очевидно, придется несколько отступать от соглашения, но все-таки следует брать точность возможно более близкую к той, которая получается по методу равных влияний. Вряд ли может быть получено обоснование для этого метода; он является только удобным соглашением.

Вторым вариантом метода равных влияний является соглашение о том, чтобы были равны между собой все члены в выражении для предельной относительной погрешности функции через предельные относительные погрешности аргументов. Это приводит к равенствам

$$\delta_u = \left| \frac{a}{u} \right| \left| \frac{\partial U}{\partial x} \right|_{\substack{x=a \\ y=b \\ \dots \\ w=m}} \cdot \delta_a + \left| \frac{b}{u} \right| \left| \frac{\partial U}{\partial y} \right|_{\substack{x=a \\ y=b \\ \dots \\ w=m}} \cdot \delta_y + \dots; \quad (3.31)$$

$$\left. \begin{aligned} \delta_a &= \frac{\delta_u}{n \left| \frac{a}{u} \right| \left| \frac{\partial U}{\partial x} \right|}, \\ \delta_b &= \frac{\delta_u}{n \left| \frac{b}{u} \right| \left| \frac{\partial U}{\partial y} \right|}. \end{aligned} \right\} \quad (3.32)$$

Легко видеть, что выражения для предельных относительных погрешностей аргументов формально эквивалентны выражениям для предельных абсолютных погрешностей, так как элементарными преобразованиями можно получить выражение для δ_a из выражения для ϵ_a и т. д. В приближенных вычислениях, однако, нередко бывает так, что практически не все равно, как считать, поскольку две формально эквивалентные формулы могут дать различные и по величине и по точности числа вследствие того, что мы по необходимости употребляем ограниченное число знаков. Поэтому в более важных задачах нужно попробовать написать оба выражения и сравнить удобство и точность вычислений по обеим формулам; выбирать надо такую, которая дает некоторые преимущества в работе.

Пример. Объем конуса вычисляется по известной формуле

$$V = \frac{1}{3} \pi r^2 h,$$

где r — радиус основания, h — высота конуса. Пусть приближенные значения этих величин будут 3,2 см и 4,7 см соответственно. Дифференциальная формула имеет вид

$$3\epsilon_V = r^2 h \epsilon_r + 2\pi r h \epsilon_r + \pi r^2 \epsilon_h.$$

Нам задано ϵ_V ; по методу равных влияний

$$r^2 h \epsilon_r = \epsilon_V, \quad 2\pi r h \epsilon_r = \epsilon_V, \quad \pi r^2 \epsilon_h = \epsilon_V.$$

Определяя по этим равенствам предельные абсолютные погрешности аргументов, можно их уменьшать, т. е. увеличивать значения частных производных. Для расчетов поэтому примем

$$r = 3,5, \quad r^2 = 11, \quad h = 5, \quad \pi h = 16, \quad \pi r^3 = 32.$$

Тогда получим

$$\varepsilon_{\pi} = \frac{\varepsilon_V}{55}, \quad \varepsilon_r = \frac{\varepsilon_V}{112}, \quad \varepsilon_h = \frac{\varepsilon_V}{32}.$$

Если, например, $\varepsilon_V = 1 \text{ см}^3$, то $\varepsilon_{\pi} \approx 0,02$, $\varepsilon_r \approx 0,01 \text{ см}$, $\varepsilon_h \approx 0,03 \text{ см}$. Здесь мы несколько отступили от общего правила, взяв несколько больше того, что можно было ($0,02$ вместо $\frac{1}{55}$ и т. д.). Нужная точность измерений может быть достигнута.

ЧАСТЬ II

ТОЧЕЧНОЕ ИНТЕРПОЛИРОВАНИЕ

ГЛАВА 4

ОБЩИЕ СВЕДЕНИЯ

§ 15. Приближение табличных функций; понятие о точечной интерполяции

При исследовании явлений природы с помощью математического аппарата используются различные функции.

Функции могут задаваться разными способами. Простейший из них — задание аналитического выражения, которое дает возможность по любым (допустимым) значениям аргумента или аргументов вычислить значение функции. Такие случаи на практике бывают редко, однако и в этом случае операции могут быть весьма громоздкими. Поэтому использование заданного аналитического выражения для прямого вычисления значений функции при любых значениях аргумента может оказаться затруднительным.

Часто функции определяются бесконечными рядами. Вычисление значений функции с помощью бесконечного ряда является довольно громоздкой операцией, требующей исследования сходимости и определения числа членов, необходимых для обеспечения заданной точности. Поэтому пользоваться рядом для вычисления значений функции при произвольном значении аргумента затруднительно.

Функция может быть также определена неопределенным интегралом или дифференциальным уравнением. Эти интегралы или решения дифференциальных уравнений в физических задачах иногда могут быть выражены в конечном виде, но и в этом случае они могут оказаться громоздкими. В качестве примера рассмотрим функцию $f(x)$, определенную равенством:

$$f(x) = \int \frac{dx}{(x^2 + 1)^3 (x^2 + x + 1)^2}.$$

Как известно, такой интеграл выражается в конечном виде через элементарные функции, но это выражение будет столь громоздким, что вряд ли есть смысл вычислять $f(x)$ по точной формуле всякий

раз, когда это потребуется. Обычно же функции, определяемые интегралами или дифференциальными уравнениями, не выражаются через элементарные функции в конечном виде; численные значения такой функции приходится определять тем или иным приближенным способом для отдельных значений аргумента.

Во всех тех случаях, когда значения функции либо невозможно точно вычислить, либо вычисление слишком громоздко, прибегают к составлению таблиц функции, если данная функция встречается в разных задачах не очень редко.

Следует отметить еще один способ задания функции, прямо приводящий к таблице значений. Это тот случай, когда из общих физических соображений можно утверждать, что некоторая величина представляет собой функцию одного или нескольких аргументов *), но явление не настолько хорошо изучено, чтобы связь эту можно было облечь в математическую форму. В таких случаях производятся наблюдения, которые дают таблицу значений функции при разных значениях аргумента. По большей части такая таблица может быть получена только для довольно ограниченного числа значений аргумента, причем расширение таблицы либо невозможно, либо представляет большие затруднения.

Во всех перечисленных выше случаях и подобных им мы приходим в конечном счете к табличному заданию функции: функция $x = f(t)$ определена таблицей своих значений x_k при заданных значениях аргумента $t_k (k = 1, 2, \dots, n)$. Табличные значения функции и аргумента называются *узлами таблицы*. Если построить график табличной функции одного аргумента, то точки графика также называют *узлами*.

Значения t_k, x_k даются с определенным числом верных знаков, чаще с определенным числом знаков после запятой. Если таблица вычислена с помощью точного аналитического выражения или способом, допускающим оценку погрешности (например, с помощью бесконечного ряда), то все знаки в табличных значениях функции, кроме последнего, можно считать правильными; последний знак может отличаться не больше, чем на единицу, от того, который был бы, если бы вычисления велись с большим числом знаков.

*) Явления природы в общем весьма сложны, связи между величинами многообразны, и редко бывает так, что можно считать некоторую величину точно функцией одного или двух-трех аргументов. Когда наблюдаемая величина считается функцией только одного наблюдаемого аргумента, то это почти всегда означает, что этот единственный аргумент в основном определяет значение функции, влиянием же других аргументов на величину функции при той точности, с которой производятся наблюдения, можно пренебречь. Например, гелиоцентрические координаты малой планеты или кометы в течение небольшого промежутка времени можно считать функциями оскулирующих элементов и времени и вычислять эти координаты по формулам задачи двух тел, так как возмущения невелики и ими можно пренебречь.

Например, с семью знаками после запятой $\sin 22^\circ 19' = 0,4104697$; при округлении до пяти знаков после запятой получается пятый знак, отличающийся на единицу от соответствующего знака в более точном значении функции именно $\sin 22^\circ 19' = 0,41047$.

Если таблица функции получена численным решением системы дифференциальных уравнений (или одного уравнения), то положение осложняется, поскольку пока нет достаточно надежного и простого способа оценки погрешности численного решения дифференциального уравнения. Еще сложнее обстоит дело, когда табличная функция получена из наблюдений. Значения функции содержат в этом случае ряд ошибок разного происхождения, и потому функциональная связь искажается: если наблюдения дают несколько значащих цифр, то одна, а иногда и две-три последние цифры ненадежны. К этому случаю нельзя, вообще говоря, применить соглашение об определении предельной ошибки по записи приближенного числа (см. часть I*).

Итак, функция очень часто задается таблицей, т. е. бывает задана последовательность значений функции при различных значениях аргументов. Обычно таблица располагается так, что аргумент возрастает. Значения аргумента задаются с некоторыми промежутками, которые называют *шагами*. Стараются строить таблицы с постоянным шагом, но это не всегда удобно, так как требует излишнего количества работы и не приносит пользы, если есть области, в которых функция меняется медленно или почти по линейному закону. В таких случаях полную табличную область значений аргумента разбивают на части и в каждой части выбирают свой постоянный шаг. Шаги эти, вообще говоря, не могут быть малыми: иначе объем таблицы был бы слишком велик.

При решении задач естествознания, как правило, приходится иметь дело со случаями, когда нужны значения функции не только для табличных значений аргумента. Так, например, очень часто требуются координаты Солнца относительно центра Земли, но почти всегда не в 0^й мирового времени, как дается в Ежегодниках, а в самые различные моменты, промежуточные между табличными.

Поэтому имеет очень большое практическое значение следующая задача: дана табличная функция; необходимо найти способ приближенного определения значений функции для произвольных значений аргумента, не совпадающих с табличными значениями аргумента.

Если значение аргумента задано в у т р и области табличных значений аргумента, то указанную задачу называют задачей *интерполяции*

*) Соглашение о том, что предельная погрешность меньше половины единицы последнего знака, относится только к ошибкам округления при измерениях и вычислениях при условии, что единицы последнего знака определяются по шкале, а не на глаз. (Так принято говорить, если место указателя между делениями шкалы определяется глазомерной оценкой в десятых долях промежутка.)

(inter — между), если же значение аргумента задано вне табличной области, то говорят об *экстраполяции*.

Простейший способ интерполяции, применяемый иногда и сейчас, заключается в том, что через точки на графике, изображающие узлы таблицы, проводят плавную кривую «от руки». Эта кривая принимается за приближенный график функции, и по этому графику делается интерполяция очевидным способом. Точность этого способа сильно ограничена, как и точность всякого графика; кроме того, имеется произвол и неопределенность при проведении кривой. Поэтому такой способ не применяется к табличным функциям, за все знаки которых можно ручаться, ибо получилась бы «потеря точности», т. е. интерполированные значения были бы менее точными, чем табличные. Графическую интерполяцию можно применять только в тех случаях, когда таблица получена из недостаточно надежных наблюдений, и сама функциональная связь недостаточно надежна.

Разработка способа интерполяции по табличной функции уже во время Ньютона (и раньше) была сведена к приближению табличной функции другой, легко вычисляемой функцией. Построенное приближенное выражение данной функции и используется для интерполяции: заданное значение аргумента подставляется в нее и производятся вычисления. Обычно строится приближенное представление небольшой части таблицы в окрестности заданного значения аргумента.

Построение приближения табличной функции представляет неопределенную задачу и невозможно без предварительного введения двух соглашений или допущений.

Во-первых, нужно условиться относительно класса функций, используемых для приближения. В практических задачах естественным главным условием является возможность легкого вычисления функции по заданному значению аргумента. Этому условию удовлетворяют алгебраические полиномы, которые почти исключительно и используются для приближения. Если функция обладает свойствами, которым алгебраические полиномы не удовлетворяют, то используют другие функции. Чаще всего пользуются тригонометрическими полиномами, если функция периодическая и приближение нужно в области, охватывающей весь период. Если функция растет быстрее полинома, то естественнее было бы пользоваться показательными функциями, но это на практике редко делается. Когда периодическую функцию нужно приблизить только на небольшой части периода, то обыкновенно используются алгебраические полиномы. Выбранная приближающая функция должна содержать некоторое количество буквенных параметров, которые должны быть определены по заданной таблице. Если для приближения используется алгебраический полином степени n , то такими параметрами являются $n+1$ коэффициентов полинома.

Естественно требовать, чтобы приближение было возможно лучшим, но смысл слова «лучше» нужно определить. Это и будет вторым условием, накладываемым на задачу.

Вопрос о существовании интерполяционного полинома сводится к вопросу об обращении в нуль определителя этой системы, имеющего вид

$$W = \begin{vmatrix} 1 & t_0 & t_0^2 & \dots & t_0^n \\ 1 & t_1 & t_1^2 & \dots & t_1^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \dots & t_n^n \end{vmatrix} \quad (4.4)$$

Из курса высшей алгебры известно, что этот определитель называется определителем Вандермонда; его значение определяется формулой

$$W = [(t_n - t_0)(t_n - t_1) \dots (t_n - t_{n-1})] \times \\ \times [(t_{n-1} - t_0)(t_{n-1} - t_1) \dots (t_{n-1} - t_{n-2})] \times \dots \\ \dots \times [(t_2 - t_0)(t_2 - t_1)](t_1 - t_0).$$

По условию $t_k \neq t_s$, если $k \neq s$, поэтому

$$W \neq 0. \quad (4.5)$$

Так как определитель линейной системы не равен нулю, то эта система имеет единственное определенное решение a_0, a_1, \dots, a_n ; следовательно, интерполяционный полином существует и он только один.

Заметим, что доказанная теорема может считаться единственным основанием для формального применения способа точечной интерполяции к любой функции, поскольку в этой теореме t_k, x_k — произвольные числа. Вопрос об удовлетворительности приближения полиномом решается по свойствам функции с помощью оценки ошибки интерполяции. Иногда для обоснования (или оправдания) способа ссылаются на возможность приближения функции с помощью частной суммы степенного ряда или на теорему Вейерштрасса о возможности приблизить непрерывную функцию полиномом с любой степенью точности. Однако ни то, ни другое не имеет прямого отношения к задаче точечной интерполяции. В самом деле, действительно, никакая частная сумма степенного ряда не может точно представить функцию во всех узлах, ибо если коэффициенты полинома определены по принципу точечной интерполяции, то они не равняются соответствующим коэффициентам ряда.

Для иллюстрации этого утверждения предположим, что дана таблица значений периодической функции с шагом, равным периоду, например таблица $\sin t$ для $t = 0, 2\pi, 4\pi$ и т. д. Ясно, что построение интерполяционного полинома даст постоянную (нуль); мы получим начальный член ряда, но приближение нельзя назвать удовлетворительным. Таким образом, возможность построения степенного

ряда для этой функции не имеет прямого отношения к задаче интерполяции.

Теорема Вейерштрасса утверждает, что для заданной функции $x(t)$ и произвольного положительного числа ε можно найти полином $P_n(t)$, удовлетворяющий на интервале $a \leq t \leq b$ условию

$$|x(t) - P_n(t)| < \varepsilon,$$

но в этом условии и в доказательстве не предполагается, что $x(t_k) = P_n(t_k)$ при некоторой последовательности значений аргумента t_k ($k = 1, 2, \dots, n$).

§ 17. Интерполяционный полином Лагранжа

Для нахождения интерполяционного полинома достаточно решить в буквенном виде выведенную в § 14 систему уравнений, определяющую коэффициенты полинома. Если система решена с помощью определителей, то каждый коэффициент a_s определится формулой

$$a_s = \frac{W_s}{W}, \quad (4.6)$$

где

$$W_s = \begin{vmatrix} 1 & t_0 & t_0^2 & \dots & t_0^{s-1} & x_0 & t_0^{s+1} & \dots & t_0^n \\ 1 & t_1 & t_1^2 & \dots & t_1^{s-1} & x_1 & t_1^{s+1} & \dots & t_1^n \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & t_n & t_n^2 & \dots & t_n^{s-1} & x_n & t_n^{s+1} & \dots & t_n^n \end{vmatrix}, \quad (4.7)$$

($s = 0, 1, 2, \dots, n$).

а W определяется (4.4).

Если разложить W_s по элементам столбца x_0, x_1, \dots, x_n , то W_s будет линейным выражением вида

$$W_s = \sum_{k=0}^n x_k b_k, \quad (4.8)$$

где b_k — числа, зависящие от всех узловых значений t и от числа s . Тогда коэффициенты полинома примут вид

$$a_s = \sum_{k=0}^n x_k c_k, \quad (4.9)$$

где

$$c_k = \frac{b_k}{W}. \quad (4.10)$$

Подставляя все a_s согласно (4.9) в полином (4.1) и собирая члены, содержащие x_k , получим

$$P(t) = \sum_{k=0}^n x_k L_k(t), \quad (4.11)$$

где $L_k(t)$ — полиномы степени n . Лагранж показал, как можно получить величины $L_k(t)$, не решая системы уравнений для коэффициентов $a_0, a_1, a_2, \dots, a_n$.

Для выполнения условия точечной интерполяции полиномы $L_k(t)$ должны удовлетворять условиям:

$$P(t_k) = x_k \quad (k = 0, 1, \dots, n), \quad (4.12)$$

$$\left. \begin{aligned} L_k(t_k) &= 1, \\ L_k(t_s) &= 0, \quad \text{если } k \neq s; \\ k &= 0, 1, 2, \dots, n. \end{aligned} \right\} \quad (4.13)$$

Второе условие (4.13) означает, что корнями $L_k(t)$ являются все узловые значения t , кроме t_k . Поэтому

$$L_k(t) = A_k(t - t_0)(t - t_1) \dots (t - t_{k-1})(t - t_{k+1}) \dots (t - t_n),$$

где A_k — неизвестный коэффициент. Этот коэффициент легко определится из первого условия (4.13):

$$A_k(t_k - t_0)(t_k - t_1) \dots (t_k - t_{k-1})(t_k - t_{k+1}) \dots (t_k - t_n) = 1.$$

Таким образом, окончательно

$$L_k(t) = \frac{t - t_0}{t_k - t_0} \cdot \frac{t - t_1}{t_k - t_1} \dots \frac{t - t_{k-1}}{t_k - t_{k-1}} \frac{t - t_{k+1}}{t_k - t_{k+1}} \dots \frac{t - t_n}{t_k - t_n}.$$

Введем обозначение

$$L(t) = \prod_{s=0}^n (t - t_s).$$

Если продифференцировать $L(t)$ и подставить в производную $t = t_k$, то получим знаменатель последнего выражения для $L_k(t)$, поскольку производная будет состоять из произведений, из которых только одно не содержит $t - t_k$; после подстановки этот член даст знаменатель, а остальные произведения обратятся в нули. Поэтому

$$L_k(t) = \frac{L(t)}{(t - t_k) L'(t_k)}. \quad (4.14)$$

Таким образом, интерполяционный полином Лагранжа, содержащий непосредственно узловые точки (t_k, x_k) , имеет вид

$$P(t) = \sum_{k=0}^n x_k \frac{L(t)}{(t - t_k) L'(t_k)} \quad (4.15)$$

или, в развернутом виде,

$$P(t) = \sum_{k=0}^n x_k \frac{(t-t_0)(t-t_1)\dots(t-t_{k-1})(t-t_{k+1})\dots(t-t_n)}{(t_k-t_0)(t_k-t_1)\dots(t_k-t_{k-1})(t_k-t_{k+1})\dots(t_k-t_n)}. \quad (4.16)$$

Пример 1. Функция определена таблицей

t	0	1	2
x	0	1	4

Приблизить ее линейными функциями на участках t от 0 до 1 и от 1 до 2.
а) На участке от 0 до 1 используем конечные узлы этого участка. По формуле Лагранжа получим

$$P_I(t) = \frac{t-1}{0-1} \cdot 0 + \frac{t-0}{1-0} \cdot 1 = t.$$

б) На участке от 1 до 2 имеем

$$P_{II}(t) = \frac{t-2}{1-2} \cdot 1 + \frac{t-1}{2-1} \cdot 4 = 3t - 2.$$

Следовательно,

$$\begin{aligned} P(t) &= t, & \text{если } 0 \leq t \leq 1, \\ P(t) &= 3t - 2, & \text{если } 1 \leq t \leq 2. \end{aligned}$$

Таблица построена для функции $x = t^2$, поэтому можно сравнить построенное кусочно-линейное приближение с точным выражением $x = t^2$. На участке от 0 до 1 разность между точным значением и $P_I(t)$ $\xi_I = t^2 - t$.

Легко видеть, что ξ_I имеет минимум, равный $-0,25$ при $t = 0,5$; значит $|\xi_I| \leq 0,25$.

На участке от 1 до 2 $\xi_{II} = t^2 - 3t + 2$; ξ_{II} имеет минимум при $t = 1,5$ и этот минимум снова равен $-0,25$. Мы видим, что приближение заданной функции кусочно-линейной функцией, имеющей два звена, обеспечивает предельную погрешность 0,25 на всем интервале таблицы от 0 до 2.

Пример 2. Построить приближение функции $x = \sin t$ линейной функцией на участке t от 0 до $\frac{\pi}{4}$ по узлам на краях участка:

t :	0	$\frac{\pi}{4}$,
x :	0	$\frac{\sqrt{2}}{2}$.

По формуле Лагранжа получим

$$P(t) = \frac{t - \frac{\pi}{4}}{0 - \frac{\pi}{4}} \cdot 0 + \frac{t - 0}{\frac{\pi}{4} - 0} \cdot \frac{\sqrt{2}}{2} = \frac{2\sqrt{2}}{\pi} \cdot t$$

или

$$P(t) = 0,90031t.$$

Используя это приближение для интерполяции при $t = \frac{\pi}{6}$, получим $P\left(\frac{\pi}{6}\right) = \frac{\sqrt{2}}{3} = 0,471$ вместо точного значения 0,5.

Здесь также можно в общем виде исследовать точность приближения. Разность между функцией и полиномом

$$\xi = x - P(t) = \sin t - \frac{2\sqrt{2}}{\pi} t.$$

Приравнявая нулю производную, получим, что ξ имеет максимум при $t = 25^\circ 48' 05'' = 0,45031$ рад. Этот максимум ξ равен $0,43525 - 0,40542 = 0,02983$. Можно принять, что предельная погрешность приближения равна 0,03 на всем интервале от 0 до $\frac{\pi}{4}$:

Пример 3. Построить приближение функции $x = \sin t$ по трем узлам на промежутке от 0 до $\frac{\pi}{4}$.

$$\begin{array}{c|ccc} t & 0 & \frac{\pi}{6} & \frac{\pi}{4} \\ x & 0 & 0,5 & \frac{\sqrt{2}}{2} \end{array}.$$

По формуле Лагранжа

$$P(t) = \frac{\left(t - \frac{\pi}{6}\right)\left(t - \frac{\pi}{4}\right)}{\left(0 - \frac{\pi}{6}\right)\left(0 - \frac{\pi}{4}\right)} \cdot 0 + \frac{t\left(t - \frac{\pi}{4}\right)}{\frac{\pi}{6} \cdot \left(\frac{\pi}{6} - \frac{\pi}{4}\right)} \cdot 0,5 + \frac{t\left(t - \frac{\pi}{6}\right)}{\frac{\pi}{4} \left(\frac{\pi}{4} - \frac{\pi}{6}\right)} \cdot \frac{\sqrt{2}}{2}.$$

Введем вместо t аргумент τ , связанный с t равенством $t = \frac{\pi}{4} \tau$ (τ есть значение угла, если за единицу принят угол $\frac{\pi}{4}$). После упрощений полином принимает вид

$$P(\tau) = \left(\frac{9}{4} - \sqrt{2}\right)\tau + \left(\frac{3\sqrt{2}}{2} - \frac{9}{4}\right)\tau^2.$$

Прямой проверкой можно убедиться в том, что $P(\tau)$ точно представляет все узлы. Используя $P(\tau)$ для интерполяции на значение $t = \frac{\pi}{12}$, чему соответствует $\tau = 1/3$, получаем

$$P\left(\frac{1}{3}\right) = 0,5 - \frac{\sqrt{2}}{6} = 0,264.$$

Значение $\sin 15^\circ$ с тремя знаками после запятой равно 0,259, значит ошибка интерполяции равна $-0,005$.

§ 18. Оценка ошибки точечной интерполяции

Если данная функция не является полиномом, то интерполяционный полином дает значения, совпадающие со значениями функции только в узлах и, может быть, еще в нескольких изолированных точках *). При интерполяции для заданных промежуточных значений аргумента полином только приближенно представляет функцию. Весьма существенно определить предельную погрешность точечной

*) Отметим, что то же обстоятельство будет иметь место, если данная функция является полиномом, степень которого больше степени интерполяционного полинома.

интерполяции, т. е. по возможности меньшую верхнюю границу модуля ошибки. Это может быть сделано следующим образом.

Пусть для функции $x(t)$ по узлам $(t_k, x_k; k=0, 1, 2, \dots, n)$ построен интерполяционный полином $P_n(t)$ n -й степени. Построим вспомогательную функцию

$$F(z) = x(z) - P_n(z) - KL(z),$$

где $L(z) = (z - t_0)(z - t_1) \dots (z - t_n)$ и K — пока неопределенное число. По условию точечной интерполяции

$$x(t_k) = P_n(t_k) \quad (k=0, 1, 2, \dots, n),$$

по определению

$$L(t_k) = 0.$$

Поэтому $F(z)$ имеет $n+1$ корней t_0, t_1, \dots, t_n . Воспользуемся неопределенностью K для того, чтобы функция $F(z)$ имела еще один корень $t \neq t_k$, т. е. введем условие

$$x(t) - P_n(t) - KL(t) = 0.$$

Заметим, что прямое определение K из этого равенства невозможно, поскольку $x(t)$ неизвестно. Применим теперь к функции $F(z)$ теорему Ролля. Согласно этой теореме, если функция $f(z)$ обращается в нуль при двух последовательных значениях аргумента t_r и t_{r+1} , то при известных условиях (непрерывность и существование производной) производная $\frac{df}{dz}$ по крайней мере один раз обращается в нуль

при каком-то значении z между t_r и t_{r+1} . Наша функция $F(z)$ имеет $n+1$ последовательных корней t_k и дополнительный корень t , который при интерполяции находится внутри одного из промежутков между смежными узлами, т. е. один из промежутков делится на две части. При интерполяции всего получается $n+1$ промежутков, на границах которых $F(z)$ обращается в нуль, имея $n+2$ корня.

При экстраполяции, если функция определена и вне табличной области аргумента, значение t либо будет меньше t_0 , либо больше t_n , т. е. к n промежуткам между последовательными узлами прибавится еще один промежуток между t и t_0 или между t_n и t . Согласно теореме Ролля, примененной к каждому из $n+1$ промежутков, производная $\frac{dF}{dz}$ обратится в нуль по крайней мере $n+1$ раз, причем эти нули будут в разных промежутках, т. е. не могут слиться. Предположим, что $x(t)$ имеет производные до $(n+1)$ -го порядка во всей области таблицы, (а также и в расширенной области, если производится экстраполяция). К функции $\frac{dF}{dz}$ применим снова теорему Ролля в промежутках между нулями; теперь этих промежутков будет n . Получим n значений, при которых $\frac{d^2F}{dz^2}$ обратится в нули.

Если будем продолжать эту операцию с производными последовательных порядков, то придем к выводу, что производная $(n+1)$ -го порядка от функции $F(z)$ обращается в нуль по крайней мере при одном значении z , которое мы обозначим через τ (это значение при интерполяции находится в области таблицы, при экстраполяции — в расширенной области). Из определения $F(z)$ следует

$$\frac{d^{n+1}F(z)}{dz^{n+1}} = \frac{d^{n+1}x(z)}{dz^{n+1}} - \frac{d^{n+1}P_n(z)}{dz^{n+1}} - K \frac{d^{n+1}L(z)}{dz^{n+1}}.$$

В этом равенстве в правой части второй член равен нулю, так как полином $P_n(z)$ степени n дифференцируется $n+1$ раз; полином $(n+1)$ степени $L(z)$, продифференцированный $n+1$ раз, даст $(n+1)!$, так как коэффициент при высшей степени аргумента в $L(z)$ равен единице.

Поэтому

$$\frac{d^{n+1}F(z)}{dz^{n+1}} = \frac{d^{n+1}x(z)}{dz^{n+1}} - K(n+1)!$$

Если положим $z = \tau$, то слева будет нуль, и мы получим

$$K = \left(\frac{d^{n+1}x}{dz^{n+1}} \right)_{z=\tau} \cdot \frac{1}{(n+1)!}.$$

По условию $z = t$ есть корень $F(z)$; значит, подстановка $z = t$ в $F(z)$ даст

$$0 = x(t) - P_n(t) - \frac{L(t)}{(n+1)!} \left(\frac{d^{n+1}x}{dt^{n+1}} \right)_{t=\tau}.$$

Отсюда видно, что ошибка интерполяции $x - P_n$ определяется формулой

$$x(t) - P_n(t) = L(t) \frac{x^{(n+1)}(\tau)}{(n+1)!}, \quad (4.17)$$

где $x^{(n+1)}$ обозначает производную $(n+1)$ порядка, причем

$$\left. \begin{array}{l} t_0 \leq \tau \leq t_n \text{ при интерполяции,} \\ t \leq \tau \leq t_n \\ t_0 \leq \tau \leq t \end{array} \right\} \text{ при экстраполяции.}$$

Приведенное рассуждение не дает возможности определить τ , поэтому на практике возможно только нахождение предельной погрешности, если окажется возможным оценить сверху производную $n+1$ -го порядка заданной функции. Обозначим через M_{n+1} максимум модуля производной $(n+1)$ -го порядка в области таблицы или в расширенной области.

Тогда для всякого заданного t получим

$$|x(t) - P_n(t)| \leq \frac{M_{n+1}}{(n+1)!} |L(t)|. \quad (4.18)$$

При желании можно построить равномерную оценку ошибки интерполяции, т. е. оценку, не зависящую от t . Для этого нужно найти такое число N , что $|L(t)| \leq N$. Тогда

$$|x(t) - P(t)| \leq \frac{NM_{n+1}}{(n+1)!}$$

при всех значениях t в табличной или расширенной области.

Пример 1. Для $x = \sqrt{t}$ построить интерполяционный полином по двум узлам:

$$\begin{aligned} t_0 &= 0,25 & x_0 &= 0,5 \\ t_1 &= 1, & x_1 &= 1. \end{aligned}$$

По формуле Лагранжа получим линейную интерполяционную формулу:

$$P(t) = \frac{t-1}{0,25-1} \cdot 0,5 + \frac{t-0,25}{1-0,25} \cdot 1 = \frac{2}{3}t + \frac{1}{3}.$$

Чтобы составить себе представление о качестве приближения, найдем

$$P(0,49) = \frac{1,98}{3} = 0,66, \text{ вместо } 0,70, \text{ ошибка } +0,04;$$

$$P(0,81) = \frac{2,62}{3} = 0,87 \text{ вместо } 0,90; \text{ ошибка } +0,03.$$

Найдем оценку ошибки интерполяции для всех t в рассматриваемой области. По общей формуле имеем

$$x(t) - P_1(t) = (t-0,25)(t-1) \frac{x''(\tau)}{2!},$$

$$x'(t) = \frac{1}{2\sqrt{t}}, \quad x''(t) = -\frac{1}{4\sqrt{t^3}},$$

где штрихи обозначают производные первого и второго порядков. Величина $|x''(t)|$ монотонно убывает в нашей области, поэтому $|x''(t)| \leq \frac{1}{4\sqrt{0,25^3}} = 2$,

$$|\sqrt{t} - P(t)| \leq |(t-0,25)(t-1)|.$$

При $t = 0,49$ правая часть равна 0,12; при $t = 0,81$ правая часть равна 0,11. Обе полученные величины заметно больше действительной ошибки. Полином, определяющий оценку ошибки, имеет вид

$$L(t) = (t-0,25)(t-1) = t^2 - 1,25t + 0,25.$$

Значение t , при котором $L(t)$ имеет экстремум, равно 0,625. Подставим $t = 0,625$ в $L(t)$, чтобы получить максимум модуля $L(t)$:

$$|L(t)| \leq 0,375^2 \approx 0,14.$$

Отсюда получаем оценку ошибки, общую для всех t :

$$|\sqrt{t} - P_1(t)| \leq 0,14.$$

Пример 2. Для функции $x = \sin t$ построить интерполяционный полином по трем узлам:

$$\begin{array}{c|ccc} t & 0 & \frac{\pi}{6} & \frac{\pi}{2} \\ \hline x & 0 & 0,5 & 1 \end{array}.$$

По формуле Лагранжа имеем

$$P(t) = \frac{\left(t - \frac{\pi}{6}\right)\left(t - \frac{\pi}{2}\right)}{\left(0 - \frac{\pi}{6}\right)\left(0 - \frac{\pi}{2}\right)} \cdot 0 + \frac{(t-0)\left(t - \frac{\pi}{2}\right)}{\left(\frac{\pi}{6} - 0\right)\left(\frac{\pi}{6} - \frac{\pi}{2}\right)} \cdot 0,5 + \frac{(t-0)\left(t - \frac{\pi}{6}\right)}{\left(\frac{\pi}{2} - 0\right)\left(\frac{\pi}{2} - \frac{\pi}{6}\right)} \cdot 1.$$

Для упрощения введем обозначение $t = \pi\alpha$; тогда

$$P(t) = -\frac{9}{2} \alpha(2\alpha - 1) + \alpha(6\alpha - 1) = \frac{7}{2} \alpha - 3\alpha^2.$$

Контроль

t	α	$P(\alpha)$	$\sin t$
0	0	0	0
$\frac{\pi}{6}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{\pi}{2}$	$\frac{1}{2}$	1	1

Проверим качество приближения в частных случаях:

t	α	$P(\alpha)$	$\sin t$	Ошибка
$\frac{\pi}{4}$	$\frac{1}{4}$	$\frac{11}{16} = 0,69$	0,71	+ 0,02
$\frac{\pi}{3}$	$\frac{1}{3}$	$\frac{5}{6} = 0,83$	0,87	+ 0,04

Оценим ошибку интерполяции. Здесь

$$L(t) = t\left(t - \frac{\pi}{6}\right)\left(t - \frac{\pi}{2}\right).$$

$$|x'''(t)| = |\cos t| \leq 1,$$

$$|\sin t - P(t)| \leq \left|t\left(t - \frac{\pi}{6}\right)\left(t - \frac{\pi}{2}\right)\right| \cdot \frac{1}{6} = \frac{\pi^3}{72} \cdot \alpha(6\alpha - 1)(2\alpha - 1).$$

Сравним ошибки, полученные оценкой по этой формуле, с действительными ошибками в только что указанных частных случаях:

t	$ L(t) $	Оценка ошибки
$\frac{\pi}{4}$	$\frac{\pi^3}{192} \approx 0,16$	0,03
$\frac{\pi}{3}$	$\frac{\pi^3}{108} \approx 0,30$	0,05

В обоих случаях ошибка, полученная оценкой с точным полиномом $L(t)$, только немного больше точной ошибки.

Построим равномерную оценку для всех t . Для этого нужно исследовать поведение $L(t)$ во всей области t от 0 до $\frac{\pi}{2}$. В рассматриваемом случае

$$L(t) = \frac{\pi^8}{12} \alpha (2\alpha - 1)(6\alpha - 1);$$

так как $t = \pi\alpha$, то α имеет значения в области от 0 до $1/2$. В этой области исследуем на максимум модуля полином

$$M(\alpha) = 12\alpha^3 - 8\alpha^2 + \alpha.$$

Найдем производные:

$$\frac{dM}{d\alpha} = 36\alpha^2 - 16\alpha + 1; \quad \frac{d^2M}{d\alpha^2} = 72\alpha - 16.$$

Приравнявая первую производную нулю

$$36\alpha^2 - 16\alpha + 1 = 0,$$

получим значения α , при которых $M(\alpha)$ имеет максимум и минимум:

$$\alpha_{1,2} = \frac{4 \pm \sqrt{7}}{18}.$$

Первое из них дает минимум, второе максимум. Простым преобразованием можно привести выражения для экстремального значения $M(\alpha)$ к виду

$$M(\alpha_{1,2}) = -\frac{2}{27}(7\alpha_{1,2} - 1).$$

Для этого нужно исключить $\alpha_{1,2}^2$ и $\alpha_{1,2}^3$, учитывая уравнение, определяющее $\alpha_{1,2}$. Подстановка экстремальных значений α_1 и α_2 дает:

$$M(\alpha_1) = \frac{\sqrt{343} - 10}{243} = 0,0351; \quad M(\alpha_2) = -\frac{\sqrt{343} + 10}{243} = -0,117.$$

Наибольшее по модулю значение $M(\alpha)$ во всей области есть 0,117; соответствующее предельное значение разности $|\sin t - P(t)|$ равно $\frac{\pi^8}{72} \cdot 0,117 = 0,0504$; $|\sin t - P(t)| \leq 0,0504$. Это значение достигается при $\alpha = \frac{4 + \sqrt{7}}{18} = 0,37$ или *) $t = 0,37\pi \approx 67^\circ$ *).

*) В рассмотренных примерах построение полинома Лагранжа не вызывало трудностей, так как было задано малое количество узлов, и значения функции и аргументов имели мало знаков.

ГЛАВА 5

ИНТЕРПОЛИРОВАНИЕ ПО ТАБЛИЦЕ С ПЕРЕМЕННЫМ ШАГОМ

§ 19. Разделенные разности табличной функции

Если таблица значений функции дана не с постоянным шагом, т. е. промежутки между смежными значениями аргумента различны в разных местах таблицы, то разности между смежными значениями функции не могут служить для описания изменения функции. Для этого используются величины, которые называют разделенными разностями (употребляются иногда и другие термины, например, «подъемы»).

Пусть дана табличная функция:

$$\begin{array}{c} t \\ x \end{array} \parallel \begin{array}{ccccccc} t_0 & t_1 & t_2 & \dots & t_n \\ x_0 & x_1 & x_2 & \dots & x_n \end{array}.$$

Разделенной разностью 1-го порядка двух табличных значений называется отношение разности значений функции к разности соответствующих значений аргумента. Это определение применимо для любой пары значений аргумента, но обычно используется для смежных значений. Обозначения разделенных разностей 1-го порядка строятся так, чтобы были указаны взятые табличные значения аргумента. Так, для приведенной выше таблицы для разделенных разностей введем обозначения:

$$\begin{aligned} x(t_1, t_0) &= \frac{x_1 - x_0}{t_1 - t_0}, & x(t_2, t_1) &= \frac{x_2 - x_1}{t_2 - t_1}, \\ x(t_3, t_2) &= \frac{x_3 - x_2}{t_3 - t_2} & \text{и т. д.} \end{aligned}$$

Разделенная разность $x(t_r, t_s) = \frac{x_r - x_s}{t_r - t_s}$ может быть записана в симметричном виде:

$$x(t_r, t_s) = \frac{x_r}{t_r - t_s} + \frac{x_s}{t_s - t_r}.$$

Из этой записи видно, что при вычислении разностей безразличен порядок взятых табличных значений, т. е.

$$x(t_r, t_s) = x(t_s, t_r).$$

Разделенной разностью 2-го порядка трех табличных значений называется отношение разности двух разделенных разностей первого порядка (3-го, 2-го) и (2-го, 1-го) табличных значений к разности между 3-м и 1-м значениями аргумента. Формально определение применимо к произвольным трем табличным значениям, но на практике обычно берут три смежных значения. Введя обозначения, аналогичные тем, которые были даны для разностей 1-го порядка, выпишем несколько разностей 2-го порядка:

$$x(t_2, t_1, t_0) = \frac{x(t_2, t_1) - x(t_1, t_0)}{t_2 - t_0},$$

$$x(t_3, t_2, t_1) = \frac{x(t_3, t_2) - x(t_2, t_1)}{t_3 - t_1},$$

$$x(t_4, t_3, t_2) = \frac{x(t_4, t_3) - x(t_3, t_2)}{t_4 - t_2} \text{ и т. д.}$$

Если воспользоваться симметричным выражением разностей первого порядка, то получим, например,

$$x(t_3, t_2, t_1) = \frac{x_3}{(t_3 - t_2)(t_3 - t_1)} + \frac{x_2}{(t_2 - t_3)(t_2 - t_1)} - \\ - \frac{x_2}{(t_2 - t_1)(t_3 - t_1)} - \frac{x_1}{(t_1 - t_2)(t_3 - t_1)};$$

после простых преобразований найдем:

$$x(t_3, t_2, t_1) = \frac{x_3}{(t_3 - t_2)(t_3 - t_1)} + \frac{x_2}{(t_2 - t_3)(t_2 - t_1)} + \frac{x_1}{(t_1 - t_3)(t_1 - t_2)}.$$

Отсюда видно, что и разделенные разности 2-го порядка симметричны относительно использованных табличных значений, т. е. не изменяются при произвольной их перестановке.

Аналогично определяются разделенные разности любого порядка, например, разности 3-го порядка

$$x(t_3, t_2, t_1, t_0) = \frac{x(t_3, t_2, t_1) - x(t_2, t_1, t_0)}{t_3 - t_0},$$

$$x(t_4, t_3, t_2, t_1) = \frac{x(t_4, t_3, t_2) - x(t_3, t_2, t_1)}{t_4 - t_1},$$

и т. д.

Независимость от порядка табличных значений есть общее свойство разделенных разностей любого порядка.

Если дано $n + 1$ значение табличной функции, то можно по этой таблице построить n разностей 1-го порядка, $n - 1$ разностей 2-го порядка, и только одну разность n -го порядка. Сами табличные значения иногда называются *разделенными разностями нулевого порядка*.

Пример. Построим таблицу разделенных разностей для функции $x = \sin t$ в области от 0 до $\frac{\pi}{2}$ для значений

$$t = 0, \quad \frac{\pi}{6}, \quad \frac{\pi}{4}, \quad \frac{\pi}{3}, \quad \frac{\pi}{2}.$$

На стр. 69 приведена полная схема вычисления разделенных разностей, в которую внесены все промежуточные операции.

В столбце, обозначенном $x^{(1)}$, выписаны разности последовательных значений функции, в столбце $t^{(1)}$ — разности аргумента.

Деление числа из столбца $x^{(1)}$ на рядом стоящее число из столбца $t^{(1)}$ дает разделенную разность 1-го порядка. Их полные обозначения:

$$x(t_1, t_0), \quad x(t_2, t_1) \dots \text{и т. д.}$$

В столбце $x^{(2)}$ выписаны разности последовательных разделенных разностей первого порядка, в столбце $t^{(2)}$ — разности $t_2 - t_0, t_3 - t_1, t_4 - t_2$. Деление числа из $x^{(2)}$ на стоящее рядом число $t^{(2)}$ дает соответствующую разделенную разность 2-го порядка. Аналогично построена схема последующих вычислений. Так как числа x_k даны с тремя знаками, то в разделенных разностях до 3-го порядка берем три знака, что совпадает с числом знаков после запятой. Во второй разности 3-го порядка это дает четыре знака после запятой, поэтому первая разность 3-го порядка взята тоже с четырьмя знаками после запятой, что не совсем законно, ибо получаются четыре знака при трехзначном делимом и четырехзначном делителе.

§ 20. Способ построения разностных интерполяционных формул

Вычисления по формуле Лагранжа довольно громоздки, по крайней мере при ручных расчетах. Действительно, если имеется $n+1$ узел, то формула состоит из $n+1$ члена. В каждом из них n сомножителей в числителе, столько же в знаменателе и еще множитель — табличное значение функции. Всего в каждом члене $2n$ вычитаний, $2(n-1)+1$ умножение и одно деление. Считая умножение и деление эквивалентными операциями, сложение и вычитание тоже, имеем полное число операций:

$$\text{сложений} \quad 2n(n+1) + n = 2n^2 + 3n,$$

$$\text{умножений} \quad 2n(n+1) = 2n^2 + 2n.$$

Если $n=4$, то имеем 44 сложения и 40 умножений. Существенно также, что очень большое число сложений может приводить к накоплению погрешностей. Кроме того, организация вычислений по формуле Лагранжа довольно хлопотлива.

От указанных недостатков свободны так называемые *разностные интерполяционные формулы*, к изложению способа вывода которых мы и приступаем.

Пусть дана таблица функции $x = f(t)$

$$\begin{array}{c|cccccc} t & t_0 & t_1 & t_2 & \dots & t_n \\ \hline x & x_0 & x_1 & x_2 & \dots & x_n \end{array}$$

и требуется вывести выражение для интерполяционного полинома n -й степени (или низшей, в частных случаях при надлежащем подборе чисел x_0, x_1, \dots, x_n).

Условие точечной интерполяции дает для определения коэффициентов a_0, a_1, \dots, a_n интерполяционного полинома

$$P(t) = a_0 + a_1 t + \dots + a_n t^n$$

$n + 1$ основных уравнений:

$$x_0 = P(t_0), \quad x_1 = P(t_1), \quad \dots, \quad x_n = P(t_n).$$

Из этих основных уравнений видно, что значения соответствующих разделенных разностей табличной функции и интерполяционного полинома должны быть одинаковы. Приравнявая эти значения, получим еще n уравнений для коэффициентов интерполяционного полинома, содержащих разделенные разности первого порядка, $n - 1$ уравнений с разностями второго порядка и т. д. и, наконец, одно уравнение с разностью n -го порядка.

Полное число уравнений расширенной таким образом системы уравнений будет равно сумме чисел натурального ряда от 1 до $n + 1$, т. е. $\frac{(n+1)(n+2)}{2}$.

При выводе формулы Лагранжа (§ 17) для определения коэффициентов a_0, a_1, \dots, a_n использовались $n + 1$ основных уравнений. Однако эти коэффициенты можно определить, выбрав из расширенной системы любые $n + 1$ независимых уравнений. Обычно берут одно из основных уравнений и по одному из каждого столбца разделенных разностей всех порядков.

Определяемые из такой системы коэффициенты будут выражены через разделенные разности от нулевого до n -го порядка. Подставляя эти коэффициенты в полином $P(t)$, мы получим формулу, которую принято называть *разностной интерполяционной формулой*.

Остановимся теперь на вопросе о выборе степени интерполяционного полинома. Рассмотрим сначала случай, когда сама табличная функция является полиномом $P(t)$ степени n . Будем строить его разности при переменном шаге, т. е. разделенные разности.

Задавая одно определенное значение аргумента t_0 , а другое произвольное t , получим

$$x(t, t_0) = \frac{P(t) - P(t_0)}{t - t_0}.$$

По теореме Безу деление полинома $P(t)$ на разность $t - t_0$ дает в остатке $P(t_0)$. Поэтому выражение $P(t) - P(t_0)$ делится на $t - t_0$ без остатка, причем получается полином $(n - 1)$ -й степени. Отсюда следует, что разделенная разность 1-го порядка полинома n -й степени есть полином $n - 1$ степени. Аналогичным способом можно показать, что разделенная разность 2-го порядка есть полином степени $n - 2$ и т. д. Разделенная разность n -го порядка есть полином нулевой степени, т. е. постоянное число; легко показать, что это число равно коэффициенту старшего члена полинома $P(t)$. Разделенные разности порядка большего, чем n , равны нулю.

Отмеченное свойство полинома имеет существенное значение в задаче интерполяции. Задаваемая табличная функция только в редких случаях может оказаться точным полиномом, и мы это обнаружим после составления таблицы разностей, если окажется, что разности некоторого порядка — постоянные числа. В общем случае данная функция не является полиномом, и точно постоянных разностей не будет. Однако в весьма многих прикладных задачах табличные функции таковы, что они могут быть приближены полиномом некоторой степени. В таблице разностей подобных функций *разности некоторого порядка будут почти постоянны, разности следующего высшего порядка очень малы*. В этом случае можно с уверенностью применять точечную интерполяцию и строить полином, степень которого равна порядку почти постоянных разностей. Построение таблицы разностей поэтому есть единственный способ выяснить нужную степень интерполяционного полинома, т. е. число узлов, необходимое для удовлетворительного приближения.

В последнее время в связи с механизацией вычислений нередко предпочитают пользоваться формулой Лагранжа, но для решения вопроса о необходимом числе узлов все-таки нужно составить таблицу разностей.

§ 21. Интерполяционная формула Ньютона для таблицы с переменным шагом

Еще Ньютоном и Грегори была выведена разностная интерполяционная формула, отличная от формулы Лагранжа, выведенной позже, и содержащая разделенные разности. В этом параграфе мы выведем формулу Ньютона, причем при выводе ограничимся частным случаем, когда число узлов равно четыре, т. е. $n = 3$.

Расширенную систему уравнений для этого случая можно привести к виду:

t	x
t_0	$x_0 = a_0 + a_1 t_0 + a_2 t_0^2 + a_3 t_0^3$
	$x(t_1, t_0) = a_1 + a_2(t_1 + t_0) + a_3(t_1^2 + t_1 t_0 + t_0^2)$
t_1	$x_1 = a_0 + a_1 t_1 + a_2 t_1^2 + a_3 t_1^3$ $x(t_2, t_1, t_0) = a_2 + a_3(t_2 + t_1 + t_0)$
	$x(t_2, t_1) = a_1 + a_2(t_2 + t_1) + a_3(t_2^2 + t_2 t_1 + t_1^2)$ $x(t_3, t_2, t_1, t_0) = a_3$
t_2	$x_2 = a_0 + a_1 t_2 + a_2 t_2^2 + a_3 t_2^3$ $x(t_3, t_2, t_1) = a_2 + a_3(t_3 + t_2 + t_1)$
	$x(t_3, t_2) = a_1 + a_2(t_3 + t_2) + a_3(t_3^2 + t_3 t_2 + t_2^2)$
t_3	$x_3 = a_0 + a_1 t_3 + a_2 t_3^2 + a_3 t_3^3$

Разделенные разности первого порядка полинома, написанные справа, получаются после сокращений на $t_1 - t_0$, $t_2 - t_1$, $t_3 - t_2$; все разделенные разности 2-го порядка от полинома получаются в написанном виде после сокращения на $(t_2 - t_0)$, $(t_3 - t_1)$; разность 3-го порядка сокращается на $(t_3 - t_0)$.

Из написанных 10 уравнений можно выбрать любые четыре. Возьмем уравнения, выписанные на верхней стороне условного равнобедренного треугольника, который получится, если вместо каждого уравнения поставить точки и соединить их прямыми. На нашей схеме используемые уравнения подчеркнуты. Из них получим значения коэффициентов:

$$a_3 = x(t_3, t_2, t_1, t_0),$$

$$a_2 = x(t_2, t_1, t_0) - (t_2 + t_1 + t_0) x(t_3, t_2, t_1, t_0),$$

$$a_1 = x(t_1, t_0) - (t_1 + t_0) x(t_2, t_1, t_0) + (t_2 t_0 + t_1 t_0 + t_2 t_1) x(t_3, t_2, t_1, t_0),$$

$$a_0 = x_0 - t_0 x(t_1, t_0) + t_1 t_0 x(t_2, t_1, t_0) - t_2 t_1 t_0 x(t_3, t_2, t_1, t_0).$$

Подстановка коэффициентов в искомый полином дает

$$\begin{aligned}
 P(t) = & x_0 - t_0 x(t_1, t_0) + t_1 t_0 x(t_2, t_1, t_0) - \\
 & - t_2 t_1 t_0 x(t_3, t_2, t_1, t_0) + t x(t_1, t_0) - \\
 & - t(t_1 + t_0) x(t_2, t_1, t_0) + t(t_1 + t_0)(t_2 + t_1 + t_0) x(t_3, t_2, t_1, t_0) + \\
 & + t^2 x(t_2, t_1, t_0) - t^2(t_2 + t_1 + t_0) x(t_3, t_2, t_1, t_0) + \\
 & + t^3 x(t_3, t_2, t_1, t_0).
 \end{aligned}$$

Полученный полином расположим не по степеням t , а по разностям последовательных порядков, причем заданное значение считаем разделенной разностью нулевого порядка. Коэффициенты при разностях представляют полиномы, степени которых равны порядкам разностей.

Эти коэффициенты легко разлагаются на множители способом группировки. После преобразований интерполяционный полином примет вид:

$$P(t) = x_0 + (t - t_0)x(t_1, t_0) + (t - t_1)(t - t_0)x(t_2, t_1, t_0) + \\ + (t - t_2)(t - t_1)(t - t_0)x(t_3, t_2, t_1, t_0).$$

Аналогично строится полином при любом числе узлов.

Для вычислений удобно представить $P(t)$ в следующем виде:

$$P(t) = x_0 + (t - t_0)\{x(t_1, t_0) + (t - t_1)[x(t_2, t_1, t_0) + \\ + (t - t_2)x(t_3, t_2, t_1, t_0)]\}.$$

Вычисления выполняются «с конца»: разность 3-го порядка умножается на $t - t_2$, к произведению прибавляется разность 2-го порядка, сумма умножается на $t - t_1$, к произведению прибавляется разность 1-го порядка, сумма умножается на $t - t_0$ и, наконец к произведению прибавляется разность нулевого порядка.

Формула и способ вычисления легко обобщается для любого числа узлов. При работе на арифмометре при составленной таблице разностей можно почти ничего не записывать, кроме чисел $t - t_0$, $t - t_1$, ...

Пример. Дана таблица значений синуса

t	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\sin t$	0,000	0,500	0,707	0,866	1,000

Требуется построить интерполяционную формулу и по ней вычислить значение синуса при

$$t = 36^\circ = \frac{\pi}{5} = 0,628.$$

Выпишем таблицу разделенных разностей:

t	$x = \sin t$	1-е разности	2-е разности	3-и разности	4-е разности	$t - t_0$
0,000	<u>0,000</u>					0,628
		<u>0,954</u>				
0,524	0,500		<u>-0,205</u>			0,104
		0,793		<u>-0,1442</u>		
0,785	0,707		-0,356		<u>+ 0,0365</u>	-0,157
		0,607		-0,0869		
1,047	0,866		-0,447			-0,419
		0,256				
1,571	1,000					

В таблице подчеркнуты разности, используемые в формуле. В последнем столбце выписаны числа $t-t_0$, $t-t_1$, $t-t_2$, $t-t_3$. Формула для произвольных t имеет вид

$$P(t) = t \{0,954 + (t - 0,524)[(-0,205) + \\ + (t - 0,785)(-0,1442 + (t - 1,047)(+0,0365))]\}.$$

При заданном значении t на арифмометре делаются следующие операции:

1) Число $+0,0365$ умножается на $-0,419$ и к произведению прибавляется $-0,1442000$, результат округляется на результатном счетчике; получается $-0,1595$;

2) Это число переносится на корпус арифмометра, умножается на $-0,157$ и к произведению прибавляется $-0,2050000$; после округления получается $-0,1800$;

3) Полученное число переносится на корпус, умножается на $0,104$ и к результату прибавляется $0,9540000$; сумма округляется, получается $0,9353$;

4) Число $0,9353$ переносится на корпус и умножается на $0,628$. Прибавлять ничего не нужно, так как начальное значение равно нулю.

В полученном таким образом числе $0,5874$ взят запасной (4-й) знак, как и в промежуточных вычислениях. По таблице получается $0,5878$, следовательно три знака оказываются верными.

Сделаем при помощи формул (4.18) оценку погрешности интерполяции. В рассматриваемом случае модуль производной любого порядка не превышает единицы. Поэтому

$$|\sin t - P(t)| < \frac{|L(t)|}{5!},$$

где

$$L(t) = 0,628 \cdot 0,104 \cdot (-0,157) \cdot (-0,419) \cdot (-0,943).$$

Нет надобности точно вычислять $L(t)$; можно несколько увеличить все сомножители, чтобы оценку ошибки сделать попроще:

$$|L(t)| < 0,7 \cdot 0,1 \cdot 0,2 \cdot 0,5 \cdot 1,0 = 0,007.$$

Таким образом,

$$|\sin t - P(t)| < \frac{0,007}{5!} \approx 0,00006.$$

Погрешность получилась значительно меньшей, чем фактическая ошибка. Это объясняется очень просто. Формула для определения погрешности не содержит значений функции и, следовательно, не учитывает точности заданных значений функции. По формуле учитывается лишь погрешность собственно интерполяции в предположении, что заданные значения функции точны.

В нашем случае некоторые значения функции даны с тремя верными знаками и поэтому точнее получить результат интерполяции нельзя. Выведенная оценка показывает, что для данной функции нельзя получить результат интерполяции точнее, чем до $0,5 \cdot 10^{-4}$, т. е. с четырьмя знаками после запятой. Для обеспечения такой точности надо было дать и табличную функцию с четырьмя знаками после запятой.

ГЛАВА 6

ИНТЕРПОЛИРОВАНИЕ ПО ТАБЛИЦЕ С ПОСТОЯННЫМ ШАГОМ

§ 22. Обыкновенные и центральные разности табличной функции с постоянным шагом

Предположим, что табличная функция построена с постоянным шагом или выбрана часть таблицы, в которой шаг не меняется:

$$\begin{array}{c} t \\ x \end{array} \parallel \begin{array}{cccccc} t_{-2} & t_{-1} & t_0 & t_1 & t_2 & \dots \\ x_{-2} & x_{-1} & x_0 & x_1 & x_2 & \dots \end{array}$$

где $t_k = t_0 + hk$, h — шаг аргумента, т. е. разность двух смежных значений t (безразлично каких, поскольку шаг постоянный). Для такой таблицы вводятся три типа разностей: 1) *обыкновенные разности*, 2) *центральные разности*, 3) *разности отрицательных порядков*.

1. Обыкновенные разности

Запишем таблицу в виде двух столбцов t и x , в которых пропускается строка между двумя последовательными значениями. Оба столбца ставятся у левого края листа. В столбце функции сделаем следующую операцию — из каждого числа вычтем предыдущее и разность запишем в следующем столбце на пропущенной строке между теми значениями, из которых получена разность:

$$\begin{array}{cccccccccccc} \dots & \dots & & & & & & & & & & \\ t_{-3} & x_{-3} & & & & & & & & & & \\ & & x_{-2\frac{1}{2}}^{1} & & & & & & & & & \\ t_{-2} & x_{-2} & & x_{-2}^{2} & & & & & & & & \\ & & x_{-1\frac{1}{2}}^{1} & & x_{-1\frac{1}{2}}^{3} & & & & & & & \\ t_{-1} & x_{-1} & & x_{-1}^{2} & & x_{-1}^{4} & & & & & & \\ & & x_{-1\frac{1}{2}}^{1} & & x_{-1\frac{1}{2}}^{3} & & x_{-1\frac{1}{2}}^{5} & & & & & \\ t_0 & x_0 & & x_0^{2} & & x_0^{4} & & x_0^{6} & & & & \\ & & x_{+1\frac{1}{2}}^{1} & & x_{+1\frac{1}{2}}^{3} & & x_{+1\frac{1}{2}}^{5} & & & & & \\ t_1 & x_{+1} & & x_1^{2} & & x_1^{4} & & & & & & \\ & & x_{\frac{1}{2}}^{1} & & x_{\frac{1}{2}}^{3} & & & & & & & \\ t_2 & x_{+2} & & x_2^{2} & & & & & & & & \\ & & x_{\frac{1}{2}}^{1} & & & & & & & & & \\ t_3 & x_{+3} & & & & & & & & & & \\ \dots & \dots & & & & & & & & & & \end{array}$$

Получается столбец разностей первого порядка. Для этих разностей мы будем пользоваться обозначениями, указанными в третьем столбце помещенной здесь же схемы разностей:

$$x_{-1/2}^1 = x_{-2} - x_{-3},$$

$$x_{-3/2}^1 = x_{-1} - x_{-2} \quad \text{и т. д.}$$

Здесь нижний значок есть среднее арифметическое значков функции, из которых разность получена; цифра 1 вверху обозначает разность первого порядка. Со столбцом разностей первого порядка сделаем такую же операцию. Получим столбец разностей второго порядка и т. д. По всей таблице применена одна и та же система обозначений: разность порядка k отмечается значком k наверху; нижний значок — среднее арифметическое значков тех разностей порядка $k-1$, из которых рассматриваемая разность образована. При такой системе записи и обозначений все разности нечетных порядков имеют дробные нижние значки со знаменателем 2, все разности четных порядков имеют целые нижние значки. Разности последовательных порядков, стоящие в одной строке, снабжены одинаковыми нижними значками. Все разности разных порядков, в том числе и значения функции (разности нулевого порядка), располагаются в равнобедренном треугольнике, с пропусками (через строчку) в каждой строке и в каждом столбце. Пусть имеем разность, обозначенную через x_s^k , где k — целое положительное число, а s — либо целое при k четном, либо целое, деленное на 2, если k — нечетное; из определения разности вытекает равенство, связывающее эту разность с разностями предыдущего порядка:

$$x_s^k = x_{s+1/2}^{k-1} - x_{s-1/2}^{k-1}.$$

Если задано $n+1$ значение функции, то в таблице содержится n разностей 1-го порядка, $n-1$ разностей 2-го порядка — и т. д. Столбец разностей n -го порядка будет содержать только одно число, и на этом таблица разностей заканчивается.

Высший порядок разности в таблице разностей на единицу меньше количества значений функции.

2. Центральные разности

В некоторых формулах оказывается удобным ввести средние арифметические смежных обыкновенных разностей, стоящих в одном столбце, которые и называются центральными или центрированными. Для них принята такая же система нижних и верхних значков, как и для обыкновенных разностей, т. е. нижний значок центральной

разности равен среднему арифметическому нижних значков тех разностей, из которых центральная получена:

$$x_{s+1/2}^k = (x_s^k + x_{s+1}^k) \cdot \frac{1}{2}.$$

При этом центральные разности нечетного порядка имеют целые нижние значки, четного порядка — дробные. Это избавляет от смешения центральных разностей с обыкновенными, ибо у обыкновенных будет обратное соотношение между знаками, и однотипные обозначения не приведут к путанице. Центральные разности записываются на пропущенных местах между теми обыкновенными, из которых они получены.

Легко показать, что для центральных разностей всех порядков сохраняется формула связи между разностью порядка k и разностями порядка $k-1$. Действительно, пусть $x_{s+1/2}^k$ — центральная разность порядка k , определенная последней формулой, приведенной выше. Выражая входящие в эту формулу k -го порядка обыкновенные разности через обыкновенные разности порядка $(k-1)$, получим

$$x_{s+1/2}^k = \frac{1}{2} [(x_{s+1/2}^{k+1} - x_{s+1/2}^{k-1}) + (x_{s+1/2}^{k-1} - x_{s-1/2}^{k-1})].$$

Если мы переставим члены внутри квадратной скобки и введем множитель $1/2$ в квадратную скобку, то по определению центральных разностей получим

$$\frac{1}{2} (x_{s+1/2}^{k-1} + x_{s+1/2}^{k-1}) = x_{s+1}^{k-1},$$

$$\frac{1}{2} (x_{s+1/2}^{k-1} + x_{s-1/2}^{k-1}) = x_s^{k-1}.$$

Поэтому

$$x_{s+1/2}^k = x_{s+1}^{k-1} - x_s^{k-1},$$

а это и есть нужная связь.

Число центральных разностей в каждом столбце на единицу меньше числа обыкновенных разностей того же порядка, так как они вставляются в промежутках между обыкновенными разностями, а число промежутков между m точками равно $m-1$. Соответственно с этим, если в столбце нулевого порядка значений функции $n+1$ чисел, то центральных разностей нулевого порядка будет n ; высший порядок центральных разностей будет $n-1$, таких разностей будет одна. Если вставить в таблицу разностей все центральные разности, то получится плотно заполненный равнобедренный треугольник.

Заметим, что для обозначения и записи обыкновенных разностей употребляются также другие символы и иная схема записи.

Например,

$$\Delta x_0 = x_1 - x_0; \quad \Delta x_1 = x_2 - x_1 \dots \text{ и т. д.};$$

$$\Delta^2 x_0 = \Delta x_1 - \Delta x_0; \quad \Delta^2 x_1 = \Delta x_2 - \Delta x_1 \text{ и т. д.}$$

В этом случае в одной строке пишут разности с одинаковыми нижними значками и все разности расположатся в прямоугольном треугольнике. Это обозначение для обыкновенных разностей нельзя распространить на центральные разности; запись в виде прямоугольного треугольника без пропусков не позволяет вставить в таблицу центральные разности. Поэтому при употреблении этих обозначений в формулах, содержащих центральные разности, для них не вводят обозначений, а просто вносят их выражения через обыкновенные разности, что усложняет выкладки и вычисления.

3. Разности отрицательных порядков

Если написана таблица разностей функции, то из двух рядом стоящих столбцов правый содержит разности первого порядка левого столбца. Если учитывать только эту связь двух столбцов, то можно было бы ко всем членам левого столбца прибавить одно и то же число, не изменяя правого столбца. Если идти по таблице разностей справа налево, то упомянутая связь двух смежных столбцов обрывается на столбце значений функции (разности нулевого порядка). В некоторых задачах (например, при численном решении дифференциальных уравнений) удобно строить налево от столбца нулевого порядка столбец таких чисел, для которых значения функции являются разностями; назовем такие числа *разностями минус первого порядка*. Из определения вытекает, что одну из разностей минус первого порядка можно задать произвольно, например, первую; она будет стоять в своем столбце строчкой выше, чем первое значение функции. Сложив это число с первым значением функции, получим второе значение разности минус первого порядка; сложение этого числа со вторым значением функции даст третье значение разности первого отрицательного порядка и т. д.

После построения столбца разностей минус первого порядка можно таким же способом построить налево от него разности минус второго порядка и т. д. Они обозначаются по той же системе, что и разности положительных порядков. Если первое число в столбце нулевого порядка имеет нижний значок s , то первое число в столбце минус первого отрицательного порядка имеет значок $s - \frac{1}{2}$, минус второго порядка $s - 1$ и т. д. Нижние значки последних чисел каждого столбца слева на $\frac{1}{2}$ больше, чем значки в соседнем столбце справа. Так как после выбора первого числа в каждом столбце

отрицательного порядка остальные разности этого же порядка получаются последовательными сложениями с разностями следующего высшего порядка, то столбцы разностей отрицательного порядка иногда называют *столбцами сумм*.

§ 23. Основные свойства обыкновенных разностей

1. Выражение разностей разных порядков через табличные значения функции

Пусть дана таблица обыкновенных разностей:

...	...			
t_{-1}	x_{-1}			
		$x_{-1/2}^1$		
t_0	x_0		x_0^2	
		$x_{1/2}^1$		$x_{1/2}^3$
t_1	x_1		x_1^2	x_1^4
		$x_{3/2}^1$		$x_{3/2}^3$
t_2	x_2		x_2^2	
		$x_{5/2}^1$		
t_3	x_3			
...	...			

Ее можно продолжить как вниз, так и вверх, используя отрицательные нижние значки у значений функции и разностей.

По определению, разности определенного порядка выражаются через разности порядка на единицу меньше, но в некоторых задачах нужно иметь разности любого порядка, непосредственно выраженные через табличные значения функции. Такие выражения легко получаются методом индукции из рассмотрения частных случаев.

Для разностей второго порядка имеем:

$$x_0^2 = x_{1/2}^1 - x_{-1/2}^1 = (x_1 - x_0) - (x_0 - x_{-1}) = x_1 - 2x_0 + x_{-1};$$

аналогично

$$x_1^2 = x_2 - 2x_1 + x_0,$$

$$x_2^2 = x_3 - 2x_2 + x_1,$$

$$x_k^2 = x_{k+1} - 2x_k + x_{k-1}.$$

где k — целое число, положительное или отрицательное. Далее,

$$\begin{aligned}x_{1/2}^3 &= x_1^2 - x_0^2 = x_2 - 2x_1 + x_0 - (x_1 - 2x_0 + x_1) = \\&= x_2 - 3x_1 + 3x_0 - x_{-1}, \\x_{3/2}^3 &= x_2^2 - x_1^2 = x_3 - 2x_2 + x_1 - (x_2 - 2x_1 + x_0) = \\&= x_3 - 3x_2 + 3x_1 - x_0, \\&\dots \dots \dots \\x_{k+1/2}^3 &= x_{k+2}^2 - 3x_{k+1} + 3x_k - x_{k-1}.\end{aligned}$$

Из этих частных примеров видно, что связь между обыкновенными разностями и табличными значениями функции определяется формулами с биномиальными коэффициентами:

$$\begin{aligned}x_k^{2p} &= x_{k+p} - C_{2p}^1 x_{k+p+1} + C_{2p}^2 x_{k+p-2} - \dots \\&\quad \dots + (-1)^p C_{2p}^p x_k + \dots + x_{k-p}, \\x_{k+1/2}^{2p+1} &= x_{k+p+1} - C_{2p+1}^1 x_{k+p} + C_{2p+1}^2 x_{k+p-1} - \dots \\&\quad \dots + (-1)^p C_{2p+1}^p (x_{k+1} - x_k) + \dots - x_{k-p}\end{aligned}$$

(k — целое число, p — целое положительное).

Эти формулы пригодны только для разностей положительных порядков. Для разностей отрицательных порядков таких формул написать нельзя, ибо эти разности определены с точностью до произвольного слагаемого.

2. Обращение таблицы

Пусть все числа таблицы значений функции написаны в обратном порядке и определены разности последовательных порядков с соблюдением условия: из последующего числа вычитается предыдущее. Легко видеть, что при этом все разности первого порядка сохраняют свои модули, но их знаки заменяются на обратные. Все разности второго порядка не изменятся ни по величине, ни по знаку. Действительно, выражение

$$x_k^2 = x_{k+1} - 2x_k + x_{k-1}$$

симметрично относительно x_k и не изменится, если числа x_{k-1} , x_k , x_{k+1} заменить на x_{k+1} , x_k , x_{k-1} . Из общих выражений разностей нечетного и четного порядков видно, что все разности четного порядка не изменятся, а разности нечетных порядков изменят знаки на обратные. Отсюда следует, что всегда нужно сохранять принятый порядок вычитания при вычислении разностей (из последующего вычитать предыдущее), так как нарушение этого порядка эквивалентно обращению таблицы с последующим изменением части разностей.

3. Сумма разностей одного порядка

Если записать все обыкновенные разности табличной функции, содержащей $n + 1$ узел по указанной выше схеме, то получим фигуру равнобедренного треугольника, пересеченного прямыми, параллельными основанию. На этих прямых расположены все разности одного порядка. Выделим столбец разностей порядка s . Если узел со значком нуль находится в начале таблицы, то первое (верхнее) число столбца разностей имеет нижний значок $\frac{s}{2}$, последняя разность — значок $n - \frac{s}{2}$. Сложим все разности порядка s , заменяя каждую из них разностью разностей порядка $s - 1$. Тогда получим

$$\sum_{k=\frac{s}{2}}^{n-\frac{s}{2}} x_k^s = \sum_{k=\frac{s}{2}}^{n-\frac{s}{2}} \left(x_{k+\frac{1}{2}}^{s-1} - x_{k-\frac{1}{2}}^{s-1} \right).$$

Если расписать правую часть подробно, то получим

$$\begin{aligned} \sum_{k=\frac{s}{2}}^{n-\frac{s}{2}} x_k^s &= \left(x_{n-\frac{s-1}{2}}^{s-1} - x_{n-1-\frac{s-1}{2}}^{s-1} \right) + \left(x_{n-1-\frac{s-1}{2}}^{s-1} - x_{n-2-\frac{s-1}{2}}^{s-1} \right) + \dots \\ &\dots + \left(x_{\frac{s+1}{2}}^{s-1} - x_{\frac{s}{2}}^{s-1} \right) + \left(x_{\frac{s+1}{2}}^{s-1} - x_{\frac{s-1}{2}}^{s-1} \right). \end{aligned}$$

Все члены, кроме первого и последнего, взаимно уничтожаются, так что

$$\sum_{k=\frac{s}{2}}^{n-\frac{s}{2}} x_k^s = x_{n-\frac{s-1}{2}}^{s-1} - x_{\frac{s-1}{2}}^{s-1}.$$

Числа, стоящие справа, представляют последнее и первое из чисел в столбце разностей порядка $s - 1$. Таким образом, сумма всех разностей некоторого порядка равна разности между последним и первым числом в столбце смежных разностей на единицу меньшего порядка. Это свойство используется при некоторых выкладках, а также для контроля составления таблицы разностей. Хотя операция составления разностей предельно проста, но все-таки иногда возможны грубые ошибки («зевки») как при ручном, так и при машинном счете. В этом случае контрольное вычисление должно давать точное равенство суммы всех разностей порядка s и разности крайних чисел в предыдущем столбце разностей.

4. Разности полинома

Тем же способом, что и в § 20, легко показать, что для полинома степени n обыкновенные разности n -го порядка постоянны, а разности более высокого порядка равны нулю. Так же, как и в случае таблицы с переменным шагом, для таблицы с постоянным шагом это свойство можно использовать для нахождения степени интерполяционного полинома: эту степень следует брать равной порядку тех разностей, которые почти постоянны для данной табличной функции.

5. Влияние ошибки в таблице на разности

Предположим, что в «среднем» числе x_0 таблицы значений функции сделана ошибка ε , т. е. вместо x_0 взято $x_0 + \varepsilon$. Поместим на одной схеме разности точной таблицы и разности таблицы с ошибкой:

t	x				
$t_0 - 3h$	x_{-3}				
		$x^1_{-1/2}$			
$t_0 - 2h$	x_{-2}		x^2_{-2}		
		$x^1_{-1/2}$	$x^3_{-1} + \varepsilon$		
$t_0 - h$	x_{-1}		$x^2_{-1} + \varepsilon$	$x^4_{-1} - 4\varepsilon$	
		$x^1_{-1/2} + \varepsilon$	$x^3_{-1/2} - 3\varepsilon$	$x^5_{-1/2} + 10\varepsilon$	
t_0	$x_0 + \varepsilon$		$x^3_0 - 2\varepsilon$	$x^4_0 + 6\varepsilon$	$x^6_0 - 20\varepsilon$
		$x^1_{+1/2} - \varepsilon$	$x^3_{1/2} + 3\varepsilon$	$x^5_{1/2} - 10\varepsilon$	
t_1	x_1		$x^3_{+1} + \varepsilon$	$x^4_1 - 4\varepsilon$	
		$x^1_{1/2}$	$x^3_{1/2} - \varepsilon$		
t_2	x_2		x^2_2		
		$x^1_{1/2}$			
t_3	x_3				

Из схемы видно, что ошибка в одном числе таблицы значений функции влияет на две разности 1-го порядка, на три разности 2-го порядка и т. д. В разностях одного порядка ошибки входят с биномиальными коэффициентами, причем порядок бинома равен порядку разности. В разностях четного порядка максимум ошибки достигается на той линии, где находится основная ошибка, а в нечетных разностях по обе стороны этой линии знаки ошибок в последовательных разностях чередуются.

Эти свойства следующим образом используются для поисков ошибок в таблице. В обычных таблицах функций, используемых

в естествознании, почти всегда разности некоторого порядка становятся достаточно малыми *). В случае наличия грубой ошибки в одном из табличных чисел эта ошибка прибавляется (с биномиальными коэффициентами и с чередующимися знаками) к малым разностям. Это приводит к сильным колебаниям величин разностей, со сменой знаков, начиная с разностей некоторого порядка. В таких случаях говорят, что разности «скачут». Такие «скачки» разностей и могут служить указанием на ошибку. Искать ее следует на той линии, на которой скачки наибольшие (или возле нее). Величину ошибки примерно можно определить, разделив «скачок» на максимальный коэффициент бинома, степень которого равна порядку разности. За «скачок» можно принять наибольшее по модулю значение разности.

Задача о влиянии ошибок таблицы на ошибки разностей довольно сложна, так как во всякой таблице почти все числа содержат ошибки округления; каждая из этих ошибок вносит ошибки в разности всех порядков, и они суммируются. Точный учет ошибок этого рода был бы весьма громоздким. Поэтому имеет смысл применить вероятностный способ.

Пример. Имеем таблицу значений полинома 3-й степени. Составим его разности:

t	x					x				
-2	-22					-22				
-1	-5	17				-5	17			
			-12					-12		
0	0	5		+6		5			+7	
			-6		0	0		-5		-4
		-1		+6		0			+3	+10
+1	-1		0		0	0		-2		+6
		-1		+6			-2		+9	
+2	-2		6		6	-2			+7	
		+5				5				
+3	+3					+3				

Рядом напишем почти ту же таблицу, сделав ошибку -1 при $t = +1$. Скачок разностей намечается в разностях 3-го порядка (минимум) и становится заметным в разности 4-го порядка. Согласно предыдущей таблице ошибка возможна на линии максимума скачка, т. е. при $t = +1$. В разностях 4-го порядка максимальный биномиальный коэффициент равен 6 и ошибка входит с тем же знаком, что и основная. Считая, что те разности, где есть скачки, должны бы быть приблизительно нулями, получим, что ошибка разности 4-го порядка равна -6 , следовательно, ошибка таблицы равна $-6:6 = -1$.

Конечно, определить ошибку, как правило, бывает труднее, чем в приведенном примере, но место и порядок ее могут быть определены таким же образом.

*) Отметим, что, строго говоря, только к таким функциям применим принцип точечной интерполяции (подробнее об этом см. § 20 гл. V).

§ 24. Способ построения интерполяционных формул для таблиц с постоянным шагом

Пусть задано некоторое промежуточное, т. е. не содержащееся в таблице значение аргумента t , и для него требуется вычислить приближенное значение функции, определенной таблицей с постоянным шагом, с помощью интерполяционного полинома. Примем за начальное то табличное значение аргумента t_0 , которое меньше всего отличается от заданного t . Это значит, что $|t - t_0| \leq \frac{1}{2}h$, где h — шаг таблицы. Может случиться, что при этом условии $t > t_0$. Так как обычно значения аргумента растут от начала таблицы к концу, то в этом случае говорят, что следует *интерполировать вперед* (от начального значения). Интерполяция в случае $t < t_0$ называется *интерполяцией назад*, в случае же $|t - t_0| = \frac{1}{2}h$ мы имеем *интерполяцию на середину*.

В задаче интерполяции мы имеем, таким образом, два параметра — t_0 и h ; первый зависит от заданного t и шага, второй, шаг h , дается таблицей. Интерполяционные формулы естественно строить так, чтобы они не содержали явно этих параметров. Это достигается весьма просто введением нормированного аргумента

$$\tau = \frac{t - t_0}{h}.$$

Табличным значениям аргумента

$$\dots t_0 - 2h, \quad t_0 - h, \quad t_0, \quad t_0 + h, \quad t_0 + 2h$$

соответствуют целочисленные значения нормированного аргумента

$$\dots -2, \quad -1, \quad 0, \quad +1, \quad +2, \dots$$

Для таблицы с аргументом τ и значениями функции

$$\dots x_{-2}, \quad x_{-1}, \quad x_0, \quad x_1, \quad x_2 \dots$$

строятся интерполяционные формулы, пригодные при всех t_0 и h . Для нормированной таблицы мы будем строить интерполяционный полином вида

$$P(\tau) = a_0 + a_1\tau + a_2\tau^2 + \dots$$

удовлетворяющий условиям точечной интерполяции

$$\dots P(-2) = x_{-2}, \quad P(-1) = x_{-1}, \quad P(0) = x_0, \quad P(1) = x_1, \quad P(2) = x_2 \dots$$

Эти условия приводят к системе уравнений для определения неизвестных коэффициентов; назовем эти уравнения основными. Как было указано в предыдущей главе, прямое решение основных уравнений является громоздкой операцией. Составим дополнительные уравнения, являющиеся следствиями основных уравнений. Для этого напишем выражения для разностей разных порядков взятого полинома через его неизвестные коэффициенты. Приравняв эти выражения разностям табличной функции, получим расширенную систему совместных уравнений, записанную в схеме, приведенной на стр. 85.

τ	x
-3	$x_{-3} = a_0 - 3a_1 + 9a_2 - 27a_3 + 81a_4 - \dots$
-2	$x_{-2} = a_0 - 2a_1 + 4a_2 - 8a_3 + 16a_4 - \dots$
-1	$x_{-1} = a_0 - a_1 + a_2 - a_3 + a_4 - \dots$
	$x_{-1/2} = a_1 - \frac{1}{2}a_1 + \frac{1}{2}a_2 - \frac{1}{2}a_3 + \frac{1}{2}a_4 - \dots$
0	$x_0 = a_0$
	$x_{1/2} = a_0 + \frac{1}{2}a_1 + \frac{1}{2}a_2 + \frac{1}{2}a_3 + \frac{1}{2}a_4 - \dots$
+1	$x_1 = a_0 + a_1 + a_2 + a_3 + a_4 + \dots$
+2	$x_2 = a_0 + 2a_1 + 4a_2 + 8a_3 + 16a_4 + \dots$
+3	$x_3 = a_0 + 3a_1 + 9a_2 + 27a_3 + 81a_4 + \dots$
	$x_{-1/2}^1 = a_1 - 5a_2 + 19a_3 - 65a_4 + \dots$
	$x_{-1/2}^2 = 2a_2 - 12a_3 + 50a_4 + \dots$
	$x_{-1/2}^3 = 6a_3 - 36a_4 + \dots$
	$x_{-1/2}^4 = 24a_4 + \dots$
	$x_{-1/2}^5 = 6a_5 - 36a_6 + \dots$
	$x_{-1/2}^6 = 6a_6 - 12a_7 + \dots$
	$x_{-1/2}^7 = 6a_7 - 12a_8 + \dots$
	$x_{-1/2}^8 = 6a_8 - 12a_9 + \dots$
	$x_{-1/2}^9 = 6a_9 - 12a_{10} + \dots$
	$x_{-1/2}^{10} = 6a_{10} - 12a_{11} + \dots$
	$x_{-1/2}^{11} = 6a_{11} - 12a_{12} + \dots$
	$x_{-1/2}^{12} = 6a_{12} - 12a_{13} + \dots$
	$x_{-1/2}^{13} = 6a_{13} - 12a_{14} + \dots$
	$x_{-1/2}^{14} = 6a_{14} - 12a_{15} + \dots$
	$x_{-1/2}^{15} = 6a_{15} - 12a_{16} + \dots$
	$x_{-1/2}^{16} = 6a_{16} - 12a_{17} + \dots$
	$x_{-1/2}^{17} = 6a_{17} - 12a_{18} + \dots$
	$x_{-1/2}^{18} = 6a_{18} - 12a_{19} + \dots$
	$x_{-1/2}^{19} = 6a_{19} - 12a_{20} + \dots$
	$x_{-1/2}^{20} = 6a_{20} - 12a_{21} + \dots$
	$x_{-1/2}^{21} = 6a_{21} - 12a_{22} + \dots$
	$x_{-1/2}^{22} = 6a_{22} - 12a_{23} + \dots$
	$x_{-1/2}^{23} = 6a_{23} - 12a_{24} + \dots$
	$x_{-1/2}^{24} = 6a_{24} - 12a_{25} + \dots$
	$x_{-1/2}^{25} = 6a_{25} - 12a_{26} + \dots$
	$x_{-1/2}^{26} = 6a_{26} - 12a_{27} + \dots$
	$x_{-1/2}^{27} = 6a_{27} - 12a_{28} + \dots$
	$x_{-1/2}^{28} = 6a_{28} - 12a_{29} + \dots$
	$x_{-1/2}^{29} = 6a_{29} - 12a_{30} + \dots$
	$x_{-1/2}^{30} = 6a_{30} - 12a_{31} + \dots$
	$x_{-1/2}^{31} = 6a_{31} - 12a_{32} + \dots$
	$x_{-1/2}^{32} = 6a_{32} - 12a_{33} + \dots$
	$x_{-1/2}^{33} = 6a_{33} - 12a_{34} + \dots$
	$x_{-1/2}^{34} = 6a_{34} - 12a_{35} + \dots$
	$x_{-1/2}^{35} = 6a_{35} - 12a_{36} + \dots$
	$x_{-1/2}^{36} = 6a_{36} - 12a_{37} + \dots$
	$x_{-1/2}^{37} = 6a_{37} - 12a_{38} + \dots$
	$x_{-1/2}^{38} = 6a_{38} - 12a_{39} + \dots$
	$x_{-1/2}^{39} = 6a_{39} - 12a_{40} + \dots$
	$x_{-1/2}^{40} = 6a_{40} - 12a_{41} + \dots$
	$x_{-1/2}^{41} = 6a_{41} - 12a_{42} + \dots$
	$x_{-1/2}^{42} = 6a_{42} - 12a_{43} + \dots$
	$x_{-1/2}^{43} = 6a_{43} - 12a_{44} + \dots$
	$x_{-1/2}^{44} = 6a_{44} - 12a_{45} + \dots$
	$x_{-1/2}^{45} = 6a_{45} - 12a_{46} + \dots$
	$x_{-1/2}^{46} = 6a_{46} - 12a_{47} + \dots$
	$x_{-1/2}^{47} = 6a_{47} - 12a_{48} + \dots$
	$x_{-1/2}^{48} = 6a_{48} - 12a_{49} + \dots$
	$x_{-1/2}^{49} = 6a_{49} - 12a_{50} + \dots$
	$x_{-1/2}^{50} = 6a_{50} - 12a_{51} + \dots$
	$x_{-1/2}^{51} = 6a_{51} - 12a_{52} + \dots$
	$x_{-1/2}^{52} = 6a_{52} - 12a_{53} + \dots$
	$x_{-1/2}^{53} = 6a_{53} - 12a_{54} + \dots$
	$x_{-1/2}^{54} = 6a_{54} - 12a_{55} + \dots$
	$x_{-1/2}^{55} = 6a_{55} - 12a_{56} + \dots$
	$x_{-1/2}^{56} = 6a_{56} - 12a_{57} + \dots$
	$x_{-1/2}^{57} = 6a_{57} - 12a_{58} + \dots$
	$x_{-1/2}^{58} = 6a_{58} - 12a_{59} + \dots$
	$x_{-1/2}^{59} = 6a_{59} - 12a_{60} + \dots$
	$x_{-1/2}^{60} = 6a_{60} - 12a_{61} + \dots$
	$x_{-1/2}^{61} = 6a_{61} - 12a_{62} + \dots$
	$x_{-1/2}^{62} = 6a_{62} - 12a_{63} + \dots$
	$x_{-1/2}^{63} = 6a_{63} - 12a_{64} + \dots$
	$x_{-1/2}^{64} = 6a_{64} - 12a_{65} + \dots$
	$x_{-1/2}^{65} = 6a_{65} - 12a_{66} + \dots$
	$x_{-1/2}^{66} = 6a_{66} - 12a_{67} + \dots$
	$x_{-1/2}^{67} = 6a_{67} - 12a_{68} + \dots$
	$x_{-1/2}^{68} = 6a_{68} - 12a_{69} + \dots$
	$x_{-1/2}^{69} = 6a_{69} - 12a_{70} + \dots$
	$x_{-1/2}^{70} = 6a_{70} - 12a_{71} + \dots$
	$x_{-1/2}^{71} = 6a_{71} - 12a_{72} + \dots$
	$x_{-1/2}^{72} = 6a_{72} - 12a_{73} + \dots$
	$x_{-1/2}^{73} = 6a_{73} - 12a_{74} + \dots$
	$x_{-1/2}^{74} = 6a_{74} - 12a_{75} + \dots$
	$x_{-1/2}^{75} = 6a_{75} - 12a_{76} + \dots$
	$x_{-1/2}^{76} = 6a_{76} - 12a_{77} + \dots$
	$x_{-1/2}^{77} = 6a_{77} - 12a_{78} + \dots$
	$x_{-1/2}^{78} = 6a_{78} - 12a_{79} + \dots$
	$x_{-1/2}^{79} = 6a_{79} - 12a_{80} + \dots$
	$x_{-1/2}^{80} = 6a_{80} - 12a_{81} + \dots$
	$x_{-1/2}^{81} = 6a_{81} - 12a_{82} + \dots$
	$x_{-1/2}^{82} = 6a_{82} - 12a_{83} + \dots$
	$x_{-1/2}^{83} = 6a_{83} - 12a_{84} + \dots$
	$x_{-1/2}^{84} = 6a_{84} - 12a_{85} + \dots$
	$x_{-1/2}^{85} = 6a_{85} - 12a_{86} + \dots$
	$x_{-1/2}^{86} = 6a_{86} - 12a_{87} + \dots$
	$x_{-1/2}^{87} = 6a_{87} - 12a_{88} + \dots$
	$x_{-1/2}^{88} = 6a_{88} - 12a_{89} + \dots$
	$x_{-1/2}^{89} = 6a_{89} - 12a_{90} + \dots$
	$x_{-1/2}^{90} = 6a_{90} - 12a_{91} + \dots$
	$x_{-1/2}^{91} = 6a_{91} - 12a_{92} + \dots$
	$x_{-1/2}^{92} = 6a_{92} - 12a_{93} + \dots$
	$x_{-1/2}^{93} = 6a_{93} - 12a_{94} + \dots$
	$x_{-1/2}^{94} = 6a_{94} - 12a_{95} + \dots$
	$x_{-1/2}^{95} = 6a_{95} - 12a_{96} + \dots$
	$x_{-1/2}^{96} = 6a_{96} - 12a_{97} + \dots$
	$x_{-1/2}^{97} = 6a_{97} - 12a_{98} + \dots$
	$x_{-1/2}^{98} = 6a_{98} - 12a_{99} + \dots$
	$x_{-1/2}^{99} = 6a_{99} - 12a_{100} + \dots$

Из схемы видно, что в каждом из написанных столбцов разностей полинома первые члены одинаковы. Схему можно еще дополнить, вставив центральные разности между обыкновенными. Мы вставим только те, которые будут у нас использованы для вывода наиболее употребительных в астрономии формул.

Различные формулы для заданного числа узлов можно получить, выбирая разными способами столько уравнений, сколько задано узлов. (Количество узлов подбирается, исходя из указанного выше признака почти постоянства разностей.)

Принято брать по одному уравнению из каждого столбца разностей, включая и разности нулевого порядка, т. е. таблицу значений функции. Следует еще отметить, что разностные интерполяционные полиномы всегда принято располагать по разностям возрастающего порядка, а не по степеням аргумента.

§ 25. Формулы Ньютона для интерполяции вперед и назад

1. Формула Ньютона для интерполяции вперед

Выделим из таблицы значений функции ее часть, начинающуюся со значения x_0 , соответствующего значению $\tau=0$. Разности, образованные из этой части таблицы, образуют на схеме предыдущего параграфа равнобедренный треугольник. Для определения коэффициентов интерполяционного полинома возьмем уравнения, расположенные на верхней стороне этого треугольника (на схеме они подчеркнуты сплошными линиями). Ограничиваясь для определенности случаем четырех узлов со значениями τ 0, 1, 2, 3, мы будем иметь четыре уравнения:

$$\begin{aligned} a_0 &= x_0, & a_1 + a_2 + a_3 &= x_{1/2}^1, \\ 2a_2 + 6a_3 &= x_1^2, & 6a_3 &= x_{3/2}^3. \end{aligned}$$

Обозначая интерполяционный полином через $N^+(\tau)$ (значок «плюс» обозначает интерполяцию вперед), в данном случае имеем

$$N^+(\tau) = a_0 + a_1\tau + a_2\tau^2 + a_3\tau^3.$$

Выражая коэффициенты через разности, получим

$$\begin{aligned} N^+(\tau) &= x_0 + \tau \left(x_{1/2}^1 - \frac{1}{2} x_1^2 + \frac{1}{3} x_{3/2}^3 \right) + \\ &\quad + \tau^2 \left(\frac{1}{2} x_1^2 - \frac{1}{2} x_{3/2}^3 \right) + \tau^3 \cdot \frac{1}{6} x_{3/2}^3. \end{aligned}$$

Расположив правую часть по разностям последовательных порядков, получим интерполяционный полином Ньютона в окончательном виде:

$$N^+(\tau) = x_0 + \tau x_{1/2}^1 + \frac{\tau(\tau-1)}{2!} x_1^2 + \frac{\tau(\tau-1)(\tau-2)}{3!} x_{3/2}^3.$$

Этот полином естественным образом обобщается на случай любого числа узлов: последовательные члены содержат разности возрастающих порядков с нижними значками, увеличивающимися на половину при переходе от каждого члена к следующему; они умножаются на последовательные коэффициенты биномиального ряда. Так, в случае $n + 1$ узлов формула Ньютона для интерполяции вперед имеет вид

$$N^+(\tau) = x_0 + \sum_{k=1}^n \frac{\tau(\tau-1)(\tau-2)\dots(\tau-k+1)}{k!} x_{k/2}^k.$$

2. Формула Ньютона для интерполяции назад

Выделим часть таблицы, заканчивающуюся значением аргумента t_0 , т. е. будем строить формулу для таблицы с узлами, имеющими значки $\dots, -3, -2, -1, 0$. Ограничимся для простоты четырьмя узлами. Уравнения для определения коэффициентов возьмем из нижней стороны равнобедренного треугольника схемы на стр. 85 (на схеме они подчеркнуты пунктиром):

$$a_0 = x_0,$$

$$a_1 - a_2 + a_3 = x_{-1/2}^1,$$

$$2a_2 - 6a_3 = x_{-1}^2,$$

$$6a_3 = x_{-3/2}^3.$$

Решение этих уравнений дает:

$$a_3 = \frac{1}{6} x_{-3/2}^3,$$

$$a_2 = \frac{1}{2} x_{-1}^2 + \frac{1}{2} x_{-3/2}^3,$$

$$a_1 = x_{-1/2}^1 + \frac{1}{2} x_{-1}^2 + \frac{1}{3} x_{-3/2}^3,$$

$$a_0 = x_0.$$

Подставим эти коэффициенты в интерполяционный полином, который мы обозначим через $N^-(\tau)$ (значок «минус» соответствует интерполяции назад). Расположив полином сначала по степеням τ , а затем по разностям возрастающих порядков, получим

$$N^-(\tau) = x_0 + \tau x_{+1/2}^1 + \frac{\tau(\tau+1)}{2!} x_{-1}^2 + \frac{\tau(\tau+1)(\tau+2)}{3!} x_{-3/2}^3.$$

При интерполяции назад величина τ всегда отрицательна, поскольку

$t < t_0$; введем вместо τ модуль этого числа $u = -\tau$. Тогда получим:

$$N^-(u) = x_0 - ux_{-1/2} + \frac{u(u-1)}{2!} x_{-1}^2 - \frac{u(u-1)(u-2)}{3!} x_{-3/2}^3.$$

Формула в этом виде отличается от формулы для $N^+(\tau)$ только чередованием знаков при последовательных членах.

В случае $n+1$ узлов формула Ньютона для интерполяции назад имеет вид

$$N^-(u) = x_0 + \sum_{k=1}^n (-1)^k \frac{u(u-1)\dots(u-k+1)}{k!} x_{-k/2}^k.$$

Формулу для $N^-(\tau)$ называют часто формулой *по восходящим разностям*, а для $N^+(\tau)$ — *по нисходящим разностям*. Заметим, что обе формулы Ньютона не изменяют своей структуры, если начальное значение выделенной части таблицы имеет значок «г», отличный от нуля. В этом случае в правой части ко всем нижним значкам нужно прибавить g .

Пример 1. Определить прямое восхождение Луны 2 января 1950 г., 6 ч. 30 м. мирового времени.

Из Астрономического Календаря на 1950 г. выписываем α_ζ для $t = 2, 3, 4, 5, 6, 7$ января и составляем таблицу разностей:

t	α_ζ			
Январь	2	4 ^h	47 ^m	4 ^s
				3376 ^s
	3	5	43	20 + 111 ^s
				3487
				— 99 ^s
	4	6	41	27 + 12 + 7 ^s
				3499
				— 32 + 27 ^s
	5	7	39	46 — 80 + 34 — 12 ^s
				3419
				— 58 + 15 — 11 ^s
	6	8	36	45 — 138 + 49 — 23
				3281
				— 9 — 8
	7	9	31	26 — 147 + 41
				3134
				+ 32
	8	10	23	40
				3019 — 115
	9	11	13	59

Чтобы выяснить, достаточно ли шести узлов, грубо определяем коэффициент при пятой разности в формуле Ньютона. У нас $\tau \approx \frac{1}{4}$, поэтому коэффициент при пятой разности равен

$$\frac{\frac{1}{4} \left(-\frac{3}{4}\right) \cdot \left(-\frac{7}{4}\right) \cdot \left(-\frac{11}{4}\right) \cdot \left(-\frac{15}{4}\right)}{120} < \frac{3 \cdot 2 \cdot 3 \cdot 4}{16 \cdot 120} \approx 0,04;$$

можно принять, что коэффициент приблизительно равен 0,04, поэтому член с пятой разностью дает больше 1°.

Так как предельная погрешность таблицы равна 0°,5, то шести узлов недостаточно, ибо член с разностью 6-го порядка может дать число, большее, чем 0°,5. Поэтому дополним таблицу седьмым и, для контроля, восьмым узлами. Коэффициент при шестой разности приблизительно равен произведению $0,04 \cdot \frac{5}{6}$, что можно принять грубо за 0,03, и член с шестой разностью даст число порядка 0,4, что меньше погрешности таблиц.

Вычисления располагаем по следующей схеме:

		Коэффициенты	Разности	Члены формулы
$\tau = 0,27083$		$N_1 = 0,27083$	$+ 3376^s$	$\Delta_1 = + 914^s,3$
$\tau - 1 = -0,72917$	$\frac{\tau - 1}{2} = -0,3646$	$N_2 = -0,0987$	$+ 111$	$\Delta_2 = - 11,0$
$\tau - 2 = -1,72917$	$\frac{\tau - 2}{3} = -0,5764$	$N_3 = + 0,0569$	$- 99$	$\Delta_3 = - 5,6$
$\tau - 3 = -2,72917$	$\frac{\tau - 3}{4} = -0,682$	$N_4 = -0,039$	$+ 7$	$\Delta_4 = - 0,3$
$\tau - 4 = -3,72917$	$\frac{\tau - 4}{5} = -0,747$	$N_5 = + 0,029$	$- 27$	$\Delta_5 = + 0,8$
$\tau - 5 = -4,72917$	$\frac{\tau - 5}{6} = -0,788$	$N_6 = -0,023$	$- 12^s$	$\Delta_6 = + 0^s,3$
				$\Delta = + 898^s,5$
				$\Delta = 14^m 58^s$
				$a_0 = 4^h 47^m 4^s$
				$a = 5^h 2^m 2^s$

Для упражнения определим еще, можно ли в приведенных выше вычислениях остановиться на довольно малой нисходящей разности 4-го порядка. Оценка коэффициента дает

$$\left| \frac{\frac{1}{4} \cdot \left(-\frac{3}{4}\right) \cdot \left(-\frac{7}{4}\right) \cdot \left(-\frac{11}{4}\right)}{24} \right| < \frac{3}{16} \cdot \frac{2 \cdot 3}{24} \approx 0,05;$$

член с четвертой разностью дает 0,35, что меньше табличной погрешности. Этот расчет показывает, что не всегда можно доверять первоначальной оценке. Нужно выяснить, устойчива ли малость разности, так как малость может вызываться тем, что мы попали на смену знака разностей данного порядка в таблице. Поэтому, получив малую разность некоторого порядка и прикинув, что она мало влияет на результат, нужно проверить следующий член. В рассмотренном примере мы довели счет до разности 7-го порядка, она почти такая же, как и разность 6-го порядка; можно поэтому остановиться на члене с разностью 6-го порядка.

К схеме вычислений полезно сделать следующие пояснения. Дана полная схема расчета, которая может быть несколько сокращена при пользовании арифмометром. В первом столбце подготовлены множители в числителях коэффициентов. В следующем столбце подготовлены сомножители полных коэффициентов, которые представлены в форме

$$N_1 = \tau,$$

$$N_2 = N_1 \frac{\tau - 1}{2};$$

$$N_3 = N_2 \frac{\tau - 2}{3} \text{ и т. д.}$$

По этим формулам получены последовательные коэффициенты формулы N_1, N_2, \dots, N_6 . Коэффициенты умножаются на разности того же порядка; получаются последовательные члены (начиная со второго) формулы, обозначенные $\Delta_1, \Delta_2, \dots, \Delta_6$.

Вычисления ведутся с запасным знаком (десятые доли секунды), чтобы уменьшить накопление ошибок от отдельных членов формулы. Поэтому вычисляем $\tau = N_1$ с пятью знаками после запятой, чтобы после умножения на четырехзначную первую разность можно было ручаться за десятые доли секунды. Эта операция законна, так как τ задается, и, следовательно, это число можно считать точным. Поскольку разность 2-го порядка трехзначная, а разность 3-го порядка мало от нее отличается, то коэффициенты при них подсчитаны с четырьмя знаками после запятой (с «запасом точности»). Разности остальных порядков двузначные, поэтому коэффициенты при них достаточно взять с тремя знаками после запятой. В четвертом столбце выписаны для удобства последовательные разности. Этот столбец можно и не заполнять, если таблица разностей расположена вблизи схемы (впрочем, если вычисление поручается техническому исполнителю, этот столбец лучше выписать, ибо столбец вторых сомножителей в интерполяционной формуле расположен по наклонной линии, а столбец первых сомножителей по вертикали, и это может служить источником ошибок). Последний столбец содержит последовательные члены формулы; сложив их с учетом десятых долей и округлив сумму до секунд, получим число, которое нужно прибавить к начальному значению. Отдельные члены не нужно вычислять в таких задачах, в которых по предыдущему опыту известно достаточное число узлов для обеспечения нужной точности. Если это число неизвестно, то последний столбец необходим, так как надо знать отдельные последовательные члены формулы, чтобы найти наибольший порядок разностей, влияющих на результат интерполяции. Это, как выше сказано, и определяет степень полинома, т. е. число узлов. В нашем случае член с разностью 4-го порядка не влияет на результат, но следующий член дает почти единицу, поэтому нужно взять еще разность 6-го порядка и, следовательно, семь узлов.

Пример 2. Определить склонение Луны 29 декабря 1950 г., в 17 ч. 45 м. мирового времени. За начальное значение аргумента принимаем момент декабрь 30,0, ближайший к заданному. В данном случае интерполировать следует «назад». Выписываем из Астрономического Календаря таблицу t , $\delta\zeta$ и образуем разности:

t	$\delta\zeta$
Декабрь 24	$+28^{\circ}23',0$
	$-27',1$
25	27 55,9
	$-80',0$
	$-107,1$
	$+6',0$
26	26 8,8
	$-74,0$
	$+3',6$
	$-181,1$
	$+9,6$
	$-1',5$
27	23 7,7
	$-64,4$
	$+2,1$
	$+0',2$
	$-245,5$
	$+11,7$
	$-1',3$
28	19 2,2
	$-52,7$
	$+0',8$
	$-298,2$
	$+12',5$
29	14 4,0
	$-40',2$
	$-338',4$
30	$8^{\circ}25',6$

Разность 6-го порядка достаточно мала, поэтому можно ограничиться семью узлами.

У нас $t - t_0 = -6^h 15^m$; $\tau = -0,26042$. Вычисления производим по схеме, указанной в примере 1:

τ	Коэффициенты	Разности	Члены формулы
$\tau = -0,26042$	$N_1 = -0,26042$	$-338,4$	$\Delta_1 = +88',13$
$\tau + 1 = 0,7396$	$\frac{\tau + 1}{2} = 0,3698$	$N_2 = -0,0963$	$-40,2$
		$\Delta_2 = +3,87$	
$\tau + 2 = 1,7396$	$\frac{\tau + 2}{3} = 0,580$	$N_3 = -0,0559$	$+12,5$
		$\Delta_3 = -0,70$	
$\tau + 3 = 2,7396$	$\frac{\tau + 3}{4} = 0,68$	$N_4 = -0,04$	$+0,8$
		$\Delta_4 = +0,03$	
$\tau + 4 = 3,7396$	$\frac{\tau + 4}{5} = 0,75$	$N_5 = -0,03$	$-1,3$
		$\Delta_5 = +0,04$	
$\tau + 5 = 4,7396$	$\frac{\tau + 5}{6} = 0,79$	$N_6 = -0,02$	$+0,2$
		$\Delta_6 = -0,00$	
$\delta_0 = 8^{\circ}25',6$			$\Delta = +91',3$
$\Delta = 1\ 31,3$			
$\delta = 9^{\circ}56',9$			

§ 26. Формула Стирлинга

Назовем *центральной линией* таблицы разностей ту, на которой находится начальное значение и обыкновенные разности четного порядка со знаком нуль. Дополним таблицу центральными разностями нечетного порядка, стоящими на той же линии, и возьмем уравнения, расположенные на центральной линии. Ограничиваясь пятью узлами, т. е. учитывая лишь разности 4-го порядка, получим следующую систему уравнений для определения коэффициентов полинома:

$$\begin{aligned}a_0 &= x_0, \\a_1 + a_3 &= x_0^1, \\2a_2 + 2a_4 &= x_0^2, \\6a_3 &= x_0^3, \\24a_4 &= x_0^4.\end{aligned}$$

Решение этой системы дает следующие значения коэффициентов:

$$\begin{aligned}a_0 &= x_0, \\a_1 &= x_0^1 - \frac{1}{6} x_0^3, \\a_2 &= \frac{1}{2} x_0^2 - \frac{1}{24} x_0^4, \\a_3 &= \frac{1}{6} x_0^3, \quad a_4 = \frac{1}{24} x_0^4.\end{aligned}$$

Интерполяционный полином имеет вид

$$S(\tau) = x_0 + \tau \left(x_0^1 - \frac{1}{6} x_0^3 \right) + \tau^2 \left(\frac{1}{2} x_0^2 - \frac{1}{24} x_0^4 \right) + \tau^3 \frac{1}{6} x_0^3 + \tau^4 \frac{1}{24} x_0^4.$$

Располагая его по разностям последовательных порядков, получим *формулу Стирлинга* в обычном виде:

$$S(\tau) = x_0 + \frac{\tau}{1!} x_0^1 + \frac{\tau^2}{2!} x_0^2 + \frac{\tau(\tau^2 - 1)}{3!} x_0^3 + \frac{\tau^2(\tau^2 - 1)}{4!} x_0^4.$$

Для формулы Стирлинга при полном соблюдении принципа точечной интерполяции всегда берется нечетное число узлов: начальный и симметричные относительно него узлы, предшествующие и последующие. Поэтому полином Стирлинга всегда имеет четную степень. Закон образования последовательных членов полинома, начиная с третьего, таков: коэффициент при разности четного порядка получается из предыдущего коэффициента (при разности

нечетного порядка) увеличением на единицу аргумента факториала в знаменателе и добавлением множителя τ в числителе; коэффициент при разности нечетного порядка $2k+1$ получается из предыдущей разности тоже нечетного порядка увеличением на две единицы аргумента факториала в знаменателе и добавлением в числителе множителя $\tau^2 - k^2$. Так, например, члены с пятой и шестой разностями имеют вид:

$$\frac{\tau(\tau^2-1)}{(3+2)!} \cdot (\tau^2-2^2) x_0^{3+2} = \frac{\tau(\tau^2-1)(\tau^2-2^2)}{5!} x_0^5,$$

$$\frac{\tau(\tau^2-1)(\tau^2-2^2)\tau}{(5+1)!} x_0^{5+1} = \frac{\tau^2(\tau^2-1)(\tau^2-2^2)}{6!} x_0^6.$$

Закон образования коэффициентов в формуле Стирлинга таков, что коэффициенты изменяются неравномерно. Например, при $\tau = \frac{1}{2}$ последовательные коэффициенты равны

$$\frac{1}{2}, \frac{1}{8}, -\frac{1}{16}, -\frac{1}{128}, \frac{3}{256}, \frac{1}{1024}, \dots$$

Знаки коэффициентов чередуются парами: два положительных, два отрицательных, два положительных и т. д. Коэффициент при нечетной разности, начиная с пятой, больше предшествующего коэффициента при четной разности. Поэтому при использовании формулы Стирлинга не следует останавливаться на разности нечетного порядка, так как соответствующий член при медленном убывании разностей может оказаться больше предшествующего.

Кроме того, формула Стирлинга, оканчивающаяся нечетной разностью, не будет точно представлять крайние из используемых узлов.

Практически, однако, часто дело будет сводиться к формуле, заканчивающейся на нечетной разности, так как любой коэффициент при четной разности гораздо меньше предшествующего коэффициента при нечетной разности. Если, например, член с пятой разностью давал 5 единиц последнего знака, то при $\tau = \frac{1}{2}$ член с шестой разностью даст меньше половины единицы последнего знака и, следовательно, может повлиять только на запасный знак и изменить результат не больше, чем на единицу последнего знака (если во всех членах учитывается запасный знак).

Вследствие симметрии формулы Стирлинга относительно линии, проходящей через начальный узел, эта формула без всяких изменений применяется как при интерполяции вперед ($\tau > 0$), так и при интерполяции назад ($\tau < 0$).

Пример. Определить прямое восхождение Луны 13 декабря 1950 г. в 9 ч. 36 м. мирового времени.

За начальный момент примем 13 декабря, 0 ч. мирового времени; выпишем α в три предшествующих и три последующих момента и образуем таблицу обыкновенных разностей:

t	α							
Декабрь 10	17 ^h 41 ^m 57 ^s							
		4111 ^s						
11	18 50 28	— 205 ^s						
		3906						
		— 111 ^s						
12	19 55 34	— 316						
		+ 97 ^s						
		3590						
		— 14						
		— 30 ^s						
13	20 55 24	+ 3425	— 330	+ 20	+ 67	— 38	— 15	
		3260		+ 53		— 45		
14	21 49 44		— 277		+ 22			
		2983		+ 75				
15	22 39 27		— 202					
		2781						
16	23 25 48							

Затем вписываем на линии, проходящей через начало, центральные разности нечетного порядка. Вычисление коэффициентов формулы Стирлинга

$$S(\tau) = x_0 + S_1 x_0^1 + S_2 x_0^2 + S_3 x_0^3 + S_4 x_0^4 + S_5 x_0^5 + S_6 x_0^6$$

производим при помощи соотношений

$$S_1 = \tau, \quad S_2 = S_1 \frac{\tau}{2}, \quad S_3 = S_1 \frac{\tau^2 - 1}{6}, \quad S_4 = S_3 \frac{\tau}{4},$$

$$S_5 = S_3 \frac{\tau^2 - 4}{20}, \quad S_6 = S_5 \frac{\tau}{6},$$

Весь расчет, необходимый для интерполяции, расположим в следующей схеме:

		Коэффициенты	Разности	Члены формулы
$\tau = 0,40000$	$\frac{\tau}{4} = 0,1000$	$S_1 = 0,40000$	+ 3425 ^s	+ 1370,0 ^s
$\tau^2 = 0,1600$	$\frac{\tau^2}{2} = 0,0800$	$S_2 = 0,0800$	— 330	— 26,4
$\tau^2 - 1 = -0,8400$	$(\tau^2 - 1):6 = -0,1400$	$S_3 = -0,056$	+ 20	— 1,1
$\tau^2 - 4 = -3,8400$	$(\tau^2 - 4):20 = -0,192$	$S_4 = -0,006$	+ 67	— 0,4
	$\tau:6 = 0,067$	$S_5 = +0,001$	— 38	— 0,4
		$S_6 = +0,001$	— 15 ^s	— 0,0 ^s
$\alpha_0 = 20^h 55^m 24^s$				
$\Delta = 22^m 22^s, \quad \alpha = 21^h 17^m 46^s$				$\Delta = +1342^s$

Через Δ в схеме обозначена сумма членов формулы Стирлинга, содержащих разности от 1-го до 6-го порядков. Это число нужно прибавить к начальному значению a_0 , чтобы получить искомое значение прямого восхождения.

§ 27. Формула Бесселя (два варианта)

1. Первый вариант

Заданное для интерполяции значение аргумента заключено между двумя табличными значениями, которые на схеме § 24 (стр. 84) обозначены через t_0 и $t_0 + h = t_1$. Примем за начало отсчета среднее значение аргумента $\frac{1}{2}(t_0 + t_1)$. Соответственно с этим введем

аргумент $\tau' = \frac{t - t_0}{h} - \frac{1}{2}$, который в узлах принимает значения

$$\begin{aligned} \tau' & \parallel \dots - \frac{5}{2}, -\frac{3}{2}, -\frac{1}{2}, +\frac{1}{2}, +\frac{3}{2}, \dots \\ t & \parallel \dots t_{-2}, t_{-1}, t_0, t_1, t_2, \dots \end{aligned}$$

при этом $t_k = t_0 + kh$, где k — целые отрицательные и положительные числа.

Ограничимся четырьмя узлами t_{-1}, t_0, t_1, t_2 , т. е. будем строить интерполяционный полином 3-й степени вида

$$B(\tau') = b_0 + b_1\tau' + b_2\tau'^2 + b_3\tau'^3.$$

Составим систему уравнений, аналогичную той, которая была приведена в § 19, но с аргументом τ' (см. стр. 96).

Соберем уравнения, расположенные на средней линии между узлами t_1 и t_0 :

$$b_0 + \frac{1}{4}b_2 = x_{1/2}, \quad b_1 + \frac{1}{4}b_3 = x_{1/2}^1, \quad 2b_2 = x_{1/2}^2, \quad 6b_3 = x_{1/2}^3.$$

Решение системы дает

$$b_3 = \frac{1}{6}x_{1/2}^3, \quad b_2 = \frac{1}{2}x_{1/2}^2, \quad b_1 = x_{1/2}^1 - \frac{1}{24}x_{1/2}^3, \quad b_0 = x_{1/2} - \frac{1}{8}x_{1/2}^3.$$

Подставляя эти коэффициенты в полином

$$B(\tau') = b_0 + b_1\tau' + b_2\tau'^2 + b_3\tau'^3$$

и располагая его затем по разностям возрастающих порядков, получим формулу Бесселя:

$$B(\tau') = x_{1/2} + \tau'x_{1/2}^1 + \frac{\tau'^2 - \frac{1}{4}}{2!}x_{1/2}^2 + \frac{\tau'(\tau'^2 - \frac{1}{4})}{3!}x_{1/2}^3.$$

Формула довольно легко обобщается на случай любого четного числа

τ'

$$-1,5 \quad x_{-1} = b_0 - \frac{3}{2}b_1 + \frac{9}{4}b_2 - \frac{27}{8}b_3$$

$$x_{1/2}^1 = b_1 - 2b_2 + \frac{13}{4}b_3$$

$$-0,5 \quad x_0 = b_0 - \frac{1}{2}b_1 + \frac{1}{4}b_2 - \frac{1}{8}b_3$$

$$x_0^2 = 2b_2 - 3b_3$$

$$x_{1/2} = b_0 + \frac{1}{4}b_2$$

$$x_{1/2}^1 = b_1 + \frac{1}{4}b_3$$

$$x_{1/2}^2 = 2b_2$$

$$x_{1/2}^3 = 6b_3$$

$$+0,5 \quad x_1 = b_0 + \frac{1}{2}b_1 + \frac{1}{4}b_2 + \frac{1}{8}b_3$$

$$x_1^2 = 2b_2 + 3b_3$$

$$x_{3/2}^1 = b_1 + 2b_2 + \frac{13}{4}b_3$$

$$+1,5 \quad x_2 = b_0 + \frac{3}{2}b_1 + \frac{9}{4}b_2 + \frac{27}{8}b_3$$

узлов. Например, интерполяционный полином 5-й степени имеет вид

$$B(\tau') = x_{1/2} + \tau' x_{1/2}^1 + \frac{\tau'^2 - 0,25}{2!} x_{1/2}^2 + \\ + \frac{\tau'(\tau'^2 - 0,25)}{3!} x_{1/2}^3 + \frac{(\tau'^2 - 0,25)(\tau'^2 - 2,25)}{4!} x_{1/2}^4 + \\ + \frac{\tau'(\tau'^2 - 0,25)(\tau'^2 - 2,25)}{5!} x_{1/2}^5.$$

Закон же образования коэффициентов довольно прост: коэффициент при разности нечетного порядка получается из предыдущего коэффициента (при четной разности) увеличением на единицу аргумента факториала в знаменателе и умножением на τ' числителя; коэффициент при разности четного порядка $2k$ получается из коэффициента при предыдущей разности четного порядка увеличением на две единицы аргумента факториала в знаменателе и добавлением множителя $\tau'^2 - \frac{(2k-1)^2}{4}$ в числителе. Отметим, что в общей формуле свободные члены биномов в числителе имеют вид

$$\frac{1}{4}, \frac{3^2}{4} = \frac{9}{4}, \frac{5^2}{4} = \frac{25}{4}, \frac{7^2}{4} = \frac{49}{4} \text{ и т. д.}$$

Легко написать в общем виде члены с разностями четного и нечетного порядков. Коэффициент при $x_{1/2}^{2k}$ равен

$$\frac{\left(\tau'^2 - \frac{1}{4}\right)\left(\tau'^2 - \frac{9}{4}\right) \dots \left[\tau'^2 - \frac{(2k-1)^2}{4}\right]}{(2k)!};$$

коэффициент при $x_{1/2}^{2k+1}$ равен

$$\frac{\tau' \left(\tau'^2 - \frac{1}{4}\right)\left(\tau'^2 - \frac{9}{4}\right) \dots \left[\tau'^2 - \frac{(2k-1)^2}{4}\right]}{(2k+1)!}.$$

Из формул видно, что особенно удобна формула Бесселя при $\tau' = 0$, т. е. при $t = \frac{t_0 + t_1}{2}$. Интерполяция в этом случае называется *интерполяцией на середину* (в середину). При этом в формуле Бесселя обращаются в нуль коэффициенты при всех нечетных разностях, что вдвое уменьшает число сохраняющихся членов, благодаря чему сокращается количество вычислений и уменьшается погрешность. Формула без изменений применяется при t , близком к t_0 ($\tau' < 0$), и при t , близком к t_1 ($\tau' > 0$). Если, например, $t = t_0 + 0,25h$, то $\tau' = -0,25$, если же $t = t_1 - 0,25h = t_0 + 0,75h$, то $\tau' = +0,25$.

2. Второй вариант

Заменим в формуле Бесселя τ' на $\tau - 0,5$, т. е. будем отсчитывать нормированный аргумент от начального узла. В коэффициентах при разностях последовательных порядков (кроме первого) есть множители вида

$$\tau'^2 - \frac{(2m-1)^2}{4}.$$

Преобразуя эти множители к аргументу τ , получим

$$\tau'^2 - \frac{(2m-1)^2}{4} = \tau^2 - \tau - m^2 + m = (\tau - m)(\tau + m - 1).$$

Полином Бесселя с новым аргументом записывается так:

$$\begin{aligned} B(\tau) = & x_{1/2} + \left(\tau - \frac{1}{2}\right) x_{1/2}^1 + \frac{\tau(\tau-1)}{2!} x_{1/2}^2 + \\ & + \frac{\tau(\tau-1)\left(\tau - \frac{1}{2}\right)}{3!} x_{1/2}^3 + \frac{(\tau+1)\tau(\tau-1)(\tau-2)}{4!} x_{1/2}^4 + \\ & + \frac{(\tau+1)\tau(\tau-1)(\tau-2)\left(\tau - \frac{1}{2}\right)}{5!} x_{1/2}^5. \end{aligned}$$

Эта формула может быть получена и из схемы § 19, если отобрать уравнения, расположенные на средней линии между начальным и следующим узлами. Читателю рекомендуется сделать это самостоятельно и таким образом проверить только что написанную формулу.

Отметим еще изменение второго варианта формулы Бесселя, часто применяемое на практике. Сделаем простые преобразования, не требующие пояснений:

$$x_{1/2} = \frac{x_0 + x_1}{2} = \frac{x_0 + x_0 + x_{1/2}^1}{2} = x_0 + \frac{1}{2} x_{1/2}^1.$$

Замена первого члена по этому равенству и приведение членов дает формулу, отличающуюся только в первых двух членах от предыдущей:

$$B(\tau) = x_0 + \tau \cdot x_{1/2}^1 + \frac{\tau(\tau-1)}{2!} x_{1/2}^2 + \dots$$

Пример. Определить прямое восхождение Луны 13 декабря 1950 г. в 9 ч. 36 м. мирового времени. Указанный момент заключен между таблич-

ными моментами 13,0 и 14,0. Выпишем табличные значения, симметричные относительно средней линии, и вычислим нужные для интерполяции разности:

t	α						
Декабрь 11	18 ^h 50 ^m 28 ^s			3906 ^s			
12	19 55 34			— 316 ^s			
				3590		— 14 ^s	
13	20 55 24		— 330			+ 67 ^s	
	21 22 34	3260 ^s	— 304 ^s	+ 53 ^s	+ 44 ^s	— 45 ^s	
14	21 49 44		— 277		+ 22 ^s		
		2983		+ 75 ^s			
15	22 39 27		— 202 ^s				
		2781 ^s					
16	23 25 48						

Определяем τ и τ' : $\tau = 0,4$, $\tau' = 0,4 - 0,5 = -0,1$. Вычисления можно вести по следующей схеме:

	Коэффициенты	Разности	Члены формулы
$\tau' = -0,1$	$B_1 = -0,1$	+ 3260 ^s	$\Delta_1 = -326,0$
$\tau'^2 = 0,25 = -0,24$	$B_2 = -0,1200$	— 304	$\Delta_2 = + 36,5$
$\tau' : 3 = -0,0333$	$B_3 = + 0,004$	+ 53	$\Delta_3 = + 0,2$
$\tau'^2 - 2,25 = -2,24$	$B_4 = + 0,022$	+ 44	$\Delta_4 = + 1,0$
$(\tau'^2 - 2,25) : 12 = -0,187$	$B_5 = -0,0004$	— 45 ^s	$\Delta_5 = + 0,0$
$\tau' : 5 = -0,02$			$\Delta = -288^s = -4^m 48^s$
$\alpha = 21^h 17^m 46^s$			$\alpha_0 = 21^h 22^m 34^s$

В нашем примере вычисления по формуле Бесселя оказались несколько более простыми, чем по формуле Стирлинга.

Здесь также обращает на себя внимание то обстоятельство, что коэффициент B_2 несколько больше, чем B_1 , коэффициент B_4 в пять раз больше, чем B_3 . Можно показать, что коэффициенты при нечетных разностях всегда значительно меньше, чем при предыдущих четных разностях.

Действительно, из общей формулы Бесселя (1-й вариант) имеем

$$\frac{B_{2m+1}}{B_{2m}} = \frac{\tau'}{2m+1},$$

отсюда

$$\left| \frac{B_{2m+1}}{B_{2m}} \right| < \frac{1}{2(2m+1)},$$

так как $|\tau'| < 0,5$ (равенства не может быть, поскольку $\tau' = 0,5$ означает, что берется табличное значение аргумента). Далее,

$$\frac{B_{2m}}{B_{2m-1}} = \frac{\tau'^2 - \frac{(2m-1)^2}{4}}{2m\tau'}.$$

Обозначим для краткости это отношение буквой u ; эта величина есть функция τ' и зависит от параметра m . Так как

$$\frac{du}{d\tau'} = \frac{\tau'^2 + \frac{(2m-1)^2}{4}}{2m\tau'^2} > 0$$

при всех τ' , то отношение u монотонно растет при всех τ' . Если $\tau' \leq 0$, то τ' принимает значения от $-0,5$ до 0 . Поэтому при $\tau' = -0,5$ имеем минимальную величину u :

$$u_{\min} = \frac{\frac{1}{4} - \frac{(2m-1)^2}{4}}{-m} = m-1; \quad \text{при } \tau' = 0 \quad u = \infty.$$

Если $\tau' \geq 0$, то $u < 0$ и принимает значения от $-\infty$ до $|m-1|$. Следовательно, отношение коэффициента при четной разности к предыдущему превышает $m-1$, независимо от величины τ' .

§ 28. Общие замечания о применении разностных интерполяционных формул

Если заданное значение аргумента находится вблизи начала таблицы, то применять можно только формулу Ньютона для интерполяции вперед, так как для других формул требуется знание узлов, предшествующих тому, который принимается за начальный. По той же причине можно пользоваться только формулой Ньютона для интерполяции назад, если заданное значение аргумента находится около конца таблицы. Если же заданное значение таково, что принятое начальное значение находится внутри таблицы, то можно выбирать любую из разностных формул.

Могут быть указаны два критерия, которыми при этом пользуются. Простейший критерий — величины коэффициентов при разностях последовательных порядков в разных формулах. При заданном t и соответствующем τ выгоднее та формула, которая дает меньшие коэффициенты при разностях, так как при этом уменьшаются неизбежные ошибки от округлений в таблице и в раз-

ностях. Для применения указанного критерия приведем таблицу коэффициентов для различных интерполяционных формул.

Порядок разностей	τ	Интерполяционные формулы		
		Ньютона (интерполяция вперед)	Стирлинга	Бесселя (I вариант)
2-й	0,1	— 0,045	+ 0,005	— 0,045
	0,2	— 0,080	+ 0,020	— 0,080
	0,3	— 0,105	+ 0,045	— 0,105
	0,4	— 0,120	+ 0,080	— 0,120
	0,5	— 0,125	+ 0,125	— 0,125
3-й	0,1	+ 0,0285	— 0,0165	+ 0,006
	0,2	+ 0,0480	— 0,0320	+ 0,008
	0,3	+ 0,595	— 0,0455	+ 0,007
	0,4	+ 0,0640	— 0,0560	+ 0,004
	0,5	+ 0,0625	— 0,0625	+ 0,000
4-й	0,1	— 0,0207	— 0,0004	+ 0,0078
	0,2	— 0,0336	— 0,0016	+ 0,0144
	0,3	— 0,0402	— 0,0034	+ 0,0193
	0,4	— 0,0416	— 0,0056	+ 0,0224
	0,5	— 0,0391	— 0,0078	+ 0,0234

Как сказано выше, достаточно задать значения τ от 0 до 0,5, так как при $\tau > 0,5$ нужно просто взять другое начальное значение. Нам понадобятся только модули коэффициентов, но мы поставили также и знаки. Сравнение коэффициентов показывает, что во всех случаях коэффициенты формулы Ньютона не меньше коэффициентов двух других формул, а в ряде случаев и больше них.

Другим критерием качества интерполяционных формул может служить оценка ошибки интерполяции. Эта оценка зависит от свойств интерполируемой функции, от числа узлов и от значений аргумента. Комбинации узлов различны при пользовании разными формулами; максимум модуля производной порядка $n + 1$ берется в разных областях аргумента. Поэтому оценки ошибки интерполяции получатся несколько различными по разным формулам. Если можно выполнить сравнение оценок ошибок интерполяции по разным формулам, то выбирается формула с наименьшей оценкой ошибки интерполяции.

* * *

В заключение этой главы дадим краткую сводку правил применения формул точечной интерполяции.

1. Выбираем начальное значение t_0 так, чтобы $|t - t_0| \leq \frac{1}{2}h$. Если $t > t_0$, то интерполируем вперед, если $t < t_0$ — интерполируем назад.

2. Составляем таблицу разностей и по порядку разностей, остающихся приблизительно постоянными, определяем степень интерполяционного полинома.

3. Вычисляем аргумент $\tau = \frac{t - t_0}{h}$, если используются формулы Ньютона, Стирлинга и Бесселя (II вариант) или $\tau' = \tau - \frac{1}{2}$ при пользовании формулой Бесселя в I варианте.

4. После этого вычисляем последовательные члены выбранной формулы до тех пор, пока они влияют на результат. Если в таблице m знаков после запятой, то часто последовательные члены считают с $m + 1$ знаком. Лишний знак учитывается при сложении и отбрасывается после вычисления суммы с округлением.

ЧАСТЬ III

СВЕДЕНИЯ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

ГЛАВА 7

СЛУЧАЙНЫЕ СОБЫТИЯ; ОСНОВНЫЕ ПОНЯТИЯ И ТЕОРЕМЫ

§ 29. Случайные явления

Понятие вероятности вводится при изучении таких явлений, для которых невозможно точно предсказать результаты будущих наблюдений, если даже и известны условия, в которых явление будет происходить.

Примеры. а) При всякого рода измерениях вследствие несовершенства инструментов и разных случайных причин появляются случайные ошибки измерений. Приступая к измерениям, мы не можем заранее точно указать ни величины ошибки, ни достаточно узких границ, в которых ошибка будет заключаться;

б) владелец лотерейного билета не может до розыгрыша предсказать размер выигрыша;

в) при случайном выборе звезды из каталога нельзя заранее предсказать ее основные характеристики;

г) при случайном выборе человека из некоторой совокупности людей, например новобранцев, нельзя заранее указать цвет глаз или другую подобную характеристику;

д) если в урне находятся белые и черные шары, то при извлечении шара, не глядя в урну, мы не можем заранее сказать, какой цвет будет у вынутого шара.

Такого рода явления называют *случайными*.

Обсуждая ожидаемые результаты предстоящего наблюдения случайного явления, возможно зачастую указать не один определенный результат, а несколько; при этом *априорные суждения* (суждения, высказанные до опыта) зависят до некоторой степени от тех условий, в которых явление будет происходить.

Примеры. а) Если известно, что модуль случайной ошибки при измерении угломерным инструментом не превышает $10'$, то на вопрос о величине ошибки при предстоящем наблюдении можно дать ответ, например, в такой форме: возможны ошибки с модулями от $0'$ до $2'$ или от $2'$ до $4'$, или от $4'$ до $6'$ и т. д.;

б) если из каталога выбирается звезда определенного спектрального класса, то можно примерно указать возможные границы для ее абсолютной величины, более узкие, чем при отсутствии сведений о спектральном типе;

в) если в урне три белых шара и два черных, то до извлечения шара можно сказать, что возможные результаты будут: белый шар или черный шар;

г) если свободно подбрасывается монета, то возможно выпадение надписи или герба;

д) если свободно бросается игральная кость, то возможны следующие результаты:

1, 2, 3, 4, 5, 6 очков.

Многократные наблюдения случайных явлений дают некоторые практические критерии, с помощью которых можно, принимая во внимание известные условия, ожидать, что из всех предполагаемых результатов наблюдений одни более возможны, чем другие.

Практика измерений показывает, что малые случайные ошибки встречаются чаще, чем большие. На основании этого, приступая к измерению, мы с большей уверенностью ожидаем малой ошибки, чем большой, если при измерении будут соблюдаться все необходимые технические условия. Во всех подобных случаях в практике употребляется слово «возможно», «вероятно», «почти наверно» и т. п., которые качественно оценивают степень уверенности в некотором результате предстоящего наблюдения.

При введении количественной характеристики такой «степени уверенности» условимся сначала рассматривать только такие случайные явления, которые удовлетворяют следующим условиям:

1. Число возможных результатов наблюдений n конечно; обозначим эти результаты буквами

$$A_1, A_2, \dots, A_k, A_{k+1}, \dots, A_n.$$

Назовем *случаем* любой возможный результат случайного явления. В частности, результаты A_k суть случаи.

Слово «случай» следует понимать достаточно широко. Если рассматриваются осадки, то дождь есть «случай», отсутствие осадков также есть «случай». Если говорят о возможных физических характеристиках случайно выбранной звезды, то принадлежность ее к тому или иному спектральному классу есть случай; разные значения показателя цвета также являются случаями. Для правильного понимания последнего примера следует учесть, что речь идет об указании физических характеристик *до* исследования звезды.

2. Все случаи, перечисленные в пункте 1, образуют *полную группу случаев*, т. е. хотя бы один из них обязательно произойдет *).

3. Перечисленные в пункте 1 случаи *несовместны*, т. е. появление одного из них исключает появление всех остальных (употребляется также термин: попарно несовместимые события).

Если в примере (а) на стр. 103 ответ на вопрос о величине модуля ошибки дать в форме — возможна ошибка с модулем от $0'$ до $5'$ или от $5'$ до $10'$, то эти случаи несовместны, если только условиться, к какому из случаев отнести ошибку, модуль которой точно равен $5'$.

*) В литературе употреблялось и употребляется также название «единственно возможные» случаи, чтобы указать на полноту группы.

4. Случаи, перечисленные в пункте 1, *равновозможны*. Понятие о равновозможности случаев мы будем считать *первичным*, отражающим свойства случаев. Использование понятия равновозможности в разных задачах мотивируется главным образом симметрией явления относительно случаев. В некоторых задачах предположение о равной возможности является гипотезой о свойствах явления, которая подлежит проверке наблюдениями. Только в простейших задачах легко указать, допустимо ли предположение о равновозможности случаев, перечисленных в списке, подобном тому, который дан в пункте 1.

Для иллюстрации предположим, что бросают игральную кость так, что сверху может оказаться любая из граней с 1, 2, ..., 6 точками. Если кость представляет точный однородный куб, то выпадение той или иной грани сверху есть случайное явление, для которого полная группа случаев имеет такой вид: выпадает 1, 2, 3, 4, 5, 6, и все эти случаи попарно несовместны и равновозможны вследствие симметрии граней относительно центра тяжести, вокруг которого вращается кость при бросании и падении. Если бы кость состояла из двух частей, деревянной и железной, то центр тяжести переместился бы к железной грани, и при падении кости было бы, по здравому смыслу, более вероятно, что сверху окажется число на деревянной грани, чем число на железной грани. Утверждение о равной возможности граней не было бы оправданным по сведениям о строении кости, так как отсутствует симметрия относительно центра тяжести. Если бы таких сведений не было, то первоначально можно было бы сделать гипотезу о равновозможности всех чисел 1, 2, ..., 6. Положим для определенности, что на целиком железной грани поставлена единица, а на деревянной — шестерка. В большой серии бросаний выпадение шестерки гораздо чаще, чем единицы, заставило бы пересмотреть гипотезу о равной возможности выпадения всех чисел от единицы до шестерки.

Рассмотрим еще пример, который в дальнейшем будет часто встречаться. Бросают монету; ставится вопрос, какая сторона может оказаться наверху. Единственно возможных и несовместных случаев два: надпись или герб. Если считать, что монета представляет идеальный тонкий однородный диск, то можно назвать перечисленные случаи равновозможными, так как нет оснований думать, что одна сторона чем-нибудь отличается от другой и имеет больше шансов оказаться наверху.

Значительно труднее установить равновозможность событий в явлениях природы. С формальной точки зрения, считая некоторые случаи равновозможными, мы вводим в задачу дополнительное предположение, которое не всегда может быть строго оправданным с физической точки зрения.

Рассмотрим другой пример. В ящике лежат три белых шара и два черных. Предполагается, что шары на ощупь неразличимы и тщательно перемешаны. Не глядя в ящик, вынимают первый попавшийся шар. Спрашивается (до опыта), какого цвета может быть вынутый шар. Ответ может быть дан в такой форме: белый шар, черный шар. Эти два случая можно назвать несовместными, так как по условию вынимается только один шар. Согласно нашему определению случаи нельзя назвать равновозможными, так как белых шаров больше. Чтобы перечислить равновозможные случаи, предположим, что шары имеют номера: белые от № 1 до 3, черные от № 4 до 5. Мы можем теперь иначе подсчитать случаи: 1) белый № 1, 2) белый № 2, 3) белый № 3, 4) черный № 4, 5) черный № 5. Эти пять случаев можно считать единственно возможными, несовместными и равновозможными, так как все шары эквивалентны по условию.

§ 30. Классическое определение вероятности

Первой задачей теории вероятностей является введение понятия вероятности, т. е. числа, характеризующего степень уверенности в том или ином результате предстоящего наблюдения.

Предположим, что перечисленные при выяснении возможных результатов наблюдения некоторого явления случаи

$$A_1, A_2, \dots, A_k, A_{k+1}, \dots, A_n$$

являются единственно возможными, несовместными и равновероятными. Тот результат, вероятность которого требуется определить, назовем *событием*.

Пусть C — событие, которое может произойти или не произойти в предстоящих наблюдениях. Пусть, далее, при перечислении возможных случаев A_1, A_2, \dots, A_n мы можем выделить те случаи, при которых событие происходит; предположим, что это случаи A_1, \dots, A_k , где $k \leq n$. Назовем эти случаи *благоприятными* для события C .

Дадим следующее определение:

Вероятностью случайного события называется отношение числа благоприятных для данного события C случаев к числу единственно возможных, равновероятных и несовместных случаев:

$$= P(C) = \frac{k}{n}. \quad (7.1)$$

Для вычисления вероятности нужно сначала перечислить все возможные случаи, затем из них выделить благоприятные. Заметим, что в некоторых задачах возможны разные способы подсчета единственно возможных, равновероятных и несовместных случаев. Например, список случаев при бросании игральной кости может быть дан в двух формах:

- а) нечетное число (1, 3, 5); четное число (2, 4, 6),
- б) числа 1, 2, 3, 4, 5, 6.

Выведем некоторые следствия из определения вероятности.

1. Пусть $k=0$; тогда по определению $p=0$. Так как нет случаев, благоприятных событию, то событие невозможно. Таким образом, *вероятность невозможного события равна нулю*.

2. Пусть $k=n$; на основании определения получаем $p=1$. Так как все случаи благоприятны событию, то оно наверно произойдет; такое событие будем называть *достоверным*. Таким образом, *вероятность достоверного события равна единице*.

3. В общем случае $0 \leq k \leq n$, следовательно, $0 \leq p \leq 1$. Вероятность может принимать значения от 0 до 1; согласно классическому определению p принимает только рациональные значения.

4. Если k есть число благоприятных событию случаев, то $n - k$ — число неблагоприятных случаев. Обозначая через q вероятность того, что событие не произойдет, получим по определению

$$q = \frac{n - k}{n}, \quad (7.2)$$

так как случаи, неблагоприятные событию, благоприятны непоявлению события. Складывая равенства (7.1) и (7.2), получаем

$$p + q = 1. \quad (7.3)$$

§ 31. Примеры вычисления вероятности

Пример 1. Бросают монету; определить вероятность выпадения надписи (сверху).

Список единственно возможных, равновозможных и несовместных случаев будет: «надпись», «герб», т. е. $n = 2$; из них благоприятен случай «надпись», т. е. $k = 1$. Отсюда вероятность $p = 1/2$.

Пример 2. Бросают игральную кость; определить вероятность выпадения пятерки. Единственно возможные, равновозможные и несовместные случаи будут 1, 2, 3, 4, 5, 6 (может выпасть). Благоприятен из них один случай — выпадение пятерки; следовательно, $p = 1/6$.

Пример 3. Бросают игральную кость; вычислить вероятность выпадения нечетного числа.

1-й способ

Единственно возможные, равновозможные и несовместные случаи: 1, 2, 3, 4, 5, 6. Из них благоприятные случаи: 1, 3, 5. Поэтому $p = 3/6 = 1/2$.

2-й способ

Единственно возможные, равновозможные, несовместные случаи: «четное число», «нечетное число». Из них благоприятен один случай (нечетное число). Следовательно,

$$p = \frac{1}{2}.$$

Пример 4. Бросают две монеты; определить вероятность выпадения надписей на обеих монетах.

Запишем в таблице:

1-я монета	(надпись)	(надпись)	(герб)	(герб)
2-я монета	(надпись)	(герб)	(надпись)	(герб)

Из таблицы видно, что $k = 1$, $n = 4$, поэтому $p = 1/4$.

Пример 5. В ящике три белых и пять черных шаров; шары перемешаны и на ощупь неразличимы. Не глядя в ящик, вынимают сразу два шара; вычислить вероятность того, что оба вынутые шара окажутся черными.

Число всех случаев есть число способов, какими можно вынуть два шара из восьми шаров, т. е. $n = C_8^2$. Число благоприятных случаев есть

число способов, которыми можно вынуть два черных шара из пяти черных шаров, т. е. $k = C_5^2$. Отсюда получаем

$$k = 10; \quad n = 28; \quad p = \frac{10}{28} = \frac{5}{14}.$$

Пример 6. В урне 12 белых и 18 черных шаров; вынимают сразу 10 шаров. Вычислить вероятность вынуть четыре белых и шесть черных шаров.

Полное число случаев есть число способов, которыми можно вынуть 10 шаров из тридцати, т. е. $n = C_{30}^{10}$. Число благоприятных случаев определяется так: четыре белых шара из 12 можно вынуть числом способов, равным C_{12}^4 ; каждой четверке белых соответствует ряд шестерок черных шаров; их число есть C_{18}^6 , поэтому число благоприятных случаев равно $C_{12}^4 \cdot C_{18}^6$. Вероятность вынуть четыре белых и шесть черных шаров

$$p = \frac{C_{12}^4 \cdot C_{18}^6}{C_{30}^{10}} = \frac{495 \cdot 18\,564}{30\,045\,015} \approx 0,306.$$

Только что рассмотренная задача является примером из *теории выборки*. В различных областях знания приходится иметь дело со следующей задачей. Имеется полная совокупность объектов («генеральная совокупность»); требуется дать такие характеристики, которые бы в некоторой степени могли ее описать.

Примеры: а) в совокупности звезд, отобранных по признаку спектрального класса, определить долю спектрально-двойных звезд;

б) в совокупности малых планет определить среднее значение наклоности;

в) в партии одинаковых изделий определить процент брака.

Для полного решения задачи надо обследовать всю совокупность, что может быть весьма громоздким делом, а иногда и невозможным (например, определение процента неразрывающихся артиллерийских снарядов). В подобных случаях отбирают некоторую долю генеральной совокупности, и определяют намеченные числовые характеристики. После этого появляется задача о вероятности получить по выборке величину характеристики, близкую к характеристике генеральной совокупности. Рассмотренный пример и содержит одну из задач теории выборки. Генеральной совокупностью является собрание шаров в полной урне, в которой отношение числа белых шаров к числу черных равно 2:3. Ставится вопрос, какова вероятность в выборке, состоящей из 10 шаров, получить то же отношение, т. е. по выборке судить о составе полной урны.

Из примеров 4), 5) и 6) видно, что определение числа всех случаев и числа благоприятных случаев не всегда можно выполнить так просто, как в первых трех примерах. В частности, если бы в четвертом примере было пять монет вместо двух, то всех случаев было бы 32, и их прямое перечисление было бы хлопотливым. Поэтому одной из задач теории вероятностей является вывод правил для вычисления вероятностей одних событий по известным вероятностям других. В следующих двух параграфах будут указаны такие правила. Другой, весьма важной задачей является установление условий, при которых вероятность близка к 1 или к 0.

Эта задача имеет связь с приложениями, так как событие можно считать практически почти достоверным, если его вероятность очень мало отличается от единицы, и практически почти невозможным или редким, если его вероятность близка к нулю.

§ 32. Теорема сложения вероятностей

Теорема. *Вероятность того, что произойдет одно из двух несовместных событий C и D (безразлично, какое именно), равняется сумме вероятностей этих событий.*

Доказательство. Пусть список единственно возможных, равновероятных и несовместных случаев имеет вид

$$A_1, A_2, \dots, A_k, A_{k+1}, \dots, A_{k+l}, A_{k+l+1}, \dots, A_n.$$

Пусть

$$A_1, A_2, \dots, A_k$$

суть все случаи, благоприятные событию C . Вследствие несовместности событий C и D ни один из этих k случаев не может быть благоприятен событию D . Положим, что $A_{k+1}, A_{k+2}, \dots, A_{k+l}$ суть все случаи, благоприятные событию D . Обозначим вероятности событий C и D через $P(C)$ и $P(D)$. Так как всех случаев n , из них k случаев благоприятны событию C и l случаев благоприятны событию D , то

$$P(C) = \frac{k}{n}, \quad P(D) = \frac{l}{n}.$$

Случаи $A_1, A_2, \dots, A_k, A_{k+1}, \dots, A_{k+l}$, вместе взятые, благоприятны либо событию C , либо событию D , поэтому

$$P(C \text{ или } D) = \frac{k+l}{n}.$$

Сравнивая последнее равенство с предыдущими двумя равенствами, получаем *)

$$P(C \text{ или } D) = \frac{k}{n} + \frac{l}{n},$$

или

$$P(C \text{ или } D) = P(C) + P(D). \quad (7.4)$$

Следствие 1. Если события C_1, C_2, \dots, C_s несовместны и их вероятности p_1, p_2, \dots, p_s , то

$$P(C_1 \text{ или } C_2, \dots, \text{или } C_s) = p_1 + p_2 + \dots + p_s. \quad (7.5)$$

*) В некоторых пособиях появление события C или события D обозначается $C + D$ (символическое сложение), и теорема сложения вероятностей записывается так:

$$P(C + D) = P(C) + P(D),$$

если C и D несовместны.

Следствие 2. Появление одного из несовместных событий C_1, C_2, \dots, C_s (безразлично, какого именно) можно назвать событием C и условиться говорить, что события C_1, \dots, C_s представляют несовместные виды события C . Тогда теорему сложения для s событий можно сформулировать следующим образом:

Вероятность события C , имеющего несовместные виды C_1, \dots, C_s , равняется сумме вероятностей этих видов.

Следствие 3. Пусть при некотором наблюдении обязательно произойдет какое-нибудь из несовместных событий C_1, C_2, \dots, C_n , вероятности которых равны p_1, p_2, \dots, p_n . Тогда

$$p_1 + p_2 + \dots + p_n = 1. \quad (7.6)$$

Действительно, по теореме сложения вероятность какого-нибудь из несовместных событий C_1, C_2, \dots, C_n равна сумме вероятностей этих событий, а по условию какое-нибудь из событий наверное произойдет.

Формула (7.3) представляет частный случай формулы (7.6). Действительно, мы можем утверждать, что событие либо произойдет (вероятность равна p), либо не произойдет (вероятность равна q). Эти результаты несовместны. Поэтому, применяя формулу (7.6), можем написать $p + q = 1$.

Пример 1. В ящике пять белых, семь зеленых и восемь красных шаров; определить вероятность вынуть зеленый или красный шар при извлечении шара.

Вероятность вынуть зеленый шар

$$P(\text{зел.}) = \frac{7}{20};$$

вероятность вынуть красный шар

$$P(\text{кр.}) = \frac{8}{20}.$$

По теореме сложения

$$P(\text{зел. или кр.}) = \frac{7}{20} + \frac{8}{20} = \frac{15}{20} = \frac{3}{4}.$$

Результат может быть проверен простым подсчетом. Так как определяется вероятность события «вынуть зеленый или красный шар», то число благоприятных случаев $k = 7 + 8 = 15$; $n = 20$. Поэтому

$$P(\text{зел. или кр.}) = \frac{15}{20} = \frac{3}{4}.$$

В этом примере можно было бы и не указывать числа шаров; достаточно было задать вероятности.

Пример 2. Бросают игральную кость; определить вероятность выпадения пятерки или четного числа.

По результатам примера 2) из предыдущего параграфа имеем

$$P(5) = \frac{1}{6};$$

легко видеть, что

$$P(\text{четн. число}) = \frac{1}{2}.$$

Так как выпадение 5 и выпадение четного числа представляют несовместные события, то

$$P(5 \text{ или четн. число}) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}.$$

Пример 3. Бросают игральную кость; определить вероятность выпадения пятерки или нечетного числа.

Теорему сложения нельзя применить, так как указанные события совместны. Однако, так как первое событие есть вид второго, то

$$P(5 \text{ или нечетн. число}) = P(\text{нечетн. число}) = 1/2.$$

§ 33. Теорема умножения вероятностей

Определение 1. Событие называется *сложным*, если оно заключается в том, что происходит два или несколько событий; эти события называются *составляющими* сложного события. Употребляется также выражение: «сложное событие есть совмещение событий». (Слово «совмещение» следует понимать достаточно широко; можно говорить о событиях, происходящих в разное время и в разных местах.)

Определение 2. Вероятность события C , вычисленная в предположении, что событие D произошло, называется *условной вероятностью события C при условии D* и обозначается $P(C/D)$. Вероятности вида $P(C)$ называют иногда *безусловными*.

Теорема. Вероятность сложного события, представляющего совмещение двух событий, равняется произведению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие произошло.

Доказательство. Пусть полная группа равновозможных и несовместных случаев имеет вид

$$A_1, A_2, \dots, A_k, A_{k+1}, \dots, A_{k+l+1}, \dots, A_n.$$

Положим, что только первые k случаев благоприятны обоим составляющим событиям C и D рассматриваемого сложного события. Пусть случаи A_1, A_2, \dots, A_{k+l} представляют все случаи благоприятные событию C . Случаи, благоприятные событию D , но неблагоприятные вместе с тем событию C , могут быть среди случаев A_{k+l+1}, \dots, A_n , но они нас не интересуют.

Запишем изложенное в следующей схеме:

$$\begin{array}{c} \text{Благоприятные } C \\ \hline \underbrace{A_1, A_2, \dots, A_k}_{\text{Благоприятные } C \text{ и } D} \quad \underbrace{A_{k+1}, \dots, A_{k+l}}_{\text{Неблагоприятные } D} \quad \underbrace{A_{k+l+1}, \dots, A_n}_{\text{Неблагоприятные } C} \end{array}$$

По определению вероятности имеем:

$$P(C) = \frac{k+l}{n}, \quad P(C \text{ и } D) = \frac{k}{n}.$$

Для вычисления вероятности D в предположении, что событие C произошло, будем рассуждать так.

Если событие C произошло, то это значит, что произошел только один из случаев A_1, A_2, \dots, A_{k+l} ; поэтому здесь число всех случаев равно $k+l$. Из них событию D благоприятны случаи A_1, A_2, \dots, A_k , следовательно,

$$P(D/C) = \frac{k}{k+l}. \quad (7.7)$$

Теперь легко проверить теорему *):

$$P(C)P(D/C) = \frac{k+l}{n} \frac{k}{k+l} = \frac{k}{n} = P(C \text{ и } D), \quad (7.8)$$

$$P(C \text{ и } D) = P(C)P(D/C). \quad (7.9)$$

Следствие 1. Меняя взаимно обозначения событий, получим

$$P(D \text{ и } C) = P(D)P(C/D); \quad (7.10)$$

так как $P(D \text{ и } C) = P(C \text{ и } D)$, то

$$P(C)P(D/C) = P(D)P(C/D). \quad (7.11)$$

Эта формула устанавливает связь между двумя безусловными ($P(C)$, $P(D)$) и двумя условными вероятностями двух событий. Мы видим, что только три из них могут быть заданы произвольно; четвертая определяется по формуле. Нет, однако, полного произвола и в задании трех вероятностей, так как вычисляемая четвертая вероятность не должна быть больше единицы.

Следствие 2. Теорема умножения для нескольких событий может быть записана так:

$$\begin{aligned} P(C_1 \text{ и } C_2 \text{ и } \dots \text{ и } C_s) = \\ = P(C_1)P(C_2/C_1)P(C_3/C_1, C_2) \dots P(C_s/C_1, C_2, \dots, C_{s-1}), \end{aligned} \quad (7.12)$$

где $P(C_3/C_1, C_2)$ обозначает «вероятность события C_3 , вычисленная в предположении, что произошли события C_1 и C_2 »; аналогичен смысл других подобных обозначений.

Следствие 3. Теорема деления вероятностей: из формулы

$$P(C \text{ и } D) = P(C)P(D/C) = P(D)P(C/D)$$

следуют равенства

$$P(D/C) = \frac{P(C \text{ и } D)}{P(C)}, \quad P(C/D) = \frac{P(C \text{ и } D)}{P(D)}. \quad (7.13)$$

*) Сложное событие (совмещение событий) обозначают нередко символически знаком умножения, т. е. вместо C и D пишут $C \times D$.

т. е. условная вероятность одного события, вычисленная в предположении о появлении другого события, равна частному от деления вероятности совмещения событий на вероятность второго события.

Определение 3. События называются *взаимно независимыми*, если вероятность каждого из них не зависит от того, произошли или не произошли другие события.

Если это условие не выполнено, то события называют *связанными* или *зависимыми*. Формально условие взаимной независимости двух событий можно записать так:

$$P(C/D) = P(C), \quad P(D/C) = P(D),$$

причем из определения условной вероятности следует, что второе равенство есть следствие первого.

Доказанная выше теорема умножения применима к связанным событиям. В случае независимых событий теорема умножения может быть сформулирована проще:

Вероятность совмещения нескольких независимых событий равна произведению вероятностей составляющих событий:

$$P(C_1 \text{ и } C_2 \text{ и } \dots \text{ и } C_s) = P(C_1) P(C_2) \dots P(C_s). \quad (7.14)$$

Справедливость теоремы ясна из того, что в силу определения независимости вероятность события C_2 не зависит от появления или не появления события C_1 ; поэтому

$$P(C_1 \text{ и } C_2) = P(C_1) P(C_2/C_1) = P(C_1) P(C_2). \quad (7.15)$$

Аналогично проводится доказательство при любом числе составляющих событий.

Пример 1. Бросают две монеты; определить вероятность выпадения двух надписей.

Выпадение надписей на монетах — события независимые, так как вероятность выпадения надписи на второй монете, очевидно, не зависит от того, какой стороной упала первая монета. Вероятность надписи (н.) для каждой из монет равна $\frac{1}{2}$. По теореме умножения вероятность выпадения двух надписей, т. е. события, представляющего совмещение двух событий, можно вычислить так:

$$P(\text{н. и н.}) = P(\text{н.}) \cdot P(\text{н.}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Пример 2. Из ящика, содержащего пять белых и четыре черных шара, вынимают один шар, отмечают его цвет, вынутый шар кладут обратно, шары перемешивают и снова вынимают шар. Вычислить вероятность того, что оба вынутые шара — черные.

Два события — извлечение черного шара в первый раз и извлечение черного шара во второй раз — в силу условий опыта можно считать независимыми. Поэтому

$$P(\text{ч. ш. и ч. ш.}) = P(\text{ч. ш.}) \cdot P(\text{ч. ш.}) = \frac{4}{9} \cdot \frac{4}{9} = \frac{16}{81}.$$

Пример 3. Из ящика, содержащего пять белых шаров и четыре черных шара, вынимают один за другим два шара, но, в отличие от предыдущего примера, не кладут обратно шара, вынутого в первый раз. Определить вероятность вынуть два черных шара.

Здесь вероятность вынуть черный шар во второй раз зависит от того, какой шар будет фактически вынут первый раз, черный или белый. В первом случае эта вероятность равна $\frac{8}{9}$, во втором случае — $\frac{1}{2}$. Наши два события связаны, поэтому

$$P(\text{ч. 1-й и ч. 2-й}) = P(\text{ч. 1-й}) \cdot P(\text{ч. 2-й/ч. 1-й})$$

или

$$P(\text{ч. 1-й и ч. 2-й}) = \frac{4}{9} \cdot \frac{3}{8} = \frac{1}{6}.$$

Здесь $P(\text{ч. 2-й/ч. 1-й})$ означает вероятность вынуть черный шар второй раз, вычисленную в предположении, что первый раз вынут черный шар.

§ 34. Полная вероятность; гипотезы

Пусть событие C может происходить при осуществлении одного или нескольких из несовместных случайных условий H_1, H_2, \dots, H_n , которые назовем *гипотезами*. Так как гипотезы случайны, то необходимо задать их вероятности P_1, P_2, \dots, P_n . Список гипотез должен быть полным, поэтому по теореме сложения должно выполняться равенство

$$\sum_{k=1}^n P_k = 1.$$

Предположим еще, что заданы условные вероятности события при каждой из гипотез; введем для них обозначения: $p_k = P(C/H_k)$, $k = 1, 2, \dots, n$.

Полной вероятностью события C при указанных условиях называется вероятность, вычисленная в предположении, что событие появится по какой-нибудь из гипотез (безразлично какой).

Для вычисления полной вероятности рассуждаем так: событие — появление C — имеет виды $(H_k \times C)$, т. е. совмещение осуществления гипотезы H_k и появления события C ; $k = 1, 2, \dots, n$. Вероятность каждого вида вычисляется по теореме умножения вероятностей для связанных событий, так как вероятность события C зависит от того, какая из гипотез осуществится. Имеем

$$P(H_k \times C) = P(H_k)P(C/H_k) = P_k p_k.$$

Виды $(H_k \times C)$ несовместны, поэтому по теореме сложения вероятностей

$$P(C) = \sum_{k=1}^n P(H_k \times C).$$

Полная вероятность вычисляется по формуле

$$P(C) = \sum_{k=1}^n P_k p_k. \quad (7.16)$$

Для пояснения вывода общей формулы повторим на частном примере все рассуждения.

Пример. Имеем три красных ящика, в каждом из которых по четыре белых и пять черных шаров, и семь зеленых ящиков, в каждом из которых по два белых и три черных шара. Все ящики и шары одинаковы и тщательно перемешаны. Вынимаем шар из первого попавшегося ящика; опыт по условию поставлен так, что мы не можем заранее сказать, в какой из ящиков попадает рука. Определим вероятность вынуть белый шар.

Для решения надо принять в первую очередь во внимание, что рассматриваемое событие — извлечение белого шара — имеет два несовместимых вида, так как белый шар можно вынуть из какого-нибудь зеленого ящика или из какого-нибудь красного ящика. Поэтому по теореме сложения вероятностей

$$P(\text{бел.}) = P(\text{бел. из зел. ящ.}) + P(\text{бел. из красн. ящ.}).$$

Извлечение белого шара из какого-нибудь зеленого ящика есть сложное событие, состоящее из двух событий, так как для его осуществления необходимо попасть в какой-нибудь зеленый ящик и вместе с тем вынуть белый шар. То же можно сказать и об извлечении белого шара из красного ящика. По теореме умножения (для связанных событий) имеем

$$P(\text{бел. из зел. ящ.}) = P(\text{зел. ящ.}) P(\text{бел./зел. ящ.}),$$

$$P(\text{бел. из красн. ящ.}) = P(\text{красн. ящ.}) P(\text{бел./красн. ящ.});$$

здесь $P(\text{зел.})$ есть вероятность попасть в зеленый ящик, равная $7/10$, $P(\text{красн.})$ — вероятность попасть в какой-нибудь из красных ящиков, равная $3/10$, $P(\text{бел./зел. ящ.})$ — вероятность вынуть белый шар, если известно, что попали в зеленый ящик, а $P(\text{бел./красн. ящ.})$ — вероятность вынуть белый шар, если известно, что попали в красный ящик. Так как

$$P(\text{бел./зел. ящ.}) = \frac{2}{5}, \quad P(\text{бел./красн. ящ.}) = \frac{4}{9},$$

то искомая вероятность вынуть белый шар будет равна:

$$P(\text{бел.}) = \frac{7}{10} \cdot \frac{2}{5} + \frac{3}{10} \cdot \frac{4}{9} = \frac{31}{75}.$$

Решение можно записать в следующей таблице:

	Число ящиков	Число черных шаров	Число белых шаров	Всего шаров в ящике	Условная вероят- ность вынуть бел. шар	Вероят- ность попасть в ящик
Зеленый	7	3	2	5	$\frac{2}{5}$	$\frac{7}{10}$
Красный	3	5	4	9	$\frac{4}{9}$	$\frac{3}{10}$
Сумма	10					1

Для решения нашей задачи не было надобности указывать состав ящиков; достаточно было знать вероятность вынуть белый шар, если известно, что попали в красный ящик, и такую же вероятность для зеленых ящиков; вместо количества ящиков можно было бы указать доли ящиков различных цветов. Отметим, что в этой задаче было бы неправильно

подсчитывать общее число всех белых и черных шаров во всех ящиках и делить число белых шаров на число всех шаров, т. е. пользоваться непосредственно определением вероятности. Это можно делать только в том случае, если извлечение любого из шаров одинаково возможно. В рассматриваемой же задаче мы имеем некоторое преобладание белых шаров в тех местах, где стоят красные ящики, по сравнению с местами, где находятся зеленые ящики. Поэтому нельзя считать, что шары тщательно перемешаны, и мы имеем одинаковые шансы захватить любой из шаров.

§ 35. Априорные и апостериорные вероятности гипотез

Пусть появление или не появление события связано с некоторыми предположениями о причинах этого события или об условиях, при которых оно произойдет. Каждому из этих предположений соответствует определенная вероятность события. Кроме того, предположения (гипотезы) считаются также случайными, следовательно, для определенности задачи должны быть заданы вероятности гипотез. Если, например, извлекается шар из урны, в которой находится пять шаров и о которой известно только, что в ней находятся белые и черные шары, то мы можем сделать ряд предположений (гипотез) о составе урны; каждой из этих гипотез будет соответствовать определенная вероятность появления белого шара. В нашем примере имеем следующие гипотезы и соответствующие им вероятности появления белого шара:

			P (бел.)
1)	4 бел.	и 1 черн.	0,8
2)	3 »	2 »	0,6
3)	2 »	3 »	0,4
4)	1 »	4 »	0,2

Если мы произведем опыт — извлечение шара — и шар окажется белым, то представляет интерес выяснение вопроса, какая из перечисленных гипотез более вероятна. Разрешение такого вопроса может иметь значение для суждения о вероятности случайных причин, вызвавших при опыте появление события, если таких причин несколько. Возможность решения такой задачи (при надлежащей постановке) ясна из следующих простых соображений.

Если бы в нашем примере не было указано, что в урне есть белые и черные шары, то мы должны были бы прибавить еще две гипотезы:

- 5) 5 бел. и 0 черн.,
- 6) 0 » 5 ».

Появление при опыте белого шара приводит нас по необходимости к тому, что вероятность шестой гипотезы равна нулю.

Вероятности гипотез в условиях задачи до проведения опытов называются *априорными*, а вероятности, вычисленные после опыта, в котором событие произошло, — *апостериорными*.

Рассмотрим теперь следующую задачу. Предположим, что появление события C может быть объяснено n несовместными и единственно возможными гипотезами:

$$H_1, H_2, \dots, H_n.$$

Предположим, что известна вероятность каждой гипотезы до опыта:

P_1, P_2, \dots, P_n , т. е. априорные вероятности (как было указано,

$\sum_{k=1}^n P_k = 1$). Наконец, предположим, что для каждой из гипотез известна вероятность события C при условии, что его появление объясняется этой гипотезой; обозначим эти вероятности p_1, p_2, \dots, p_n . Из наших предположений вытекает, что p_k есть условная вероятность события C , найденная в предположении, что на появление события повлияла именно та причина (или те условия), которая соответствует гипотезе H_k ($k = 1, 2, \dots, n$). Требуется определить вероятность каждой из гипотез, если известно, что событие произошло.

Приступим к решению. Условная вероятность гипотезы H_k ($k = 1, 2, \dots, n$) в предположении, что событие C произошло, может быть вычислена по теореме деления (7.13):

$$P(H_k/C) = \frac{P(C \times H_k)}{P(C)},$$

где $P(C \times H_k)$ — вероятность совмещения двух событий — C и H_k , т. е. появления C вследствие действия причин, соответствующих гипотезе H_k . Чтобы вычислить нужное нам число $P(H_k/C)$, необходимо вычислить $P(C \times H_k)$ и $P(C)$.

Первое из них можно определить по теореме умножения для связанных событий, если считать «первым» событием появление гипотезы H_k :

$$P(C \times H_k) = P(H_k)P(C/H_k) = P_k p_k.$$

Вероятность события C можно определить, как полную вероятность появления C , по формуле (7.16):

$$P(C) = \sum_{s=1}^n P_s p_s;$$

поэтому

$$P(H_k/C) = \frac{P_k p_k}{\sum_{s=1}^n P_s p_s}; \quad k = 1, 2, \dots, n. \quad (7.17)$$

По этой формуле и можно вычислять апостериорные вероятности гипотез, т. е. вероятности в предположении, что событие C произошло.

Совокупность формул (7.17) обычно называется *теоремой Бейеса*.

Следствие. Если априорные (до опыта) вероятности гипотез неизвестны, то гипотезы приходится считать равновероятными (за отсутствием сведений). Следовательно,

$$P_1 = P_2 = \dots = P_n = \frac{1}{n};$$

тогда по формуле Бейеса

$$P(H_k/C) = \frac{P_k}{\sum_{s=1}^n P_s}. \quad (7.18)$$

Таким образом, если до опыта гипотезы равновероятны, то после опыта, если событие появилось, вероятности (апостериорные) гипотез пропорциональны условным вероятностям события (вычисленным в предположении, что на появление события повлияли соответствующие гипотезы).

Пример 1. Имеется урна, в которой находится четыре шара. О составе урны ничего неизвестно, кроме того, что в ней находятся белые и черные шары. Вынут один шар, который оказался белым. Требуется определить вероятный состав урны на основании результата опыта.

Относительно состава урны можно по условию задачи сделать следующие предположения:

- I. В урне 3 белых и 1 черный шар,
- II. » 2 » 2 » »
- III. » 1 » 3 » »

Так как других условий нет, то эти три предположения мы будем считать равновероятными. Обозначая вероятности этих гипотез P_1, P_2, P_3 , имеем

$$P_1 = P_2 = P_3 = \frac{1}{3}.$$

Вероятность вынуть белый шар по I гипотезе равна $\frac{3}{4}$, по II — $\frac{1}{2}$ и по III — $\frac{1}{4}$. По формулам Бейеса получаем вероятности гипотез после опыта:

$$P(\text{I/бел. ш.}) = \frac{\frac{1}{3} \cdot \frac{3}{4}}{\frac{1}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{4}} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2},$$

$$P(\text{II/бел. ш.}) = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3},$$

$$P(\text{III/бел. ш.}) = \frac{\frac{1}{3} \cdot \frac{1}{4}}{\frac{1}{2}} = \frac{1}{6}.$$

$$\text{Контроль. } \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1.$$

Мы видим, что результат опыта заставляет нас изменить суждения о вероятностях гипотез. Гипотезы, считавшиеся до опыта равновероятными, после опыта имеют разные вероятности, и наиболее вероятной оказывается первая.

Пример 2. Рассмотрим другой вариант примера 1. Пусть до опыта известно только, что в урне могут быть белые шары и шары других цветов. Тогда придется построить следующие гипотезы:

- I. 4 белых шара и 0 шаров другого цвета, $p_1 = 1$,
 II. 3 » » » 1 » » » , $p_2 = \frac{3}{4}$,
 III. 2 » » » 2 » » » , $p_3 = \frac{1}{2}$,
 IV. 1 » » » 3 » » » , $p_4 = \frac{1}{4}$,
 V. 0 » » » 4 » » » , $p_5 = 0$.

Априорную вероятность каждой из гипотез придется считать равной $1/5$.

Если при извлечении шара появился белый шар, то апостериорные вероятности каждой из гипотез, вычисленные по следствию из формул Бейеса, будут:

$$P(I/\text{бел. ш.}) = \frac{1}{1 + \frac{3}{4} + \frac{1}{2} + \frac{1}{4} + 0} = \frac{1}{2\frac{1}{2}} = 0,4,$$

$$P(II/\text{бел. ш.}) = \frac{\frac{3}{4}}{2\frac{1}{2}} = 0,3; \quad P(III/\text{бел. ш.}) = \frac{\frac{1}{2}}{2\frac{1}{2}} = 0,2,$$

$$P(IV/\text{бел. ш.}) = \frac{\frac{1}{4}}{2\frac{1}{2}} = 0,1; \quad P(V/\text{бел. ш.}) = \frac{0}{2\frac{1}{2}} = 0.$$

Как и следовало ожидать, пятая гипотеза на основании результата опыта оказывается невозможной.

ГЛАВА 8

ЗАДАЧА О ПОВТОРЕНИИ ИСПЫТАНИЙ

§ 36. Формулировка задачи и вывод основной формулы

Определение 1. Если в случайном явлении может появиться или не появиться случайное событие C , то *испытанием* будем называть осуществление условий, при которых станет известным, появилось или не появилось данное событие.

Всякое наблюдение, фактически произведенное, представляет собой испытание.

Определение 2. Повторяющиеся испытания, в каждом из которых может произойти или не произойти случайное событие C , называются *независимыми по отношению к событию C* , если вероятность события во всяком испытании не зависит от результатов других испытаний, т. е. от того, сколько раз фактически появилось событие при других испытаниях.

Пример 1. Монету бросают 100 раз. Вероятность выпадения надписи в испытании № 91 равна $\frac{1}{2}$, независимо от того, выпала ли надпись при 90 предыдущих бросаниях 0 раз, 10 раз или даже 90 раз.

Может показаться, что этот пример противоречит обыденным представлениям. Если 90 раз подряд выпадала надпись, то играющий, собираясь бросать 91-й раз, скорее склонен ожидать выпадения герба, чем надписи, т. е. как бы приписывать большую вероятность выпадению герба. Отсутствие оснований для этого легко обнаружить, воспользовавшись замечанием одного автора: «монета не имеет памяти». Поэтому, когда монету собираются бросать 91-й раз, то вероятность выпадения надписи зависит только от свойств монеты, как и при любом другом бросании. Как будет видно из дальнейшего, дело здесь несколько сложнее. Выпадение надписи 90 раз подряд есть событие, вероятность которого очень мала, весьма мала также и вероятность того, что надпись выпадает 91 раз подряд, но это относится к вероятности совокупности результатов, а не к вероятности выпадения при отдельном опыте, которая остается равной половине, независимо от результата предыдущих опытов.

Заметим, что весьма малая вероятность выпадения надписи 90 раз подряд, получающаяся в предположении, что вероятность выпадения надписи равна $\frac{1}{2}$ в каждом отдельном опыте, может заставить нас сомневаться в правильности этого предположения, если действительно надпись выпадет 90 раз подряд. Возможно, например, что монета по той или иной причине настолько несимметрична, что вероятность выпадения надписи при отдельном опыте близка к единице, однако нельзя считать это утверждение

достоверным, так как маловероятное событие может все-таки произойти. Только в том случае, если последовательные серии опытов (например, по 100 бросаний монеты в каждой серии) неизменно будут приводить к аналогичным результатам (значительное преобладание выпадения надписи), будет основание для пересмотра предположения о равновероятности выпадения надписи и герба в одном опыте.

Задача о повторении испытаний формулируется следующим образом. Производится n независимых по отношению к событию C испытаний, в каждом из которых вероятность события равна p (и постоянна). Вычислить вероятность того, что событие повторится k раз ($0 \leq k \leq n$); чередование появлений и не-появлений безразлично.

Ниже мы выведем основную формулу для решения этой задачи. Обозначим буквой C появление события в каком-нибудь испытании и буквой \bar{C} непоявление события; вероятность \bar{C} обозначим буквой q . Тогда

$$p + q = 1.$$

Представим себе одну из возможных последовательностей появлений и непоявлений события в n испытаниях, в которых событие C происходит k раз:

$$C, C, \bar{C}, \bar{C}, \bar{C}, C, \bar{C}, \dots, \bar{C}, \bar{C};$$

в этом списке знак C должен встречаться k раз, а число знаков \bar{C} равно $n - k$. Если обозначить указанную последовательность буквой A_1 , то по теореме умножения вероятностей для независимых событий будет:

$$P(A_1) = p \cdot p \cdot q \cdot q \cdot q \cdot p \cdot q \dots q \cdot q = p^k q^{n-k}.$$

Очевидно, что вероятность всякой другой последовательности результатов испытаний, в которой событие происходит k раз, также равна $p^k q^{n-k}$, так как соответствующее произведение вероятностей будет отличаться только порядком сомножителей. Вероятность повторения события k раз в каком угодно порядке, которую мы обозначим $P_{k,n}$, вычисляется по теореме сложения вероятностей:

$$P_{k,n} = \sum_{s=1}^m P(A_s) = m p^k q^{n-k},$$

где m — число возможных последовательностей. Это число есть число способов, которыми в нашем списке букву C можно поместить k раз на n местах; поэтому m равна числу сочетаний из n элементов по k : C_n^k .

Таким образом получаем основную формулу:

$$P_{k,n} = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}. \quad (8.1)$$

При не очень малых n и k для вычисления $P_{n,k}$ можем пользоваться таблицами биномиальных коэффициентов и факториалов.

Пример 2. Игральную кость бросают пять раз; определить вероятность того, что единица выпадает три раза.

В нашем случае $p = \frac{1}{6}$, $q = 1 - \frac{1}{6} = \frac{5}{6}$, $n = 5$, $k = 3$

$$P_{5,3} = C_5^3 \cdot \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 = \frac{5 \cdot 4}{1 \cdot 2} \cdot \frac{1}{216} \cdot \frac{25}{36} = \frac{125}{3888}.$$

§ 37. Распределение вероятностей чисел повторений события

Если испытания производятся n раз, то до испытаний можно только сказать, что событие повторится либо 0 раз, либо 1 раз, либо 2 раза, ..., либо n раз. По формуле (8.1), полагая $k = 0, 1, 2, \dots, n$, получаем

$$\left. \begin{aligned} P_{0,n} &= q^n; \quad P_{1,n} = npq^{n-1}; \\ P_{2,n} &= \frac{n(n-1)}{1 \cdot 2} p^2 q^{n-2}, \dots, P_{n-1,n} = np^{n-1}q, \quad P_{n,n} = p^n. \end{aligned} \right\} \quad (8.2)$$

Легко видеть, что эти вероятности представляют последовательные члены разложения бинома Ньютона:

$$\begin{aligned} (q+p)^n &= \\ &= q^n + npq^{n-1} + \dots + p^n. \end{aligned} \quad (8.3)$$

Вероятность повторения события k раз равняется тому члену разложения, который содержит p^k . Поэтому таблицу значений $P_{k,n}$ при всех k называют *распределением вероятностей*.

Так как $p+q=1$, то из последнего равенства следует $P_{0,n} + P_{1,n} + \dots + P_{n,n} = 1$;

(8.4)

то же равенство следует из формулы (7.6), так как случаи повторения события 0 раз, 1 раз, 2 раза, ..., n раз исчерпывают все возможные случаи, и один из них наверно произойдет.

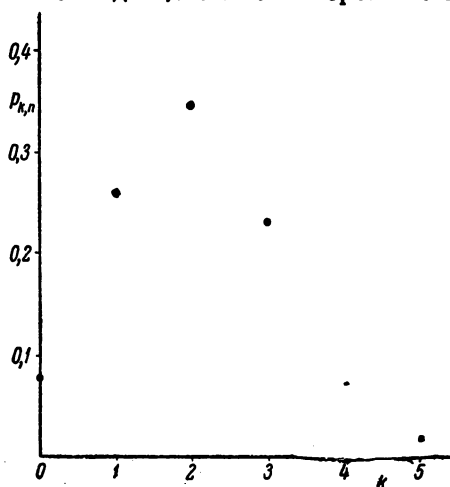


Рис. 1. Распределение вероятностей в задаче о повторении испытаний (число испытаний $n=5$, вероятность события $p=0,4$).

Пример 1. Пусть $p = 0,4$, $q = 0,6$, $n = 5$. Следовательно, k может принимать значения 0, 1, 2, 3, 4, 5. Вычисляя $P_{k,n}$, представим распределение вероятностей в виде таблицы

k	0	1	2	3	4	5
$P_{k,n}$	0,07776	0,25920	0,34560	0,23040	0,07680	0,01024

Для контроля сложим все числа $P_{k,n}$; получается единица, как и должно быть.

Распределение вероятностей можно представить на графике; на оси абсцисс отложим числа повторений от 0 до n , на оси ординат —

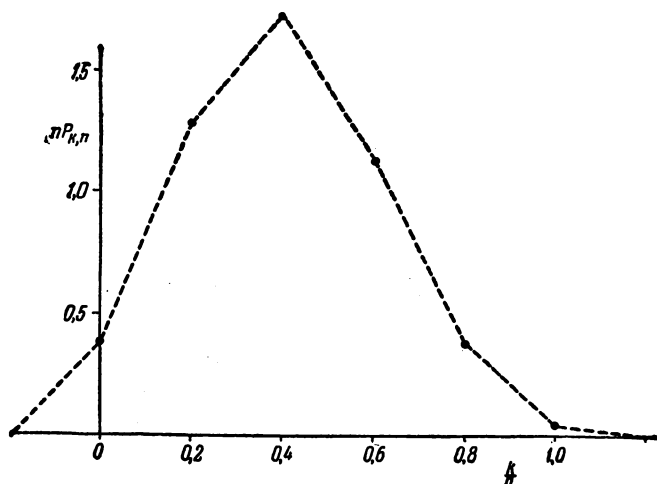


Рис. 2. Распределение вероятностей в задаче о повторении испытаний (число испытаний $n = 5$, вероятность $p = 0,4$).

вероятности. Следует обратить внимание на то, что график состоит из отдельных точек, так как число повторений не изменяется непрерывно, а принимает только последовательные целые значения от 0 до n . В подобных задачах говорят, что k принимает дискретные значения (в нашей задаче — целые, положительные). На рис. 1 представлен график для примера, рассмотренного выше. Тот же график можно строить и несколько иначе, откладывая на оси абсцисс вместо числа повторений его отношение к числу испытаний k/n , а на оси ординат — произведения вероятностей на число испытаний (рис. 2).

Точки на графике соответствуют различным числам повторений от 0 до n . Точки соединены пунктиром, а не сплошными линиями, чтобы подчеркнуть, что график состоит из отдельных точек.

Этот способ построения графика удобен тем, что основание графика всегда равно единице длины (по оси абсцисс), ординаты не будут слишком малыми. Построив вторым способом график (рис. 2),

отложим по обе стороны основания (отрезок длиной единица от начала координат) по одному отрезку длиной $1/n$ и построим замкнутую ломаную, изображенную на рис. 2. Легко показать, что площадь, ограниченная этой ломаной и осью абсцисс, равна единице. Действительно, мы имеем сумму площадей двух треугольников и n трапеций, поэтому:

$$S = \frac{1}{n} \left[\frac{1}{2} n P_{0,n} + \frac{1}{2} n (P_{0,n} + P_{1,n}) + \frac{1}{2} n (P_{1,n} + P_{2,n}) + \dots \right. \\ \left. \dots + \frac{1}{2} n (P_{n-1,n} + P_{n,n}) + \frac{1}{2} n P_{n,n} \right] = \\ = P_{0,n} + P_{1,n} + \dots + P_{n,n} = 1.$$

Пример 1. Составить распределение вероятностей при пяти испытаниях, если $p = \frac{1}{2}$.

k	0	1	2	3	4	5
$P_{k,5}$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Пример 2. Составить распределение вероятностей при четырех испытаниях, если $p = \frac{2}{3}$. Возвышая $1/3 + 2/3$ в четвертую степень, получим

k	0	1	2	3	4
$P_{k,4}$	$\frac{1}{81}$	$\frac{8}{81}$	$\frac{24}{81}$	$\frac{32}{81}$	$\frac{16}{81}$

Выясним некоторые общие свойства полученного нами *биномиального распределения* вероятностей. Так как это распределение дискретное, то обычный аппарат анализа не пригоден для исследования изменения чисел $P_{k,n}$ в зависимости от изменения k . Чтобы провести такое исследование, построим отношения смежных значений вероятностей:

$$\frac{P_{1,n}}{P_{0,n}}, \frac{P_{2,n}}{P_{1,n}}, \frac{P_{3,n}}{P_{2,n}}, \dots, \frac{P_{k,n}}{P_{k-1,n}}, \frac{P_{k+1,n}}{P_{k,n}}, \dots, \frac{P_{n-1,n}}{P_{n-2,n}}, \frac{P_{n,n}}{P_{n-1,n}}.$$

Вычисляя при помощи основной формулы общий член этой последовательности, получим

$$\frac{P_{k+1,n}}{P_{k,n}} = \frac{n-k}{k+1} \frac{p}{q}.$$

Из этого выражения следует, что члены в последовательности отношений монотонно убывают при увеличении k . Поэтому можно отметить следующие основные случаи:

1. Первое из отношений меньше единицы или равно единице

Тогда все остальные отношения меньше единицы:

$$\frac{P_{1,n}}{P_{0,n}} \leq 1, \frac{P_{2,n}}{P_{1,n}} < 1, \dots, \frac{P_{k+1,n}}{P_{k,n}} < 1, \dots, \frac{P_{n,n}}{P_{n-1,n}} < 1;$$

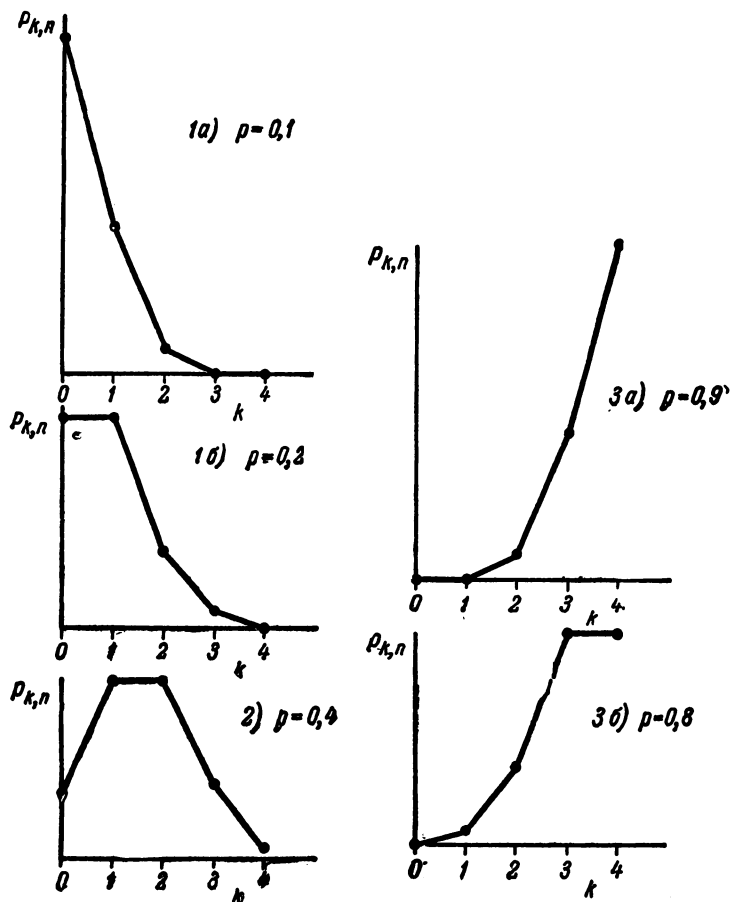


Рис. 3. Виды распределения вероятностей в задаче о повторении испытаний при $n=4$ и разных p .

k	0	1	2	3	4
1a	0,66	0,29	0,05	0,0036	0,0001
1б	0,41	0,41	0,15	0,03	0,0016
2	0,13	0,35	0,35	0,15	0,02
3a	0,0016	0,03	0,15	0,41	0,41
3б	0,0001	0,0036	0,05	0,29	0,66

отсюда

$$P_{0,n} \geq P_{1,n} > P_{2,n} > \dots > P_{n-1,n} > P_{n,n}.$$

1а) Если $P_{1,n} : P_{0,n} < 1$, то вероятности биномиального распределения монотонно убывают и наибольшая вероятность будет при $k=0$, т. е. наиболее вероятное число повторений есть 0. График распределения для этого случая дан на рис. 3 (1а).

1б) Если первое отношение равно единице, то в биномиальном распределении первые две вероятности равны между собой, последующие монотонно убывают. Здесь мы имеем два наиболее вероятных числа повторений — 0 и 1 с одинаковыми вероятностями. График распределения дан на рис. 3 (1б).

2. Первое из отношений больше единицы, последнее меньше единицы

Тогда, вследствие монотонного убывания отношений, где-то внутри последовательности должен быть один переход от отношений, превышающих единицу, к отношениям, меньшим единицы. Так как вероятности суть дискретные числа, то и отношения их тоже дискретные. Поэтому могут быть два подслучая:

2а) Некоторое отношение больше единицы, следующее за ним меньше единицы.

2б) Некоторое из отношений точно равно единице (предыдущее тогда больше единицы, последующее меньше единицы).

Рассмотрим их отдельно.

2а) Имеем

$$\dots \frac{P_{x-1,n}}{P_{x-2,n}} > \frac{P_{x,n}}{P_{x-1,n}} > 1; \frac{P_{x+1,n}}{P_{x,n}} < 1, \frac{P_{x+2,n}}{P_{x+1,n}} < 1, \dots;$$

отсюда

$$\dots P_{x-2,n} < P_{x-1,n} < P_{x,n} > P_{x+1,n} > P_{x+2,n}.$$

Из неравенств видно, что распределение вероятностей имеет максимум при числе повторений, равном x . Чтобы определить наиболее вероятное число повторений x , подставим в два неравенства

$$\frac{P_{x,n}}{P_{x-1,n}} > 1, \quad \frac{P_{x+1,n}}{P_{x,n}} < 1$$

вместо вероятностей из выражения их основной формулы. Тогда получим

$$\frac{n-x+1}{x} \frac{p}{q} > 1, \quad \frac{n-x}{x+1} \frac{p}{q} < 1$$

или

$$xp + xq < np + p, \quad xp + xq > np - q.$$

Так как $p + q = 1$, то

$$np - q < x < np + p.$$

Разность границ для x равна

$$np + p - (np - q) = p + q = 1;$$

поэтому, если $np + p$ (и $np - q$) — дробное число, то между ними заключается только одно целое число, равное целой части числа $np + p$, которое и будет наиболее вероятным числом повторений. График такого распределения приведен на рис. 3 (2).

Условием подслучая 2а), т. е. единственного максимума распределения, является дробность чисел $np + p$ и $np - q$. Если $np + p$ — целое число, то $np - q$ тоже целое, на единицу меньшее число, и задача как будто не имеет решения, но, как мы сейчас увидим, это не так.

Заметим, что подслучай 1а), именно монотонное убывание вероятностей, также охватывается полученными неравенствами. Действительно, если

$$1 > np + p > 0, \quad np - q < 0,$$

то $x = 0$.

Таким образом, получим простое условие максимума в начале распределения

$$p < \frac{1}{n+1} \quad \text{или} \quad n < \frac{q}{p}.$$

Рассмотрим подслучай 2б). Здесь мы предполагаем, что

$$\dots \frac{P_{x,n}}{P_{x-1,n}} > \frac{P_{x+1,n}}{P_{x,n}} = 1 > \frac{P_{x+2,n}}{P_{x+1,n}} > \dots$$

Отсюда

$$\dots < P_{x-1,n} < P_{x,n} = P_{x+1,n} > P_{x+2,n} > \dots$$

Таким образом, существуют два смежных числа повторений x и $x + 1$, имеющих равные вероятности, большие предыдущих и последующих. Распределение имеет плоский максимум, т. е. на графике две равные смежные ординаты будут максимальными (см. рис. 3). Так как из равенства единицы одного из отношений вытекает и весь описанный ход распределения, то для разыскания в этом подслучае наиболее вероятных чисел повторения достаточно преобразовать это основное равенство. Имеем

$$\frac{P_{x+1,n}}{P_{x,n}} = \frac{n-x}{x+1} \cdot \frac{p}{q} = 1,$$

откуда

$$x = np - q, \quad x + 1 = np + p.$$

Так как число повторений должно быть целым, то из написанных равенств следует, что подслучай 2б) имеет место при условии, что $np + p$ и $(np - q)$ — целые числа.

Случай 1б) охватывается нашими равенствами. Действительно, если $np - q = 0$, $np + p = 1$, то $x = 0$ и $x = 1$.

3. Последнее из отношений больше или равно единице

Тогда

$$\frac{P_{n,n}}{P_{n-1,n}} \geq 1.$$

Так как отношения убывают, то и все отношения больше единицы, и поэтому

$$P_{0,n} < P_{1,n} < P_{2,n} < \dots < P_{n-1,n} \leq P_{n,n}.$$

3а) В случае $P_{n-1,n} < P_{n,n}$ имеем монотонно возрастающее распределение вероятностей с наиболее вероятным числом повторений, равным n (числу испытаний) (см. рис. 3). Этот подслучай охватывается неравенствами для подслучая 2а. Действительно, если $np+p$ — дробное число и $np+p > n$, $np-q < n$, то $x=n$, следовательно,

$$p > \frac{n}{n+1}, \text{ или } n < \frac{p}{q}.$$

3б) Если $P_{n-1,n} = P_{n,n}$, то распределение на конце имеет два смежных одинаковых максимума, которые будут иметь место при

$$p = \frac{n}{n+1} \text{ или } n = \frac{p}{q}$$

(см. кривую 3б, рис. 3).

Подслучай 2б охватывается случаем 3б. В самом деле, если $np+p=n$, то $np-q=n-1$. Таким образом, условия для случая 2 можно считать самыми общими, так как они автоматически включают также случаи 1 и 3. Из сказанного вытекает следующее правило определения наиболее вероятного (или наиболее вероятных) числа повторений.

Нужно вычислить величину:

$$np+p.$$

Если это число дробное, то его целая часть и равна наиболее вероятному числу повторений. Если же $np+p$ — целое число, то наибольшие одинаковые вероятности имеют числа повторений $np+p$ и $np+p-1$ (или $np+p$ и $np-q$).

Наиболее вероятное число повторений есть число повторений, вероятность которого больше, чем вероятность всякого другого числа повторений, взятого отдельно. Заметим, что из этого отнюдь не следует, что весьма вероятно появление наиболее вероятного числа появлений. Наоборот, почти во всех задачах получается, что появление именно этого числа повторений (наиболее вероятного) имеет меньшую вероятность, чем появление какого-нибудь одного из других чисел повторений, безразлично какого.

Пример 1. Определить наиболее вероятное число повторений при $n=15$; $p=\frac{2}{3}$. Вычислим границы, между которыми заключается наиболее вероятное число повторений:

$$np - q = 9\frac{2}{3}, \quad np + p = 10\frac{2}{3};$$

так как границы — смешанные числа, то имеем только одно наиболее вероятное число повторений:

$$10\frac{2}{3} > x > 9\frac{2}{3}.$$

Следовательно, $x=10$.

Пример 2. Определить наиболее вероятное число повторений при $n=9$; $p=0,6$.

Вычислим границы

$$np + p = 6, \quad np - q = 5.$$

Так как границы — целые числа, то здесь наиболее вероятных чисел повторения два — 5 и 6.

Вычислим в этом примере вероятность одного из наиболее вероятных чисел повторений

$$P_{5,9} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4} (0,6)^5 \cdot (0,4)^4 \approx 0,25.$$

Из последнего примера видно, что наиболее вероятное число повторений имеет малую вероятность уже при сравнительно небольшом числе испытаний.

Более вероятно (0,75), что число повторений будет не 5, а какое-нибудь другое. Пример подтверждает общее замечание, сделанное несколько ранее.

Можно показать, что появление одного из нескольких чисел повторения, близких к наиболее вероятному, гораздо вероятнее, чем появление группы других чисел повторения. Для того чтобы в этом убедиться, полезно решить несколько задач следующего типа.

Число испытаний 8, вероятность события равна $\frac{1}{4}$. Сравнить вероятность двух событий:

а) число повторений заключено в пределах от 1 до 3 (группа чисел, близких к наиболее вероятному числу повторений);

б) число повторений от 5 до 7.

Полезно также сравнить вероятность события (а) с вероятностью какого-нибудь (безразлично какого) из остальных чисел повторения.

§ 38. Приближенная формула Лапласа для вычисления вероятности числа повторений события

В § 36 была выведена формула (8.1)

$$P_{k,n} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

для вычисления вероятности того, что событие C повторится k раз при n испытаниях, если испытания независимы, и вероятность C в каждом испытании равна p .

Вычисления по этой формуле неудобны, если n велико. Поэтому мы выведем приближенную формулу для вычисления $P_{k,n}$, которой можно пользоваться, если n превышает 10—20.

Из анализа известна *формула Стирлинга* для приближенного вычисления факториалов:

$$m! \approx \sqrt{2\pi} m^m e^{-m} \sqrt{m}. \quad (8.5)$$

Вычисления по этой формуле дают малую относительную погрешность даже при небольших значениях m ; при увеличении m точность растет. Для иллюстрации приведем сравнение точных величин факториалов с приближенными, вычисленными по формуле Стирлинга:

	m	4	7	10	20
Точн.	$m!$	24	5 040	3 628 800	$2\,432,9 \cdot 10^{14}$
Приблиз.	$(m!)$	23,51	4 980,6	3 598 700	$2\,422,7 \cdot 10^{14}$
	$\frac{m! - (m!)}{m!}$	0,020	0,012	0,008	0,004

Подставляя величины факториалов $n!$, $k!$, $(n-k)!$ по формуле Стирлинга в формулу (8.1), после элементарных преобразований получим

$$P_{k,n} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \cdot \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.$$

Введем в эту формулу вместо числа повторений k отклонение u числа повторения от числа np , близкого к наиболее вероятному числу повторений:

$$u = k - np.$$

Тогда

$$P_{k,n} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq \left(1 + \frac{u}{np}\right) \left(1 - \frac{u}{nq}\right)}} \cdot \left(\frac{np}{np+u}\right)^{np+u} \left(\frac{nq}{nq-u}\right)^{nq-u}.$$

Разделив числитель и знаменатель третьего и четвертого множителей соответственно на np и nq , можно написать

$$P_{k,n} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{\left(1 + \frac{u}{np}\right) \left(1 - \frac{u}{nq}\right)}} \cdot \left(1 + \frac{u}{np}\right)^{-np-u} \cdot \left(1 - \frac{u}{nq}\right)^{-nq+u}.$$

До сих пор мы, основываясь на формуле Стирлинга, делали лишь тождественные преобразования. Предположим теперь, что число испытаний n велико и число повторений k мало отличается от np ; кроме того, допустим, что p не близко ни к нулю, ни к единице. Из этих предположений следует, что число u мало по сравнению с np и nq . Упростим на основании сделанных предположений выражение для $P_{k,n}$. Так как $\frac{u}{np}$ и $\frac{u}{nq}$, по условию, малые дроби, то

третий множитель можно заменить единицей; отметим, что при этом мы пренебрегаем числом $\frac{p-q}{2} \cdot \frac{u}{\sqrt{npq}}$ и вторыми и высшими степенями дробей $\frac{u}{np}$ и $\frac{u}{nq}$; это легко показать, если разложить в ряд биномы

$$\left(1 + \frac{u}{np}\right)^{-\frac{1}{2}} \quad \text{и} \quad \left(1 - \frac{u}{nq}\right)^{-\frac{1}{2}}.$$

Для упрощения четвертого множителя воспользуемся тождеством

$$\ln \left(1 + \frac{u}{np}\right)^{-np-u} = -(np+u) \ln \left(1 + \frac{u}{np}\right),$$

где \ln обозначает натуральный логарифм. Разлагая $\ln \left(1 + \frac{u}{np}\right)$ в ряд по степеням $\frac{u}{np}$, получаем

$$\begin{aligned} \ln \left(1 + \frac{u}{np}\right)^{-(np+u)} &= -np \left(1 + \frac{u}{np}\right) \left(\frac{u}{np} - \frac{u^2}{2n^2p^2} + \dots\right) = \\ &= -np \left(\frac{u}{np} + \frac{u^2}{n^2p^2} - \frac{u^2}{2n^2p^2} + \dots\right) = -u - \frac{u^2}{2np}; \end{aligned}$$

отсюда

$$\left(1 + \frac{u}{np}\right)^{-(np+u)} = e^{-u - \frac{u^2}{2np} + \dots}.$$

Точно так же преобразуется последний множитель

$$\begin{aligned} \ln \left(1 - \frac{u}{nq}\right)^{-(nq-u)} &= nq \left(1 - \frac{u}{nq}\right) \left(\frac{u}{nq} + \frac{u^2}{2n^2q^2} + \dots\right) = \\ &= nq \left(\frac{u}{nq} - \frac{u^2}{2n^2q^2} + \dots\right) = u - \frac{u^2}{2nq} + \dots \end{aligned}$$

Таким образом,

$$\left(1 - \frac{u}{nq}\right)^{-(nq-u)} = e^{u - \frac{u^2}{2nq} + \dots}.$$

Ограничиваясь в четвертом и пятом множителях вторыми степенями u , получим приближенную формулу для вычисления $P_{k,n}$:

$$P_{k,n} = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2npq}}. \quad (8.6)$$

Подставляя сюда вместо u его значение $u = k - np$, получим формулу Лапласа в окончательном виде:

$$P_{k,n} = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(k-np)^2}{2npq}}. \quad (8.7)$$

Для упрощения записи формулы Лапласа введем вместо числа повторений испытаний k случайную величину z , определяемую формулой

$$z = \frac{k - np}{\sqrt{npq}}.$$

Двум значениям k , отличающимся на единицу, согласно этой формуле соответствуют следующие значения z :

$$z = \frac{k - np}{\sqrt{npq}} \quad \text{и} \quad z + \Delta z = \frac{k + 1 - np}{\sqrt{npq}}; \quad \text{отсюда} \quad \Delta z = \frac{1}{\sqrt{npq}}.$$

Вводя в формулу (8.7) z и Δz , перепишем ее в следующем простом виде:

$$P_{k,n} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Delta z, \quad (8.8)$$

где

$$z = \frac{k - np}{\sqrt{npq}}, \quad \Delta z = \frac{1}{\sqrt{npq}}. \quad (8.9)$$

Функция $\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ встречается во многих задачах теории вероятностей; ее называют *функцией Лапласа — Гаусса*. Таблица значений этой функции для z от 0,00 до 3,00 с интервалом в 0,01 дана в конце книги.

По формулам (8.8) и (8.9) с помощью этой таблицы легко вычисляются приближенные значения вероятности числа повторений события.

Формула (8.7) дает довольно точные значения $P_{k,n}$ даже при небольших значениях n , если p близко к 0,5. В качестве примера приведем таблицу значений $P_{k,n}$ при $n = 20$, $p = 0,4$ и $q = 0,6$. При этом везде ограничимся тремя знаками после запятой *):

k	2	4	6	8	10	12	14	16
Точн. знач. $P_{k,n}$. .	0,003	0,035	0,124	0,180	0,117	0,036	0,005	0,000
Прибл.	0,004	0,035	0,120	0,182	0,118	0,035	0,004	0,000

Из этой таблицы видно, что ошибка не превышает 0,004; относительная ошибка меньше при значениях k , близких к $np = 8$, т. е. к наиболее вероятному числу повторений; при значениях k , заметно отличающихся от наиболее вероятного, приближенная формула представляет значения вероятности хуже. Так и должно быть при тех упрощениях, какие были сделаны для вывода формулы.

*) Во второй строке таблицы название «точное» $P_{k,n}$ означает, что верны все знаки, но это не значит, что выписаны абсолютно точные значения (вычисления приближенные).

По формуле (8.7) легко исследовать распределение вероятностей. $P_{k,n}$ по этой формуле имеет максимум при $k = np$. Эта приближенная величина k_m наиболее вероятного числа повторений может отличаться от точного значения на величину, не превышающую правильной дроби (p или q), так как наиболее вероятное число повторений заключено между числами $np - q$ и $np + p$ или равно каждому из них, если имеем случай двух наиболее вероятных чисел повторений.

Подставляя в (8.7) $k = np$ для приближенной величины вероятности наиболее вероятного числа повторений, получаем формулу

$$P_{m,n} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}}.$$

При $n = 9$, $p = 0,6$, $q = 0,4$ по этой формуле получим $k_m = 5,4$ (вместо точных 5 и 6) и $P_{m,n} = 0,253$ (вместо 0,25 в примере 2 на стр. 129).

§ 39. Приближенная кривая распределения вероятностей

По формуле (8.7) легко также построить приближенно график распределения вероятностей различных чисел повторения событий. По оси ординат будем откладывать величины $y = nP_{k,n}$, а по оси абсцисс — числа $x = \frac{k}{n} - p$,

т. е. примем за начало координат точку, соответствующую приближенному, наиболее вероятному числу повторений. Тогда уравнение кривой, приближенно представляющей график распределения вероятностей, можно написать в виде

$$y = \frac{1}{\sqrt{\frac{pq}{n}}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2 \frac{pq}{n}}}. \quad (8.10)$$

Кривая эта симметрична относительно оси ординат и имеет на этой оси максимум. Она имеет две точки перегиба при $x = \pm \sqrt{\frac{pq}{n}}$; ось x является для этой кривой асимптотой. Ординаты кривой легко вычислить, пользуясь табл. I функции $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$; в рассматриваемом

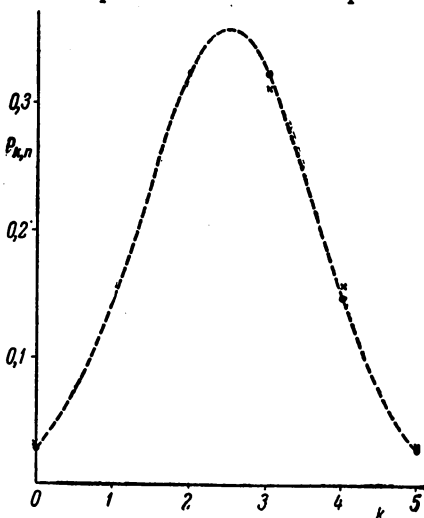


Рис. 4. Сравнение значений вероятностей $P_{k,n}$, вычисленной по точной формуле (крестики) при $n = 5$, $p = q = 0,5$ и приближенной формуле Лапласа (точки). Пунктирная кривая построена по формуле (8.7).

случае $z = \frac{x}{\sqrt{\frac{pq}{n}}}$. Кривая для случая $n=5$, $p=0,5$, $q=0,5$

изображена на рис. 4. Рисунок показывает, что даже при $n=5$ приближенная кривая удовлетворительно представляет точный график распределения.

§ 40. Распределение Пуассона (закон редких событий)

Приведенный в § 38 вывод приближенной формулы для $P_{k,n}$ непригоден, если p очень мало (или близко к 1). Для получения приближенного выражения для $P_{k,n}$ в случае малых p основную формулу перепишем в виде

$$P_{n,k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n-k},$$

где $a = np$. Знаменатель n^k заменим произведением; тогда

$$P_{k,n} = \frac{n}{n} \cdot \frac{n-1}{n} \dots \frac{n-k+1}{n} \cdot \frac{a^k}{k!} \left(1 - \frac{a}{n}\right)^n \left(1 - \frac{a}{n}\right)^{-k}.$$

Если k — определенное число, a — определенное число, то при $n \rightarrow \infty$ $p \rightarrow 0$. Поэтому

$$\lim_{n \rightarrow \infty} \frac{n-m}{n} = 1, \quad \lim_{n \rightarrow \infty} \left(1 - \frac{a}{n}\right)^n = e^{-a}, \quad \lim_{n \rightarrow \infty} \left(1 - \frac{a}{n}\right)^{-k} = 1$$

$$(0 \leq m \leq k-1).$$

Если k и a малы по сравнению с n и p — малое число по сравнению с единицей, то в последнем равенстве для $P_{k,n}$ все множители, кроме $\frac{a^k}{k!}$, можно заменить их пределами. В этом случае

$$P_{k,n} \approx \frac{a^k e^{-a}}{k!}. \quad (8.11)$$

Очевидно, что аналогичное распределение можно построить для случая, когда p близко к единице.

Распределение Пуассона (8.11) используется в задачах, относящихся к редким случаям (испускание β -лучей, слабые растворы и т. п.). В конце книги для распределения Пуассона приведена таблица по аргументам a и k .

ГЛАВА 9

ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 41. Случайные величины

При изучении явлений природы и общественной жизни приходится встречаться с величинами, числовых значений которых нельзя указать заранее (до наблюдения), если даже и известны некоторые условия, при которых будут происходить наблюдения.

Известно, например, что всякое измерение при наблюдении делается с ошибками, среди которых есть случайные. Далее, если определяются характеристики (показатель цвета, параллакс и т. п.) некоторой звезды, взятой из каталога наугад, то до измерений нельзя сказать, какие значения они имеют *).

Если рассматривается совокупность малых планет с определенными для каждой из них элементами орбит, то при изучении общих свойств такой совокупности мы считаем эти элементы случайными для произвольно взятого астероида. Это утверждение имеет такой смысл: во всей совокупности астероидов каждый из элементов, например среднее расстояние, имеет самые разнообразные значения; выбирая произвольную (обезличенную) планету, мы не можем сказать заранее, каким у нее будет, например, среднее расстояние. Величины такого рода называют *случайными*. Величины, значения которых возможно указать, условимся называть *определенными*.

О случайных величинах приходится говорить даже и в тех случаях, когда явление хорошо изучено и закон его известен. Мы можем по закону Ома предсказать силу тока, если известно напряжение и сопротивление, но действительно произведенные измерения могут показать некоторое (незначительное) отклонение от предсказанной величины. Такое отклонение является ошибкой измерения, некоторая доля которой вызывается случайными причинами и не может быть заранее предсказана. Вопрос о случайных ошибках измерений поэтому также рассматривается в теории вероятности.

*) Для правильного понимания последнего примера следует учесть, что для самой звезды ее физические характеристики вовсе не случайны, а определяются ее происхождением и эволюцией. Их приходится считать случайными, если мы выбираем звезду случайным образом.

Случайные величины могут быть разных типов. Мы будем рассматривать два типа. Назовем случайную величину *дискретной*, если она может принимать только конечное или счетное множество значений.

В задаче о повторении испытаний мы познакомились с классическим примером дискретной случайной величины, имеющей целочисленные значения. Если производится, например, 100 испытаний, то число повторений события есть число случайное, которое может иметь значения от 0 до 100. Вероятности этих значений можно вычислить по точной или приближенной формуле.

Пусть выпущена денежная лотерея, содержащая 100 билетов по 5 рублей за билет, с выигрышами в 100 руб., 50 руб. и 10 руб., причем не все билеты выигрывают. Для купившего один билет такой лотереи размер выигрыша есть дискретная случайная величина со следующими значениями:

95 руб., 45 руб., 5 руб., —5 руб.

Все числа получены вычитанием из выигрыша стоимости билета, чтобы получить чистый выигрыш, так как в подобных лотереях стоимость билета не возвращается. Четвертое значение соответствует случаю проигрыша; оно обязательно должно быть внесено в группу значений, чтобы она была полной.

Примером дискретных (тоже целочисленных) случайных величин может также быть число звезд в кратных системах звезд. Если рассматривать совокупность всех кратных систем и обращать внимание только на число звезд в системе, что оно может иметь значения 2, 3, 4, ... и будет случайным при выборе произвольного члена совокупности.

Случайной величиной *непрерывного* типа назовем такую, которая может принять любое значение из некоторого определенного интервала. Из таких величин мы рассмотрим только те, для которых возможно определить вероятность принять какое-нибудь значение в любом заданном интервале. Тот-определенный интервал, в котором находятся все возможные значения непрерывной случайной величины, назовем ее *областью*.

Простейшие примеры величин непрерывного типа: абсолютные величины звезд определенного спектрального класса, элементы орбит малых планет, ошибки измерений и т. п. В первых двух примерах элемент случайности входит при изучении всей совокупности объектов, когда рассматривается произвольно и случайно выбранный объект.

В этой главе мы будем рассматривать только дискретные случайные величины с конечным числом возможных значений. Предположим также, что эти значения случайной величины единственно возможны и несовместны, т. е. известна полная группа значений, и появление одного из них исключает появление всех остальных значений.

С точки зрения теории вероятностей вопрос о значениях дискретной величины станет определенным, если будут даны вероятности отдельных числовых значений случайной величины X .

Обозначая вероятности последовательных значений x_1, x_2, \dots, x_n через p_1, p_2, \dots, p_n , условимся записывать данные в виде так называемой таблицы распределения*):

$$\begin{array}{l} \text{числовые значения } X \parallel x_1, x_2, \dots, x_n, \\ \text{вероятности} \parallel p_1, p_2, \dots, p_n. \end{array}$$

При этих обозначениях X (без значка) есть в сущности только название величины; в то же время под x подразумевают ее произвольное значение. Так как предполагается, что перечислены все возможные значения x_k , то по теореме сложения вероятностей

$$p_1 + p_2 + \dots + p_n = 1.$$

Определенная величина, очевидно, имеет одно значение, и вероятность этого значения равна единице.

Определение. Две случайные величины, рассматриваемые в одной задаче, назовем *взаимно независимыми*, если вероятность всякого значения одной из них не зависит от того, какое значение принимает другая величина.

Это определение легко применить с достаточными основаниями только в простейших задачах. Если, например, бросают две игральные кости два игрока, то числа очков, которые выпадут на костях, будут случайными и независимыми. Значительно труднее установить независимость в физических задачах. Если в совокупности звезд рассматриваются такие характеристики, как абсолютная величина и радиальная скорость, то нет физических причин, чтобы думать о связи (вероятностной) между этими величинами, однако полной уверенности в этом не может быть. Если подобные величины рассматриваются совместно, то к предположению об их независимости надо относиться осторожно. Можно еще заметить, что в ряде задач предположение о независимости является гипотезой, которая должна быть проверена наблюдениями.

Если значения случайной величины X обозначим x_r ($r = 1, 2, \dots, m$), а значения Y через y_s ($s = 1, 2, \dots, n$), то условие независимости X и Y можно записать так:

$$P(x_r | y_s) = P(x_r), \quad P(y_s | x_r) = P(y_s) \quad (9.1)$$

при всех возможных r и s .

§ 42. Математическое ожидание дискретной случайной величины

Определение. *Математическим ожиданием* дискретной случайной величины X называется сумма произведений ее числовых значений на их вероятности. Математическое ожидание обозначается

*) С примерами подобных таблиц мы уже встречались в предыдущей главе.

через $E(X)$ или \bar{x} (употребляются еще обозначения M_X , $M(X)$). Согласно определению

$$E(X) = \sum_{k=1}^n p_k x_k. \quad (9.2)$$

Для уяснения смысла понятия математического ожидания рассмотрим следующую задачу.

Выпущена лотерея в 100 билетов стоимостью по 5 руб. с выигрышами: два по 50 руб., шесть по 25 руб., десять по 5 руб. (стоимость билета не возвращается). Определить математическое ожидание выигрыша для обладателя одного билета.

По условиям лотереи выигрыш есть величина случайная. Поэтому нужно составить таблицу всех возможных значений выигрыша (исключив стоимость билета) и найти их вероятности.

Имеем:

X	45	20	0	—5
p	0,02	0,06	0,10	0,82

(последний столбец относится к случаю, когда не будет выигрыша). По определению,

$$E(X) = 45 \text{ руб.} \cdot 0,02 + 20 \text{ руб.} \cdot 0,06 + 0 \cdot 0,10 - 5 \text{ руб.} \cdot 0,82 = -2 \text{ руб.}$$

Если подсчитать общий выигрыш по всем билетам, то получим 300 руб. — 500 руб. = —200 руб., поскольку общая сумма выигрышей 300 руб., а за билеты уплатили 500 руб. Если разделить общий выигрыш на число билетов, то получим выигрыш, приходящийся в среднем на один билет. Этот средний выигрыш точно равняется математическому ожиданию; оба числа получаются одними и теми же действиями, выполняемыми в несколько разном порядке.

Таким образом, математическое ожидание соответствует *среднему арифметическому*. Эти числа равны, если мы знаем точно все значения X и их вероятности. Поэтому математическое ожидание иногда называют *теоретическим средним значением* случайной величины.

Свойства математического ожидания

I. Если значения случайной величины числа именованные, то и ее математическое ожидание число именованное той же размерности.

Это непосредственно вытекает из определения, так как значения x умножаются на отвлеченные числа (значения вероятностей).

II. Математическое ожидание — число положительное, если все значения X положительны, и может быть положительным или отрицательным числом, если среди значений X есть отрицательные.

III. Математическое ожидание определенной величины равно ее числовому значению.

Так как X по условию может принимать только одно определенное значение c , то вероятность этого значения равна 1:

$$E(X) = c \cdot 1 = c.$$

IV. Математическое ожидание произведения случайной величины X на определенную величину c равно произведению значения этой определенной величины на математическое ожидание случайной.

Так как случайная величина не является обычным определенным числом, то необходимо сначала условиться, что означает умножение случайной величины на определенное число. Очевидным определением такого произведения будет: произведение случайной величины на определенное число есть новая случайная величина, значения которой представляют произведения значений множителя на определенный множитель, а вероятности значений произведения равняются вероятностям соответствующих значений множителя. Согласно определению, если

$$X \begin{vmatrix} x_1, & x_2, & \dots, & x_n, \\ p_1, & p_2, & \dots, & p_n. \end{vmatrix}$$

то таблица распределения cX будет такой:

$$cX \begin{vmatrix} cx_1, & cx_2, & \dots, & cx_n, \\ p_1, & p_2, & \dots, & p_n. \end{vmatrix}$$

Поэтому

$$E(cX) = \sum_{k=1}^n cx_k p_k = c \sum_{k=1}^n x_k p_k = cE(X).$$

V. Если m и M — наименьшее и наибольшее из значений величины x , то

$$m \leq E(X) \leq M.$$

Действительно,

$$E(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n.$$

Заменим в правой части все числа x_k на m ; так как числа p_1, p_2, \dots, p_n , как вероятности, все положительные, то правая часть уменьшается. Поэтому

$$\begin{aligned} E(X) &\geq mp_1 + mp_2 + \dots + mp_n = \\ &= m(p_1 + p_2 + \dots + p_n) = m \cdot 1 = m. \end{aligned}$$

Совершенно так же можно доказать, что

$$E(X) \leq M.$$

Отметим, что указанное неравенство превращается в строгое равенство только в случае, если X есть определенная величина.

§ 43. Теоремы сложения и умножения математических ожиданий

1. Теорема сложения.

Математическое ожидание суммы двух или нескольких случайных величин равняется сумме их математических ожиданий.

Имеем две случайные величины:

$$\begin{array}{l} X \parallel x_1, x_2, \dots, x_n, \\ \parallel p_1, p_2, \dots, p_n; \end{array} \quad \begin{array}{l} Y \parallel y_1, y_2, \dots, y_m, \\ \parallel P_1, P_2, \dots, P_m. \end{array}$$

В общем случае, когда X принимает какое-нибудь из своих n значений, Y может принять любое из своих m значений. Поэтому сумма $X+Y$ принимает mn значений, получаемых сложением каждого из значений X с каждым из значений Y . Таблица распределения $X+Y$ имеет вид

$$\begin{array}{l} X+Y \parallel x_1+y_1, \dots, x_1+y_m; \quad x_2+y_1, \dots, x_k+y_1, \dots, x_n+y_m, \\ \parallel p_{11}, \dots, p_{1m}; \quad p_{21}, \dots, p_{kl}, \dots, p_{nm}. \end{array}$$

где p_{kl} ($k=1, \dots, n$; $l=1, \dots, m$) — неизвестные нам вероятности значений x_k+y_l , т. е. вероятности совмещения x_k с y_l . Сумма $X+Y$ принимает значение x_k+y_l , если произойдет совмещение двух событий: X принимает значение x_k , а Y — значение y_l .

Поэтому p_{kl} есть вероятность сложного события, состоящего из двух событий — появления x_k и появления y_l . Так как в условии теоремы не оговорено, что вероятность x_k не зависит от появления или не появления y_l , иначе говоря, величины X и Y могут быть связанными, то мы должны применять теорему умножения вероятностей в общем виде. Поэтому

$$p_{kl} = p_k P(y_l/x_k) = P_l P(x_k/y_l), \quad (9.3)$$

где, как и раньше, $P(x_k/y_l)$ есть вероятность того, что X примет значение x_k , если известно, что Y получил значение y_l (число таких равенств равно mn , причем k принимает целые положительные значения от 1 до n , а l такие же значения от 1 до m). По определению математического ожидания имеем:

$$E(X+Y) = \sum_{l=1}^m \sum_{k=1}^n (x_k + y_l) p_{kl}.$$

Раскроем скобки и заменим p_{kl} их выражениями в первой форме (9.3), если они являются коэффициентами при x_k , и второй формой (9.3), если

они суть коэффициенты при y_i . Меняя, кроме того, в первом члене порядок суммирования, получим:

$$E(X+Y) = \sum_{k=1}^n p_k x_k \left\{ \sum_{i=1}^m P(y_i/x_k) \right\} + \sum_{i=1}^m P_i y_i \left\{ \sum_{k=1}^n P(x_k/y_i) \right\}.$$

Легко показать, что каждая из сумм, стоящих в фигурных скобках, равна единице. Действительно, по теореме сложения вероятностей, написанной в обратном виде,

$$\sum_{i=1}^m P(y_i/x_k) = P(y_1 + y_2 + \dots + y_m/x_k)$$

(сложение в правой части символическое, означающее появление либо y_1 , либо y_2 , ..., либо y_m). Так как список значений полный, то какое-нибудь из перечисленных в списке значений появится наверно, независимо от значения x_k ; поэтому

$$\sum_{i=1}^m P(y_i/x_k) = 1$$

при всех x_k . Совершенно аналогично получим

$$\sum_{k=1}^n P(x_k/y_i) = 1$$

при всех y_i . Поэтому

$$E(X+Y) = \sum_{k=1}^n p_k x_k + \sum_{i=1}^m P_i y_i = E(X) + E(Y). \quad (9.4)$$

Следует отметить, что при доказательстве теоремы мы не накладывали никаких условий на случайные величины X и Y .

2. Теорема умножения

Математическое ожидание произведения независимых случайных величин равняется произведению их математических ожиданий.

Имеем случайные величины:

$$\begin{array}{l} X \parallel x_1, x_2, \dots, x_n, \\ \parallel p_1, p_2, \dots, p_n; \end{array} \quad \begin{array}{l} Y \parallel y_1, y_2, \dots, y_m, \\ \parallel P_1, P_2, \dots, P_m. \end{array}$$

Произведение случайных величин есть случайная величина, значения которой получим, умножая всякое значение одной на всякое значение другой. Таблица распределения величины будет такой:

$$\begin{array}{l} XY \parallel x_1 y_1, x_2 y_1, \dots, x_n y_1; \quad x_1 y_2, \dots, x_2 y_2, \dots, x_k y_2, \dots, x_n y_2, \\ \parallel p_{11}, p_{21}, \dots, p_{n1}, \quad p_{12}, \dots, p_{n2}, \dots, p_{k2}, \dots, p_{n2}. \end{array}$$

Вследствие независимости величин вероятность совмещения значений x_k и y_l ($k=1, 2, \dots, n$; $l=1, 2, \dots, m$) равняется произведению их вероятностей:

$$p_{kl} = p_k P_l.$$

Поэтому

$$E(XY) = \sum_{k=1}^n \sum_{l=1}^m p_k P_l x_k y_l.$$

Вынося $p_k x_k$ за знак внутренней суммы, получим

$$E(XY) = \sum_{k=1}^n p_k x_k \left(\sum_{l=1}^m P_l y_l \right).$$

Сумма в скобке есть математическое ожидание величины Y , т. е. определенное число, которое можно вынести за знак первой суммы. Тогда первая сумма даст математическое ожидание величины X . Таким образом,

$$E(XY) = E(X)E(Y). \quad (9.5)$$

Доказанная теорема легко обобщается для любого числа взаимно независимых случайных величин:

$$E(XYZ \dots) = E(X)E(Y)E(Z) \dots \quad (9.6)$$

Необходимо обратить особое внимание на то, что теорема умножения верна только для независимых случайных величин. Если же величины связаны, теорема значительно изменяет свое содержание, так как должны быть известны условные вероятности каждой из величин, вычисленные в предположении, что появилось каждое из значений другой величины.

§ 44. Дисперсия случайной величины; свойства дисперсии

Пусть имеется случайная величина X , математическое ожидание которой известно; мы обозначим это ожидание через \bar{x} .

Определение. *Дисперсией* случайной величины называется математическое ожидание квадрата отклонения ее значений от ее математического ожидания.

Пусть наша величина X имеет значения x_1, x_2, \dots, x_n , а вероятности этих значений суть p_1, p_2, \dots, p_n . Составим новую случайную величину $X - \bar{x}$, значения которой будут $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$; вероятности этих значений те же, что и вероятности значений x_1, x_2, \dots, x_n , так как \bar{x} — число определенное. Значения $(x_1 - \bar{x}), \dots, (x_n - \bar{x})$ называются *отклонениями* значений случайной величины от ее математического ожидания; новая случайная величина $X - \bar{x}$ называется отклонением случайной величины от ее математического ожидания. Если мы возведем в квадрат значения

$(x_1 - \bar{x}), \dots, (x_n - \bar{x})$, то получим значения еще одной случайной величины — квадрата отклонения случайной величины от ее математического ожидания. Вероятности значений этой величины те же, что и вероятности значений X . Записывая квадрат отклонения по уже знакомой нам схеме, получим

$$(X - \bar{x})^2 \begin{matrix} \parallel \\ p_1, & p_2, & \dots, & p_n \end{matrix}$$

Дисперсию обычно обозначают $D(X)$ или σ_x^2 . По определению,

$$D(X) \equiv \sigma_x^2 = E\{(X - \bar{x})^2\} \quad (9.7)$$

или

$$D(X) = \sum_{k=1}^n p_k (x_k - \bar{x})^2.$$

Свойства дисперсии

I. Дисперсия случайной величины есть число положительное, обращающееся в нуль только в случае, когда x есть величина определенная.

Действительно, числа

$$(x_k - \bar{x})^2, \quad k = 1, 2, \dots, n$$

неотрицательны, поэтому $D(x)$ есть число неотрицательное по свойству II математического ожидания. Сумма $\sum_{k=1}^n p_k (x_k - \bar{x})^2$ может обращаться в нуль только в том случае, когда каждое из слагаемых равно нулю, т. е. при

$$x_1 = x_2 = \dots = x_n = \bar{x},$$

а в этом случае x — величина определенная, имеющая значение \bar{x} .

II. Дисперсия определенной величины равна нулю.

Действительно, в этом случае математическое ожидание равно значению определенной величины, отклонение равно нулю с вероятностью, равной единице; значит, дисперсия равна нулю.

III. Дисперсия случайной величины равна разности между математическим ожиданием квадрата этой величины и квадратом ее математического ожидания.

Преобразуя $(X - \bar{x})^2$ и применяя теорему сложения математического ожидания и его свойства IV и III, получим

$$D(X) = E\{(X - \bar{x})^2\} = E\{X^2 - 2\bar{x}X + \bar{x}^2\} = E(X^2) - 2\bar{x}E(X) + \bar{x}^2.$$

Так как $E(X) = \bar{x}$, то

$$D(X) = E(X^2) - 2\bar{x}^2 + \bar{x}^2 = E(X^2) - \bar{x}^2$$

или

$$D(X) = E(X^2) - [E(X)]^2. \quad (9.8)$$

IV. Дисперсия произведения случайной величины X на определенную величину c равна произведению квадрата определенного множителя на дисперсию случайного множителя.

Пусть $E(X) = \bar{x}$, $\sigma_X^2 = E[(X - \bar{x})^2]$.

По свойству IV математического ожидания

$$E(cX) = cE(X) = c\bar{x},$$

поэтому

$$D(cX) = E\{(cX - c\bar{x})^2\} = E\{c^2(X - \bar{x})^2\} = c^2E\{(X - \bar{x})^2\}$$

или

$$\sigma_{cX}^2 = c^2\sigma_X^2. \quad (9.9)$$

V. Дисперсия суммы взаимно независимых случайных величин равна сумме их дисперсий.

При доказательстве ограничимся двумя слагаемыми (при большем числе слагаемых доказательство будет таким же).

Обозначим через \bar{x} и \bar{y} математические ожидания случайных величин X и Y . Так как

$$E(X + Y) = \bar{x} + \bar{y}$$

по теореме сложения математических ожиданий, то, согласно (9.8),

$$D(X + Y) = E\{(X + Y)^2\} - (\bar{x} + \bar{y})^2.$$

Раскрывая $(X + Y)^2$ и $(\bar{x} + \bar{y})^2$ и применяя теорему сложения математических ожиданий и свойства IV и III математического ожидания, получим

$$D(X + Y) = E(X^2) + 2E(XY) + E(Y^2) - \bar{x}^2 - 2\bar{x}\bar{y} - \bar{y}^2$$

Так как X и Y независимы по условию, то

$$E(XY) = E(X)E(Y) = \bar{x}\bar{y}.$$

Поэтому второй и пятый члены правой части предыдущего равенства взаимно уничтожаются. Используя (9.4) из предыдущего параграфа мы получим формулу сложения дисперсий:

$$D(X + Y) = D(X) + D(Y)$$

или

$$\sigma_{(X+Y)}^2 = \sigma_X^2 + \sigma_Y^2. \quad (9.10)$$

Легко показать таким же способом, что

$$D(X - Y) = D(X) + D(Y) = \sigma_{(X-Y)}^2 = \sigma_x^2 + \sigma_y^2. \quad (9.11)$$

VI. Если

$$U = aX + bY + cZ + \dots + r;$$

где X, Y, Z — взаимно независимые случайные величины, а a, b, c, \dots, r — определенные величины, то дисперсия U определяется формулой

$$\sigma_U^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + c^2\sigma_z^2 + \dots + 0.$$

По теореме сложения дисперсий (обобщенной для алгебраической суммы нескольких слагаемых)

$$\sigma_U^2 = \sigma_{aX}^2 + \sigma_{bY}^2 + \sigma_{cZ}^2 + \dots + \sigma_r^2;$$

$$\sigma_{aX}^2 = a^2\sigma_x^2, \quad \sigma_{bY}^2 = b^2\sigma_y^2; \quad \sigma_{cZ}^2 = c^2\sigma_z^2;$$

наконец, согласно свойству II,

$$\sigma_r^2 = 0.$$

Сопоставляя последние равенства, получим

$$D(U) = a^2D(X) + b^2D(Y) + c^2D(Z) + \dots + 0. \quad (9.12)$$

§ 45. Математическое ожидание и дисперсия числа повторений

Пусть производится n независимых по отношению к событию C испытаний, в каждом из которых вероятность C равна p . Число k повторений события C есть дискретная случайная величина, которая может принимать значения $0, 1, 2, 3, \dots, n$. Как было установлено в § 36—37, таблица распределения этой величины имеет вид

k	0	1	2	...	m	...	$n-1$	n
$P_{k,n}$	q^n	npq^{n-1}	$\frac{n(n-1)}{2}p^2q^{n-2}$...	$\frac{n(n-1)\dots(n-m+1)}{m!}p^mq^{n-m}$...	$np^{n-1}q$	p^n

Выведем формулы для математического ожидания и дисперсии числа повторений события.

При помощи таблицы распределения найдем:

$$\begin{aligned} E(k) = & 0 \cdot q^n + 1 \cdot npq^{n-1} + \\ & + 2 \cdot \frac{n(n-1)}{2} p^2q^{n-2} + \dots + \frac{n(n-1)\dots(n-m+1)}{(m-1)!} p^mq^{n-m} + \dots + \\ & + n(n-1)p^{n-1}q + np^n. \end{aligned}$$

или

$$E(k) = np \left[q^{n-1} + (n-1)pq^{n-2} + \dots \right. \\ \left. \dots + \frac{(n-1) \dots (n-m+1)}{(m-1)!} p^{m-1} q^{n-m} + \dots \right. \\ \left. \dots + (n-1)p^{n-2}q + p^{n-1} \right].$$

Выражение, стоящее в квадратной скобке, есть разложение в бином Ньютона величины $(q+p)^{n-1}$. Но $q+p=1$, поэтому

$$E(k) = np.$$

Заметим, что $E(k)$ равняется наиболее вероятному числу повторений или отличается на величину, меньшую единицы.

Перейдем теперь к определению дисперсии числа повторений.

По определению, дисперсия есть математическое ожидание квадрата отклонения случайной величины от ее математического ожидания. Так как по результатам предыдущей задачи $E(k) = np$, то, обозначая дисперсию числа повторений через σ_k^2 и применяя свойство III дисперсии и (9.8), получим

$$\sigma_k^2 = E(k^2) - n^2 p^2.$$

Для вычисления $E(k^2)$ воспользуемся приемом, указанным в книге В. И. Романовского.

Из тождества

$$k^2 = k(k-1) + k$$

следует, что

$$E(k^2) = E\{k(k-1)\} + E(k).$$

Так как k есть случайная величина, принимающая значения $0, 1, 2, \dots, n$, то $k(k-1)$ — также случайная величина, значения которой

$$0, 0, 2, 3 \cdot 2, 4 \cdot 3, \dots;$$

их вероятности те же, что и вероятности соответствующих значений k . Поэтому

$$E\{k(k-1)\} = 0 \cdot q^n + 0 \cdot npq^{n-1} + 2 \frac{n(n-1)}{1 \cdot 2} p^2 q^{n-2} + \\ + 3 \cdot 2 \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} p^3 q^{n-3} + 4 \cdot 3 \frac{n(n-1)(n-2)(n-3)}{1 \cdot 2 \cdot 3 \cdot 4} p^4 q^{n-4} + \dots \\ \dots + \frac{m(m-1)n(n-1) \dots (n-m+1)}{1 \cdot 2 \cdot 3 \dots (m-2)(m-1)m} p^m q^{n-m} + \dots \\ \dots + n(n-1)(n-2)p^{n-1}q + n(n-1)p^n = \\ = n(n-1)p^2 \left[q^{n-2} + (n-2) \cdot pq^{n-3} + \frac{(n-2)(n-3)}{1 \cdot 2} p^2 q^{n-4} + \dots \right. \\ \left. \dots + \frac{(n-2)(n-3) \dots (n-m+1)}{1 \cdot 2 \dots (n-2)} p^{m-2} q^{n-m} + \dots \right. \\ \left. \dots + (n-2)p^{n-3}q + p^{n-2} \right].$$

Легко видеть, что выражение в квадратной скобке есть разложение по биному Ньютона величины $(q + p)^{n-2}$, равной единице. Таким образом,

$$E\{k(k-1)\} = n(n-1)p^2,$$

откуда

$$E(k^2) = n(n-1)p^2 + np;$$

поэтому

$$\sigma_k^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$$

или, окончательно,

$$\sigma_k^2 = npq.$$

Следует отметить, что и математическое ожидание и дисперсия числа повторений фактически входят в приближенную формулу Лапласа (8.7) задачи о повторении испытаний. Заменяя \sqrt{npq} через σ_k и np через \bar{k} , запишем эту формулу в виде

$$P_{k,n} \approx \frac{1}{\sigma_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{(k-\bar{k})^2}{2\sigma_k^2}}. \quad (9.13)$$

ГЛАВА 10

ЗАКОН БОЛЬШИХ ЧИСЕЛ

Законом больших чисел называется совокупность теорем о вероятностях событий при большом числе испытаний и о вероятностях значений сумм случайных величин, когда число слагаемых велико.

§ 46. Лемма Чебышева — Маркова

Лемма. Если случайная величина X принимает только неотрицательные значения и ее математическое ожидание равно \bar{x} , то вероятность того, что X примет значение, меньшее, чем $t^2\bar{x}$ (t^2 — произвольное положительное число), больше, чем $1 - \frac{1}{t^2}$, т. е.

$$P(X < t^2\bar{x}) > 1 - \frac{1}{t^2}.$$

Обозначим через M наибольшее значение величины X . Если t^2 настолько велико, что $t^2\bar{x} \geq M$, то лемма очевидно верна, ибо в этом случае

$$P(X < t^2\bar{x}) = 1.$$

Случай $0 < t^2 < 1$ также не представляет интереса, так как тогда

$$1 - \frac{1}{t^2} < 0$$

и мы получим очевидный результат — вероятность больше отрицательного числа.

Таким образом, в дальнейшем можно считать, что

$$\bar{x} \leq t^2\bar{x} < M.$$

Из этого неравенства и из свойства V математического ожидания (см. стр. 139) следует, что среди значений X найдутся как бóльшие,

так и меньшие, чем $t^2\bar{x}$. Выделим из значений X те, которые больше или равны $\bar{x}t^2$, и обозначим их x_1, x_2, \dots, x_k . Тогда таблицу распределения можно будет записать в виде

$$X \left\| \begin{array}{c|c} x_1, x_2, \dots, x_k & x_{k+1}, \dots, x_n \\ \hline p_1, p_2, \dots, p_k & p_{k+1}, \dots, p_n \end{array} \right.$$

По определению,

$$E(X) = \bar{x} = x_1p_1 + x_2p_2 + \dots + x_kp_k + x_{k+1}p_{k+1} + \dots + x_np_n.$$

Отбросим в правой части все члены с индексами $k+1$ и больше. Так как все x_s по условию неотрицательны, а p_s положительны ($s=1, 2, \dots, n$), то правая часть от этого может только уменьшиться. Поэтому

$$\bar{x} \geq x_1p_1 + x_2p_2 + \dots + x_kp_k.$$

В правой части все x_r ($r=1, 2, \dots, k$) заменим через $t^2\bar{x}$, которое меньше заменяемых чисел или равно одному из них. Вследствие того, что все $p_r > 0$, правая часть уменьшится; поэтому

$$\bar{x} > t^2\bar{x}(p_1 + p_2 + \dots + p_k)$$

или, так как $t^2 > 1$,

$$p_1 + p_2 + \dots + p_k < \frac{1}{t^2}.$$

По теореме сложения вероятностей левая часть неравенства есть вероятность того, что X примет какое-нибудь из значений x_1, x_2, \dots, x_k , т. е. что значение X будет больше или равно $t^2\bar{x}$.

Обозначая через $P(X \geq t^2\bar{x})$ вероятность того, что $X \geq t^2\bar{x}$, имеем

$$P(X \geq t^2\bar{x}) < \frac{1}{t^2}.$$

Величина X , наверное, имеет либо значение, большее $t^2\bar{x}$, либо равное этому числу, либо меньшее. Поэтому

$$P(X < t^2\bar{x}) + P(X \geq t^2\bar{x}) = 1.$$

Сопоставляя последнее равенство с предыдущим неравенством, получим

$$P(X < t^2\bar{x}) > 1 - \frac{1}{t^2}, \quad (10.1)$$

что и требовалось доказать.

Следствие. Если в неравенстве

$$P(X \geq t^2 \bar{x}) < \frac{1}{t^2}$$

считать t^2 очень большим числом, то можно сказать, что весьма мало вероятны значения величины X , значительно превышающие ее математическое ожидание.

§ 47. Теорема Я. Бернулли

Определение. Отношение числа повторений k к числу испытаний n в задаче о повторении испытаний называется *относительной частотой*; если испытания (наблюдения) произведены, то эта величина есть отношение числа случаев, когда событие произошло, к общему числу испытаний.

Бернулли было показано, что при большом числе испытаний в большинстве случаев относительная частота, вообще говоря, должна быть близка к вероятности. Именно это утверждение и составляют суть теоремы Я. Бернулли; теорему эту мы докажем в двух формулировках.

I формулировка. Если производится неограниченный ряд независимых по отношению к событию C испытаний, в каждом из которых вероятность события p остается постоянной, то с вероятностью, как угодно мало отличающейся от единицы, можно ожидать, что при достаточно большом числе испытаний n разность между относительной частотой k/n и вероятностью p будет как угодно мала по абсолютной величине.

II формулировка. Если производится неограниченный ряд независимых по отношению к событию C испытаний, в каждом из которых вероятность p события остается постоянной, то как бы ни были малы заданные положительные числа ϵ и δ , можно по ним найти такое число N , что при числе испытаний большем N вероятность выполнения неравенства $\left| \frac{k}{n} - p \right| < \epsilon$ будет больше $1 - \delta$, т. е.

$$P \left[\left| \frac{k}{n} - p \right| < \epsilon \right] > 1 - \delta \quad \text{при} \quad n > N(\epsilon, \delta) = \frac{pq}{\delta \epsilon^2}. \quad (10.2)$$

Вторая формулировка отличается от первой тем, что числом ϵ оценивается отклонение относительной частоты от вероятности, число δ оценивает степень близости вероятности к 1, а определяемое по ϵ и δ число N есть то достаточное число испытаний, при котором доказываемое неравенство будет верно.

Доказательство. Случайная величина $(k - np)^2$ может иметь только положительные значения и в частном случае равно нулю.

Математическое ожидание этой величины известно; эта есть дисперсия числа повторений, равная npq (см. § 45). Применяя лемму

Чебышева к случайной величине $(k - np)^2$, находим

$$P[(k - np)^2 < t^2 npq] > 1 - \frac{1}{t^2}.$$

Вероятность выполнения неравенства не изменится, если его заменить эквивалентным неравенством. Неравенство, стоящее под знаком вероятности (в квадратной скобке), эквивалентно неравенству

$$\left| \frac{k}{n} - p \right| < t \sqrt{\frac{pq}{n}},$$

в котором t считается положительным. Поэтому

$$P\left\{\left| \frac{k}{n} - p \right| < t \sqrt{\frac{pq}{n}}\right\} > 1 - \frac{1}{t^2}$$

Определим t так, чтобы выполнялось равенство

$$t \sqrt{\frac{pq}{n}} = \varepsilon;$$

из этого условия получаем

$$t^2 = \frac{n\varepsilon^2}{pq}.$$

Тогда

$$P\left\{\left| \frac{k}{n} - p \right| < \varepsilon\right\} > 1 - \frac{pq}{n\varepsilon^2}. \quad (10.3)$$

Выводом этого неравенства теорема в первой формулировке в сущности доказана, так как при заданных p , q и любом малом ε всегда можно выбрать для n такое большое значение N , что число, стоящее в правой части неравенства, будет как угодно близко к 1 и остается таким, если $n > N$. Чтобы доказать теорему во второй формулировке, зададим ε и δ и определим N из равенств

$$N = \frac{pq}{\delta\varepsilon^2}, \quad \delta = \frac{pq}{N\varepsilon^2}. \quad (10.4)$$

При $n > N$

$$\frac{pq}{n\varepsilon^2} < \frac{pq}{N\varepsilon^2} = \delta,$$

следовательно,

$$\frac{pq}{n\varepsilon^2} < \delta, \quad 1 - \frac{pq}{n\varepsilon^2} > 1 - \delta.$$

Мы не нарушим неравенства (10.3), если заменим его первую часть на $1 - \delta$. Это и доказывает теорему во второй формулировке.

Следует отметить, что мы пользуемся при определении N и при доказательстве теоремы Бернулли леммой, которая доказана после двукратного преобразования точного равенства в неравенство; поэтому то значение N , которое мы таким образом получаем, достаточно, но его нельзя считать необходимым. Уточнение оценки показывает, что неравенства теоремы Бернулли верны и при числе испытаний N .

Определение нижней границы достаточного числа испытаний, при котором выполняются неравенства теоремы Бернулли, возможно и в том случае, если неизвестна величина вероятности. Чтобы оценить эту границу, рассмотрим величину $pq = v$, являющуюся произведением двух чисел, сумма которых равна единице. Найдем максимум функции:

$$v = p(1 - p).$$

Приравнявая первую производную нулю, находим, что v имеет экстремум $1/4$, а по знаку второй производной убеждаемся в том, что при $p = 1/2$ и $q = 1/2$ величина v имеет максимум, равный $1/4$. Поэтому, каковы бы ни были p и q , всегда имеем

$$v = pq \leq \frac{1}{4}.$$

Отсюда на основании равенства (10.4) для достаточного при всех p числа испытаний получаем следующее выражение:

$$N' = \frac{1}{4\delta\epsilon^2}. \quad (10.5)$$

Это замечание имеет значение при обращении теоремы Бернулли, по которой за приближенную величину неизвестной вероятности можно принять наблюдаемую частоту его появлений. Сказанное не представляет точного содержания обратной теоремы Бернулли, которая формулируется аналогично доказанной теореме (прямой).

Рассмотрим некоторые примеры применения теоремы Бернулли.

Пример 1. Бросают монету. Требуется определить число опытов, достаточное для того, чтобы с вероятностью, большей 0,9, можно было ожидать, что относительная частота будет отличаться от вероятности $1/2$ меньше, чем на $1/6$ по абсолютной величине.

В нашем случае $p = q = \frac{1}{2}$, $\epsilon = 0,2$, $\delta = 1 - 0,9 = 0,1$. По формуле (10.4) получаем

$$N = \frac{\frac{1}{2} \cdot \frac{1}{2}}{0,2^2 \cdot 0,1} \approx 63.$$

Таким образом, мы можем утверждать, что при $n \geq 63$ вероятность неравенства $\left| \frac{k}{n} - \frac{1}{2} \right| < 0,2$ больше 0,9:

$$P\left(\left|\frac{k}{n} - \frac{1}{2}\right| < 0,2\right) > 0,9.$$

Отметим, что это неравенство определяет довольно широкие границы для числа повторений. Действительно, неравенство $\left| \frac{k}{n} - \frac{1}{2} \right| < 0,2$ эквивалентно двум неравенствам:

$$-0,2 < \frac{k}{n} - 0,5 < 0,2$$

или

$$0,3n < k < 0,7n.$$

То же число опытов получится, если применять «более грубую» формулу

$$N' = \frac{1}{4\delta\epsilon^2}.$$

Это объясняется тем, что в данном случае $v = pq = \frac{1}{4}$. Положим $n = 100$; тогда

$$30 < k < 70.$$

Если бы мы захотели сузить границы для k , то надо было бы уменьшить ϵ , что значительно увеличит N , как видно из формулы (10.4), поскольку в знаменателе имеем ϵ^2 .

Пример 2. Монету бросают 200 раз; оценить вероятность того, что

$$\left| \frac{k}{100} - 0,5 \right| < 0,1.$$

В этой задаче $p = q = 0,5$, $\epsilon = 0,1$; за величину N можно принять 200.

Нижняя граница вероятности указанного неравенства будет известна, если определим δ ; по формуле (10.4)

$$\delta = \frac{\frac{1}{2} \cdot \frac{1}{2}}{200 \cdot 0,1^2} = \frac{1}{8}.$$

Таким образом, бросив монету 200 раз, с вероятностью, большей $1 - \frac{1}{8} = 0,875$, можно ожидать, что

$$\left| \frac{k}{200} - 0,5 \right| < 0,1$$

или

$$80 < k < 120.$$

Пример 3. Монету бросают 900 раз; какие границы для числа повторений нужно указать, если требуется, чтобы это указание было дано с вероятностью, большей 0,99.

Здесь $p = \frac{1}{2}$, $q = \frac{1}{2}$, $1 - \delta = 0,99$; $\delta = 0,01$; за N следует принять 900.

По формуле (10.4) определим ϵ :

$$\epsilon^2 = \frac{pq}{N\delta} = \frac{1}{36}, \quad \epsilon = \frac{1}{6}.$$

Следовательно,

$$\left| \frac{k}{900} - 0,5 \right| < \frac{1}{6}$$

или

$$300 < k < 600.$$

Итак, с вероятностью, большей, чем 0,99, можно ожидать, что при 900 бросаниях монеты число выпадений надписи будет заключаться между 300 и 600.

Пример 4. Игральную кость бросают 1200 раз; оценить вероятность того, что число выпадений шестерки будет заключаться между 150 и 250.

В этом примере, как и в предыдущих, границы числа появлений события симметричны относительно математического ожидания числа повторений

события, которое в нашем случае равно $1200 \cdot \frac{1}{6} = 200$. Такое задание границ объясняется той формой, в какой доказывається теорема Бернулли:

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) > 1 - \delta.$$

По заданным границам определяем сначала ε :

$$\begin{aligned} -\varepsilon &< \frac{k}{1200} - \frac{1}{6} < +\varepsilon, \\ 200 - 1200\varepsilon &< k < 200 + 1200\varepsilon, \\ 200 - 1200\varepsilon &= 150; \quad \varepsilon = \frac{1}{24}. \end{aligned}$$

Проверим правую часть:

$$200 + 1200\varepsilon = 200 + 1200 \cdot \frac{1}{24} = 250.$$

Положим $N = 1200$; для оценки вероятности следует вычислить δ . По формуле (10.4) получаем

$$\delta = \frac{\frac{1}{6} \cdot \frac{5}{6}}{1200 \cdot \frac{1}{24^2}} = \frac{1}{15}.$$

Таким образом, с вероятностью, большей чем $\frac{14}{15}$, можно ожидать, что при 1200 бросаниях кости число выпадений шестерки будет заключаться между 150 и 250.

Заметим, что если для определения δ воспользоваться «более грубой» формулой (10.5):

$$N' = \frac{1}{4\delta\varepsilon^2},$$

то получим в данном случае значительно худший результат: $\delta = \frac{3}{25}$.

§ 48. Предельная теорема Лапласа

Теорема. Пусть производятся независимые по отношению к некоторому событию испытания, в каждом из которых вероятность этого события p постоянна. Если число испытаний n неограниченно растет, то вероятность того, что число k повторений события заключается между числами k_1 и k_2 ($k_1 < k_2$), имеет своим пределом при n , стремящемся к бесконечности, величину

$$\frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{t^2}{2}} dt,$$

где

$$z_1 = \frac{k_1 - np}{\sqrt{npq}}; \quad z_2 = \frac{k_2 - np}{\sqrt{npq}}$$

или, короче,

$$\lim_{n \rightarrow \infty} P(k_1 \leq k \leq k_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{t^2}{2}} dt, \quad (10.6)$$

Доказательство. Для приближенного вычисления вероятности числа k повторений события в § 35 была выведена формула (8.8),

$$P_{k,n} = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \Delta t,$$

где

$$t = \frac{k - np}{\sqrt{npq}}, \quad \Delta t = \frac{1}{\sqrt{npq}},$$

причем величина Δt есть приращение t , соответствующее двум значениям k , отличающимся на единицу. По теореме сложения вероятностей

$$P(k_1 \leq k \leq k_2) = \sum_{k=k_1}^{k_2} P_{k,n} \quad (10.7)$$

(например, $P(6 \leq k \leq 10) = P_{6,n} + P_{7,n} + P_{8,n} + P_{9,n} + P_{10,n}$). Обозначим через z_1 и z_2 значения z , соответствующие числам k_1 и k_2 :

$$z_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad z_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

Подставляя в (10.7) вместо всех чисел $P_{k,n}$ их приближенные значения по преобразованной формуле (8.8), получим приближенное равенство:

$$P(k_1 \leq k \leq k_2) = \sum_{t=z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \Delta t, \quad (10.8)$$

где суммирование распространяется на все значения t , соответствующие целым числам k от k_1 до k_2 . Из выражения для Δt сле-

дует, что при $n \rightarrow \infty$, $\Delta t \rightarrow 0$. Функция $e^{-\frac{t^2}{2}}$, принимающая при конечном n дискретные значения, непрерывна. Поэтому, заменяя в пределе сумму в (10.8) интегралом, мы получим точное равенство:

$$\lim_{n \rightarrow \infty} P(k_1 \leq k \leq k_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{t^2}{2}} dt.$$

Теперь разобьем интеграл, стоящий в полученной основной формуле, на два интеграла, изменив в первом пределы интегрирования. Тогда

$$\int_{z_1}^{z_2} e^{-\frac{t^2}{2}} dt = \int_{z_1}^0 e^{-\frac{t^2}{2}} dt + \int_0^{z_2} e^{-\frac{t^2}{2}} dt = \int_0^{z_2} e^{-\frac{t^2}{2}} dt - \int_0^{z_1} e^{-\frac{t^2}{2}} dt.$$

Введем функцию $\Phi(z)$, определяемую равенством

$$\Phi(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (10.9)$$

Теперь основную формулу можно переписать в виде

$$\lim_{n \rightarrow \infty} P(k_1 \leq k \leq k_2) = \Phi(z_2) - \Phi(z_1). \quad (10.10)$$

Значения функции $\Phi(z)$ при не очень больших z можно вычислять разложением интеграла (10.9) в ряд Маклорена и нахождением затем его частной суммы.

Функцию $\Phi(z)$ достаточно вычислять для положительных значений z , поскольку эта функция нечетна. Действительно, заменяя в интеграле (10.9) z на $-z$ и t на $-\tau$, получим

$$\Phi(-z) = \int_0^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = - \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau.$$

т. е.

$$\Phi(-z) = -\Phi(z). \quad (10.11)$$

В таблице III в конце книги приведены значения функции $\Phi(z)$ для z от 0,00 до 5,00.

Практически теорему Лапласа используют для нахождения приближенного значения указанной вероятности при больших n . В этом случае

$$P(k_1 \leq k \leq k_2) \approx \Phi(z_2) - \Phi(z_1),$$

где

$$z_1 = \frac{k_1 - np}{\sqrt{npq}},$$

$$z_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

Теорему Лапласа можно представить еще и в другом виде. Неравенства $k_1 \leq k \leq k_2$ эквивалентны неравенствам

$$\alpha_1 \leq \frac{k}{n} - p \leq \alpha_2,$$

где

$$\alpha_1 = \frac{k_1}{n} - p, \quad \alpha_2 = \frac{k_2}{n} - p.$$

Тогда

$$z_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{\alpha_1}{\sqrt{\frac{pq}{n}}}; \quad z_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{\alpha_2}{\sqrt{\frac{pq}{n}}}.$$

Числа α_1 и α_2 можно задать вместо чисел k_1 и k_2 ; они представляют границы, между которыми заключается отклонение относительной частоты $\frac{k}{n}$ от вероятности p и могут быть как отрицательными, так и положительными.

Вероятность неравенства не изменится, если это неравенство заменить эквивалентным ему неравенством. Поэтому можно написать:

$$\lim_{n \rightarrow \infty} P\left(\alpha_1 < \frac{k}{n} - p < \alpha_2\right) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{t^2}{2}} dt = \Phi(z_2) - \Phi(z_1),$$

причем

$$z_1 = \frac{\alpha_1}{\sqrt{\frac{pq}{n}}}, \quad z_2 = \frac{\alpha_2}{\sqrt{\frac{pq}{n}}}.$$

В частности, если $\alpha_1 < 0$, $\alpha_2 = -\alpha_1 = \alpha$, то в силу нечетности функции $\Phi(z)$

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{k}{n} - p\right| < \alpha\right) = 2\Phi\left(\frac{\alpha}{\sqrt{\frac{pq}{n}}}\right).$$

Если n — большое число, то можно написать приближенное равенство

$$P\left(\left|\frac{k}{n} - p\right| < \alpha\right) \approx 2\Phi\left(\frac{\alpha}{\sqrt{\frac{pq}{n}}}\right). \quad (10.12)$$

Функция $\Phi(z)$ принимает значения, мало отличающиеся от 0,5, когда z превышает 3; поэтому предел P близок к единице, если α превышает $3\sqrt{\frac{pq}{n}}$ ($\Phi(z) = 0,49865$, если $z = 3$). Пусть, например, $n = 100$, $p = q = 0,5$; тогда $3\sqrt{\frac{pq}{n}} = 0,15$, и мы можем на основании теоремы Лапласа сказать: почти наверно (т. е. с вероятностью, близкой к единице) абсолютное значение отклонения $\frac{k}{n}$ от p меньше 0,15 (утверждение верно только приблизительно). Хотя теорема Лапласа верна только в пределе, ею часто пользуются для

вычисления вероятности и при конечном значении n . Вполне удовлетворительные результаты получаются, если число npq порядка 100 или больше.

Пример 1. Игральная кость бросается 100 раз; вычислить вероятность того, что число появлений герба будет заключаться между 40 и 60.

Здесь $n = 100$, $p = q = 0,5$, $\sqrt{npq} = 5$, $k_1 = 40$, $k_2 = 60$, поэтому $z_1 = -2$, $z_2 = 2$. Получаем приближенно

$$P(40 < k < 60) = \Phi(2) - \Phi(-2) = 2\Phi(2).$$

По таблице находим $\Phi(2) = 0,4772$; поэтому

$$P(40 < k < 60) = 0,95.$$

Пример 2. Игральную кость бросают 2400 раз; определить вероятность того, что абсолютное значение разности между частотой и вероятностью выпадения шестерки не превышает $\frac{1}{24}$.

Здесь $p = \frac{1}{6}$, $q = \frac{5}{6}$, $npq = 333$,

$$-a_1 = a_2 = a = \frac{1}{24}, \quad -z_1 = z_2 = 2,74;$$

следовательно,

$$P\left(\left|\frac{k}{2400} - \frac{1}{6}\right| < \frac{1}{24}\right) \approx 2\Phi(2,74) = 0,994.$$

§ 49. Неравенства и теорема Чебышева

Рассмотрим n попарно независимых случайных величин $X^{(1)}$, $X^{(2)}$, ..., $X^{(n)}$ (здесь значки обозначают номера этих величин, а не значения, которые может принимать случайная величина); каждая из величин может принимать ряд значений, число которых может быть различным для различных величин.

Обозначим математические ожидания этих величин через \bar{x}_1 , \bar{x}_2 , ..., \bar{x}_n и их дисперсии σ_1^2 , σ_2^2 , ..., σ_n^2 . Рассмотрим сумму данных случайных величин

$$X = X^{(1)} + X^{(2)} + \dots + X^{(n)};$$

для ее математического ожидания и дисперсии введем обозначения;

$$\bar{x} = E(X),$$

$$\sigma^2 = E\{(X - \bar{x})^2\}.$$

По теоремам сложения математических ожиданий и дисперсий имеем

$$\bar{x} = \bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n,$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2.$$

Случайная величина $(X - \bar{x})^2$ принимает только положительные (или нулевые) значения, и ее математическое ожидание известно. Поэтому

к этой величине можно применить лемму Чебышева — Маркова:

$$P[(X - \bar{x})^2 < \sigma^2 t^2] > 1 - \frac{1}{t^2} *).$$

Неравенство, стоящее в прямых скобках, заменим эквивалентным

$$|X - \bar{x}| < \sigma t$$

(вероятность осуществления неравенства от этого не изменится). Тогда получим

$$P\{|X - \bar{x}| < t\sigma\} > 1 - \frac{1}{t^2}. \quad (10.13)$$

Это неравенство назовем *первым неравенством Чебышева*, если под X подразумевать одну случайную величину. Оно позволяет оценивать вероятность заданного отклонения от \bar{x} при любом законе распределения.

Заменяя в (10.13) X , \bar{x} , σ^2 их выражениями, получим

$$P\{|X^{(1)} + X^{(2)} + \dots + X^{(n)} - (\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n)| < t\sqrt{\sigma_1^2 + \dots + \sigma_n^2}\} > 1 - \frac{1}{t^2}. \quad (10.14)$$

Это неравенство назовем *вторым неравенством Чебышева*. Деля на n обе части неравенства, стоящего в левой части, получим второе неравенство Чебышева в другой форме:

$$P\left\{\left|\frac{X^{(1)} + X^{(2)} + \dots + X^{(n)}}{n} - \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n}\right| < \frac{t\sqrt{\sigma_1^2 + \dots + \sigma_n^2}}{n}\right\} > 1 - \frac{1}{t^2}. \quad (10.15)$$

Здесь $\frac{X^{(1)} + X^{(2)} + \dots + X^{(n)}}{n}$ есть среднее арифметическое случай-

ных значений данных величин, а $\frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n}$ — среднее арифметическое их математических ожиданий. Неравенство Чебышева, написанное во второй форме, оценивает вероятность того, что абсолютная величина отклонения первого среднего от второго среднего меньше величины, зависящей от суммы дисперсий данных величин.

Перейдем теперь к доказательству теоремы Чебышева.

Т е о р е м а. Если случайные величины $X^{(1)}, X^{(2)}, \dots, X^{(n)}$, число которых может быть как угодно большим, попарно независимы, если они имеют определенные математические ожидания $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ и определенные равномерно ограниченные дисперсии, то с вероятностью, как угодно близкой к единице, можно ожидать, что абсолютная величина разности между средним арифметическим данных

*) Напомним, что t^2 — произвольное положительное число.

величин и средним арифметическим их математических ожиданий как угодно мала по абсолютной величине, когда число слагаемых достаточно велико.

Обозначим через ε верхнюю границу абсолютной величины разности между средними арифметическими, указанными в формулировке теоремы, и будем считать ε сколь угодно малым положительным числом. Тогда теорема Чебышева может быть записана так:

Как бы ни были малы заданные положительные числа ε и δ , неравенство

$$P \left\{ \left| \frac{X^{(1)} + X^{(2)} + \dots + X^{(n)}}{a} - \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n} \right| < \varepsilon \right\} > 1 - \delta;$$

$$n > N(\varepsilon, \delta).$$

будет верным при всех значениях $n > N(\varepsilon, \delta)$, где N — определенное положительное число, которое можно вычислить по заданным ε и δ .

Доказательство. Напишем неравенство Чебышева во второй форме:

$$P \left\{ \left| \frac{X^{(1)} + X^{(2)} + \dots + X^{(n)}}{n} - \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n} \right| < \frac{t \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}}{n} \right\} > 1 - \frac{1}{t^2}.$$

Чтобы привести это неравенство в соответствие с тем, которое нужно доказать, выберем t так, чтобы

$$\frac{t \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}}{n} = \varepsilon; \quad t = \frac{n\varepsilon}{\sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}}.$$

При этом t имеем

$$P \left\{ \left| \frac{X^{(1)} + X^{(2)} + \dots + X^{(n)}}{n} - \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n} \right| < \varepsilon \right\} > 1 - \frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n^2 \varepsilon^2}.$$

По условию теоремы дисперсии равномерно ограничены некоторым числом B , т. е.

$$\sigma_1^2 < B, \quad \sigma_2^2 < B, \quad \dots, \quad \sigma_n^2 < B,$$

откуда

$$\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2 < nB.$$

Если мы в правой части неравенства для P заменим $\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$ на nB , то, очевидно, неравенство не нарушится. Следо-

вательно,

$$P\left\{\left|\frac{X^{(1)} + X^{(2)} + \dots + X^{(n)}}{n} - \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n}\right| < \varepsilon\right\} > 1 - \frac{B}{n\varepsilon^2}.$$

Выберем теперь число N так, чтобы удовлетворялось равенство $\frac{B}{N\varepsilon^2} = \delta$, т. е. $N = \frac{B}{\varepsilon^2\delta}$; тогда $\frac{B}{n\varepsilon^2} < \delta$, если $n > N$. Заменяя в правой части неравенства для P число $\frac{B}{n\varepsilon^2}$ на δ , мы не нарушим неравенства. Таким образом,

$$P\left\{\left|\frac{X^{(1)} + X^{(2)} + \dots + X^{(n)}}{n} - \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{n}\right| < \varepsilon\right\} > 1 - \delta, \quad (10.16)$$

если

$$n > N = \frac{B}{\varepsilon^2\delta}. \quad (10.17)$$

что и требовалось доказать.

Следствие. Если случайная величина X имеет определенное математическое ожидание a и определенную дисперсию σ^2 , и производится ряд независимых наблюдений для определения значений этой величины, то с вероятностью, сколь угодно мало отличающейся от единицы, можно ожидать, что при достаточно большом числе наблюдений среднее арифметическое из наблюдаемых значений как угодно мало отличается по абсолютной величине от математического ожидания этой величины.

Для доказательства отметим, что те случайные значения величины X , которые получаются при независимых наблюдениях, можно считать независимыми случайными величинами $\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}$. Так как каждая из них может иметь только те значения, которые имеет X , и притом с теми же вероятностями, то математическое ожидание и дисперсия каждой из этих величин такая же, как и для X . Употребляя обозначения теоремы Чебышева, можем написать

$$\begin{aligned} \bar{\xi}_1 &= \bar{\xi}_2 = \dots = \bar{\xi}_n = a, \\ \sigma_1^2 &= \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2. \end{aligned}$$

Отсюда вытекает, что выполняется важное условие ограниченности дисперсий, входившее в теорему Чебышева, причем за величину B можно принять σ^2 . Применяя к величинам $\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}$ теорему Чебышева, получим

$$P\left\{\left|\frac{\xi^{(1)} + \xi^{(2)} + \dots + \xi^{(n)}}{n} - a\right| < \varepsilon\right\} > 1 - \delta, \quad n > N_1 = \frac{\sigma^2}{\varepsilon^2\delta}. \quad (10.18)$$

что и доказывает следствие; достаточное число наблюдений есть N_1 .

Следствие из теоремы Чебышева имеет большое значение для приложений. Согласно ему при достаточно большом числе наблюдений можно ожидать с вероятностью, близкой к единице, т. е. «почти наверно», что среднее из наблюдаемых значений будет как угодно мало отличаться от математического ожидания, т. е. от теоретического ординатного значения.

Рассмотрим пример приложения следствия из теоремы Чебышева *).

Определенная, но неизвестная нам величина a (например, длина) измеряется n раз; вследствие случайных ошибок измерений мы получаем разные числа: a_1, a_2, \dots, a_n . Спрашивается, сколько измерений достаточно произвести, чтобы с вероятностью, большей, чем 0,99, можно было ожидать, что среднее из полученных чисел отличается от точного значения не больше, чем на 0,1.

Так как, по условию, a — определенная величина, то ее математическое ожидание равно a . Для решения задачи необходимо знать верхнюю границу дисперсии случайных измерений. В конкретных задачах такого типа, какой мы рассматриваем, она не может быть определена, так как неизвестны вероятности разных значений. Если же аналогичные измерения производились и раньше, то можно указать некоторую верхнюю границу дисперсии на основании результатов предыдущих измерений (этот вопрос будет рассмотрен в части IV). Мы просто предположим, что верхняя граница дисперсии задана и равна 0,02.

Так как в нашем случае $\epsilon = 0,1$, $\delta = 1 - 0,99$, $\sigma^2 = 0,02$, то

$$N_1 = \frac{0,02}{0,01 \cdot 0,01} = 200.$$

Таким образом, если число наблюдений превышает 200, то с вероятностью, большей 0,99, можно ожидать, что среднее арифметическое измеренных значений a_1, a_2, \dots, a_n отличается от a по абсолютной величине не больше, чем на 0,1.

§ 50. Замечания о законе больших чисел. Статистические вероятности

Кратко формулируя теорему Бернулли, можно сказать, что при достаточно большом числе наблюдений относительная частость будет мало отличаться от вероятности. Существенная особенность такого утверждения заключается в том, что не только вероятно, хотя вероятность и может сделаться как угодно близкой к единице.

Поэтому, если бы мы проделали опыты, описанные в рассмотренных выше примерах, то нельзя ожидать, что эти опыты наверно покажут согласие с результатом расчетов. Рассмотрим для иллюстрации пример 2, § 47. Результаты вычислений показали, что с вероятностью, большей 0,875, можно ожидать, что число выпадений надписи будет заключаться между 30 и 120. Отдельный опыт (мо-

*) Следует отметить, что такого рода примеры служат в основном для уяснения содержания теоремы. Оценки достаточного числа наблюдений по этой теореме получаются преувеличенными, и, кроме того, условия независимости наблюдений не всегда можно считать выполненными.

нету бросают 200 раз) может дать число выпадений надписи, выходящее из этих границ, тем более, что полученная граница вероятности не очень близка к единице. Чтобы выяснить практический смысл результата решения этого примера, надо представить себе, что производится большой ряд таких опытов (всякий опыт заключается в том, что монету бросают 200 раз). Применяя вторично теорему Бернулли, мы можем ожидать с вероятностью, близкой к единице, что в большинстве из этих опытов число выпадений надписи будет заключаться в указанных границах (грубо говоря, наше предсказание о границах будет выполняться в среднем в каждом семи опытах из восьми).

Значительное отклонение результатов действительно произведенных опытов от указанных может свидетельствовать о невыполнении в рассматриваемом явлении условий теоремы Бернулли, из которых наиболее жестким является требование независимости опытов.

В теореме Бернулли вероятность события считалась известной, и речь шла о предсказании (с некоторой вероятностью) границ, между которыми будет заключаться число повторений события. Уже и в этом виде теорема указывает на возможность получения неизвестной вероятности случайного события из большого количества наблюдений над появлением и непоявлением событий.

Такая возможность вытекает из *обратной теоремы Бернулли*: если можно считать, что в неограниченном ряде независимых опытов событие имеет постоянную вероятность (неизвестную нам), то с вероятностью, как угодно близкой к единице, можно ожидать, что отношение числа фактических появлений событий к числу наблюдений будет как угодно мало отличаться от вероятности, если число наблюдений достаточно велико. Поэтому

$$P\left\{\left|p - \frac{k}{n}\right| < \varepsilon\right\} > 1 - \delta \quad (10.19)$$

при $n > N(\varepsilon, \delta)$, причем N можно определить по заданным ε и δ . При больших n получаем

$$p \approx \frac{k}{n}. \quad (10.20)$$

Вычисленные на основании результатов наблюдений вероятности называют *статистическими* или *эмпирическими*.

Теорему Чебышева можно сформулировать так: если число случайных величин достаточно велико, то среднее арифметическое случайных значений этих величин мало отличается от среднего арифметического их математических ожиданий. Аналогично можно сформулировать и следствие из теоремы Чебышева. Эти формулировки показывают, что при достаточно большом числе независимых наблюдений можно с вероятностью, близкой к единице, получить из наблюдений приближенную величину среднего арифметического из математических ожиданий или математического ожидания одной величины.

Следует отметить, что те оценки необходимого числа наблюдений, какие получаются при применении теорем Бернулли и Чебышева и вытекают из способов доказательств этих теорем, сильно преувеличены. В работах С. Н. Бернштейна, А. Я. Хинчина и других указаны некоторые способы уточнения оценок.

Применение рассмотренных теорем закона больших чисел на практике ограничено условием независимости опытов или случайных величин. В ряде работ по теории вероятностей было показано, что снятие этого ограничения допустимо, если связь между изучаемыми случайными величинами слаба; были установлены также некоторые условия применимости закона больших чисел к связанным случайным величинам, но этот вопрос не изучен в достаточной степени.

Теорему Лапласа обычно называют *предельной теоремой*, так как в ней устанавливается предел, к которому стремится вероятность того, что число повторений заключается между заданными границами или (в другой форме) вероятность того, что отклонение частоты от вероятности заключено в данных границах. Чтобы избежать преувеличения оценок достаточного числа наблюдений, какие получаются по теореме Бернулли, нередко пользуются теоремой Лапласа. Надо, однако, заметить, что это дает менее определенные результаты, чем применение теоремы Бернулли, так как неизвестна оценка ошибки, которую мы совершаем, заменяя формулу, верную в пределе при $n \rightarrow \infty$, такой же формулой при определенном значении n .

Для проверки закона больших чисел неоднократно производились различные опыты (например, бросание монеты). Результаты этих опытов показывают, что те «предсказания», какие можно делать на основании теоремы, оказываются верными почти во всех опытах (опыты, конечно, ставились так, чтобы выполнялись условия теорем).

Вопрос о приложимости закона больших чисел к явлениям природы довольно сложен. Утверждения, подобные тем, какие делались нами в некоторых примерах, должны даваться с надлежащими оговорками и проверяться систематическими наблюдениями.

ГЛАВА II

НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 51. Функция распределения непрерывной случайной величины

В этой главе мы рассмотрим такие случайные величины, которые могут принимать любое значение в некоторой области. Будем обозначать через X непрерывную случайную величину, принимающую любое вещественное значение в области от a до b . В этом случае нельзя задать вероятности отдельных значений и нельзя ставить вопрос о вероятности определенных значений, так как их бесконечное множество. Для непрерывных случайных величин имеет смысл только вопрос о вероятности принять значение в некотором интервале, который может включать в себя область задания, а иногда может и частично перекрывать область (a, b) . Чаше всего ставится задача — вычислить вероятность того, что случайная величина X , заданная в области (a, b) , примет какое-нибудь значение в интервале (α, β) , составляющем часть области (a, b) . Эту задачу будем считать основной. Область (a, b) , в частности, может простираться от 0 до ∞ или от $-\infty$ до $+\infty$.

Для решения задачи о вычислении вероятности того, что случайная величина примет какое-либо значение в заданном интервале в случае дискретной величины, нужно было задать таблицу значений величины и вероятностей этих значений. Если величина непрерывна, то вместо таблицы (распределения вероятностей) нужно задать некоторые функции (функции распределения).

Для дальнейшего представим основную задачу графически. Пусть отрезок AB на числовой оси изображает область возможных значений случайной величины. Событию — принятию случайной величиной X некоторого значения x — поставим в соответствие попадание в точку x при случайном бросании точки на прямую. Если α, β изображает заданный интервал, то графически основная задача состоит в вычислении вероятности попасть на отрезок $\alpha\beta$ (рис. 5). Такая вероятность есть функция координат концов отрезка α и β , но задание такой функции двух аргументов излишне усложнило бы теорию. Вместо этого задают функцию распределения, которую можно определить так:

Функцией распределения непрерывной случайной величины называется функция одного аргумента, значение которой равно вероятности того, что случайная величина примет значение, меньшее аргумента функции распределения *). Обозначим функцию распределения $F(x)$; тогда по определению имеем равенство

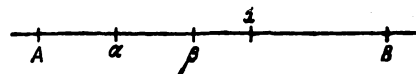


Рис. 5. К определению основной задачи для непрерывных случайных величин.

$$P(X < x) = F(x), \quad (11.1)$$

в котором левая часть читается так: вероятность того, что случайная величина X примет значение, меньшее, чем x . Из определения вытекают следующие свойства функции распределения $F(x)$:

$$\left. \begin{aligned} F(x) &= 0, & \text{если } x \leq a, \\ F(x) &= 1, & \text{если } x \geq b. \end{aligned} \right\} \quad (11.2)$$

Если x монотонно растет от a до b , то $F(x)$ также монотонно растет от 0 до 1, так как увеличение x расширяет область возможных значений величины, что увеличивает вероятность попадания.

На основании этих свойств можно утверждать, что схематический график функции распределения имеет вид, изображенный на рис. 6.

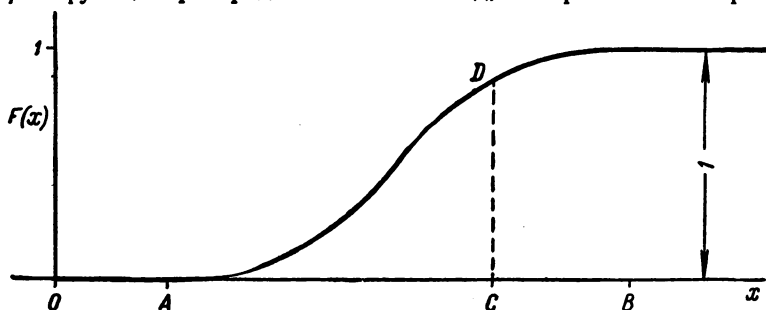


Рис. 6. График функции распределения непрерывной случайной величины. $OA = a$, $OC = x$, $OB = b$, $CD = P(X < x)$.

Слева от области задания — прямая линия, совпадающая с осью x , между a и b — монотонно поднимающаяся кривая с ординатами от 0 до 1; справа от точки B — прямая, параллельная оси абсцисс, отстоящая на расстоянии единицы от нее **).

Чтобы доказать целесообразность задания функции распределения, покажем, что знание ее позволяет легко решить сформулированную выше основную задачу.

*) Заметим, что понятие функции распределения без изменения можно распространить и на дискретные величины.

**) Нетрудно установить, что для дискретной случайной величины график функции распределения есть ступенчатая линия.

Пусть точки α и β ограничивают область, для которой требуется определить вероятность попасть в нее. Тогда

$$F(\beta) = P(X < \beta), \quad F(\alpha) = P(X < \alpha).$$

Так как область от α до β состоит из двух непересекающихся частей — от α до α и от α до β , то случайное событие (попадание левее точки β) имеет два несовместных вида — попадание левее α и попадание между α и β . По теореме сложения вероятностей можно написать:

$$P(X < \alpha) + P(\alpha < X < \beta) = P(X < \beta).$$

Поэтому

$$P(\alpha < X < \beta) = P(X < \beta) - P(X < \alpha)$$

или

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha). \quad (11.3)$$

Последнее равенство показывает, что знание функции распределения непрерывной случайной величины дает весьма простой способ решения основной задачи: *вероятность принять значение в некотором интервале равна разности значений функции распределения в верхней и нижней границах интервала.*

Одна из задач учения о случайных величинах заключается в выборе суммарных числовых характеристик величины и в установлении способа их вычисления. С понятием о функции распределения тесно связана одна из таких характеристик — медиана.

Медианой случайной величины называется ее значение, определяемое условием: значения величины, большие и меньшие, чем медиана, равновероятны, т. е.

$$P(X < m) = P(X > m) = \frac{1}{2}. \quad (11.4)$$

Если задана функция распределения $F(x)$, то медиана определяется уравнением $F(x) = 0,5$, которое имеет единственное решение вследствие монотонности $F(x)$. На графике $F(x)$ медиана определяется очевидным способом: проводим прямую, параллельную оси абсцисс на расстоянии 0,5 от нее, и определяем абсциссу точки пересечения ее с кривой.

§ 52. Плотность вероятности

Пусть для случайной величины X , определенной в области (a, b) , известна функция распределения $F(x)$. Вычисляя вероятность случайной величины попасть в часть области от x до $x + \Delta x$ ($\Delta x > 0$), получим

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x).$$

Разделим обе части на Δx :

$$\frac{P(x < X < x + \Delta x)}{\Delta x} = \frac{F(x + \Delta x) - F(x)}{\Delta x}. \quad (11.5)$$

Величина, стоящая слева, есть отношение вероятности попасть на отрезок длиной Δx к этой длине; такую величину можно назвать *средней плотностью вероятности* на отрезке Δx с началом в точке x . Величина справа есть отношение соответствующего приращения функции распределения к приращению аргумента. Перейдем в равенстве к пределу, полагая, что $\Delta x \rightarrow 0$ и что $F(x)$ — дифференцируемая функция. Предел левой части можно назвать *плотностью вероятности* в точке x , по аналогии с понятием о линейной плотности в точке; предел правой части равен производной функции распределения. Обозначая плотность вероятности через $p(x)$, имеем простую связь между плотностью вероятности и функцией распределения: *плотность вероятности равна производной от функции распределения по аргументу распределения* *):

$$p(x) = \frac{dF(x)}{dx}. \quad (11.6)$$

Отметим некоторые свойства плотности вероятности:

I. Плотность вероятности есть неотрицательная величина при всех значениях аргумента.

Это следует из того, что $p(x)$ есть производная нигде не убывающей функции $F(x)$.

II. Если случайная величина задана в конечной области от a до b , то $p(x) = 0$, если $x < a$ или $x > b$.

Действительно, значения x вне области (a, b) невозможны, вероятность попасть на любой отрезок вне области равна нулю, а потому и плотность вероятности равна нулю.

На основании этого свойства можно формально всегда считать, что случайная величина задана на всей числовой прямой от $-\infty$ до $+\infty$, но $p(x)$ имеет тройное «кусочное» определение: $p(x)$ равно нулю налево от a , положительно от a до b , и равно нулю направо от b .

III. Значения $p(x)$ на границах области и в области могут быть какими угодно, в частности и бесконечными:

Это следует из того, что производная монотонно возрастающей в области функции $F(x)$ ничем не ограничена. В связи с этим следует помнить, что не нужно смешивать плотность вероятности с вероятностью, значения которой не превышают единицы.

IV. Из определения плотности вероятности вытекает приближенное равенство

$$P(x < X < x + \Delta x) \approx p(x) \Delta x, \quad (11.7)$$

где Δx — достаточно малое положительное число.

*) В силу именно этого равенства плотность вероятности иногда называют *дифференциальной функцией распределения*, а функцию распределения, определенную выше, *интегральной функцией распределения*.

Это равенство часто записывается в выкладке как точное. Такая запись не вносит ошибки, если после этого по смыслу задачи должен быть осуществлен переход к пределу при $\Delta x \rightarrow 0$.

V. Решение основной задачи дается формулой

$$P(a < X < b) = \int_a^b p(x) dx. \quad (11.8)$$

Эта формула очевидна, если вспомнить решение основной задачи с помощью функции распределения $F(x)$ и учесть связь между $p(x)$ и $F(x)$ [$F(x)$ — первообразная по отношению к $p(x)$].

VI. Если случайная величина задана в конечной области от a до b , то

$$\int_a^b p(x) dx = 1. \quad (11.9)$$

т. е. площадь, заключенная между осью абсцисс и графиком плотности вероятности, равна 1.

Действительно, по формуле (11.8) левая часть есть вероятность попасть в область от a до b , а по условию величина может иметь значения только в этой области; значит, вероятность попасть в область (a, b) равна единице. Написанное выше равенство иногда называют *условием нормирования плотности вероятности*. Это значит, что на плотность вероятности должно быть наложено условие (11.9), если эта функция выбирается до некоторой степени произвольно.

График плотности вероятности обычно называют *кривой распределения*. Кривая распределения может иметь один или несколько максимумов в разных местах области. Значение случайной величины, соответствующее наибольшей ординате этой кривой, называется *модой*.

§ 53. Математическое ожидание, дисперсия и моменты

В главе 9 математическое ожидание случайной величины было определено только для случайных величин, которые могут принимать лишь конечное число дискретных значений (его можно распространить и на величины, имеющие бесконечное число дискретных значений).

Введение понятия о функциях распределения непрерывных случайных величин дает возможность ввести понятие *математического ожидания непрерывной случайной величины*.

Определение 1. Если X — непрерывная случайная величина, принимающая значения x в промежутке от a до b ($a < b$), и $p(x)$ — плотность вероятности этой величины, то ее математическим

ожиданием называется выражение:

$$E(X) = \bar{x} = \int_a^b xp(x) dx. \quad (11.10)$$

Можно считать, что формула (11.10) определяет *теоретическое среднее значение* непрерывной случайной величины.

Определение 2. Математическое ожидание случайной величины называется *центром* ее распределения.

Точка на оси абсцисс на графике плотности вероятности, имеющая абсциссой математическое ожидание величины, также называется центром распределения. (Такое название объясняется тем, что центр тяжести неоднородного отрезка, абсциссы концов которого равны a и b , определяется по такой же формуле, как и математическое ожидание, если плотность в точке равна $p(x)$.)

Определение 3. Математическое ожидание произвольной однозначной непрерывной функции $\varphi(X)$ непрерывной случайной величины X , для которой плотность вероятности есть $p(x)$, определяется формулой

$$E\{\varphi(X)\} = \int_a^b \varphi(x)p(x) dx. \quad (11.11)$$

Допустимость такого определения вытекает из того, что вероятность для $\varphi(X)$ принять значения, заключенные между $\varphi(x_1)$ и $\varphi(x_2)$, равна вероятности того, что X примет значение между x_1 и x_2 , каковы бы ни были x_1 и x_2 .

Формула (11.11) определяет *теоретическое среднее значение* функции $\varphi(X)$.

Все свойства математических ожиданий, выведенные для дискретных величин, верны и для непрерывных величин.

Перейдем к определению понятия *моментов* случайной величины.

Определение 4. *Начальным моментом порядка s* случайной величины X , определенной в области от a до b и имеющей плотность вероятности $p(x)$, называется число

$$\nu_s = \int_a^b x^s p(x) dx, \quad (11.12)$$

т. е. математическое ожидание s -й степени случайной величины (s — целое положительное число).

В силу условия нормирования и определения математического ожидания

$$\begin{aligned} \nu_0 &= 1, \\ \nu_1 &= E(x) = \bar{x}. \end{aligned}$$

Определение 5. *Центральным моментом порядка s случайной величины X с плотностью вероятности $p(x)$ называется число*

$$\mu_s = \int_a^b (x - \bar{x})^s p(x) dx, \quad (11.13)$$

т. е. математическое ожидание s -й степени отклонения величины от ее математического ожидания.

Из определения следует, что

$$\mu_0 = 1, \quad (11.14)$$

$$\mu_1 = 0, \quad (11.15)$$

так как

$$\mu_0 = \int_a^b p(x) dx = \nu_0 = 1,$$

$$\mu_1 = \int_a^b x p(x) dx - \bar{x} \int_a^b p(x) dx = \bar{x} - \bar{x} \cdot 1 = 0.$$

Далее,

$$\mu_2 = \int_a^b (x - \bar{x})^2 p(x) dx = \sigma_x^2, \quad (11.16)$$

т. е. центральный момент второго порядка есть математическое ожидание квадрата отклонения величины от ее математического ожидания. По определению, которое было дано в главе о дискретных величинах, такое математическое ожидание называется *дисперсией*. Распространяя это определение на непрерывные величины, мы можем сказать, что *центральный момент второго порядка есть дисперсия непрерывной случайной величины*. Выведенные в § 44 свойства дисперсии случайной величины остаются справедливыми и для непрерывных величин, ибо мы при доказательстве этих свойств фактически не пользовались дискретностью величин. Выведем формулы, выражающие центральные моменты через начальные. Для этого множитель $(x - \bar{x})^s$ разложим по биному Ньютона; получим

$$\begin{aligned} \mu_s = & \int_a^b x^s p(x) dx - s\bar{x} \int_a^b x^{s-1} p(x) dx + \\ & + \frac{s(s-1)}{1 \cdot 2} \bar{x}^2 \int_a^b x^{s-2} p(x) dx + \dots + (-1)^k C_s^k \bar{x}^k \int_a^b x^{s-k} p(x) dx + \dots \\ & \dots + \frac{s(s-1)}{1 \cdot 2} \bar{x}^{s-2} (-1)^{s-2} \int_a^b x^2 p(x) dx + \\ & + (-1)^{s-1} s \bar{x}^{s-1} \int_a^b x p(x) dx + (-1)^s \bar{x}^s \int_a^b p(x) dx. \end{aligned}$$

Все интегралы равны начальным моментам:

$$\nu_s, \nu_{s-1}, \dots, \nu_2, \nu_1, \nu_0.$$

Как указано выше, $\nu_0 = 1$, $\nu_1 = \bar{x}$, поэтому последние два члена приводятся и, следовательно,

$$\mu_s = \sum_{k=0}^{s-2} (-1)^k C_s^k \bar{x}^k \nu_{s-k} + (-1)^{s-1} (s-1) \bar{x}^s. \quad (11.17)$$

В частности,

$$\left. \begin{aligned} \mu_2 &= \nu_2 - \bar{x}^2, \\ \mu_3 &= \nu_3 - 3\nu_2 \bar{x} + 2\bar{x}^3, \\ \mu_4 &= \nu_4 - 4\nu_3 \bar{x} + 6\nu_2 \bar{x}^2 - 3\bar{x}^4. \end{aligned} \right\} \quad (11.18)$$

Моментами можно было бы пользоваться как суммарными (сводными) числовыми характеристиками случайной величины, но размерность момента порядка s равна размерности X в степени s , что неудобно. Более естественно считать числовыми характеристиками числа $\sqrt[s]{\mu_s}$, в частности $\sigma = \sqrt{\mu_2}$; использование именно центральных моментов исключает влияние начала отсчета величины X , но не исключает влияния масштаба при измерении этой величины. По этой причине за *моментные числовые характеристики* следует принять числа

$$m_s = \frac{\sqrt[s]{\mu_s}}{\sigma} \quad \text{при } s > 2.$$

Числа m_s не зависят от способа измерения величины x , т. е. не зависят ни от выбора начала отсчета, ни от выбора единиц. Это отвлеченные числа, и их можно использовать для сравнения разнообразных случайных величин между собою.

Моментные характеристики дополняют основные характеристики, зависящие и от начала отсчета, и от масштаба: среднее значение \bar{x} , медиану m и среднее квадратичное отклонение σ_x . Малость числа m_s при нечетных s может служить указанием на то, что распределение (плотность вероятности) близко к симметрии относительно центра распределения.

§ 54. Равномерное распределение вероятностей

Для иллюстрации общих положений рассмотрим наиболее простой случай, когда плотность вероятности в промежутке от a до b есть величина постоянная. В этом случае график плотности вероятности есть отрезок, параллельный оси абсцисс (рис. 7, а). Чтобы площадь, ограниченная всей кривой распределения, равнялась единице (условие нормирования), необходимо, чтобы эта постоянная равнялась $\frac{1}{b-a}$.

Функцию распределения $F(x)$ можно найти интегрированием плотности вероятности при начальном условии $F(a) = 0$:

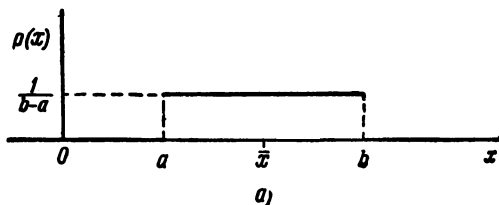
$$F(x) = \int_a^x \frac{dt}{b-a}.$$

отсюда

$$F(x) = \frac{x-a}{b-a}.$$

График функции распределения есть отрезок прямой, пересекающей ось абсцисс в точке $X=a$ и имеющей ординату, равную единице при $X=b$ (рис. 7, б). Вероятность того, что X имеет значение между α и β , получим по формуле

$$P(\alpha < X < \beta) = \frac{\beta - \alpha}{b - a}.$$



Найдем математическое ожидание

$$\bar{x} = \int_a^b \frac{x dx}{b-a}$$

или

$$\bar{x} = \frac{b+a}{2}.$$

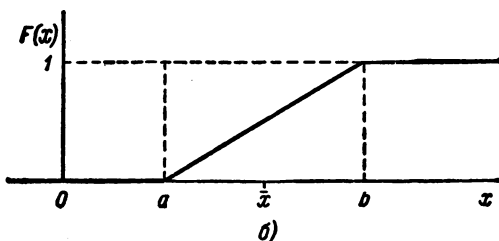


Рис. 7. Случай равномерного распределения: а) график плотности вероятности; б) график функции распределения.

Этот результат можно считать очевидным; распределение симметрично, поэтому центр распределения должен быть в геометрическом центре области. Медиана для равномерного распределения совпадает с математическим ожиданием.

Перенеся начало отсчета равномерно распределенной случайной величины в центр распределения, получим новую величину

$$Y = X - \frac{b+a}{2} = X - \bar{x}.$$

Очевидно,

$$\bar{y} = \bar{x} - \bar{x} = 0.$$

Плотность вероятности Y можно записать в форме

$$p(y) = \frac{1}{2c},$$

где $2c = b - a$ — размер области.

Дисперсия равномерно распределенной величины Y равна дисперсии величины X , так как обе величины отличаются только перемещением начала отсчета, что не влияет на дисперсию. Найдем эту дисперсию

$$D(X) = D(Y) = \int_{-c}^c \frac{y^2 dy}{2c} = \frac{c^2}{3}.$$

$$\sigma_{\omega} = \frac{c}{\sqrt{3}} = 0,58c.$$

В равномерном распределении величина c есть модуль предельного отклонения от центра; он связан с σ_{ω} равенством $c = \sigma_{\omega} \sqrt{3}$.

§ 55. Формулировка теоремы Ляпунова. Нормальное распределение вероятностей

При изучении явлений природы приходится иметь дело со случайными величинами, которые можно рассматривать как суммы большого числа независимых случайных слагаемых. Предельная теорема Ляпунова характеризует (при некоторых ограничениях) закон распределения таких сумм. Эта теорема формулируется следующим образом.

Пусть

$$Z = X^{(1)} + X^{(2)} + \dots + X^{(n)},$$

где $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ — взаимно независимые случайные величины, имеющие определенные математические ожидания, дисперсии и «абсолютные» центральные моменты порядка $2 + \alpha$:

$$E(X^{(k)}) = \bar{x}_k, \quad D(X^{(k)}) = \sigma_k^2, \quad E(|X^{(k)} - \bar{x}_k|^{2+\alpha}) = \gamma_k,$$

где α — некоторое положительное число, причем выполняется условие

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \gamma_k}{\sigma^{2+\alpha}} = 0.$$

Тогда вероятность неравенства

$$t_1 \sigma < Z - \bar{z} < t_2 \sigma, \quad (11.19)$$

в котором $\bar{z} = \sum_{k=1}^n \bar{x}_k$, $\sigma^2 = \sum_{k=1}^n \sigma_k^2$, при неограниченном возрастании n имеет пределом величину

$$\frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{\tau^2}{2}} d\tau. \quad (11.20)$$

Утверждение этой теоремы можно приближенно сформулировать и несколько иначе: при достаточно большом числе n случайных слагаемых вероятность неравенства $\alpha < Z < \beta$ сколь угодно близка к величине

$$\frac{1}{\sigma \sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}} dx. \quad (11.20a)$$

Из этой формулировки предельной теоремы Ляпунова видно, что при неограниченном увеличении числа слагаемых случайных величин плотность вероятности суммы переходит в пределе в функцию

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad (11.21)$$

если только слагаемые удовлетворяют условиям теоремы. Закон распределения с плотностью вероятности, определяемой формулой (11.21), называется *нормальным* (говорят, что случайная величина распределена нормально). Числа a и σ следует считать *параметрами* закона.

Аналогичный закон распределения дает приближенная формула Лапласа для задачи о повторении испытаний. В теории и практике случайных величин пользовались нормальным законом, распространяя закон, выведенный для случайной величины определенного типа (число повторений) на произвольные случайные величины. Теорема Ляпунова выясняет условия применимости нормального закона распределения: он пригоден в тех задачах, в которых изучаемую случайную величину можно считать суммой большого числа случайных слагаемых; существенно при этом, что слагаемые могут подчиняться любым законам распределения.

Дополнительное условие относительно существования чисел γ_k и стремления к нулю отношения суммы этих чисел к степени величины σ обозначает, согласно акад. С. Н. Бернштейну, что среди слагаемых не должно быть сильно выделяющихся по величинам и по значениям дисперсий. Делались опыты построения закона распределения суммы равномерно распределенных слагаемых. Эти опыты показали, что уже при числе слагаемых порядка 20 получается хорошее приближение к точному нормальному закону.

§ 56. Приближенный вывод нормального закона

Доказательство теоремы Ляпунова слишком громоздко для того, чтобы воспроизводить его в этом курсе. Поэтому мы рассмотрим упрощенный вывод нормального закона, предложенный Пирсоном. Введем следующие условия. 1) значения случайной величины X представляют собой отклонения от некоторой постоянной величины a ; 2) каждое из отдельных отклонений есть результат случайного действия большого числа причин, каждая из которых вызывает малое случайное отклонение; 3) указанные причины действуют независимо.

Эти условия были впервые введены в теории ошибок измерений, где измеряемая величина имеет какое-то вполне определенное числовое значение, а те разные числа, какие мы получаем при измерениях, появляются благодаря действию разных случайных причин (случайная струя воздуха, толчок от стука дверью вдали от лаборатории и т. п.). Разница между наблюдаемым и истинным (неизвестным нам) числом есть случайная ошибка наблюдений.

Когда речь идет о распределении такой величины, как наклонности совокупности астероидов и т. п., можно говорить об отклонении от некоторой постоянной величины, но эти отклонения нельзя назвать ошибками. Второе из условий в этом случае менее оправдано, так как отклонение отдельного значения от какого-то постоянного уровня является результатом действия не только случайных причин, но и систематических (например, возмущения от больших планет в примере с астероидами). Поэтому второе условие в этом случае следует считать только упрощением при построении теоретической схемы.

Указанным условиям мы дадим более четкую формулировку в виде следующих предположений:

1) отклонение U величины X от a вызывается действием n причин, каждая из которых обязательно вызывает малое по абсолютной величине случайное отклонение $+\epsilon$ или $-\epsilon$;

2) одинаковые по абсолютной величине элементарные отклонения одинаково вероятны:

$$P(\epsilon) = P(-\epsilon) = \frac{1}{2};$$

3) причины, вызывающие элементарные отклонения, взаимно независимы, т. е. вероятность того, что какая-нибудь причина даст отклонение $+\epsilon$ (или $-\epsilon$), не зависит от того, какие отклонения вызывают фактически остальные причины.

Отметим, что предположения 2) и 3) налагают довольно существенное ограничение на схему; впрочем, их можно считать достаточно близкими к действительности в теории ошибок. Ограничение же, что модуль отклонения один и тот же для всех величин, не является существенным, оно вводится лишь для упрощения дальнейших выкладок.

Исходя из принятых трех предположений, приступим к выводу плотности вероятности величины U .

Пусть k причин дали положительные отклонения ϵ , а остальные $n-k$ причин — отрицательные. Тогда значение суммарного отклонения U , которое мы обозначим через u_k , будет

$$u_k = k\epsilon + (n-k)(-\epsilon) = (2k-n)\epsilon.$$

Вероятность такого значения вычисляется по формуле (8.1) задачи о повторении испытаний, так как действие каждой из причин можно уподобить испытанию, в котором событие (отклонение $+\epsilon$) может произойти или не произойти. Согласно второму предположению $p = q = \frac{1}{2}$. Поэтому

$$P_{k,n} = \frac{n!}{k!(n-k)!} \cdot \left(\frac{1}{2}\right)^n.$$

Так как наша задача заключается в установлении связи между величиной U и плотностью вероятности, то предположим теперь, что не k , а $k+1$ причина дали значения $+\epsilon$. Тогда

$$u_{k+1} = (2k+2-n)\epsilon, \quad P_{k+1,n} = \frac{n!}{(k+1)!(n-k-1)!} \left(\frac{1}{2}\right)^n.$$

Длину интервала (u_k, u_{k+1}) обозначим через Δu , а его середину — буквой u . Согласно предыдущим формулам

$$\Delta u = u_{k+1} - u_k = 2\varepsilon,$$

$$u = \frac{u_{k+1} + u_k}{2} = (2k - n + 1)\varepsilon.$$

Условимся считать вероятность того, что случайная величина U принимает значения в интервале (u_k, u_{k+1}) , приближенно равной среднему из $P_{k,n}$ и $P_{k+1,n}$. Тогда, обозначая через y плотность вероятности величины U , отнесенную к середине u указанного интервала, будем иметь

$$y = \frac{P}{\Delta u},$$

где

$$P = \frac{P_{k,n} + P_{k+1,n}}{2}, \quad \Delta u = u_{k+1} - u_k.$$

Отсюда

$$\Delta y = \frac{\Delta P}{\Delta u},$$

где

$$\Delta P = P_{k+1,n} - P_{k,n}$$

и, следовательно,

$$\frac{\Delta y}{y} = \frac{\Delta P}{P}.$$

Вычисляя ΔP и P , имеем

$$\Delta P = \left(\frac{1}{2}\right)^n \frac{(n-2k-1)! n!}{(k+1)!(n-1)!},$$

$$P = \left(\frac{1}{2}\right)^n \frac{(n+1)! n!}{(k+1)!(n-k)!}.$$

Поэтому

$$\frac{\Delta y}{y} = \frac{2(n-2k-1)}{n+1} = \frac{(n-2k-1)\varepsilon \cdot 2\varepsilon}{(n+1)\varepsilon^2}.$$

Для получения плотности вероятности y надо найти связь между y и u . Так как $(n-2k-1)\varepsilon = -u$, $2\varepsilon = \Delta u$, то

$$\frac{\Delta y}{y} = -\frac{u \Delta u}{(n+1)\varepsilon^2}.$$

При конечном n последнее равенство дает приближенную связь между величиной U , плотностью вероятности y и их приращениями. Положим теперь, что $n \rightarrow \infty$; тогда естественно считать, что $\varepsilon \rightarrow 0$, поскольку иначе были бы допустимы бесконечно большие отклонения, вероятности которых не были бы бесконечно малы.

Мы предположим, кроме того, что $(n+1)\varepsilon^2$ стремится к конечному, не равному нулю пределу σ^2 . В пределе получим дифференциальное уравнение кривой плотности вероятности

$$\frac{dy}{y} = -\frac{u du}{\sigma^2}.$$

Интегрирование этого уравнения дает

$$y = C e^{-\frac{u^2}{2\sigma^2}}.$$

где C — произвольная постоянная, а σ — параметр. Плотность вероятности должна удовлетворять условию нормирования; из полученного уравнения следует, что u не ограничено; поэтому имеем

$$\int_{-\infty}^{\infty} C e^{-\frac{u^2}{2\sigma^2}} du = 1.$$

Это уравнение дает возможность выразить C через σ . Делая подстановку $\frac{u}{\sigma} = t$, получим

$$C\sigma \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1$$

Из курса анализа известно, что

$$\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi},$$

значит,

$$C = \frac{1}{\sigma\sqrt{2\pi}}.$$

Окончательно плотность вероятности нормального распределения запишется так:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}. \quad (11.22)$$

В этом уравнении u есть отклонение значения случайной величины от некоторого постоянного числа a . Если вместо отклонений введем в уравнение значение самой величины x ($x = u + a$), то получим плотность вероятности нормального распределения в общем виде:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (11.23)$$

§ 57. Параметры нормального закона. Кривая Гаусса

Чтобы выяснить смысл параметров a и σ нормального закона, найдем математическое ожидание и дисперсию случайной величины, подчиненной этому закону.

По формуле (11.10) получаем

$$E(X) = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Сделаем замену переменной под знаком интеграла

$$\frac{x-a}{\sigma} = t, \quad dx = \sigma dt.$$

Тогда получим

$$E(X) = \int_{-\infty}^{+\infty} \frac{\sigma t}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + a \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

В первом члене под знаком интеграла стоит нечетная функция аргумента t ; как известно из курса анализа, интеграл от такой функции в пределах от $-\infty$ до $+\infty$ равен нулю. Интеграл во втором члене выражает условие нормирования нормального закона частного вида ($a=0$, $\sigma=1$), поэтому этот интеграл равен единице. Следовательно,

$$E(X) = \bar{x} = a.$$

Таким образом, параметр a есть среднее значение случайной величины X .

Найдем теперь дисперсию нормально распределенной величины. По формуле (11.16) находим

$$D(X) = \int_{-\infty}^{\infty} \frac{(x-a)^2}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Заменяя опять x на $\frac{x-a}{\sigma} = t$, получим

$$D(X) = \sigma^2 \int_{-\infty}^{\infty} \frac{t^2}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

В известной формуле интегрирования по частям

$$\int_a^b u dv = uv \Big|_a^b - \int_a^b v du$$

положим

$$u = \frac{t}{\sqrt{2\pi}}, \quad dv = t e^{-\frac{t^2}{2}} dt,$$

откуда

$$v = -e^{-\frac{t^2}{2}}.$$

Применение интегрирования по частям дает

$$D(X) = -\sigma^2 \frac{te^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \Big|_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Первый член равен нулю, так как показательная функция $e^{-\frac{t^2}{2}}$ убывает при возрастании t быстрее, чем любая степень t , в частности,

быстрее, чем t . Поэтому

$$D(X) = \sigma^2, \quad (11.24)$$

т. е. параметр σ в нормальном законе распределения есть среднее

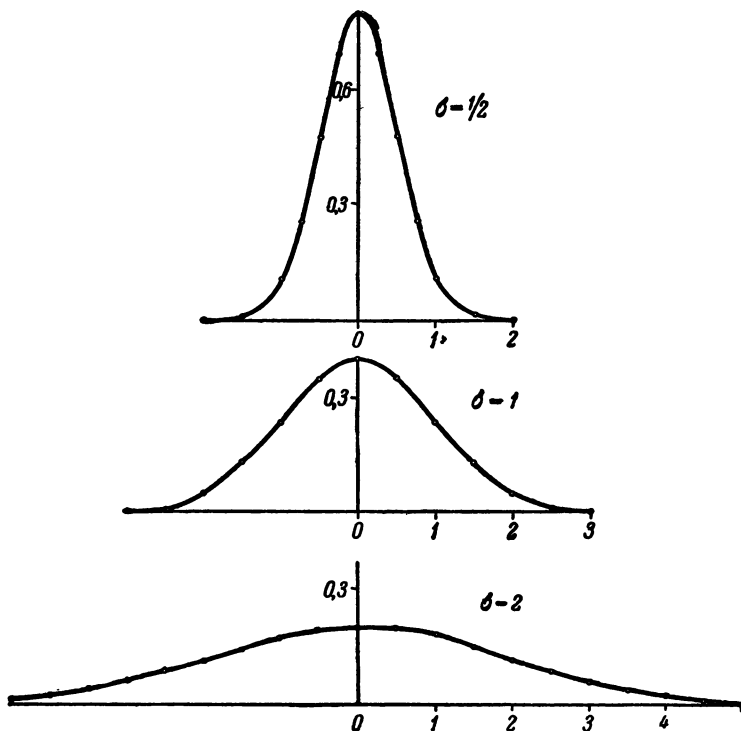


Рис. 8. Кривые нормального распределения для различных значений дисперсии σ^2 .

квадратичное отклонение. Полагая в формуле нормального закона $a = \bar{x}$, запишем его в виде

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}. \quad (11.25)$$

Если среднее значение \bar{x} равно нулю, а среднее квадратичное отклонение σ — единице, то мы имеем так называемый *стандартный нормальный закон* распределения:

$$w = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (11.26)$$

Для плотности вероятности w такого стандартного закона составлены таблицы (табл. I в конце книги). Отметим, что уравнение (11.25) может быть приведено к стандартному виду (11.26) подстановкой:

$$\left. \begin{aligned} w &= \sigma y, \\ z &= \frac{x - \bar{x}}{\sigma}. \end{aligned} \right\} \quad (11.27)$$

Исследуем вид кривой нормального распределения (11.25), которую нередко называют *кривой Гаусса*.

Элементарными средствами анализа выясняется следующее.

Вся кривая расположена по одну сторону от оси абсцисс и симметрична относительно центра распределения $x = \bar{x}$. Ось абсцисс является асимптотой кривой. Кривая имеет максимум, равный $\frac{1}{\sigma\sqrt{2\pi}}$, в центре распределения $x = \bar{x}$ и две точки перегиба $x = \bar{x} \pm \sigma$.

Согласно формулам (11.27) кривую Гаусса (11.25) при произвольном σ можно получить из стандартной кривой (11.26), соответствующей $\sigma = 1$, $x = 0$, деля ординаты на σ и уменьшая абсциссы на \bar{x} и умножая их на σ . Таким образом, при $\sigma > 1$ мы получим более вытянутую вдоль оси абсцисс кривую с меньшей максимальной ординатой, чем у стандартной; при $\sigma < 1$ максимальная ордината больше, чем при $\sigma = 1$. На прилагаемом рис. 8 даны кривые Гаусса для $\sigma = \frac{1}{2}$, $\sigma = 1$, $\sigma = 2$ (масштаб осей координат один и тот же для всех кривых).

§ 58. Функция нормального распределения. Вычисление вероятностей

Как было отмечено в начале настоящей главы, основной задачей при изучении непрерывной случайной величины является вычисление вероятности

$$P(\alpha < X < \beta).$$

Проще всего эта задача решается по формуле (11.3) при помощи функции распределения $F(x)$.

Найдем эту функцию для нормального закона. Так как $F(x)$ является первообразной по отношению к плотности вероятности $p(x)$, причем $F(-\infty) = 0$, то

$$F(x) = \int_{-\infty}^x p(t) dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\bar{x})^2}{2\sigma^2}} dt. \quad (11.28)$$

В силу симметрии нормального закона относительно центра распределения $F(\bar{x}) = \frac{1}{2}$. Поэтому

$$F(x) = \frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}} \int_{\bar{x}}^x e^{-\frac{(t-\bar{x})^2}{2\sigma^2}} dt.$$

Полагая для стандартного нормального закона $\bar{x} = 0$, $\sigma = 1$, $x = z$, имеем

$$F(z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}} dt. \quad (11.29)$$

Второе слагаемое в этой формуле обозначается через $\Phi(z)$ и называется *интегралом вероятности*. При таком обозначении

$$F(z) = \frac{1}{2} + \Phi(z).$$

Производя в $F(x)$ замену переменной интегрирования $\frac{t-\bar{x}}{\sigma} = z$, найдем

$$F(x) = \frac{1}{2} + \Phi\left(\frac{x-\bar{x}}{\sigma}\right). \quad (11.30)$$

Таким образом, при помощи интеграла вероятностей $\Phi(z)$ легко найти функцию распределения для нормального закона с заданными значениями \bar{x} и σ . Как уже упоминалось в § 48, для функции $\Phi(z)$ составлены подробные таблицы (см. табл. III в конце книги)*).

Согласно формулам (11.3) и (11.30) вероятность того, что случайная величина примет какое-нибудь значение в интервале (α, β) , выражается формулой:

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta-\bar{x}}{\sigma}\right) - \Phi\left(\frac{\alpha-\bar{x}}{\sigma}\right). \quad (11.31)$$

Если вместо случайной величины X рассматривать ее отклонение от среднего значения $U = X - \bar{x}$, то эта формула примет вид

$$P(A < U < B) = \Phi\left(\frac{B}{\sigma}\right) - \Phi\left(\frac{A}{\sigma}\right), \quad (11.32)$$

*) Иногда вместо таблиц функции $\Phi(z)$ даются таблицы либо $2\Phi(z)$, либо $\int_0^z \frac{1}{\sqrt{\pi}} e^{-t^2} dt$; надлежащей простой подстановкой можно переходить от одной из этих функций к другой.

где $A = \alpha - \bar{x}$ и $B = \beta - \bar{x}$ — границы интервала, в котором заключены те значения величины U , для которых ищется вероятность.

Формуле (11.32) можно придать более простой вид в случае, если интервал (A, B) симметричен относительно центра распределения. Действительно, как было указано в § 44, функция $\Phi(z)$ является нечетной. Поэтому из (11.32) при $B > 0$ и $A = -B$ следует:

$$P(|U| < B) = 2\Phi\left(\frac{B}{\sigma}\right). \quad (11.33)$$

Выясним теперь, насколько вероятны различные отклонения величины, подчиненной нормальному закону, от ее теоретического среднего значения.

Вычислим при помощи формулы (11.33) вероятность того, что U по абсолютной величине не превышает σ :

$$P(|U| < \sigma) = 2\Phi(1).$$

По таблицам $\Phi(1) = 0,3413$; значит,

$$P(|U| < \sigma) = 0,6826 \approx \frac{2}{3}. \quad (11.34)$$

Считая (в случае большого числа наблюдений), что вероятность мало отличается от относительной частоты, можно сказать: если распределение близко к нормальному, то приблизительно в двух третях случаев абсолютная величина отклонения не превышает σ . Это утверждение часто называют *правилом сигм*. В теории ошибок число σ называют *средней ошибкой*, в статистике — *средним квадратичным отклонением* или просто *средним отклонением*; для σ употребляется еще название «стандарт».

Выведем еще правило «трех сигм». Вычислим вероятность того, что U по абсолютной величине не превышает 3σ :

$$P(|U| < 3\sigma) = 2\Phi(3) = 0,9973. \quad (11.35)$$

Отсюда вытекает *правило трех сигм*: если распределение близко к нормальному, то мало вероятно, чтобы отклонение превышало по абсолютной величине 3σ .

Точно так же можно показать, что

$$P(|U| < 4\sigma) = 0,9994; \quad P(|U| < 2\sigma) \approx 0,95.$$

Вероятное отклонение определяется из следующего условия: вероятность того, что U не превышает по абсолютной величине свое вероятное отклонение, равна $1/2$. Обозначим вероятное отклонение через r . По определению,

$$P(|U| < r) = 0,5,$$

но

$$P(|U| < r) = 2\Phi\left(\frac{r}{\sigma}\right),$$

поэтому $\Phi\left(\frac{r}{\sigma}\right) = 0,25$.

Из таблиц получаем

$$\frac{r}{\sigma} = 0,6745 \quad \text{или} \quad r \approx \frac{2}{3} \sigma.$$

Заменяя вероятность относительной частотой, можно сказать, что приблизительно в половине случаев $|U|$ меньше вероятного отклонения, если число наблюдений велико.

Следует подчеркнуть еще раз, что такого рода заключения выведены для теоретического распределения. Перенос этих заключений на наблюдаемые распределения может быть сделан только тогда, когда наблюдаемое распределение мало отличается от нормального.

§ 59. Моменты нормального распределения

Записывая уравнение нормального распределения в форме (11.22), получим начальный момент первого порядка:

$$\nu_1 = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

В § 57 показано, что

$$\nu_1 = a = \bar{x}. \quad (11.36)$$

Можно сказать, что нормальное распределение, написанное в форме (11.25), есть распределение отклонений случайной величины от теоретического среднего значения. Поэтому центральные моменты нормального распределения можно получить, пользуясь следующим видом плотности вероятности

$p(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}$. Отсюда находим:

$$\mu_2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2\sigma^2}} du$$

но μ_2 есть дисперсия, а в § 51 показано, что дисперсия равна σ^2 ; следовательно,

$$\mu_2 = \sigma^2. \quad (11.37)$$

Имеем, далее,

$$\mu_3 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} u^3 e^{-\frac{u^2}{2\sigma^2}} du = 0 \quad (11.38)$$

ибо под интегралом нечетная функция u . Интегрированием по частям легко показать, что

$$\mu_4 = 3\sigma^4 = 3\mu_2^2. \quad (11.39)$$

Выражения для μ_3 и μ_4 можно использовать как предварительный критерий для решения вопроса, можно ли наблюдаемое распределение считать нормальным или близким к нему. Ответ будет отрицательным, если μ_3 заметно отличается от нуля, или μ_4 заметно отличается от $3\sigma^4$, или имеет место и то и другое *).

Легко получить общее выражение для центрального момента любого порядка нормального распределения. Момент любого нечетного порядка равен нулю, так как под интегралом

$$\mu_{2k+1} = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} u^{2k+1} e^{-\frac{u^2}{2\sigma^2}} du \quad (11.40)$$

стоит нечетная функция u , а интеграл берется в пределах от $-\infty$ до $+\infty$. Для момента произвольного четного порядка выведем рекуррентную формулу. По определению,

$$\mu_{2k} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^{2k} \frac{1}{\sigma} e^{-\frac{u^2}{2\sigma^2}} du = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^{2k-1} \frac{u}{\sigma^2} e^{-\frac{u^2}{2\sigma^2}} du.$$

Применим к этому интегралу формулу интегрирования по частям

$$\int_a^b w dv = wv \Big|_a^b - \int_a^b v dw,$$

принимая

$$w = u^{2k-1}, \quad dv = \frac{u}{\sigma^2} e^{-\frac{u^2}{2\sigma^2}} du.$$

Отсюда

$$v = -e^{-\frac{u^2}{2\sigma^2}}, \quad dw = (2k-1) u^{2k-2} du.$$

Поэтому

$$\mu_{2k} = -\frac{1}{\sqrt{2\pi}} u^{2k-1} e^{-\frac{u^2}{2\sigma^2}} \Big|_{-\infty}^{\infty} + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (2k-1) u^{2k-2} e^{-\frac{u^2}{2\sigma^2}} du.$$

Первый член обращается в нуль при $u = -\infty$ и $u = +\infty$, что легко показать по правилу Лопиталя, применяя его k раз; второй член преобразуем следующим образом:

$$\begin{aligned} \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (2k-1) u^{2k-2} e^{-\frac{u^2}{2\sigma^2}} du &= \\ &= (2k-1) \sigma^2 \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} u^{2k-2} e^{-\frac{u^2}{2\sigma^2}} du = (2k-1) \sigma^2 \mu_{2k-2}. \end{aligned}$$

*) Метод вычисления моментов наблюдаемого распределения будет рассмотрен в части V.

Таким образом,

$$\mu_{2k} = (2k-1) \sigma^2 \mu_{2k-2} = (2k-1)(2k-3) \sigma^4 \mu_{2k-4} = \dots;$$

отсюда

$$\mu_{2k} = (2k-1)(2k-3) \dots 3 \cdot 1 \cdot (\sigma^2)^k \mu_0, \quad (11.41)$$

или

$$\mu_{2k} = 1 \cdot 3 \cdot 5 \dots (2k-3)(2k-1) \sigma^{2k}, \quad (11.42)$$

так как $\mu_0 = 1$. В частности, при $k = 2, 3, 4$

$$\left. \begin{aligned} \mu_4 &= 1 \cdot 3 \mu_2^2 = 3\sigma^4; \\ \mu_6 &= 1 \cdot 3 \cdot 5 \mu_2^3 = 15\sigma^6, \\ \mu_8 &= 1 \cdot 3 \cdot 5 \cdot 7 \mu_2^4 = 105\sigma^8. \end{aligned} \right\} \quad (11.43)$$

§ 60. Понятие о распределениях, отличных от нормального

Если значения случайной величины получаются по схеме, близкой к схеме теоремы Ляпунова, то нормальный закон должен удовлетворительно представлять материал, полученный наблюдениями значений случайной величины. Вопрос о сравнении теоретического распределения с эмпирическим будет рассмотрен подробно в части V этого курса. Здесь мы отметим один из способов сравнения: а) подбираются числовые значения параметров выбранного теоретического закона распределения; б) для ряда выбранных значений величины определяются значения теоретической функции распределения; в) вычисляются соответствующие эмпирические вероятности, т. е. отношения чисел наблюдавшихся значений, не превышающих тех значений, которые выбраны в пункте б), к числу всех наблюдений. Сравнение теоретических и эмпирических вероятностей и позволяет судить о большем или меньшем согласии теории с наблюдениями.

Накопленный опыт изучения эмпирических распределений разных величин показал, что не всегда нормальный закон удовлетворительно представляет наблюдения. Кроме величин, которые показывают недостаточную близость к нормальному закону, есть величины, которые по своим физическим свойствам не могут подчиняться нормальному закону.

Рассмотрим, например, распределение параллаксов звезд. Это распределение не может подчиняться нормальному закону по двум очевидным причинам. Во-первых, нормальный закон определен на всей числовой оси, т. е. от $-\infty$ до $+\infty$, тогда как параллакс есть величина положительная, следовательно, кривая распределения должна быть ограничена слева; поскольку практически, если речь идет только о звездах, больших параллаксов не может быть, то кривая распределения ограничена и справа. Во-вторых, число звезд растет с уменьшением параллакса, значит, максимальная плотность

вероятности будет при значении параллакса, равном нулю. Так как звезды с большими параллаксами редки, то кривая должна асимптотически стремиться к оси абсцисс при увеличении параллакса. Получается, что кривая распределения должна напоминать букву *J*, т. е. сильно отличается от кривой Гаусса.

Вторым примером может служить распределение модулей скорости членов некоторой совокупности движущихся тел, например астероидов. Опять-таки эта величина принимает только неотрицательные значения, и потому кривая распределения ограничена слева; так как бесконечные скорости не имеют физического смысла, то кривая должна быть ограничена и справа, хотя здесь значительно труднее наметить границу, чем слева.

Из сказанного следует, что и для решения основной задачи и для описания случайных величин нельзя ограничиться только нормальным законом. Необходимо строить и другие теоретические распределения, исходя из каких-либо вероятностных схем или других теоретических соображений. Рассмотрим кратко некоторые виды кривых распределения, вошедших в практику.

1. Кривые Шарлье типа А

Вид функции, представляющей плотность вероятности в том случае, когда непригоден нормальный закон, можно выбрать из следующих соображений. Если нормальный закон

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

оказывается недостаточно удовлетворительным, то можно задать плотность вероятности в виде

$$\varphi(x) = p(x) \Pi(x),$$

где $\Pi(x)$ — многочлен не выше четвертой степени, который Шарлье назвал *пертурбационным (возмущающим) многочленом* *). Заменим значения x их отклонениями u от среднего значения \bar{x} ; тогда

$$\varphi(u) = p(u) \Pi(u).$$

Для определения пяти коэффициентов полинома четвертой степени можно получить пять уравнений, определяя моменты от нулевого до четвертого порядка включительно.

Система имеет в общем случае одно решение; коэффициенты полинома $\Pi(x)$ выражаются через центральные моменты порядков 2,

*) Термин заимствован из небесной механики и означает, что множитель $\Pi(x)$ изменяет $p(x)$ так, чтобы получилась плотность вероятности, более удовлетворительно представляющая наблюдения.

3 и 4. Вместо моментов 3-го и 4-го порядков введем *асимметрию* A и *эксцесс* E по формулам

$$A = \frac{\mu_3}{\sigma^3}, \quad (11.44)$$

$$E = \frac{\mu_4}{\sigma^4} - 3. \quad (11.45)$$

Из § 59 следует, что для нормального закона $A = E = 0$; поэтому можно считать, что A и E характеризуют отклонение распределения от нормального. Коэффициенты полинома $\Pi(u)$ a_0, a_1, a_2, a_3, a_4 выражаются через A, E, σ :

$$\begin{aligned} a_0 &= 1 + \frac{1}{8} E, & a_1 &= \frac{1}{2} \frac{A}{\sigma}, & a_2 &= -\frac{1}{4} \frac{E}{\sigma^2}, \\ a_3 &= \frac{1}{6} \frac{A}{\sigma^3}, & a_4 &= \frac{1}{24} \frac{E}{\sigma^4}. \end{aligned}$$

Если в полиноме $P(u)$ собрать члены с асимметрией и отдельно с эксцессом, то оказывается, что коэффициент при A равен произведению 3-й производной нормальной плотности $p(x)$ на $-\frac{1}{6} \sigma^3$, а коэффициент при E равен произведению 4-й производной $\Pi(u)$ на $\frac{1}{24} \sigma^4$.

В результате плотность вероятности закона Шарлье примет вид

$$\varphi(u) = p(u) - \frac{1}{6} A \sigma^3 p^{III}(u) + \frac{1}{24} E \sigma^4 p^{IV}(u). \quad (11.46)$$

Из этого выражения видно, что при малых A и E члены, содержащие асимметрию и эксцесс, тоже малы, и потому основным членом выражения является плотность вероятности нормального распределения.

Для функций $p^{III}(u)$ и $p^{IV}(u)$ составлены таблицы при $\sigma = 1$, с помощью которых легко вычисляется $\varphi(u)$.

Решение основной задачи выполняется весьма просто:

$$P(\alpha < U < \beta) = \int_{\alpha}^{\beta} p(u) du - \frac{1}{6} A \sigma^3 \int_{\alpha}^{\beta} p^{III}(u) du + \frac{1}{24} E \sigma^4 \int_{\alpha}^{\beta} p^{IV}(u) du.$$

Если положить $\sigma = 1$, т. е. измерять u , принимая σ за единицу, то интеграция выполняется совсем просто:

$$\begin{aligned} P(\alpha < U < \beta) &= \Phi(\beta) - \Phi(\alpha) - \frac{1}{6} A [p^{II}(\beta) - p^{II}(\alpha)] + \\ &+ \frac{1}{24} E [p^{III}(\beta) - p^{III}(\alpha)]. \end{aligned}$$

Вместо $p^{II}(u)$ можно написать $\Phi^{III}(u)$ и, аналогично, $\Phi^{IV}(u)$ вместо $p^{III}(u)$. Для функций $\Phi^{III}(u)$ и $\Phi^{IV}(u)$ составлены таблицы.

2. Кривые Пирсона

Кривые распределения, уравнения которых вывел Пирсон, были получены им из приближенного рассмотрения более общей задачи теории вероятностей, чем задача о повторении испытаний при равных p и q , к которой сводится упрощенный вывод нормального распределения. Рядом авторов было отмечено, что кривые Пирсона можно получить, формально обобщая то дифференциальное уравнение, которое было получено для кривой Гаусса.

При выводе нормального закона (см. § 56, стр. 177) в уравнении

$$\frac{dy}{y} = -\frac{u du}{\sigma^2}$$

в знаменателе правой части было постоянное число; мы получим кривые более общего вида, если в знаменателе вместо постоянного поставим функцию от u . Полагая, что эта функция разлагается в ряд Маклорена и ограничиваясь первыми тремя членами, получим дифференциальные уравнения кривых Пирсона:

$$\frac{dy}{y} = \frac{(x-a) dx}{b_0 + b_1 x + b_2 x^2} \quad (11.47)$$

Это уравнение содержит четыре параметра a , b_0 , b_1 , b_2 ; при интегрировании войдет еще произвольная постоянная, но она выразится через эти четыре параметра. Действительно, по условию нормировки площадь, ограниченная кривой, должна равняться единице. Это условие и дает уравнение для определения произвольной постоянной, как это было и для нормального распределения.

Интегрирование уравнения (11.47) дает целый ряд функций распределения (среди них есть функции, дающие U -образные и J -образные кривые), которыми можно пользоваться как интерполяционными, для представления наблюдаемого материала. Для облегчения пользования кривыми Пирсона составлены таблицы. С кривыми Пирсона можно познакомиться по книге Л. К. Лахтина «Кривые распределения и построение для них интерполяционных формул по способам Пирсона и Брунса» (М., 1922). В качестве примера приведем уравнение кривой Пирсона 3-го типа, которая применяется в некоторых задачах:

$$y = y_0 e^{-\gamma x} \left(1 + \frac{x}{l}\right)^m, \quad (11.48)$$

где

$$\gamma = \frac{2\mu_2}{\mu_3}, \quad l = \mu_2 \gamma, \quad m = \frac{4}{\beta_1} - 1, \quad \beta_1 = \frac{\mu_3}{\mu_2^2},$$

$$y_0 = \frac{\gamma^{m+1} l^m \cdot e^{-l\gamma}}{\Gamma(m+1)}.$$

Здесь

$$\Gamma(m+1) = \int_0^{\infty} x^m e^{-x} dx$$

— эйлеров интеграл 2-го рода, который не сводится к элементарным функциям, если m не является целым положительным числом. Для этого интеграла также составлены таблицы.

Кроме кривых Пирсона строились некоторые другие кривые распределения частного вида. Однако следует отметить, что на практике предпочитают не брать функций с числом параметров, превышающим 4. В противном случае пришлось бы при составлении уравнений для определения параметров пользоваться вычисленными моментами выше четвертого порядка. Моменты высоких порядков, как легко видеть из определения и способа вычисления, зависят главным образом от «краев» распределения, которые мало характерны для распределения. По этой же причине в дифференциальном уравнении кривых Пирсона в знаменателе берутся только три члена.

Особо следует отметить нередко встречающиеся случаи, когда эмпирическая кривая распределения имеет два максимума. Для таких распределений даже среднее значение величины нельзя считать характерным. В ряде таких случаев можно думать, что статистический материал представляет сумму двух распределений, каждое из которых нормальное или близко к нему. Для таких распределений разработан вопрос о разложении на два нормальных распределения.

3. Распределение Максвелла

Распределением Максвелла называют распределение модулей скоростей молекул или других частиц или материальных точек, полученное в предположении, что компоненты скоростей по осям координат (прямоугольных) имеют одинаковые дисперсии. Можно показать, что плотность вероятности распределения Максвелла имеет вид

$$p(v) = \frac{2}{\sqrt{2\pi}} \frac{v^2}{\sigma_0^3} e^{-\frac{v^2}{2\sigma_0^2}}, \quad (11.49)$$

где v — модуль скорости, σ_0 — единственный параметр распределения, представляющий среднее квадратичное отклонение по каждой из осей координат. Приведем без вывода основные числовые характеристики этого распределения:

среднее значение

$$\bar{v} = \frac{4}{\sqrt{2\pi}} \sigma_0 \approx 1.596 \sigma_0; \quad p(\bar{v}) = \frac{16}{\pi \sqrt{2\pi}} e^{-\frac{4}{\pi}} \cdot \frac{1}{\sigma_0} \approx 0.569 \cdot \frac{1}{\sigma_0};$$

мода

$$v_m = \sigma_0 \sqrt{2} \approx 1,414\sigma_0; \quad p(v_m) = \frac{4}{e\sqrt{2\pi}} \cdot \frac{1}{\sigma_0} \approx 0,588 \cdot \frac{1}{\sigma_0};$$

среднее квадратичное отклонение

$$\sigma = \sqrt{3 - \frac{8}{\pi}} \sigma_0 \approx 0,674\sigma_0 \approx \frac{2}{3} \sigma_0;$$

функция распределения

$$F(v) = P(V < v) = 2\Phi\left(\frac{v}{\sigma_0}\right) - 2\frac{v}{\sigma_0} \Phi'\left(\frac{v}{\sigma_0}\right).$$

Значения функции распределения приводятся в таблице:

v	$F(v)$	v	$F(v)$
0	0,0000	2,0	0,7384
0,5	0,0309	2,5	0,9001
1,0	0,1986	3,0	0,9709
1,5	0,4779	3,5	0,9932
2,0	0,7384	4,0	0,9991

Приведем еще результаты вычисления некоторых вероятностей:

$$P(V < v_m) = 0,427,$$

$$P(V < \bar{v}) = 0,535,$$

$$P(\bar{v} - \sigma < V < \bar{v} + \sigma) = 0,677,$$

$$P(\bar{v} - 2\sigma < V < \bar{v} + 2\sigma) = 0,995.$$

Последние две вероятности мало отличаются от аналогичных вероятностей для нормального закона. Если построить нормальное распределение с центром v и средним квадратичным отклонением σ , то при $\sigma_0 = 1$ ординаты кривых распределений отличаются не больше, чем на 0,08.

4. Распределение Стьюдента

Подобное распределение используется для оценки вероятности отклонений выборочных средних от генеральной средней совокупности, подчиняющейся нормальному закону распределения. В. И. Романовский использовал распределение Стьюдента в теории ошибок в задачах с малым количеством наблюдений. Плотность вероятности в распределении Стьюдента имеет вид

$$p(t) = C(n) \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}, \quad (11.50)$$

где n — число объектов выборки (в частности, число наблюдений).
Остальные величины определяются формулами:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n},$$

$$\bar{\sigma} = \sqrt{\frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n-1}}, \quad \sigma_{\bar{x}} = \frac{\bar{\sigma}}{\sqrt{n}},$$

$$t = \frac{\bar{x} - x_0}{\sigma_{\bar{x}}},$$

S зависит только от n, x_1, x_2, \dots, x_n — выборочные значения величины, x_0 — среднее значение во всей совокупности (генеральное).
Распределение Стьюдента есть смысл использовать при числе наблюдений, не превышающем 20, так как уже при $n = 20$ распределение мало отличается от нормального.

ГЛАВА 12

РАСПРЕДЕЛЕНИЕ СОВОКУПНОСТИ ДВУХ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 61. Плотность вероятности совокупности двух величин

Рассмотрим совместно две непрерывные случайные величины X и Y . Основной задачей для нас явится вычисление вероятности того, что X будет иметь какое-нибудь значение в заданном интервале (x_1, x_2) , а Y — в интервале (y_1, y_2) , т. е.

$$x_1 < X < x_2; \quad y_1 < Y < y_2.$$

Если эти события (а именно, осуществления неравенств) независимы, то величины называются *независимыми*, и интересующая нас вероятность равняется произведению вероятностей обоих неравенств:

$$P \left\{ \begin{matrix} x_1 < X < x_2 \\ y_1 < Y < y_2 \end{matrix} \right\} = P(x_1 < X < x_2) P(y_1 < Y < y_2).$$

Если вероятность того, что какая-нибудь из величин примет значение в произвольном интервале, зависит от значения другой величины, то такие величины называются *коррелятивно связанными* (употребляются также выражения «находятся в корреляции», «между X и Y есть корреляции» и т. д.). Корреляционная связь вполне определена с точки зрения теории вероятностей, если известен закон определения вероятности совмещения написанных выше неравенств при любых значениях x_1, x_2, y_1, y_2 . Таким законом может быть либо функция распределения совокупности двух случайных величин, либо плотность вероятности. Можно ввести понятие о плотности вероятности следующим образом.

Пусть известна вероятность

$$P \left\{ \begin{matrix} x < X < x + \Delta x \\ y < Y < y + \Delta y \end{matrix} \right\}$$

того, что X примет значение в интервале от x до $x + \Delta x$ и одновременно Y — от y до $y + \Delta y$ ($\Delta x > 0, \Delta y > 0$). Эта вероятность

зависит в общем случае от x , y и Δx , Δy . Число

$$\frac{P\left\{\begin{array}{l} x < X < x + \Delta x \\ y < Y < y + \Delta y \end{array}\right\}}{\Delta x \Delta y} \quad (12.1)$$

можно назвать *средней плотностью вероятности* в соответствующих друг другу интервалах $(x, x + \Delta x)$ и $(y, y + \Delta y)$. Найдем

$$\lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{P\left\{\begin{array}{l} x < X < x + \Delta x \\ y < Y < y + \Delta y \end{array}\right\}}{\Delta x \Delta y}. \quad (12.2)$$

Если этот предел существует, то он в общем случае есть функция x и y , которую мы обозначим $f(x, y)$ и назовем *плотностью вероятности совокупности двух случайных величин* *). Из определения вытекает приближенное равенство:

$$P\left\{\begin{array}{l} x < X < x + \Delta x \\ y < Y < y + \Delta y \end{array}\right\} = f(x, y) \Delta x \Delta y, \quad (12.3)$$

если Δx и Δy — достаточно малые числа.

Если задана плотность вероятности совокупности двух случайных величин, то вероятность того, что они примут значения в промежутках от x_1 до x_2 и от y_1 до y_2 соответственно, определяется по очевидной точной формуле, представляющей обобщение аналогичной формулы одномерной задачи:

$$P\left\{\begin{array}{l} x_1 < X < x_2 \\ y_1 < Y < y_2 \end{array}\right\} = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y) dx dy. \quad (12.4)$$

Можно поставить и более общий вопрос о вероятности величинам принять значения в некоторой двумерной области S . Эта задача решается формулой

$$P\left\{\begin{array}{l} X \\ Y(S) \end{array}\right\} = \int_S \int f(x, y) dx dy. \quad (12.5)$$

Пусть все значения случайных величин X и Y заключены в области (двухмерной) Σ . Так как наши величины наверно примут какие-нибудь значения из области Σ , то по последней формуле

$$\int_{\Sigma} \int f(x, y) dx dy = 1. \quad (12.6)$$

Эта формула представляет основное свойство двумерной плотности вероятности, которое по аналогии с соответствующим свойством

*) Эту функцию называют также *дифференциальной функцией распределения*.

в одномерной задаче можно назвать *условием нормирования плотности вероятности*.

Из определения плотности вероятности следует, что $f(x, y) = 0$ вне области Σ . Поэтому условие нормирования можно всегда записывать в виде

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1. \quad (12.7)$$

Для каждой из величин X, Y рассматриваемой совокупности определим теоретические средние значения (математические ожидания) следующими формулами:

$$\bar{x} = E(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y) dx dy, \quad (12.8)$$

$$\bar{y} = E(Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yf(x, y) dx dy. \quad (12.9)$$

Точка с координатами (\bar{x}, \bar{y}) называется *центром распределения*. Обобщением этих понятий является понятие *математического ожидания функции* $\Psi(X, Y)$ рассматриваемых величин:

$$E[\Psi(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Psi(x, y) f(x, y) dx dy. \quad (12.10)$$

Читатель самостоятельно может ввести также понятия начальных и центральных моментов различных порядков для совокупности двух величин.

§ 62. Условные плотности вероятности

В этом параграфе мы исследуем связь между плотностью вероятности $f(x, y)$ совокупности двух величин X и Y и плотностями вероятностей $f_1(x)$ и $f_2(y)$ каждой из этих величин в отдельности.

Наиболее просто эта связь находится для независимых величин. Действительно, если X и Y независимы, то при любых $\Delta x > 0$ и $\Delta y > 0$

$$P\left\{ \begin{matrix} x < X < x + \Delta x \\ y < Y < y + \Delta y \end{matrix} \right\} = P(x < X < x + \Delta x) P(y < Y < y + \Delta y).$$

Разделив это равенство на произведение $\Delta x \Delta y$ и перейдя к пределу при $\Delta x \rightarrow 0$ и $\Delta y \rightarrow 0$, получим в левой части плотность вероятности совокупности X и Y , а в правой — произведение плотностей вероятностей каждой из величин X и Y :

$$f(x, y) = f_1(x) f_2(y). \quad (12.11)$$

Перейдем теперь к рассмотрению двух связанных величин.

По теореме умножения вероятностей имеем точное равенство

$$P \left\{ \begin{array}{l} x < X < x + \Delta x \\ y < Y < y + \Delta y \end{array} \right\} = \\ = P(x < X < x + \Delta x) P(y < Y < y + \Delta y | x < X < x + \Delta x),$$

в котором второй множитель справа есть условная вероятность того, что величина Y примет значение в промежутке от y до $y + \Delta y$, если известно, что величина X приняла какое-нибудь значение от x до $x + \Delta x$. Разделим обе части написанного равенства на $\Delta x \Delta y$ и перейдем к пределу, полагая, что величины Δx и Δy стремятся к нулю. Левая часть даст плотность вероятности $f(x, y)$ совокупности величин X и Y . Первый множитель справа, деленный на Δx , дает в пределе плотность вероятности $f_1(x)$ величины X . Второй множитель справа, деленный на Δy , даст функцию от y , содержащую x параметром и не содержащую Δx и Δy . Обозначим эту функцию $\varphi_2(y | x)$ и назовем ее *условной плотностью вероятности величины Y при заданном значении x* . Формальным определением этой условной вероятности является равенство

$$\varphi_2(y | x) = \lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{P(y < Y < y + \Delta y | x < X < x + \Delta x)}{\Delta y}.$$

После перехода к пределу в выражении вероятности получается равенство

$$f(x, y) = f_1(x) \cdot \varphi_2(y | x). \quad (12.12)$$

Совершенно аналогично, изменяя порядок величин, получим равенство

$$f(x, y) = f_2(y) \cdot \varphi_1(x | y). \quad (12.13)$$

в котором $\varphi_1(x | y)$ есть условная плотность вероятности величины X при условии, что Y имеет значение y . Эта функция определяется формулой

$$\varphi_1(x | y) = \lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{P(x < X < x + \Delta x | y < Y < y + \Delta y)}{\Delta x}.$$

Из написанных связей между условными и безусловными плотностями вероятностей следует, что только три из пяти могут быть заданы до некоторой степени произвольно; остальные две вычисляются по полученным формулам. Произвол, кроме того, ограничен условиями нормирования пяти плотностей вероятностей:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1, \quad \int_{-\infty}^{\infty} f_1(x) dx = 1; \quad \int_{-\infty}^{\infty} f_2(y) dy = 1; \\ \int_{-\infty}^{\infty} \varphi_1(x | y) dx = 1, \quad \int_{-\infty}^{\infty} \varphi_2(y | x) dy = 1,$$

Равенства (12.12), (12.13) дают возможность получить простое дифференциальное соотношение между двумя условными плотностями вероятности. Из этих равенств имеем

$$f_1(x) \varphi_2(y|x) = f_2(y) \varphi_1(x|y). \quad (12.14)$$

Логарифмируя это соотношение по основанию e , получим

$$\ln f_1(x) + \ln \varphi_2(y|x) = \ln f_2(y) + \ln \varphi_1(x, y).$$

Продифференцируем полученное равенство один раз по x и второй раз по y . Производные первых членов левой и правой частей равенства равны нулю, и мы получим искомое соотношение:

$$\frac{\partial^2 [\ln \varphi_1(x|y)]}{\partial x \partial y} = \frac{\partial^2 [\ln \varphi_2(y|x)]}{\partial x \partial y}. \quad (12.15)$$

Из равенств (12.12) и (12.13) следует, что плотность вероятности совокупности двух случайных величин должна быть функцией, из которой может быть выделена множителем плотность вероятности каждой из величин в отдельности. Можно, следовательно, представить $f(x, y)$ в таком виде:

$$f(x, y) = f_1(x) f_2(y) \varphi(x, y). \quad (12.16)$$

Из (12.11) и (12.16) следует, что для независимых величин функция $\varphi(x, y)$ тождественно равна единице. Нетрудно убедиться и в обратном. Действительно, если величины X и Y взаимно независимы, то при любых значениях x_1, x_2, y_1, y_2 можно написать для прямоугольной области:

$$\begin{aligned} P \left\{ \begin{array}{l} x_1 < X < x_2 \\ y_1 < Y < y_2 \end{array} \right\} &= \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_1(x) f_2(y) dx dy = \\ &= \int_{y_1}^{y_2} f_2(y) dy \int_{x_1}^{x_2} f_1(x) dx, \end{aligned} \quad (12.17)$$

или

$$P \left\{ \begin{array}{l} x_1 < X < x_2 \\ y_1 < Y < y_2 \end{array} \right\} = P(x_1 < X < x_2) \cdot P(y_1 < Y < y_2).$$

Последнее равенство означает, что наши величины независимы.

Таким образом, величины X и Y взаимно независимы тогда и только тогда, когда

$$\varphi(x, y) \equiv 1. \quad (12.18)$$

Следовательно, характер корреляционной связи полностью определяется функцией $\varphi(x, y)$.

Одной из характеристик условного распределения X при заданном значении $Y = y$ является *условное математическое ожидание*

(условное среднее значение), определяемое формулой:

$$m_1(y) = \int_{-\infty}^{+\infty} x \varphi_1(x|y) dx. \quad (12.19)$$

Аналогично определяется и другое условное среднее значение:

$$m_2(x) = \int_{-\infty}^{+\infty} y \varphi_2(y|x) dy. \quad (12.20)$$

Линии на плоскости (x, y) , определяемые уравнениями $x = m_1(y)$ и $y = m_2(x)$, называются *линиями регрессии*, а их уравнения — *уравнениями регрессии*.

Из определений средних значений \bar{x} и \bar{y} по всей совокупности и условных средних значений $m_1(y)$ и $m_2(x)$ вытекают соотношения:

$$\left. \begin{aligned} \bar{x} &= \int_{-\infty}^{+\infty} f_2(y) m_1(y) dy, \\ \bar{y} &= \int_{-\infty}^{+\infty} f_1(x) m_2(x) dx. \end{aligned} \right\} \quad (12.21)$$

§ 63. Нормальное распределение двух случайных величин

Распределение двух величин назовем *нормальным*, если каждая из них подчиняется закону Гаусса, когда значение другой заключено в произвольном малом промежутке. Это определение предполагает, что функции $\varphi_1(x|y)$ и $\varphi_2(y|x)$, которые мы назвали условными плотностями вероятностей величины X и Y , суть плотности нормальных распределений. Так как для функции $\varphi_1(x|y)$ величина y является параметром, то естественно считать в общем случае, что среднее значение и среднее квадратичное отклонение, входящие в плотность нормального распределения, зависят от y . Поэтому функция $\varphi_1(x|y)$ должна иметь вид

$$\varphi_1(x|y) = \frac{1}{\sqrt{2\pi} s_1(y)} \exp \left\{ -\frac{[x - m_1(y)]^2}{2s_1^2(y)} \right\}. \quad (12.22)$$

Здесь $m_1(y)$ есть условное среднее значение (математическое ожидание) величины X при заданном y , а $s_1(y)$ — условное среднее квадратическое отклонение X . По тем же причинам можно написать:

$$\varphi_2(y|x) = \frac{1}{\sqrt{2\pi} s_2(x)} \exp \left\{ -\frac{[y - m_2(x)]^2}{2s_2^2(x)} \right\}, \quad (12.23)$$

где m_2 и s_2 имеют такой же смысл для величины Y , как m_1 и s_1 для X .

Воспользуемся равенством (12.15) для определения зависимости между функциями $s_1(y)$, $s_2(x)$, $m_1(y)$ и $m_2(x)$. Эта зависимость даст возможность установить необходимую форму этих функций.

Введем обозначения:

$$k_1(y) = \frac{1}{s_1^2(y)}, \quad k_2(x) = \frac{1}{s_2^2(x)}. \quad (12.24)$$

Тогда

$$\begin{aligned} \ln \varphi_1(x|y) &= -\ln \sqrt{2\pi} + \ln \sqrt{k_1(y)} - \frac{1}{2} k_1(y) [x - m_1(y)]^2, \\ \ln \varphi_2(x|y) &= -\ln \sqrt{2\pi} + \ln \sqrt{k_2(x)} - \frac{1}{2} k_2(x) [y - m_2(x)]^2. \end{aligned}$$

При двукратном дифференцировании по x и y первые два члена, входящие в выражения для $\ln \varphi_1$ и $\ln \varphi_2$, дадут нули. Поэтому, согласно (12.15),

$$\begin{aligned} x k_1'(y) - m_1'(y) k_1(y) - m_1(y) k_1'(y) = \\ = y k_2'(x) - m_2'(x) k_2(x) - m_2(x) k_2'(x). \end{aligned} \quad (12.25)$$

Дифференцируем равенство (12.25) один раз по y , второй раз по x ; так как второй и третий члены левой части зависят только от y , то их производные по x равны нулю; по той же причине дадут нули второй и третий члены правой части. В результате получим

$$k_1''(y) = k_2''(x).$$

Поскольку x и y не связаны функциональной зависимостью, это тождество возможно только при условиях

$$k_1''(y) = 2C \quad \text{и} \quad k_2''(x) = 2C,$$

где C — произвольная постоянная. Интегрируя дважды каждое из последних равенств, получаем

$$\begin{aligned} k_1(y) &= Cy^2 + d_1y + f_1, \\ k_2(x) &= Cx^2 + d_2x + f_2, \end{aligned}$$

где d_1 , d_2 , f_1 , f_2 — произвольные постоянные интегрирования. Полученная форма функции $k_1(y)$ такова, что $k_1(y) \rightarrow \infty$, когда $y \rightarrow \infty$, если при этом $C \neq 0$ и $d_1 \neq 0$. Так как согласно (12.24)

$$k_1(y) = \frac{1}{s_1^2(y)},$$

то при тех же условиях $s_1(y) \rightarrow 0$.

Функция $s_1(y)$ представляет среднее квадратичное отклонение нормального распределения X при $Y = y$; если $s_1(y)$ очень мало при больших значениях y , то это значит, что весьма маловероятны незначительные отклонения от среднего $m_1(y)$.

В обычных приложениях теории корреляции такой характер связи не представляет интереса, так как это означало бы устойчивость (почти постоянство) значений X при больших значениях Y . То же самое можно сказать и о функции $k_2(x)$. Поэтому, чтобы исключить эти особые виды распределений, мы положим

$$C = d_1 = d_2 = 0.$$

откуда

$$k_1(y) = f_1, \quad k_2(x) = f_2.$$

Поэтому

$$k_1(y) = \frac{1}{s_1^3} \quad \text{и} \quad k_2(x) = \frac{1}{s_2^3},$$

где s_1 и s_2 — теперь постоянные числа. Подставив k_1 и k_2 в равенство (12.23), получим

$$\frac{m'_1(y)}{s_1^3} = \frac{m'_2(x)}{s_2^3}. \quad (12.26)$$

Так как x и y не связаны функциональной зависимостью, то последнее равенство возможно только при условии, что каждая из функций, стоящих в левой и правой частях (12.24), равняется постоянному числу, т. е.

$$\frac{m'_1(y)}{s_1^3} = m, \quad \frac{m'_2(x)}{s_2^3} = m,$$

где m — произвольная постоянная. Интегрируя последние равенства получим

$$m_1(y) = ms_1^3 y + p_1, \quad m_2(x) = ms_2^3 x + p_2, \quad (12.27)$$

где p_1 и p_2 — произвольные постоянные интегрирования. Подставляя найденные значения $m_1(y)$, $m_2(x)$ (12.27) в условные плотности вероятности (12.22) и (12.23), имеем:

$$\varphi_1(x|y) = \frac{1}{s_1 \sqrt{2\pi}} \exp \left\{ -\frac{[x - (ms_1^3 y + p_1)]^2}{2s_1^2} \right\}, \quad (12.28)$$

$$\varphi_2(y|x) = \frac{1}{s_2 \sqrt{2\pi}} \exp \left\{ -\frac{[y - (ms_2^3 x + p_2)]^2}{2s_2^2} \right\}. \quad (12.29)$$

В выражении (12.28) для $\varphi_1(x|y)$ величина $ms_1^3 y + p_1$ есть теоретическое среднее значение величины X , соответствующее заданному значению y , а в выражении (12.29) $\varphi_2(y|x)$ величина $ms_2^3 x + p_2$ есть теоретическое среднее значение Y , соответствующее заданному значению x . Средние квадратичные отклонения s_1 и s_2 условных распределений оказались постоянными.

Таким образом, в случае нормальной корреляции среднее из значений каждой величины, соответствующих определенному значению другой величины, есть линейная функция этого значения. Иными словами, линии регрессии являются прямыми. Уравнения прямых регрессии (уравнения регрессии) имеют вид

$$\left. \begin{aligned} x &= ms_1^2 y + p_1, \\ y &= ms_2^2 x + p_2. \end{aligned} \right\} \quad (12.30)$$

Коэффициенты ms_1^2 и ms_2^2 этих уравнений называют *коэффициентами регрессии*.

Обозначим через \bar{x} , \bar{y} координаты точки пересечения прямых регрессии.

Очевидно,

$$\left. \begin{aligned} p_1 &= \bar{x} - ms_1^2 \bar{y}, \\ p_2 &= \bar{y} - ms_2^2 \bar{x}. \end{aligned} \right\}$$

Поэтому условные плотности вероятности и уравнения регрессии можно записать в виде

$$\left. \begin{aligned} \varphi_1(x|y) &= \frac{1}{s_1 \sqrt{2\pi}} \exp \left\{ -\frac{[(x - \bar{x}) - ms_1^2(y - \bar{y})]^2}{2s_1^2} \right\}, \\ \varphi_2(y|x) &= \frac{1}{s_2 \sqrt{2\pi}} \exp \left\{ -\frac{[(y - \bar{y}) - ms_2^2(x - \bar{x})]^2}{2s_2^2} \right\}; \\ x - \bar{x} &= ms_1^2(y - \bar{y}), \\ y - \bar{y} &= ms_2^2(x - \bar{x}). \end{aligned} \right\}$$

§ 64. Плотность вероятности нормального распределения

Приступая к выводу формулы для плотности вероятности нормального распределения, перенесем начало координат в точку пересечения (\bar{x}, \bar{y}) прямых регрессии.

Тогда для новых (смещенных) случайных величин U, V уравнения прямых регрессии примут вид

$$\left. \begin{aligned} u &= ms_1^2 v, \\ v &= ms_2^2 u, \end{aligned} \right\} \quad (12.31)$$

а условные плотности вероятности запишутся так:

$$\left. \begin{aligned} \varphi_1(u|v) &= \frac{1}{s_1 \sqrt{2\pi}} \exp \left\{ -\frac{u^2 - 2ms_1^2 uv + m^2 s_1^4 v^2}{2s_1^2} \right\}, \\ \varphi_2(u|v) &= \frac{1}{s_2 \sqrt{2\pi}} \exp \left\{ -\frac{v^2 - 2ms_2^2 uv + m^2 s_2^4 u^2}{2s_2^2} \right\}. \end{aligned} \right\} \quad (12.32)$$

Подставляя φ_1 и φ_2 в равенство (12.14), получим

$$\begin{aligned} f_1(u) \frac{1}{s_2 \sqrt{2\pi}} \exp \left\{ -\frac{v^2}{2s_2^2} + muv - \frac{m^2 s_2^2 u^2}{2} \right\} = \\ = f_2(v) \frac{1}{s_1 \sqrt{2\pi}} \exp \left\{ -\frac{u^2}{2s_1^2} + muv - \frac{m^2 s_1^2 v^2}{2} \right\}, \end{aligned}$$

где $f_1(u)$ и $f_2(v)$ — неизвестные пока плотности вероятности распределения каждой величины в отдельности.

Сократим последнее равенство на e^{muv} и запишем в одной части множители, зависящие только от u , в другой — от v :

$$\begin{aligned} \frac{1}{s_2} f_1(u) \exp \left[-\frac{1}{2} \left(m^2 s_2^2 - \frac{1}{s_1^2} \right) u^2 \right] = \\ = \frac{1}{s_1} f_2(v) \exp \left[-\frac{1}{2} \left(m^2 s_1^2 - \frac{1}{s_2^2} \right) v^2 \right]. \end{aligned}$$

Так как u и v не связаны функциональной зависимостью, то равенство возможно только при условии

$$\left. \begin{aligned} \frac{1}{s_2} f_1(u) \exp \left[-\frac{1}{2} \left(m^2 s_2^2 - \frac{1}{s_1^2} \right) u^2 \right] &= C, \\ \frac{1}{s_1} f_2(v) \exp \left[-\frac{1}{2} \left(m^2 s_1^2 - \frac{1}{s_2^2} \right) v^2 \right] &= C, \end{aligned} \right\} \quad (12.33)$$

где C — произвольная постоянная. Следовательно,

$$\left. \begin{aligned} f_1(u) &= C s_2 \exp \left[-\frac{1}{2} \left(\frac{1}{s_1^2} - m^2 s_2^2 \right) u^2 \right], \\ f_2(v) &= C s_1 \exp \left[-\frac{1}{2} \left(\frac{1}{s_2^2} - m^2 s_1^2 \right) v^2 \right]. \end{aligned} \right\} \quad (12.34)$$

Из вида функций $f_1(u)$, $f_2(v)$ заключаем: в случае нормальной корреляции распределение каждой величины в отдельности подчиняется нормальному закону.

Запишем полученные плотности вероятности в обычной форме:

$$\left. \begin{aligned} f_1(u) &= \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma_u^2}}, \\ f_2(v) &= \frac{1}{\sigma_v \sqrt{2\pi}} e^{-\frac{v^2}{2\sigma_v^2}}. \end{aligned} \right\} \quad (12.35)$$

Сравнивая эти формулы с предыдущими, найдем:

$$\sigma_u^2 = \frac{s_1^2}{1 - m^2 s_1^2 s_2^2}, \quad \sigma_v^2 = \frac{s_2^2}{1 - m^2 s_1^2 s_2^2}. \quad (12.36)$$

Вместо величины m введем новую постоянную R , определенную равенством

$$R^2 = m^2 s_1^2 s_2^2. \quad (12.37)$$

Тогда

$$s_1^2 = \sigma_u^2 (1 - R^2), \quad s_2^2 = \sigma_v^2 (1 - R^2),$$

$$m s_1^2 = R \frac{\sigma_u}{\sigma_v}, \quad m s_2^2 = R \frac{\sigma_v}{\sigma_u}.$$

Подставляя теперь найденные для функций $f_1(u)$ и $\varphi_2(v|u)$ выражения (12.34) и (12.32) в равенство (12.12) (с заменой x , y на u , v), получим *плотность вероятности нормального распределения величин U и V* :

$$f(u, v) = \frac{1}{2\pi \sqrt{1 - R^2} \sigma_u \sigma_v} \times$$

$$\times \exp \left\{ -\frac{1}{2(1 - R^2)} \left[\frac{u^2}{\sigma_u^2} - 2R \frac{u}{\sigma_u} \frac{v}{\sigma_v} + \frac{v^2}{\sigma_v^2} \right] \right\}. \quad (12.38)$$

Уравнения регрессии с новыми обозначениями (R вместо m ; σ_u , σ_v вместо s_1 , s_2) примут вид

$$u = R \frac{\sigma_u}{\sigma_v} v, \quad v = R \frac{\sigma_v}{\sigma_u} u. \quad (12.39)$$

Найдем теперь центр нормального распределения. В нашем случае:

$$m_1(v) = R \frac{\sigma_u}{\sigma_v} v, \quad m_2(u) = R \frac{\sigma_v}{\sigma_u} u.$$

Поэтому по формулам (12.21)

$$\bar{u} = \frac{R \sigma_u}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} v e^{-\frac{v^2}{2\sigma_v^2}} dv, \quad \bar{v} = \frac{R \sigma_v}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-\frac{u^2}{2\sigma_u^2}} du.$$

Интегралы в правых частях, очевидно, равны нулю и, следовательно,

$$\bar{u} = 0, \quad \bar{v} = 0. \quad (12.40)$$

Вспоминая, что начало координат у нас расположено в точке пересечения прямых регрессии, мы приходим к следующему выводу:

при нормальном распределении двух случайных величин точка пересечения линий регрессии совпадает с центром распределения *).

Для того чтобы найти «вероятностное выражение» для величины R , вычислим математическое ожидание произведения случайных величин U и V . Подставляя $f(u, v)$ из (12.38) в равенство

$$E(UV) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} uv f(u, v) du dv,$$

получим

$$E(UV) =$$

$$= \int_{-\infty}^{\infty} \frac{u}{\sigma_u \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma_u^2}} \int_{-\infty}^{\infty} \frac{v}{\sigma_v \sqrt{1-R^2} \sqrt{2\pi}} \exp \left[-\frac{\left(v - R \frac{\sigma_v}{\sigma_u} u \right)^2}{2(1-R^2)\sigma_v^2} \right] dv du.$$

Чтобы вычислить внутренний интеграл, сделаем подстановку

$$\frac{v - R \frac{\sigma_v}{\sigma_u} u}{\sigma_v \sqrt{1-R^2}} = z.$$

Тогда

$$dv = \sigma_v \sqrt{1-R^2} dz, \quad \frac{v}{\sigma_v \sqrt{1-R^2}} = z + \frac{Ru}{\sigma_u \sqrt{1-R^2}};$$

при этом u считается постоянной величиной.

Интеграл по v примет вид

$$\int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \sigma_v \sqrt{1-R^2} dz + \int_{-\infty}^{\infty} R \frac{\sigma_v}{\sigma_u} u \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Первый из интегралов обращается в нуль, поскольку под интегралом нечетная функция z ; второй равен $R \frac{\sigma_v}{\sigma_u} u$, так как

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1.$$

Следовательно,

$$E(UV) = R \frac{\sigma_v}{\sigma_u} \int_{-\infty}^{\infty} u^2 \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma_u^2}} du.$$

*) Отметим, кстати, что в рассматриваемом случае координаты центра распределения совпадают с теоретическими средними каждой из величин в отдельности.

Так как $\bar{u} = 0$, то интеграл в правой части есть дисперсия величины u , равная σ_u^2 . Поэтому

$$E(UV) = R\sigma_u\sigma_v,$$

откуда

$$R = \frac{E(UV)}{\sigma_u\sigma_v}. \quad (12.41)$$

Перейдем в полученных выражениях для плотности вероятности (12.38), уравнений регрессии (12.39) и коэффициента R (12.41) от u и v к исходным величинам x и y . Так как замена x, y на u, v сделана изменением начала отсчета, то

$$u = x - \bar{x}, \quad v = y - \bar{y}.$$

Поэтому $\sigma_u = \sigma_x$, $\sigma_v = \sigma_y$, и выражение для плотности вероятности примет вид

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-R^2}} \exp \left\{ -\frac{1}{2(1-R^2)} \left[\frac{(x-\bar{x})^2}{\sigma_x^2} - 2R \frac{x-\bar{x}}{\sigma_x} \cdot \frac{y-\bar{y}}{\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2} \right] \right\}. \quad (12.42)$$

Уравнения регрессии запишутся так:

$$y - \bar{y} = R \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad x - \bar{x} = R \frac{\sigma_x}{\sigma_y} (y - \bar{y}), \quad (12.43)$$

где y и x , стоящие слева, представляют средние значения при определенных x и y , стоящих в правых частях. Поэтому y и x , стоящие слева, иногда обозначают \bar{y}_x и \bar{x}_y . Для коэффициента R можно получить выражение:

$$R = \frac{E\{(x-\bar{x})(y-\bar{y})\}}{\sigma_x\sigma_y} = \frac{E(XY) - \bar{x}\bar{y}}{\sigma_x\sigma_y}. \quad (12.44)$$

Обозначая через ρ_{yx} и ρ_{xy} коэффициенты регрессии, имеем

$$\rho_{yx} = R \frac{\sigma_y}{\sigma_x}, \quad \rho_{xy} = R \frac{\sigma_x}{\sigma_y}. \quad (12.45)$$

Входящие в эти формулы величины \bar{x} , \bar{y} , σ_x , σ_y являются теоретическими средними значениями и средними квадратичными отклонениями каждой из величин X, Y в отдельности. Точка (\bar{x}, \bar{y}) является центром распределения и в то же время точкой пересечения прямых регрессии. Вероятностный смысл величины R раскрывается следующей теоремой:

для того чтобы две величины, подчиненные нормальному закону, были взаимно независимы, необходимо и достаточно, чтобы $R = 0$.

Для доказательства этой теоремы представим $f(x, y)$ в форме

$$f(x, y) = \left\{ \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma_x^2} \right] \right\} \left\{ \frac{1}{\sigma_y \sqrt{2\pi}} \exp \left[-\frac{(y - \bar{y})^2}{2\sigma_y^2} \right] \right\} \times \\ \times \left\{ \frac{1}{\sqrt{1-R^2}} \exp \left\{ -\frac{R}{2(1-R^2)} \left[\frac{R(x - \bar{x})^2}{\sigma_x^2} - \frac{2(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{R(y - \bar{y})^2}{\sigma_y^2} \right] \right\} \right\}.$$

В предыдущем параграфе (см. формулу (12.16)) последний множитель был обозначен через $\varphi(x, y)$ и было указано, что функция $\varphi(x, y)$ полностью определяет характер корреляционной связи, именно, если величины независимы, то $\varphi(x, y)$ должно тождественно равняться единице, и наоборот. Из последней формулы видно, что в случае нормальной корреляции последний множитель тождественно обращается в единицу только при условии $R = 0$.

Из выражений для s_1^2 и s_2^2 следует, что $|R| \leq 1$. Если $|R| = 1$, то $s_1 = s_2 = 0$. Так как s_1 и s_2 — средние квадратичные отклонения условных распределений, то в этом случае каждому значению y соответствует только одно значение x (совпадающее со средним \bar{x}), и наоборот, т. е. x и y связаны однозначной функциональной зависимостью.

Действительно, если $s_1 \rightarrow 0$, то кривая Гаусса, характеризующая распределение X при заданном постоянном $Y = y$, стремится слиться с осью ординат, иными словами, вероятность какого-нибудь отклонения от среднего стремится к нулю. Так как теоретические линии регрессии — прямые, то при $|R| = 1$ однозначная зависимость между X и Y будет линейной.

Таким образом, введенное нами число R характеризует отклонение корреляционной связи от функциональной линейной связи. Число R называется *коэффициентом корреляции*.

Из формул (12.45) для ρ_{yx} и ρ_{xy} следует:

$$R^2 = \rho_{yx} \rho_{xy}$$

Таким образом, квадрат коэффициента корреляции равен произведению коэффициентов регрессии.

Как мы уже отмечали, в случае нормального распределения двух случайных величин среднее значение каждой из величин при фиксированном значении другой есть линейная функция последней. Корреляцию такого рода называют *линейной*. На основании выведенной линейности уравнений регрессии можно сказать, что изучение линейной корреляции между двумя величинами можно считать обоснованным, если есть основания полагать, что совокупность рассматриваемых случайных величин подчиняется двумерному нормальному закону распределения. При этом условии совокупность полностью характеризуется пятью числами: средними значениями обеих величин в отдельности, их средними квадратичными отклонениями и коэффициентом корреляции.

Эллипсы равных вероятностей

Из общего вида плотности вероятности двух случайных величин следует, что плотность вероятности постоянна во всех точках плоскости x, y , в которых

$$\frac{(x - \bar{x})^2}{\sigma_x^2} - 2r \frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} = \lambda^2,$$

где λ — произвольное постоянное число, r — коэффициент корреляции.

Кривая на плоскости x, y , определяемая написанным уравнением, есть эллипс, что легко проверить методами теории кривых второго порядка. Такой эллипс называют *эллипсом равных вероятностей*, так как вдоль него одинаковы вероятности попасть на одинаковые элементарные площадки. Назовем такой эллипс λ -*эллипсом* и обозначим область на плоскости x, y , ограниченную λ -эллипсом, через Δ , а вероятность точке попасть в область Δ обозначим через $P(\lambda)$.

Для сокращения письма перейдем к аргументам u и v , т. е. к отклонениям наших величин от их математических ожиданий. Уравнение λ -эллипса в новых переменных имеет вид

$$\frac{u^2}{\sigma_u^2} - 2r \frac{uv}{\sigma_u \sigma_v} + \frac{v^2}{\sigma_v^2} = \lambda^2;$$

$$\sigma_u = \sigma_x, \quad \sigma_v = \sigma_y.$$

По определению плотности вероятности имеем

$$P(\lambda) = \int_{\Delta} \int p(u, v) du dv,$$

где через $p(u, v)$ обозначена плотность вероятности.

Введем вместо u и v полярные координаты:

$$u = \rho \cos \theta, \quad v = \rho \sin \theta.$$

Преобразование интеграла дает:

$$P(\lambda) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-r^2}} \int_0^{2\pi} d\theta \int_0^{\frac{\lambda}{s\sqrt{1-r^2}}} e^{-\frac{\rho^2 s^2}{2}} \rho d\rho,$$

где

$$s^2 = \left[\frac{\cos^2 \theta}{\sigma_u^2} - 2r \frac{\cos \theta \sin \theta}{\sigma_u \sigma_v} + \frac{\sin^2 \theta}{\sigma_v^2} \right] \cdot \frac{1}{1-r^2}$$

Интегрирование по ρ дает

$$P(\lambda) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-r^2}} \left[1 - e^{-\frac{\lambda^2}{2(1-r^2)}} \right] \int_0^{2\pi} \frac{d\theta}{s^2}.$$

Для сокращения вычисления интеграла по θ учтем, что при $\lambda \rightarrow \infty$ эллипс охватывает всю плоскость; значит, $P(\infty) = 1$. С другой стороны, по этой

формуле получим

$$P(\infty) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-r^2}} \int_0^{2\pi} \frac{d\theta}{s^2}.$$

Значит,

$$P(\lambda) = 1 - \exp \frac{\lambda^2}{2(1-r^2)}.$$

В частности, если величины взаимно независимы, то $r = 0$. В этом случае, например,

$$P(2) = 0,865, \quad P(3) = 0,989;$$

полуоси соответствующих эллипсов равны удвоенным и утроенным средним квадратичным отклонениям. Такие эллипсы можно назвать *доверительными* по аналогии с понятием одномерной задачи, и соответствующие вероятности можно тоже назвать *доверительными*.

ЧАСТЬ IV

ОСНОВЫ ТЕОРИИ СЛУЧАЙНЫХ ОШИБОК ИЗМЕРЕНИЙ (СПОСОБ НАИМЕНЬШИХ КВАДРАТОВ)

ГЛАВА 13

ОБЩИЕ СВЕДЕНИЯ ОБ ОШИБКАХ ИЗМЕРЕНИЙ

§ 65. Виды ошибок измерений

Любой результат измерения при наблюдениях содержит ряд ошибок различного происхождения. Ошибки принято делить на *систематические, случайные, личные и грубые*.

1. Систематические ошибки

К систематическим ошибкам относятся в первую очередь ошибки, которые часто называют *инструментальными*. Те приборы, которыми производят измерения, не могут быть сделаны идеально точно. Если в простейшем случае прямого измерения имеется шкала со штрихами, то промежутки между штрихами несколько отличаются от номинала (например, 0,999 мм в одном месте линейки, 1,002 мм в другом и т. д.). Ошибки шкалы можно выявить, сделав измерения ее промежутков прибором высокой точности (прецизионным). После этого обследуемый прибор будет снабжен паспортом, в котором указывается ошибка в зависимости от измеряемой величины (отрезка, угла и т. п.).

Иногда вместо измерения ошибки можно организовать измерения так, чтобы ошибка была исключена. Хорошо известный пример — ошибка эксцентриситета угломерного прибора. Редко удается сделать прибор так, чтобы геометрический центр круга, разделенного на градусы, совпадал достаточно точно с осью вращения круга; поэтому дуга, отсчитанная по кругу, не измеряет нужного угла («ошибка эксцентриситета»). Измерить и учесть ошибку эксцентриситета было бы затруднительно. Однако, ошибка эксцентриситета легко исключается, если повернуть круг на 180° и снова отсчитать дугу. Из геометрических соображений ясно, что среднее арифметическое обоих отсчетов дает точное значение угла, независимо от величины эксцентриситета. К простейшим инструментальным ошибкам следует отнести также ошибку от неправильной отметки начала отсчета.

Из-за этой ошибки все результаты измерений отличаются от точных на одну и ту же величину, положительную или отрицательную.

Все инструментальные ошибки должны быть изучены в теории астрономических инструментов, измерены в каждом отдельном приборе и исключены из результатов измерений. Если возможно, то измерения производят так, чтобы ошибка исключалась комбинированием двух или нескольких измерений.

Инструментальные ошибки входят во всякий результат измерений, причем они либо постоянны, либо определенным образом зависят от других величин, в частности, от самой измеряемой величины. Это и является причиной того, что подобные ошибки называют *систематическими*.

Инструментальные ошибки не единственный вид систематических ошибок. Например, при дифференциальном определении координат небесных тел (например, малых планет) по фотографиям измеряют их положения относительно звезд сравнения. Так как координаты звезд сравнения, взятые из каталогов, содержат систематические ошибки (ошибки каталога), то и определяемые координаты небесных тел содержат такие же ошибки. Эти ошибки должны быть, как и инструментальные, исследованы и исключены из результатов измерений.

Общим признаком систематических ошибок можно считать принципиальную возможность изучить их и исключить из результатов измерений. Методы учета систематических ошибок рассматриваются в разных разделах астрономии, имеющих дело с результатами наблюдений.

2. Случайные ошибки

Опыт показывает, что многократные измерения одной и той же определенной величины, произведенные со всей возможной тщательностью, дают различные числовые значения и после учета всех известных систематических ошибок. Это явление показывает, что на результаты измерений влияют какие-то физические причины, не поддающиеся учету. Пусть, например, производится взвешивание на достаточно точных и чувствительных весах. Если в момент измерения в том же доме хлопнули дверь, то указатель отклонится в случайном направлении и получится число, отличное от точного. Если в момент другого измерения по улице мимо пройдет тяжелый грузовик, но новый толчок будет отличен от предыдущего, и результат измерения снова изменится. Если измерения не механизированы полностью, если в них принимает участие человек, то на результат могут влиять случайные изменения состояния его органов, участвующих в измерениях.

Подобного рода случайных причин, дающих отклонения от точного значения, может быть целый ряд. Каждая из этих причин дает мало заметное отклонение, так как в противном случае оно было бы

отмечено и исследовано. Однако суммарное воздействие ряда причин может дать заметные отклонения.

Под теорией ошибок подразумевается обычно именно *теория случайных ошибок*. По самому характеру случайных ошибок для построения теории привлекается аппарат теории вероятностей.

3. Личные ошибки

Практика астрономических наблюдений показала, что результаты измерений зависят в некоторой степени от физических особенностей наблюдателя при прочих равных условиях. Так, например, при регистрации момента явления один наблюдатель может систематически отмечать то или иное явление несколько раньше, чем другой. Неоднократное изучение личных ошибок разных наблюдателей показало, что эти ошибки следует отнести и к систематическим, и к случайным. Известно, что наблюдателю присуща некоторая средняя величина личной ошибки, которую и следует считать систематической и учитывать при обработке наблюдений. Однако личная ошибка в отдельных наблюдениях оказывается величиной случайной, варьирующей по разным причинам (физическое состояние наблюдателя, внешние условия и т. п.). Для выявления личных ошибок делают наблюдения, результаты наблюдений обрабатываются почти так же, как и результаты наблюдений со случайными ошибками, чтобы получить среднее значение личных ошибок.

4. Грубые ошибки

При обработке наблюдений приходится считаться с возможностью промахов или внешних влияний, дающих совершенно неверные результаты. Один из простейших промахов, например, заключается в том, что наблюдатель прочтет отсчет 20, а запишет 30. Простейший пример внешних причин грубой ошибки — не замеченный наблюдателем толчок, исказивший результат. Наличие грубых ошибок сказывается в том, что среди ряда сравнительно близких результатов есть одно или несколько значений, заметно отличающихся по величине от общего уровня, как говорят, есть «выскакивающие наблюдения». Если это отличие настолько велико, что с уверенностью можно говорить о грубой ошибке, то это измерение сразу отбрасывается. Однако так просто дело обстоит редко. Если считать, что случайная ошибка представляет сумму большого числа малых случайных ошибок, то можно представить себе такую неблагоприятную комбинацию составляющих малых ошибок, которая дает большую по модулю случайную ошибку. Правда, вероятность такой неблагоприятной комбинации очень мала, но она не равна нулю, т. е. такая комбинация в принципе возможна. Поэтому по большей части нельзя сразу признать наблюдение грубым только по признаку «выскакиваний». Ниже будет рассмотрен вероятностный критерий грубых ошибок.

§ 66. Основная гипотеза теории случайных ошибок. Способы оценки ошибок

Необходимость привлечения аппарата теории вероятностей к исследованию случайных ошибок достаточно очевидна. Так как случайная ошибка измерения есть непрерывная случайная величина, то для построения теории должна быть задана ее плотность вероятности или функция распределения.

Обозначим неизвестное точное значение измеряемой величины через a , произвольный результат измерения — через x . Ошибкой назовем величину

$$\delta = x - a. \quad (13.1)$$

Ошибка считается случайной, следовательно, и значения x также будут случайны.

Сделаем следующее предположение: случайные ошибки подчиняются нормальному закону распределения с центром, равным нулю. Согласно этому предположению, плотность вероятности случайных ошибок определяется формулой

$$p(\delta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\delta^2}{2\sigma^2}}, \quad (13.2)$$

откуда непосредственно следует, что плотность вероятности случайных результатов измерений выражается следующим образом:

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (13.3)$$

В выражении (13.3), как известно, σ^2 есть дисперсия случайных ошибок. Величину σ будем в дальнейшем называть *средней квадратичной ошибкой одного измерения*. Она является числовой характеристикой качества совокупности измерений, для которых она задана или вычислена. Вероятностный смысл величины σ известен из общей теории нормального распределения. Как мы видели,

$$P(|\delta| < \sigma) \approx 0,68; \quad P(|\delta| < 3\sigma) \approx 0,9973 \text{ и т. п.}$$

Чем больше σ , тем хуже качество измерений.

Для описания качества измерений пользуются еще тремя величинами:

а) *вероятной ошибкой*

$$r \approx \frac{2}{3} \sigma,$$

которая определяется условием $P(|\delta| < r) = 0,5$;

б) *абсолютной ошибкой*

$$m = E(|\delta|)$$

— математическим ожиданием модуля ошибки (можно показать, что $m \approx 0,8\sigma$ *);

в) *меры точности*:

$$h = \frac{1}{\sigma\sqrt{2}} \approx \frac{0,7}{\sigma}.$$

Вероятная и абсолютная ошибки тем больше, чем хуже качество измерений; мера точности растет с повышением точности измерений.

Предположение о нормальном законе распределения может быть оправдано предельной теоремой Ляпунова, если принять указанное выше представление о суммировании большого числа малых ошибок. Из условий теоремы Ляпунова следует также, что принятие нормального закона означает, что составляющие ошибки должны быть примерно одного порядка. Если можно по физическим соображениям ожидать, что есть выделяющаяся преобладающая ошибка, то нужно сделать дополнительное предположение о законе распределения преобладающей ошибки, сохранив предположение о нормальном законе для совокупности остальных составляющих ошибок, и вывести закон распределения суммы величин **).

Предположение о нормальном законе может быть проверено наблюдениями. Предположим, что найдено приближенное значение величины с такой точностью, что его можно принять за точное значение (с очень малой ошибкой). На практике такое значение может быть выведено из очень большого числа наблюдений. Величина σ может быть определена как корень квадратный из суммы квадратов отклонений «точного» значения от результатов измерений, деленной на число измерений.

Будем, следовательно, считать, что a и σ — точные. По свойству нормального закона можно вычислить вероятность того, что модуль

*) Действительно, по определению математического ожидания

$$m = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} |\delta| e^{-\frac{\delta^2}{2\sigma^2}} d\delta$$

или

$$m = \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} \delta e^{-\frac{\delta^2}{2\sigma^2}} d\delta.$$

Производя замену переменной интегрирования; $\frac{\delta^2}{2\sigma^2} = z$, получим

$$m = \sqrt{\frac{2}{\pi}} \sigma \int_0^{\infty} e^{-z} dz = \sqrt{\frac{2}{\pi}} \sigma \approx 0,8\sigma.$$

**) Подробнее этот вопрос рассмотрен в курсе В. Л. Гончарова.

ошибки находится в любых заданных границах,

$$P(|\delta| < \alpha\sigma) = 2\Phi(\alpha), \quad \alpha > 0,$$

где $\Phi(z)$ — табличная функция (см. гл. 11) (здесь границы берутся в единицах σ). Давая разные значения α в пределах от 0 до 3, подсчитаем число n_α значений δ , удовлетворяющих написанному выше неравенству. Можно вычислить также теоретическое число таких значений

$$n'_\alpha = n \cdot 2\Phi(\alpha),$$

где n — общее число наблюдений. Сравнение таблицы значений n'_α и n_α и даст представление о приемлемости или неприемлемости нормального закона. Такая проверка делалась, ее результаты приведены в книге А. Н. Крылова «Лекции о приближенных вычислениях». Там же можно познакомиться со своего рода выводом нормального закона на основании ряда гипотез.

Как и во всякой математической теории, можно исходить из других основных предположений, взамен указанного предположения о нормальном законе. Так, например, в одной из работ Гаусса в основу теории случайных ошибок кладется предположение, что наиболее вероятным значением измеряемой величины, полученным из ряда равноточных измерений, является среднее арифметическое. Приняв такой постулат, можно показать, что ошибки подчиняются нормальному закону распределения.

Рассмотрим теперь способ исключения грубых ошибок, основываясь на нормальном законе распределения случайных ошибок. Предположим, что в результате нескольких измерений мы нашли приближенное значение измеряемой величины \bar{x} и средней квадратичной ошибки σ одного измерения. Определим приближенную величину ошибки каждого измерения:

$$e_k = x_k - \bar{x} \approx \delta_k.$$

По свойству нормального распределения

$$P(|\delta| < 3\sigma) = 0,9973,$$

следовательно,

$$P(|\delta| > 3\sigma) = 0,0027.$$

Обычно считают маловероятным, чтобы модуль ошибки превысил 3σ . Поэтому, если найдется какое-нибудь из e_k , модуль которого превышает 3σ , то соответствующее измерение считается содержащим грубую ошибку и отбрасывается. В некоторых видах задач вводят условие отбрасывать измерения, у которых $|e_k| > 2\sigma$. После отбрасывания некоторых наблюдений, необходимо перевычислить и приближенное значение \bar{x} и число σ .

ГЛАВА 14

ОБРАБОТКА РАВНОТОЧНЫХ ИЗМЕРЕНИЙ ОПРЕДЕЛЕННОЙ ВЕЛИЧИНЫ

§ 67. Задача обработки измерений определенной величины

Пусть для определения неизвестной величины a сделано n измерений, давших числа x_1, x_2, \dots, x_n . Требуется только из этой совокупности чисел вывести наиболее приемлемое в каком-то смысле приближенное значение a и приближенное значение средней квадратичной ошибки одного измерения. За «наиболее приемлемое» значение принимается обычно наиболее вероятное. Та величина, которая будет принята за a , есть функция случайных чисел x_1, x_2, \dots, x_n , поэтому ее тоже следует считать случайной. Но в таком случае появляется задача о вероятностной оценке точности результата. Иными словами, кроме средней квадратичной ошибки одного измерения следует еще вычислить *среднюю квадратичную ошибку наиболее вероятного значения измеряемой величины*.

Однако прежде всего нам надо ввести понятие о равноточности измерений, являющемся вероятностным выражением «обычного» понятия одинаковой точности всех результатов x_1, x_2, \dots, x_n измерений определенной величины. Каждое измерение номер k ($k=1, 2, \dots$) дает случайный результат x_k ; случайность проявляется в том, что при повторении такой же серии измерений k -е измерение дало бы новое значение x'_k , отличное, вообще говоря, от x_k . Совокупность возможных результатов k -го измерения в общем случае определяется плотностью вероятности с параметром σ_k . Если все измерения *равноточные*, то формальным признаком этого является равенство всех σ_k одному и тому же числу σ независимо от значка k .

§ 68. Наиболее вероятное значение измеряемой величины. Способ наименьших квадратов

Пусть произведено n равноточных измерений x_1, x_2, \dots, x_n определенной величины a . Предположим временно, что величина a известна. Тогда можно определить ошибки отдельных измерений

$$\delta_k = x_k - a. \quad (14.1)$$

Будем считать также (в этом параграфе), что известна средняя квадратичная ошибка одного измерения σ . Пользуясь свойствами плотности вероятности, можно написать приближенное выражение вероятности того, что в k -м измерении ошибка близка к полученной:

$$P_k(\delta_k < \delta < \delta_k + \Delta\delta) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\delta_k^2}{2\sigma^2}} \Delta\delta \quad (k = 1, 2, \dots, n), \quad (14.2)$$

где $\Delta\delta$ — произвольное малое положительное число. Условимся записывать это равенство несколько короче:

$$P_k(\delta \approx \delta_k) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\delta_k^2}{2\sigma^2}} \Delta\delta.$$

Теперь откажемся от предположения, что величина a точно известна, и будем делать разные гипотезы относительно значения величины a . Всякой такой гипотезе будет соответствовать некоторая совокупность чисел δ_k и соответствующие вероятности P_k . Условная вероятность P_k события « $\delta \approx \delta_k$ в k -м измерении» при произвольной гипотезе о величине a определится равенством

$$P_k(\delta \approx \delta_k | a) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\delta_k^2}{2\sigma^2}} \Delta\delta, \quad (14.3)$$

где левую часть следует читать так: вероятность того, что ошибка k -го измерения приблизительно равна δ_k (заключена между δ_k и $\delta_k + \Delta\delta$) при сделанной гипотезе относительно величины a .

Аналогичные равенства можно написать для всех измерений. Каждое из них определяет вероятность только что указанного события. Предположим, что измерения взаимно независимы, т. е. вероятность каждого значения ошибки в k -м измерении не зависит от величин ошибок в остальных измерениях. Соответствующие события (появление ошибок разной величины) также взаимно независимы.

Вычислим теперь условную вероятность получить определенную совокупность ошибок $\delta_1, \delta_2, \dots, \delta_n$ при некоторой гипотезе о величине a . Эту вероятность можно обозначить так: $P(\delta \approx \delta_1, \delta_2, \dots, \delta_n | a)$. Так как мы имеем совокупность взаимно независимых событий (ошибок), то нужная вероятность может быть найдена по теореме умножения вероятностей для взаимно независимых событий; пользуясь (14.3), получаем

$$P(\delta \approx \delta_1, \delta_2, \dots, \delta_n | a) \approx \left(\frac{\Delta\delta}{\sigma\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{\sum_{k=1}^n \delta_k^2}{2\sigma^2} \right\}. \quad (14.4)$$

Поскольку, согласно (14.1), $\delta_k = x_k - a$, то

$$P(X \approx x_1, x_2, \dots, x_n | a) \approx \left(\frac{\Delta \delta}{\sigma \sqrt{2\pi}} \right)^n \exp \left\{ - \frac{\sum_{k=1}^n (x_k - a)^2}{2\sigma^2} \right\}. \quad (14.5)$$

Правая часть этого равенства есть функция величины a . Если придавать разные значения величине a , то мы будем получать различные вероятности.

Предположим теперь, что все гипотезы относительно величины a в некоторой произвольной области равновероятны. Тогда, согласно следствию из формулы Бейеса, вероятности разных гипотез после появления события пропорциональны условным вероятностям события при этих гипотезах. У нас событием является появление чисел x_1, x_2, \dots, x_n ; поэтому можно написать

$$P(a | x_1, x_2, \dots, x_n) = C \left(\frac{\Delta \delta}{\sigma \sqrt{2\pi}} \right)^n \exp \left\{ - \frac{\sum_{k=1}^n (x_k - a)^2}{2\sigma^2} \right\}, \quad (14.6)$$

где C — коэффициент пропорциональности.

Согласно равенству (14.6) вероятность гипотезы о величине a зависит от значения a , входящего в показатель правой части. Легко видеть, что вероятность P имеет наибольшее значение, если величина

$$S(a) = \sum_{k=1}^n (x_k - a)^2 \quad (14.7)$$

будет наименьшей. Задача определения наиболее вероятного значения a сводится, таким образом, к нахождению такого a , при котором $S(a)$ имеет минимум. Как известно из анализа, необходимым условием экстремума $S(a)$ является равенство

$$\frac{dS}{da} = -2 \sum_{k=1}^n (x_k - a) = 0.$$

Так как $S(a)$ есть положительно определенная квадратичная форма относительно a , то S может иметь только один экстремум и притом минимум. Приравнявая нулю первую производную, получим наиболее вероятное значение неизвестной величины a :

$$a_{\text{вер}} = \frac{\sum_{k=1}^n x_k}{n}. \quad (14.8)$$

Таким образом, мы приходим к следующему заключению: *наиболее вероятное значение определяемой величины из равноточных измерений есть среднее арифметическое.*

В дальнейшем вместо $a_{\text{вер}}$ мы будем пользоваться обозначением \bar{x} :

$$\bar{x} = \frac{\sum_{k=1}^n x_k}{n}. \quad (14.9)$$

Описанный способ вычисления приближенного значения величины a называют *способом наименьших квадратов*, поскольку в нем используется условие минимума суммы квадратов ошибок $S(a)$.

Из (14.9) вытекает следующее, полезное для дальнейшего, свойство среднего арифметического: *сумма всех отклонений от среднего арифметического равна нулю*. Действительно, согласно (14.9)

$$\sum_{k=1}^n (x_k - \bar{x}) = \sum_{k=1}^n x_k - n\bar{x} = 0. \quad (14.10)$$

§ 69. Средняя квадратичная ошибка среднего арифметического

В данной серии из n наблюдений среднее арифметическое \bar{x} есть линейная функция их результатов x_1, x_2, \dots, x_n [см. (14.8)]. Если заново сделать серию из n измерений, то вследствие влияния случайных причин значения x_k будут отличаться от полученных в первой серии, и потому значение \bar{x} будет новым. То же относится и ко всяким последующим сериям измерений. Поэтому число \bar{x} , полученное из одной серии измерений, является случайным приближенным значением искомой величины. Чтобы составить себе представление о возможных отклонениях \bar{x} от искомого точного значения, следует вычислить *среднюю квадратичную ошибку среднего арифметического*. При этом для понимания существа дела нужно уяснить себе, что средняя квадратичная ошибка среднего арифметического является вероятностной характеристикой совокупности всевозможных значений средних арифметических из n измерений.

Так как результаты измерений по условию представляют взаимно независимые случайные величины, то применение теоремы о дисперсии линейной функции таких величин дает

$$D(x) = \sum_{k=1}^n \frac{D(x_k)}{n^2}.$$

Поскольку измерения по условию равноточны, то

$$D(x_k) = \sigma^2$$

при всех k . Поэтому

$$D(\bar{x}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

и

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}. \quad (14.11)$$

Следует отметить, что эта формула совершенно точна, но сама величина σ нам неизвестна.

Исключением можно считать редкие случаи, когда производятся опытные измерения высокой точности, при которых случайные ошибки очень малы. В таком случае результаты измерений будут многозначными числами, имеющими ряд общих первых значащих цифр. Например, 3,48754, 3,48733, 3,48761 и т. д. За точное значение измеряемой величины в подобном случае можно принять общую часть всех результатов, в нашем примере 3,487 с четырьмя верными знаками. Если производятся измерения с меньшей точностью («рядовые» измерения), то отдельные измерения не будут давать все четыре верных знака, например получим 3,485, 3,488 и т. п. Отклонения от «точного» значения можно считать точными ошибками («точными» в условном смысле, именно с принятым числом значащих цифр). Средняя квадратичная ошибка одного измерения на основании n измерений при этих условиях определится как корень квадратный из суммы квадратов ошибок, деленной на число измерений. Надо, однако, заметить, что таким образом мы все-таки не получим «точное» значение σ . Величина σ есть параметр нормального закона распределения ошибок. Чтобы найти его достаточно уверенно, надо было бы сделать много подобных измерений и из них выводиться σ .

§ 70. Наиболее вероятное значение средней квадратичной ошибки одного измерения

При выводе наиболее вероятного значения измеряемой величины предполагалось, что величина σ известна. Поэтому вероятность получить некоторую совокупность ошибок, равная вероятности получить совокупность наших результатов измерений, зависела только от гипотезы о величине a .

Предположим теперь, что и величина σ также неизвестна. При рассмотрении вероятности совокупности ошибок придется тогда делать гипотезы и о значении a и о значении σ . Так же как и в § 68, вероятность получить результаты измерений x_1, x_2, \dots, x_n при гипотезах о значениях a, σ определяется формулой

$$P(X \approx x_1, x_2, \dots, x_n | a, \sigma) = \left(\frac{\Delta b}{\sqrt{2\pi}} \right)^n \sigma^{-n} \exp \left[-\frac{S(a)}{2\sigma^2} \right], \quad (14.12)$$

где

$$S(a) = \sum_{k=1}^n (x_k - a)^2. \quad (14.13)$$

Преобразуем величину $S(a)$ следующим образом:

$$\begin{aligned} S(a) &= \sum_{k=1}^n (x_k - a)^2 = \sum_{k=1}^n [(x_k - \bar{x}) + (\bar{x} - a)]^2 = \\ &= \sum_{k=1}^n (x_k - \bar{x})^2 + 2(\bar{x} - a) \sum_{k=1}^n (x_k - \bar{x}) + n(\bar{x} - a)^2. \end{aligned}$$

По свойству среднего арифметического (14.10) имеем точное равенство:

$$\sum_{k=1}^n (x_k - \bar{x}) = 0$$

(если \bar{x} точно вычислено). Поэтому

$$S(a) = \sum_{k=1}^n (x_k - \bar{x})^2 + n(\bar{x} - a)^2.$$

После этого вероятность совокупности измерений можно записать так:

$$\begin{aligned} P(X \approx x_1, x_2, \dots, x_n | a, \sigma) &= \\ &= \left(\frac{\Delta b}{\sqrt{2\pi}} \right)^n \frac{\sqrt{2\pi}}{\sqrt{n}} \sigma^{-n+1} \exp \left[- \frac{\sum_{k=1}^n (x_k - \bar{x})^2}{2\sigma^2} \right] \times \\ &\times \left\{ \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} \exp \left[- \frac{(a - \bar{x})^2}{2 \left(\frac{\sigma}{\sqrt{n}} \right)^2} \right] \right\}. \quad (14.14) \end{aligned}$$

Теперь искомая вероятность есть функция двух неизвестных параметров a и σ .

Чтобы получить правило нахождения наиболее вероятного значения σ , пригодное при всех a , положим, что все значения a равновероятны в некоторой области от $-a$ до $+a$, достаточно большой, чтобы можно было считать ее практически эквивалентной области от $-\infty$ до $+\infty$. Для этого достаточно, чтобы a было величиной порядка $3\sigma - 4\sigma$.

Умножим обе части последнего равенства на da и проинтегрируем по a от $-\infty$ до $+\infty$. По обобщенной теореме сложения вероятностей в левой части получим вероятность совокупности измерений при всех a и фиксированном σ . В правой части надо интегрировать только последний множитель, заключенный в скобку, так как лишь он содержит a . Формально этот множитель есть нормальная плотность вероятности «случайной величины» a . Поэтому интеграл от $-\infty$ до $+\infty$ даст единицу, и мы получим

$$P(X \approx x_1, x_2, \dots, x_n | \sigma) = \left(\frac{\Delta b}{\sqrt{2\pi}} \right)^n \frac{\sqrt{2\pi}}{\sqrt{n}} \exp \left(- \frac{\bar{S}}{2\sigma^2} \right) \sigma^{-n+1}, \quad (14.15)$$

где

$$\bar{S} = \sum_{k=1}^n (x_k - \bar{x})^2. \quad (14.16)$$

Применим рассуждение, подобное рассуждению § 68. Мы имеем вероятность появления наших результатов измерений при выбранной гипотезе о величине σ и любых a ; это есть условная вероятность события $(X \approx x_1, x_2, \dots, x_n)$ при гипотезе о величине σ . Предположим, что до наблюдений все гипотезы о значении σ равновероятны. По следствию из формулы Бейеса вероятности гипотез о значении σ после появления события x_1, x_2, \dots, x_n пропорциональны условным вероятностям события при гипотезах. Поэтому

$$P(\sigma | x_1, x_2, \dots, x_n) = C\sigma^{-n+1}e^{-\frac{\bar{S}}{2\sigma^2}}, \quad (14.17)$$

где в C входит и коэффициент пропорциональности, входящий в выражение следствия из формулы Бейеса и все множители, не зависящие от σ .

Теперь можно поставить задачу об определении наиболее вероятной величины σ , т. е. о наиболее вероятном значении средней квадратичной ошибки одного измерения, которое можно получить из данной совокупности измерений. Для этого нужно найти то значение σ , при котором функция

$$P(\sigma) = \sigma^{-n+1}e^{-\frac{\bar{S}}{2\sigma^2}}$$

имеет максимум. Логарифмируя $P(\sigma)$, а затем дифференцируя два раза, получим

$$\begin{aligned} \ln P &= (1-n) \ln \sigma - \frac{\bar{S}}{2\sigma^2}, \\ \frac{1}{P} \frac{dP}{d\sigma} &= \frac{1-n}{\sigma} + \frac{\bar{S}}{\sigma^3}, \\ -\frac{1}{P^2} \left(\frac{dP}{d\sigma}\right)^2 + \frac{1}{P} \frac{d^2P}{d\sigma^2} &= \frac{n-1}{\sigma^2} - \frac{3\bar{S}}{\sigma^4}. \end{aligned}$$

Приравнявая первую производную нулю, получим

$$\sigma_{\text{вер}}^2 = \frac{\bar{S}}{n-1}. \quad (14.18)$$

Подставляя эти значения σ в равенство, содержащее вторую производную, и учитывая, что при $\sigma = \sigma_{\text{вер}}$ $\left(\frac{dP}{d\sigma}\right) = 0$, имеем

$$\left(\frac{1}{P} \frac{d^2P}{d\sigma^2}\right)_{\sigma=\sigma_{\text{вер}}} = -2 \frac{(n-1)^2}{\bar{S}} < 0, \quad (14.19)$$

а это подтверждает, что полученное значение σ наиболее вероятно.

В дальнейшем вместо $\sigma_{\text{вер}}$ мы будем писать просто σ . Подставляя в (14.18) значение \bar{S} из (14.16), получаем

$$\sigma = \sqrt{\frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n-1}}. \quad (14.20)$$

После вычисления \bar{x} , σ надо просмотреть таблицу значений $\epsilon_k = x_k - \bar{x}$; если среди них найдутся такие, модуль которых превышает 3σ , их нужно отбросить и заново вычислить \bar{x} и σ .

§ 71. Второй вывод приближенного значения измеряемой величины и приближенного значения средней квадратичной ошибки одного измерения

Будем считать совокупность измерений выборкой в n экземпляров из бесчисленного множества возможных различных результатов вследствие наличия случайных ошибок. В основу теории положим два допущения:

- 1) математическое ожидание каждого результата измерения равно точному значению измеряемой величины;
- 2) дисперсия каждого результата измерения одна и та же для всех измерений.

Первое допущение можно считать естественным; второе означает, что мы имеем дело с равноточными измерениями. Никакого предположения о законе распределения случайных ошибок измерений мы делать не будем.

Согласно следствию из теоремы Чебышева при достаточно большом числе измерений с вероятностью, как угодно близкой к единице, можно утверждать, что модуль разности между точным значением и средним арифметическим из выборочных значений будет сколь угодно мал. Это дает основание для того, чтобы за приближенное значение измеряемой величины принять среднее арифметическое. Следовательно, постулируется приближенное равенство

$$a \approx \bar{x} = \frac{\sum_{k=1}^n x_k}{n}.$$

При большом числе наблюдений такое приближение в известной мере оправдано теоремой Чебышева. При малом числе наблюдений оно будет просто соглашением.

Можно, однако, применить критерий, предложенный А. А. Марковым, о приближении, не содержащем систематической погрешности. Обозначим через ξ любое приближение величины a ; это приближение не содержит систематической погрешности, если выполняется равенство: математическое ожидание приближения равно математическому ожиданию приближаемой величины (т. е. в нашей задаче — точному значению). Обозначим случайные результаты измерений $\xi_1, \xi_2, \dots, \xi_n$; их среднее арифметическое:

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}.$$

Согласно первому предположению

$$E(\bar{\xi}) = \frac{a + a + \dots + a}{n} = a.$$

Таким образом, приняв за приближенное значение среднее арифметическое измеренных значений, мы удовлетворим критерию Маркова.

Для получения приближенного значения средней квадратичной ошибки одного измерения будем рассуждать так. Найдем для нашей выборки математическое ожидание суммы квадратов отклонений отдельных выборочных значений от их среднего арифметического, т. е. величину

$$\bar{\Sigma} = E \left\{ \sum_{k=1}^n (\xi_k - \bar{\xi})^2 \right\}.$$

Преобразуем правую часть, раскрывая круглые скобки и используя равенство $\sum \xi_k = n\bar{\xi}$. Тогда получим:

$$\bar{\Sigma} = E \left\{ \sum_{k=1}^n \xi_k^2 - n\bar{\xi}^2 \right\} = \sum_{k=1}^n E(\xi_k^2) - nE(\bar{\xi}^2).$$

Так как $a = E(\xi_k) = E(\bar{\xi})$, то по свойству дисперсии для случайных величин ξ_k ($k = 1, 2, \dots, n$) имеем:

$$E(\xi_k^2) = \sigma^2 + a^2,$$

$$E(\bar{\xi}^2) = \sigma_{\bar{\xi}}^2 + a^2.$$

Величина $\sigma_{\bar{\xi}}^2$ определяется по теореме о дисперсии линейной функции, поскольку

$$\bar{\xi} = \sum \frac{\xi_k}{n};$$

значит,

$$\sigma_{\bar{\xi}}^2 = \sum \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Поэтому

$$E(\bar{\xi}^2) = \frac{\sigma^2}{n} + a^2,$$

$$\bar{\Sigma} = n(\sigma^2 + a^2) - n\left(\frac{\sigma^2}{n} + a^2\right) = (n-1)\sigma^2.$$

Таким образом, мы приходим к точной формуле: математическое ожидание суммы квадратов отклонений выборочных случайных значений величины от среднего арифметического этих значений равно произведению дисперсии одного измерения на число элементов выборки без единицы.

Математическое ожидание суммы квадратов отклонений от среднего арифметического есть среднее значение этой величины, которое можно было бы получить из всевозможных выборок. Вместо среднего значения суммы квадратов отклонений подставим в точную формулу то ее значение, которое получается из одной выборки. Тогда получим приближенную формулу для вычисления дисперсии одного измерения:

$$\sigma^2 = \frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n-1}$$

(вместо ξ_k и $\bar{\xi}$ написано x_k и \bar{x} из одной выборки). Отметим, что при выводе этой формулы не делалось никаких предположений о законе распределения случайных ошибок.

§ 72. Пример и схема обработки равноточных измерений одной величины

Географическая широта Ташкентской обсерватории определялась по наблюдениям 14 звезд. Результаты даны в столбце II приведенной ниже схемы.

I	II	III	IV	V	VI
N п/п	φ_k	x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$	
1	41°19'33",10	+ 3",10	+ 1",68	2,82	$\varphi_0 = 41^\circ 19' 30''$; $x_k = \varphi_k - \varphi_0$; $\bar{x} = 19'',87 : 14 = 1'',42$; $\bar{\varphi} = 41^\circ 19' 30'' + 1'',42 =$ $= 41^\circ 19' 31'',42$; $\sigma^2 = 50,28 : (14 - 1) =$ $= 3,87$; $\sigma = 1'',97$; $\sigma_{\bar{\varphi}} = 1'',97 : \sqrt{14} = 0'',53$; $\bar{\varphi} = 41^\circ 19' 31'',42 \pm 0'',53$; $\varphi = 41^\circ 19' 31'',4 \pm 0'',5$
2	32,96	+ 2,96	+ 1,54	2,37	
3	28,95	— 1,05	— 2,47	6,10	
4	30,08	+ 0,08	— 1,34	1,80	
5	28,81	— 1,19	— 2,61	6,81	
6	31,16	+ 1,16	— 0,26	0,068	
7	31,38	+ 1,38	— 0,04	0,002	
8	32,15	+ 2,15	+ 0,73	0,53	
9	34,89	+ 4,89	+ 3,47	12,04	
10	33,34	+ 3,34	+ 1,92	3,69	
11	32,39	+ 2,39	+ 0,97	0,94	
12	31,51	+ 1,51	+ 0,09	0,008	
13	31,35	+ 1,35	— 0,07	0,005	
14	27",80	— 2",20	— 3",62	13,10	
		+ 19",87	(— 0",01)	50,28	

Поскольку наблюдения производились примерно в одинаковых условиях, то их можно считать равноточными. В столбце III даны отклонения полученных значений широты от значения $41^\circ 19' 30''$, обозначенные буквой x_k ($k = 1, 2, \dots, 14$). Это сделано лишь в целях упрощения вычислений. Введение этих чисел взамен φ_k эквивалентно изменению начала отсчета широт, что нужно учесть только при выводе наиболее вероятного значения широты из всех наблюдений; на вычислении средних квадратичных ошибок введение нового начала отсчета не скажется, ибо, как легко убедиться, при переносе начала отсчета отклонение от среднего не изменится. В столбце IV даны отклонения значений x_k от среднего \bar{x} . По только что указанным причинам эти же числа представляют собой отклонения чисел φ_k от их среднего значения. Сумма чисел столбца IV вычисляется для контроля: если бы среднее \bar{x} было вычислено формально точно (когда $\sum x_k$ точно делится на n при том числе знаков после запятой, какое берется в вычислениях), то эта сумма должна точно равняться нулю. Обычно среднее вычисляется

с предельной погрешностью, равной половине единицы последнего знака. Числа x_k в этой контрольной операции следует полагать точными. Поэтому сумма чисел IV столбца не должна превышать произведения половины единицы последнего знака в \bar{x} на число измерений.

В столбце V даны квадраты отклонений; их сумма нужна для вычисления средних квадратичных ошибок. На практике можно не выписывать квадратов отдельных отклонений, а получить всю сумму квадратов отклонений на арифмометре способом «накопления». Так как мы лишаемся при этом контроля суммы квадратов, то контроль следует выполнить другим способом, воспользовавшись тождеством, являющимся следствием формулы для вычисления дисперсии

$$\sum_{k=1}^n (x_k - \bar{x})^2 = \sum_{k=1}^n x_k^2 - \bar{x}^2 n.$$

Оно будет выполняться почти точно, ибо при вычислении сумм квадратов на арифмометре обычно берутся все формально получаемые знаки, и расхождение будет только за счет неточности среднего \bar{x} . В каждом конкретном случае можно учесть предельное расхождение между обеими частями тождества.

Сделаем некоторые пояснения к вычислениям вне схемы.

Все вычисления наиболее вероятного значения широты и средних квадратичных ошибок помещены в столбце VI. Сначала выписывается число φ_0 , которое упомянуто выше. Затем вычисляем среднее значение \bar{x} . Прибавив это число к φ_0 , получим среднее арифметическое, т. е. наиболее вероятное значение широты. Разделив сумму квадратов отклонений от среднего на число измерений без единицы, находим квадрат наиболее вероятного значения средней квадратичной ошибки одного измерения. Деление σ на корень квадратный из числа измерений дает среднюю квадратичную ошибку результата. Результат записывают обычно так, как показано в предпоследней строке столбца VI; иногда в такой записи указывается вероятная (а не средняя квадратичная) ошибка, поэтому следует указывать, какая ошибка дается в результате.

Из величины средней квадратичной ошибки среднего значения видно, что в результате сотые доли секунды совершенно ненадежны. Поэтому окончательный результат лучше дать до десятых долей секунды, как и сделано в последней строке столбца VI. Вероятностный смысл результата вытекает из свойств нормального распределения: с вероятностью 0,68 действительное значение широты заключено между $41^\circ 19' 30'',9$ и $41^\circ 19' 31'',9$; с вероятностью 0,9973 — между $41^\circ 19' 29'',9$ и $41^\circ 19' 32'',9$.

Так как при выводе средних ошибок необходимо надежное значение суммы квадратов отклонений, то проверим его по указанному

выше тождеству. Получаем

$$\sum_{k=1}^n (x_k - \bar{x})^2 = 78,49 - 28,23 = 50,26.$$

Разница между начальным значением и контрольным ничтожна, поэтому можно быть уверенным в правильности вычислений.

Наконец, с помощью средней квадратичной ошибки одного измерения σ можно проверить, нет ли грубых ошибок в измерениях. Для этого находим отношение наиболее крупного по модулю из отклонений от среднего $x_k - \bar{x}$ ($k = 1, 2, \dots, 14$) к средней квадратичной ошибке одного измерения. В нашем примере имеем $3,62 : 1,97 = 1,84$. Это число меньше трех, поэтому можно полагать, что грубых ошибок нет.

ГЛАВА 15

ОБРАБОТКА НЕРАВНОТОЧНЫХ ИЗМЕРЕНИЙ ОПРЕДЕЛЕННОЙ ВЕЛИЧИНЫ

§ 73. Понятие о неравноточных измерениях. Веса измерений

В практической работе нередко бывают случаи, когда для возможно более надежного определения некоторой величины собирают измерения различного происхождения, т. е. сделанные разными инструментами и разными методами. Особенно это относится к фундаментальным постоянным астрономии. Одной из таких фундаментальных постоянных является среднее расстояние Земли от Солнца или, что то же, параллакс Солнца. Солнечный параллакс определяется рядом способов, дающих несколько различающиеся результаты.

Простейший случай неравноточных измерений представляет собрание не прямых измерений, а выводов из равноточных измерений, число которых различно в разных выводах.

Пусть произведено m_1 равноточных измерений, из которых выведено наиболее вероятное значение x_1 измеряемой величины. Пусть другая серия таких же измерений, сделанная на той же или другой обсерватории, содержала m_2 измерений и дала число x_2 , третья — m_3 измерений и дала результат x_3 и т. д. Ставится задача вывести наиболее вероятное значение измеряемой величины на основании результатов x_1, x_2, \dots, x_n .

Положим, что основные измерения, из которых выведены эти числа, равноточны (например, делались однотипными инструментами, хотя и на разных обсерваториях), и обозначим через σ среднюю квадратичную ошибку одного измерения. Согласно формуле (14.11) § 69 для средних квадратичных ошибок чисел x_1, x_2, \dots, x_n имеем выражения

$$\sigma_k = \frac{\sigma}{\sqrt{m_k}}, \quad k = 1, 2, \dots, n. \quad (15.1)$$

Так как средние квадратичные ошибки заданных значений x_1, x_2, \dots, x_n различны, то эти величины не могут считаться равноточными.

В других случаях заданные значения x_1, x_2, \dots, x_n должны считаться неравноточными по сведениям о различных условиях измерений, о различной точности инструментов и. т. п.

Средние квадратичные ошибки $\sigma_1, \sigma_2, \dots, \sigma_n$ тех измерений, которые дали эти значения, известны не во всех задачах такого типа. Иногда на основании общих сведений об условиях измерений при способах определения x_1, x_2, \dots, x_n известно, что измерения неравноточны, но определенных числовых критериев неравноточности нет.

Формальным признаком неравной точности измерений является то, что средние квадратичные ошибки $\sigma_1, \sigma_2, \dots, \sigma_n$ измерений тех типов, которые дали случайные значения x_1, x_2, \dots, x_n , оказываются различными. Положим, что $\sigma_1, \dots, \sigma_n$ известны.

При рассмотрении неравноточных измерений оказывается удобным ввести вместо величин $\sigma_1, \sigma_2, \dots, \sigma_n$ числа p_1, p_2, \dots, p_n , называемые *веса*ми измерений и определяемые равенствами

$$p_k = \frac{\sigma_0^2}{\sigma_k^2}, \quad (15.2)$$

где σ_0^2 — произвольное положительное число. Из определения весов следует:

а) веса неравноточных измерений обратно пропорциональны их дисперсиям (квадратам средних квадратичных ошибок);

б) веса — числа относительные, т. е. все веса можно умножить или разделить на одно и то же число (вследствие произвольности числа σ_0^2).

Коэффициенту пропорциональности σ_0^2 можно придать простой смысл: если существует такое $k = \frac{\sigma_0^2}{\sigma_k^2}$, что $\sigma_k = \sigma_0$, то $p_k = 1$; поэтому можно сказать, что σ_0 есть *средняя квадратичная ошибка измерения, имеющего единицу веса*. Иногда для сокращения говорят, что σ_0 есть *средняя квадратичная ошибка на единицу веса*. Иногда выбирают σ_0 так, чтобы сумма весов всех измерений равнялась единице; тогда все веса будут дробными числами. Чаще в астрономической практике берут σ_0 таким, чтобы все веса были целыми числами.

Существенно отметить один случай, упомянутый выше, когда веса определяются сразу. Пусть числа x_1, x_2, \dots, x_n представляют собой средние арифметические, полученные из m_1, m_2, \dots, m_n равноточных наблюдений. Из формул (15.1) для σ_k видно, что σ_k^2 обратно пропорциональны числам m_k . Но согласно (15.2) веса обратно пропорциональны σ_k^2 , поэтому веса в данном случае прямо пропорциональны числам наблюдений m_k . Обычно за веса принимают в этом случае те числа наблюдений, из которых получены средние. Это можно представить еще в такой форме: каждое

среднее, полученное из m_k наблюдений, эквивалентно m_k равно- точным значениям, поэтому m_k и есть «вес» числа x_k .

Таким образом, способ определения весов неравноточных изме- рений состоит в следующем:

1. Если известны σ_k , то выбираем произвольно σ_0 , лучше всего значение, близкое к среднему из σ_k . Веса вычисляем по формуле

$$p_k = \frac{\sigma_0^2}{\sigma_k^2}.$$

2. Если x_1, x_2, \dots, x_n — средние арифметические из m_1, m_2, \dots, m_n равноточных измерений, то эти числа и принимаются за веса.

3. Если не имеют места случаи 1) или 2) и наблюдения по ка- ким-либо причинам следует считать не равно точны ми, то веса на- значаются исследователем в известной мере произвольно, но так, чтобы более точные наблюдения имели бoльшие веса.

§ 74. Наиболее вероятное значение измеряемой величины

Пусть даны значения x_k с весами p_k . Независимо от способа получения весов мы можем по определению весов (15.2) написать

$$p_k = \frac{\sigma_0^2}{\sigma_k^2},$$

где σ_0^2 — пока неизвестный множитель (но определенный, так как выбор определенных чисел p_k означает, что числу σ_0^2 придается определенное значение). Пока будем считать, что σ_0 задано. По ве- сам и σ_0 определяем средние квадратичные ошибки измерений

$$1:\sigma_k = \frac{\sqrt{p_k}}{\sigma_0}. \quad (15.3)$$

Если заданы не веса, а числа σ_k , то задаем σ_0 и вычисляем веса по той же формуле. Нормальный закон распределения ошибок для каждого измерения может быть записан в форме

$$f_k(\delta) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp \left[-\frac{\delta^2}{2\sigma_k^2} \right] = \frac{\sqrt{p_k}}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{p_k \delta^2}{2\sigma_0^2} \right].$$

Пусть сделана некоторая гипотеза относительно искомой вели- чины a ; тогда для каждого измерения можно определить ошибку

$$\delta_k = x_k - a$$

и вычислить вероятность получить ошибку, близкую к δ_k : $P(\delta_k < \delta < \delta_k + \Delta\delta)$ или, короче, $P(\delta \approx \delta_k)$.

По нормальному закону:

$$P(\delta \approx \delta_k | a) = \frac{\sqrt{p_k} \Delta \delta}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{p_k \delta_k^2}{2\sigma_0^2} \right].$$

Полагая, как и в предыдущей главе, результаты измерений взаимно независимыми случайными величинами, найдем по теореме умножения вероятностей вероятность совокупности ошибок, близких к $\delta_1, \delta_2, \dots, \delta_n$:

$$\begin{aligned} P(\delta \approx \delta_1, \delta_2, \dots, \delta_n | a) &= \\ &= \left(\frac{\Delta \delta}{\sigma_0 \sqrt{2\pi}} \right)^n \sqrt{p_1 p_2 \dots p_n} \exp \left[-\frac{\sum p_k \delta_k^2}{2\sigma_0^2} \right]. \end{aligned} \quad (15.4)$$

При выбранной гипотезе относительно величины a ошибки $\delta_1, \delta_2, \dots, \delta_n$ однозначно определяются числами x_1, x_2, \dots, x_n . Поэтому

$$P(\delta \approx \delta_1, \delta_2, \dots, \delta_n) = P(x \approx x_1, x_2, \dots, x_n | a),$$

т. е. правая часть предыдущего равенства определяет условную вероятность совокупности результатов измерений, близких к полученным при заданном a . Эта вероятность есть функция величины a . Предположим, что до измерений все гипотезы относительно величины a равновероятны в некоторой области. По следствию из формулы Байеса апостериорные вероятности гипотез после появления события пропорциональны условным вероятностям события при гипотезах. У нас случайным событием является совокупность результатов измерений (x_1, x_2, \dots, x_n) . Поэтому можно написать

$$P(a | x_1, x_2, \dots, x_n) = C \exp \left[-\frac{\sum_{k=1}^n p_k \delta_k^2}{2\sigma_0^2} \right], \quad (15.5)$$

где в C входит и коэффициент пропорциональности, и множители

$$\left(\frac{\Delta \delta}{\sigma_0 \sqrt{2\pi}} \right)^n \sqrt{p_1 p_2 \dots p_n}.$$

Из написанного равенства следует, что наиболее вероятная гипотеза об a будет при условии, что величина

$$S(a) = \sum_{k=1}^n p_k \delta_k^2. \quad (15.6)$$

имеет минимум. Заменим δ_k на $x_k - a$ и найдем значение a , при котором $S(a)$ имеет минимум. Дифференцируя $S(a)$ и приравнявая производную нулю, имеем

$$\frac{dS(a)}{da} = - \sum_{k=1}^n 2p_k (x_k - a) = 0, \quad a \sum_{k=1}^n p_k = \sum_{k=1}^n p_k x_k. \quad (15.7)$$

Обозначим для сокращения записи

$$p = \sum_{k=1}^n p_k. \quad (15.8)$$

Для наиболее вероятного значения \bar{x}_p величины a из (15.7) получим выражение:

$$\bar{x}_p = \frac{\sum_{k=1}^n p_k x_k}{p}. \quad (15.9)$$

Эта величина называется *средним весовым из неравноточных измерений* (иногда говорят — *среднее взвешенное*). Таким образом, при неравноточных измерениях за наиболее вероятное значение измеряемой величины следует принять *среднее весовое результатов измерений*.

Формула (14.9) есть частный случай только что написанной формулы; если измерения равноточные, то можно принять вес каждого измерения равным единице, и получим простое среднее.

Легко убедиться в том, что весовое среднее обладает следующим свойством: *взвешенная сумма всех отклонений от весового среднего равна нулю*. Действительно, на основании (15.9) имеем

$$\begin{aligned} \sum_{k=1}^n p_k (x_k - \bar{x}_p) &= \sum_{k=1}^n p_k x_k - \bar{x}_p \sum_{k=1}^n p_k = \\ &= \sum_{k=1}^n p_k x_k - \bar{x}_p p = 0. \end{aligned} \quad (15.10)$$

§ 75. Средняя квадратичная ошибка среднего весового

Величины x_1, x_2, \dots, x_n — отдельные значения случайных величин — возможных результатов измерений $\xi_1, \xi_2, \dots, \xi_n$. Поэтому и среднее весовое \bar{x}_p есть одно из значений случайной величины

$$\bar{\xi}_p = \frac{\sum_{k=1}^n p_k \xi_k}{p}.$$

Так как $\bar{\xi}_p$ есть линейная функция величин ξ_k , то $\bar{\xi}_p$ подчиняется нормальному закону распределения по основной гипотезе о нормальном законе для случайных ошибок и по свойству нормального закона. Поэтому для оценки возможного рассеяния значений $\bar{\xi}_p$ достаточно определить дисперсию этой величины. По теореме о дисперсии линейной функции взаимно независимых случайных

величин имеем

$$\sigma_{\bar{\xi}_p}^2 = \frac{1}{p^3} \sum_{k=1}^n p_k^2 \sigma_k^2,$$

где σ_k^2 — дисперсии случайных результатов последовательных измерений. Вводя веса p_k вместо σ_k^2 по формулам (15.2), получим

$$\sigma_{\bar{\xi}_p}^2 = \frac{\sigma_0^2}{p^3} \sum_{k=1}^n p_k = \frac{\sigma_0^2}{p},$$

отсюда

$$\sigma_{\bar{\xi}_p} = \frac{\sigma_0}{\sqrt{p}}. \quad (15.11)$$

Эта формула точна, но средняя квадратичная ошибка на единицу веса неизвестна, кроме тех исключительных случаев, когда тщательно изучается качество измерений прецизионными измерительными приборами. То обстоятельство, что в знаменателе стоит корень квадратный из суммы весов, и является основанием для того, чтобы, пользуясь относительностью весов, выбрать их при условии, что сумма весов равна единице. При соблюдении этого условия средняя квадратичная ошибка среднего весового равна средней квадратичной ошибке на единицу веса.

Формулу (14.11) § 70 можно считать частным видом формулы (15.11); если измерения равноточные, то вес каждого из них равен единице.

§ 76. Наиболее вероятное значение средней квадратичной ошибки измерения с весом единица

В § 74 для нахождения наиболее вероятного значения измеряемой величины определялась по формуле (15.4) условная вероятность совокупности измерений при условии, что сделана гипотеза о значении a и задана величина σ_0 . Для нахождения наиболее вероятного значения средней квадратичной ошибки с весом единица воспользуемся тем же выражением (15.4), считая теперь, что величина σ_0 неизвестна, но делаемся гипотезы об этой величине. В соответствии с этим выделим в (15.4) множители σ_0 и $\frac{1}{\sqrt{2\pi}}$, обозначая остальные множители, не содержащие σ_0 , буквой C' :

$$\begin{aligned} P(x \approx x_1, x_2, \dots, x_n | a, \sigma_0) = \\ = C' \frac{1}{\sigma_0 \sqrt{2\pi}} \sigma_0^{-n+1} \exp \left[-\frac{\sum_{k=1}^n p_k \delta_k^2}{2\sigma_0^2} \right]. \end{aligned}$$

Положим, что все значения a равновероятны и проинтегрируем это выражение по a от $-\infty$ до $+\infty$. Тогда в силу обобщенной формулы полной вероятности мы найдем вероятность совокупности результатов измерений, близких к полученным, при всех значениях a и при гипотезе о значении σ_0 . Обозначая указанную вероятность через $P(x \approx x_1, x_2, \dots, x_n | \sigma_0)$, имеем

$$\begin{aligned} P(x \approx x_1, x_2, \dots, x_n | \sigma_0) = \\ = C_1 \sigma_0^{-n+1} \int_{-\infty}^{+\infty} \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{\sum_{k=1}^n p_k \delta_k^2}{2\sigma_0^2} \right] da. \end{aligned} \quad (15.12)$$

Чтобы вычислить интеграл в (15.12), преобразуем сумму, стоящую в аргументе экспонента:

$$\begin{aligned} \sum_{k=1}^n p_k \delta_k^2 &= \sum_{k=1}^n p_k (x_k - a)^2 = \\ &= \sum_{k=1}^n p_k [(x_k - \bar{x}_p) + (\bar{x}_p - a)]^2 = \\ &= \sum_{k=1}^n p_k (x_k - \bar{x}_p)^2 + 2 \sum_{k=1}^n p_k (x_k - \bar{x}_p)(\bar{x}_p - a) + \\ &\quad + \sum_{k=1}^n p_k (\bar{x}_p - a)^2. \end{aligned}$$

Во втором члене вынесем множитель $(\bar{x}_p - a)$ за знак суммы, как определенное число, хотя и неизвестное (неизвестно a); остающуюся сумму $\sum_{k=1}^n p_k (x_k - \bar{x}_p)$ заменим нулем в силу рассмотренного выше свойства весового среднего (15.10); в третьем члене определенный множитель $(\bar{x}_p - a)^2$ вынесем за знак суммы, а сумму весов $\sum_{k=1}^n p_k$ заменим на p . Тогда

$$\begin{aligned} P(x \approx x_1, x_2, \dots, x_n | \sigma_0) = \\ = C' \sigma_0^{-n+1} \exp \left[-\frac{\sum_{k=1}^n p_k (x_k - \bar{x}_p)^2}{2\sigma_0^2} \right] \times \\ \times \int_{-\infty}^{+\infty} \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{p(\bar{x}_p - a)^2}{2\sigma_0^2} \right] da. \end{aligned} \quad (15.13)$$

Сделаем еще одно очевидное преобразование:

$$\begin{aligned}
 P(x \approx x_1, x_2, \dots, x_n | \sigma_0) &= \\
 &= \frac{C'}{\sqrt{p}} \sigma_0^{-n+1} \exp \left[-\frac{\sum_{k=1}^n p_k (x_k - \bar{x}_p)^2}{2\sigma_0^2} \right] \times \\
 &\quad \times \int_{-\infty}^{+\infty} \frac{1}{\frac{\sigma_0}{\sqrt{p}} \sqrt{2\pi}} \exp \left[-\frac{(a - \bar{x}_p)^2}{2\left(\frac{\sigma_0}{\sqrt{p}}\right)^2} \right] da. \quad (15.14)
 \end{aligned}$$

Функцию, стоящую под знаком интеграла в (15.14), можно формально считать плотностью вероятности нормально распределенной «случайной» величины a с центром распределения \bar{x}_p и средним квадратичным отклонением $\frac{\sigma_0}{\sqrt{p}}$. По свойству плотности вероятности интеграл от нее по всей области равен единице. Поэтому

$$\begin{aligned}
 P(x \approx x_1, x_2, \dots, x_n | \sigma_0) &= \\
 &= C'' \sigma_0^{-n+1} \exp \left[-\frac{\sum_{k=1}^n p_k (x_k - \bar{x}_p)^2}{2\sigma_0^2} \right], \quad (15.15)
 \end{aligned}$$

где $C'' = \frac{C'}{\sqrt{p}}$.

Мы нашли условную вероятность события (x_1, x_2, \dots, x_n) при гипотезе о значении σ_0 . Положим теперь, что значения σ_0 до измерений равновероятны в некоторой области, которая может здесь считаться малой. По следствию из формулы Байеса вероятности гипотез после измерений пропорциональны условным вероятностям события (результатов измерений) при гипотезах.

Поэтому

$$\begin{aligned}
 P(\sigma_0 | x_1, x_2, \dots, x_n) &= \\
 &= K C'' \sigma_0^{-n+1} \exp \left[-\frac{\sum_{k=1}^n p_k (x_k - \bar{x}_p)^2}{2\sigma_0^2} \right], \quad (15.16)
 \end{aligned}$$

где K — коэффициент пропорциональности.

Теперь можно разыскать наиболее вероятное значение σ_0 , какое можно получить из измерений. Для этого достаточно найти то значение σ_0 , при котором правая часть (15.16) имеет максимум.

Для сокращения записи введем обозначения:

$$\bar{S}_p = \sum_{k=p}^n p_k (x_k - \bar{x}_p)^2, \quad f(\sigma_0) = \sigma_0^{-n+1} \exp \left(-\frac{\bar{S}_p}{2\sigma_0^2} \right).$$

Логарифмируя и дифференцируя $f(\sigma_0)$, получим

$$\begin{aligned} \ln f(\sigma_0) &= (1-n) \ln \sigma_0 - \frac{1}{2} \bar{S}_p \sigma_0^{-2}, \\ \frac{df(\sigma_0)}{d\sigma_0} &= f(\sigma_0) [\bar{S}_p \sigma_0^{-3} - (n-1) \sigma_0^{-1}], \\ \frac{d^2 f(\sigma_0)}{d\sigma_0^2} &= f(\sigma_0) [\bar{S}_p \sigma_0^{-3} - (n-1) \sigma_0^{-1}]^2 + \\ &\quad + f(\sigma_0) [-3\bar{S}_p \sigma_0^{-4} + (n-1) \sigma_0^{-3}]. \end{aligned}$$

Приравнявая первую производную нулю, получим значение σ_0 , соответствующее экстремуму:

$$\sigma_{0, \text{э}}^2 = \frac{\bar{S}_p}{n-1}$$

(значок «э» означает «экстремальное»). В выражении для второй производной первый член при $\sigma = \sigma_{0, \text{э}}$ обращается в нуль, второй член дает после очевидных выкладок

$$\left(\frac{d^2 f(\sigma_0)}{d\sigma_0^2} \right)_{\sigma_0 = \sigma_{0, \text{э}}} = -2(n-1) \sigma_{0, \text{э}}^{-(n+1)} e^{-\frac{n-1}{2}} < 0. \quad (15.17)$$

Из (15.17) следует, что найденное значение $\sigma_{0, \text{э}}$ является наиболее вероятным. В дальнейшем будем обозначать наиболее вероятное значение $\sigma_{0, \text{э}}$ средней квадратичной ошибки на единицу веса просто буквой σ без значков; тогда

$$\sigma = \sqrt{\frac{\sum_{k=1}^n p_k (x_k - \bar{x}_p)^2}{n-1}}. \quad (15.18)$$

Очевидно, формула (14.20) предыдущей главы является частным видом (15.18), ибо при равнооточных измерениях вес каждого можно принять равным единице.

Отметим следующее обстоятельство, существенное в практической работе. Если заданы средние квадратичные ошибки неравнооточных измерений, то для вычисления весов надо ввести «предварительное» значение средней квадратичной ошибки на единицу веса. Обозначим его через σ_0 . После вычисления весов как бы отбрасываются средние квадратичные ошибки измерений и используются только веса. Поэтому наиболее вероятное значение единичной средней квадратичной ошибки σ можно снова определить по весам и результатам измерений. Значение σ отличается от σ_0 и иногда довольно сильно. Обычно за окончательное значение принимают σ , а не σ_0 по следующим соображениям: σ_0 есть формально принятое число, никак не связанное со сделанными измерениями; при вычислении σ мы опираемся на наблюдения, поэтому это число действительно можно

считать отражающим в какой-то мере свойства измерений. Число же σ_0 есть только вспомогательное число для определения весов измерений.

В формуле (15.11) для определения средней квадратичной ошибки среднего весового после вычисления наиболее вероятного значения средней квадратичной ошибки на единицу веса надо заменить σ_0 на σ . Если есть «выскакивающие» измерения, и можно опасаться, что среди измерений имеются обремененные грубыми ошибками, то надо их исключить по правилу трех сигм. Если были заданы σ_k , то отбрасываются измерения, для которых

$$|x_k - \bar{x}| > 3\sigma_k.$$

Если σ_k не были даны (но заданы веса), то по значению σ и весам p_k нужно вычислить σ_k и воспользоваться правилом трех сигм. Вычисление σ_k нужно делать только для тех измерений, в которых x_k заметно отличается от \bar{x} («выскакивает»). Если по правилу трех сигм (иногда двух или четырех сигм) найдутся измерения с грубыми ошибками, то их либо отбрасывают, либо исправляют по каким-либо дополнительным сведениям.

§ 77. Пример и схема обработки неравноточных измерений определенной величины

При вычислении географической широты Обсерватории МГУ на Пресне были использованы наблюдения девяти звезд, произведенные Швейцером. Каждая звезда наблюдалась неоднократно, поэтому были получены средние квадратичные ошибки, указанные в третьем столбце схемы:

I	II	III	IV	V	VI	VII	VIII	IX	X
№№	x_k	σ_k	$\sigma_0 : \sigma_k$	p_k	$p_k x_k$	$x_k - \bar{x}$	$p_k (x_k - \bar{x})$	$p_k (x_k - \bar{x})^2$	$p_k x_k^2$
1	"	"			"	"	"		
2	1,29	0,25	0,80	0,64	0,83	+ 0,52	+ 0,33	0,17	1,07
3	0,39	0,27	0,74	0,55	0,21	- 0,38	- 0,21	0,08	0,08
4	1,61	0,20	1,00	1,00	1,61	+ 0,84	+ 0,84	0,71	2,59
5	1,27	0,23	0,87	0,76	0,97	+ 0,50	+ 0,38	0,19	1,23
6	0,81	0,19	1,05	1,10	0,89	+ 0,04	+ 0,04	0,00	0,72
7	0,61	0,40	0,50	0,25	0,15	- 0,16	- 0,04	0,01	0,09
8	0,22	0,21	0,95	0,90	0,20	- 0,55	- 0,50	0,28	0,04
9	0,08	0,18	1,11	1,23	0,10	- 0,69	- 0,85	0,59	0,01
9	"	"			"	"	"		
9	0,71	0,28	0,71	0,50	0,36	- 0,06	- 0,03	0,00	0,26
$\sigma_0 = 0,20$		$p = 6,93$		$5,32$		$(- 0,04)$		2,03	6,09

В первом столбце схемы указаны номера наблюдений. Так как широта определяется по наблюдениям разных звезд, то на практике часто указываются не номера наблюдений, а те звезды, которые были использованы. В нашем примере были использованы наблюдения пяти звезд из созвездия Дракона, одной звезды из Лебеда и трех звезд Большой Медведицы.

В столбце II выписаны числа $x_k = \varphi_k - \varphi_0$; $\varphi_0 = 55^\circ 45' 19''$, т. е. фактически, как и в примере гл. 13, выбрано новое начало отсчета с таким расчетом, чтобы все x_k были положительными.

В столбце III даны средние квадратичные ошибки значений, выведенные способом, описанным в предыдущей главе.

Столбцы IV и V содержат вычисление весов по средним квадратичным ошибкам с помощью формулы

$$p_k = \frac{\sigma_0^2}{\sigma_k^2},$$

в которой за σ_0 принято предварительное значение $0'',20$. Сначала вычисляются отношения σ_0/σ_k (столбец IV); возвышение этих чисел в квадрат дает столбец V. Столбец VI величин $p_k x_k$ нужен для вычисления среднего весового, т. е. наиболее вероятного значения определяемой широты, поэтому для чисел столбцов V и VI находятся суммы, выписанные в последней строке. Величины в столбцах VII, VIII и IX понятны по заголовкам; надо только учесть, что среднее весовое обозначено через \bar{x} без значка для упрощения записи.

Столбец X нужен для контроля весовой суммы квадратов отклонений от среднего весового по легко выводимой формуле

$$\sum_{k=1}^n p_k (x_k - \bar{x})^2 = \sum_{k=1}^n p_k x_k^2 - p \bar{x}^2, \quad (15.19)$$

где p определяется равенством (15.8).

Столбец VIII нужен для контроля вычисления отклонений от среднего весового и умножения их на веса. Контроль заключается в определении суммы чисел столбца VIII; если деление при вычислении среднего весового сделано точно, то сумма по свойству среднего весового должна быть точно равна нулю. Такие случаи бывают, конечно, редко, поскольку среднее весовое вычисляется приближенно (обычно с округлением). Поэтому предельная погрешность суммы равна произведению половины единицы последнего знака в числе \bar{x} на сумму весов, если сумма вычисляется на арифмометре накоплением (числа x_k и p_k при этом контроле можно формально считать точными). Если сумма чисел столбца VIII меньше этой предельной погрешности, то можно считать, что ошибок в этой части работы не было. В противном случае следует обязательно отыскать ошибку.

Описанная выше учебная схема вычислений довольно подробна; вычислитель, имеющий некоторый опыт, может ее несколько сократить и изменить: 1) для вычисления весов можно вычислить числа σ_k^2 , а затем делением на полученные числа значения σ_0^2 получать веса; 2) если не предполагается контроль весовой суммы квадратов отклонений от среднего весового, то столбец VI может быть исключен, а сумма чисел этого столбца получается на арифмометре способом накопления; 3) столбец IX можно не выписывать, а сумму получить тоже способом накопления.

Часть вычислений проводится вне схемы: определение среднего весового:

$$\bar{x} = \frac{5'',32}{6,93} = 0'',77, \quad \bar{\varphi} = 55^\circ 45' 19'' + 0'',77 = 55^\circ 45' 19'',77;$$

вычисление средней квадратичной ошибки на единицу веса:

$$\sigma^2 = \frac{2,03}{9-1} = 0,25, \quad \sigma = 0'',50;$$

вычисление средней квадратичной ошибки среднего весового:

$$\sigma_{\bar{\varphi}} = \frac{0'',50}{\sqrt{6,93}} = \frac{0'',50}{2,63} = 0'',19.$$

Контрольные вычисления:

Модуль суммы чисел столбца VIII равен $0'',4$. Чтобы выяснить допустимость такого отклонения суммы от нуля, нужно вычислить предельную погрешность этой суммы по способу ее определения (см. стр. 22). Предельная погрешность рассчитывается здесь немного сложнее, чем в пояснениях к схеме. Кроме оценки ошибки от того, что среднее весовое определено приближенно, нужно учесть погрешность от отбрасывания знаков в отдельных членах.

Так как числа p_k и x_k можно считать точными, то предельная погрешность каждого из слагаемых должна быть принята равной половине единицы последнего знака, т. е. $0,005$. Поэтому полный расчет погрешности суммы дает

$$9 \cdot 0,5 \cdot 10^{-2} + 0,5 \cdot 10^{-2} \cdot 6,93 = 8 \cdot 10^{-2}.$$

Величина суммы вдвое меньше ее предельной погрешности, поэтому нет оснований предполагать наличие ошибки в вычислениях.

Контроль весовой суммы квадратов отклонений от среднего по формуле, указанной ранее, дает

$$\bar{x}^2 = 0,59, \quad p\bar{x}^2 = 4,1, \quad \left(\sum_{k=1}^n p_k (x_k - \bar{x})^2 \right) = 6,1 - 4,1 = 2,0.$$

Расхождение между контрольным значением суммы (в скобках) и полученным в основных вычислениях ничтожно, поэтому можно считать, что и здесь ошибок не было.

Запись результата:

$$\varphi = 55^\circ 45' 19''.77 \pm 0'',19 \text{ (средн. квадр. ош.)}$$

Так как сотые доли секунды недостаточно надежны, то можно записать результат в такой форме:

$$\varphi = 55^\circ 45' 19''.8 \pm 0'',2.$$

ГЛАВА 16

ОПРЕДЕЛЕНИЕ НЕСКОЛЬКИХ НЕИЗВЕСТНЫХ ИЗ УРАВНЕНИЙ ПО СПОСОБУ НАИМЕНЬШИХ КВАДРАТОВ

§ 78. Условные и нормальные уравнения. Принцип Лежандра

В астрономической практике и ряде других прикладных областей знания довольно часто встречаются такие задачи, в которых величины, подлежащие определению, не наблюдаются непосредственно. Вместо них из наблюдений могут быть определены некоторые величины, являющиеся функциями неизвестных *).

Пример 1. Пусть наблюдения дают значения x_k , y_k величин x и y . Предполагается, что эти величины связаны функциональной зависимостью вида

$$y = a + bx + cx^2,$$

где a , b , c — коэффициенты, которые нужно определить. Каждое наблюдение даст уравнение с тремя неизвестными

$$y_k = a + bx_k + cx_k^2 \quad (k = 1, 2, \dots, n),$$

где n — число измерений.

Пример 2. При изучении некоторых инструментов ставится задача об определении ошибок их отдельных звеньев. Знание ошибок звеньев позволит вычислять ошибки всего инструмента при разных условиях. Прямое измерение ошибок звеньев потребует разборки инструмента для обмеривания звеньев. Такой способ не всегда удобен и может не дать надежных результатов, ибо при обратной сборке инструмента ошибка звеньев (например, величины зазоров) могут измениться. Будем говорить для определенности о приборе для измерения углов. Поступим так: несколько углов разных величин измерим значительно более точным прибором, чем исследуемый, причем так, чтобы результат можно было считать практически точным по сравнению с измерениями исследуемым инструментом. Измерения тех же углов исследуемым инструментом дадут значения, отличные от принятых за точные. Разности можно считать точными ошибками инструмента при разных значениях угла. Зная конструкцию инструмента, можно представить его суммарную ошибку как функцию ошибок

*) Таким образом, прямое измерение нужных величин заменяется измерением других величин, т. е. неизвестные величины определяются через посредство других величин. Поэтому говорят иногда, что производятся *косвенные* или *посредственные* измерения; эту терминологию, впрочем, нельзя считать удачной.

звеньев. Применяя эту функциональную связь при разных значениях угла, получим уравнения вида

$$\Delta a_k = f_k(a_k, \Delta a, \Delta b, \dots),$$

где Δa_k — ошибка угла, $\Delta a, \Delta b, \dots$ — ошибки звеньев, f_k — функциональная зависимость, а значок k поставлен, чтобы показать, что в некоторых случаях вид функциональной связи может зависеть от измеряемой величины.

Различные примеры подобных задач встречаются в астрометрии, небесной механике и других разделах астрономии.

В общем виде задача ставится так. Вместо подлежащих нахождению величин a, b, c, \dots из наблюдений получают величины l_k , представляющие определенные функции неизвестных. Каждое наблюдение дает *условное уравнение* вида

$$f_k(a, b, c, \dots) = l_k, \quad (16.1)$$

где k — номер наблюдения. Предполагается, что f_k — дифференцируемые функции аргументов a, b, c, \dots , которые, вообще говоря, содержат еще и параметры, изменяющиеся от наблюдения к наблюдению.

Если бы при наблюдениях не было случайных ошибок или они были бы так малы, что ими можно пренебречь, то достаточно было бы в общем случае сделать ровно столько наблюдений, сколько неизвестных, получить уравнения и решить их. Только при особом подборе величин l_k могло бы случиться, что некоторые из уравнений представляют следствия других, и число уравнений надо было бы увеличить.

Но в задачах обычного типа числа l_k содержат случайные ошибки, величинами которых пренебрегать нельзя. Если составить ровно столько независимых уравнений, сколько неизвестных, то эти ошибки целиком войдут в корни системы. Для того чтобы можно было надеяться на частичную взаимную компенсацию ошибок, принято брать число наблюдений и, следовательно, число уравнений (16.1) гораздо большим, чем число неизвестных*). Тогда получается своеобразная с алгебраической стороны система уравнений вида

$$f_k(a, b, c, \dots) = l_k \quad (k = 1, 2, \dots, n); \quad (16.2)$$

$n \gg t$, где t — число неизвестных. Уравнения такой системы называют *условными* или *начальными* **).

*) Задача о комбинировании ошибок в таких системах и о влиянии их на ошибку результата очень сложна и в случае нелинейных уравнений не исследовалась.

**) Заметим, что в этом вопросе есть разнобой в терминологии. По замечанию Н. И. Идельсона такие уравнения называются *условными* «в астрономическом смысле». Дело в том, что в геодезии «условными уравнениями» называют точные уравнения связи между неизвестными, например уравнения типа: сумма углов треугольника равна 180° (сумма измеренных углов обычно отличается от 180° на небольшую величину). Те же уравнения, о которых идет речь в этом параграфе, в геодезии называются «начальными». Мы будем употреблять название «условные».

Своеобразие системы условных уравнений (16.2) в том, что благодаря наличию случайных ошибок система несовместна, даже если функциональные связи точные (что бывает не всегда). Это означает, что не существует таких значений a, b, c, \dots , которые бы удовлетворяли одновременно всем уравнениям системы. Иными словами, подстановка любых значений неизвестных, например a', b', c', \dots , даст

$$f_k(a', b', c', \dots) - l_k = \delta_k \neq 0 \quad (k = 1, 2, \dots, n). \quad (16.3)$$

Из несовместности системы условных уравнений вытекает необходимость в соглашении о способе решения системы и в выяснении вероятностного смысла избранного способа решения. Очевидно, естественно выбрать такое соглашение, при котором модули невязок были бы возможно меньшими по модулю, но удовлетворить этому условию невозможно. Если уравнений ровно столько, сколько неизвестных, то можно сделать все невязки нулями, в обычном же случае задача об уменьшении модулей невязок неопределенна. Можно, например, отобрав столько уравнений, сколько неизвестных, решить их, и невязки в этих уравнениях будут нулями, но невязки остальных уравнений от нас уже не зависят, они могут быть и большими. Очевидно, может идти речь только о некоторых суммарных условиях, накладываемых на невязки. Из них мы отметим два условия, предложенные в разное время.

Довольно естественным было бы потребовать, чтобы сумма модулей невязок была наименьшей; это условие было предложено Эджвортом. Однако решение условных уравнений по способу Эджворта в широкое употребление не вошло.

Иное соглашение было раньше предложено Лежандром и впервые им опубликовано. После опубликования работы Лежандра Гаусс заявил, что он в течение ряда лет пользовался таким соглашением, но его, конечно, естественно назвать принципом Лежандра.

Принцип Лежандра. *Если дана система равноточных условных уравнений, то условимся искать неизвестные так, чтобы сумма квадратов невязок была наименьшей.*

Это условие быстро было принято во вполне понятным причинам. Хотя нельзя обеспечить малость отдельных невязок, но минимальность суммы квадратов обеспечивает ограниченность и отдельных невязок. Сумма квадратов невязок — аналитическая функция неизвестных, если аналитичны функции, входящие в условные уравнения. Поэтому операция составления уравнений для определения значений неизвестных, соответствующих минимуму, всегда возможна. Кроме того, как будет показано в следующем параграфе, принцип Лежандра имеет простую вероятностную интерпретацию.

Образум, согласно принципу Лежандра, сумму квадратов невязок

$$S = \sum_{k=1}^n [f_k(a, b, c, \dots) - l_k]^2 \quad (16.4)$$

для случая равноточных измерений, и напомним необходимые условия минимума этой суммы:

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = \dots = 0. \quad (16.5)$$

Полученные уравнения (16.5) для a, b, c, \dots называются *нормальными* уравнениями. Для определения неизвестных мы получаем всегда систему с числом уравнений, равным числу неизвестных, поэтому задача в общем случае становится определенной.

Если условные уравнения нелинейные, то они могут иметь несколько систем значений неизвестных и тогда нужно выбрать такую систему, которая даст «минимум минимуму» сумме квадратов невязок. Существование такого минимума обеспечено тем, что S — существенно положительная функция неизвестных, которая может обратиться в нуль только тогда, когда все невязки равны нулю. Это невозможно при наличии случайных ошибок в числах l_k . Поэтому $S(a, b, c, \dots)$ необходимо должна иметь минимум.

§ 79. Вероятностный смысл принципа Лежандра

Предположим, что случайные ошибки содержатся только в результатах измерений l_k ; функции f_k могут содержать также результаты измерений в виде параметров, но мы их будем считать не содержащими случайных ошибок. Это законно, если ошибки параметров малы по сравнению с ошибками l_k . Предположим еще, как и в предыдущих главах, что измерения l_k взаимно независимы и ошибки чисел l_k подчиняются нормальному закону распределения. При указанных условиях невязки

$$\delta_k = f_k(a, b, c, \dots) - l_k \quad (16.6)$$

представляют точные ошибки, подчиненные одному и тому же нормальному закону в соответствии с условием равной точности: дисперсии последовательных измерений одинаковы, математические ожидания ошибок равны нулю.

Предположим на минуту, что a, b, c, \dots известны и ошибки δ_k вычислены. Вероятность получить ошибки по величине, близкие к сделанным при наших измерениях, будет такова:

$$P_k(\delta_k < \delta < \delta_k + \Delta\delta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\delta_k^2}{2\sigma^2}} \Delta\delta, \quad (16.7)$$

где $\Delta\delta$ — малое положительное число (см. гл. 11). Левую часть условимся записывать кратко так: $P_k(\sim \delta_k)$. Поскольку a, b, c, \dots неизвестны, нельзя получить числа δ_k ; можно только делать разные гипотезы о величинах a, b, c, \dots , находить соответствующие δ_k

и вычислять вероятности. Чтобы отметить зависимость δ_k и P_k от a, b, c, \dots , будем писать равенство (16.7) так:

$$P_k(\sim \delta_k | a, b, c, \dots) = \frac{\Delta \delta}{\sigma \sqrt{2\pi}} e^{-\frac{\delta_k^2}{2\sigma^2}}. \quad (16.8)$$

Эта вероятность может считаться также и вероятностью получить при измерениях числа l_k , поскольку δ_k выражается через l_k и неслучайные величины $-f_k(a, b, c, \dots)$. Поэтому окончательно выражение для вероятности записывается в следующем виде:

$$P_k(\sim l_k | a, b, c, \dots) = \frac{\Delta}{\sigma \sqrt{2\pi}} e^{-\frac{\delta_k^2}{2\sigma^2}}, \quad (16.9)$$

где σ считается известной величиной, а δ_k определяется равенством (16.6).

Вероятность получить совокупность результатов измерений, близкую к полученной, как функция от a, b, c, \dots , получится отсюда по теореме умножения вероятностей для взаимно независимых событий (измерения мы считаем независимыми):

$$P(\sim l_1, l_2, \dots, l_n | a, b, c, \dots) = \left(\frac{\Delta}{\sigma \sqrt{2\pi}} \right)^n \exp \left[-\frac{\sum_{k=1}^n \delta_k^2}{2\sigma^2} \right]. \quad (16.10)$$

Делая разные гипотезы о величинах чисел a, b, c, \dots , будем получать разные вероятности фактических результатов измерений.

Положим, что все гипотезы о величинах (a, b, c, \dots) равновероятны в некоторой области, которая может быть сделана достаточно малой в окрестности неизвестных точных значений. Тогда можно применить следствие из формулы Бейеса, по которому апостериорные вероятности гипотез после появления события пропорциональны условным вероятностям событий при соответствующих гипотезах. Согласно этому следствию

$$P(a, b, c, \dots | l_1, l_2, \dots, l_n) = K \exp \left(-\frac{\sum_{k=1}^n \delta_k^2}{2\sigma^2} \right), \quad (16.11)$$

где в K включены и коэффициент пропорциональности и множитель, не зависящий от a, b, c, \dots . Таким образом, вероятность гипотез относительно значений a, b, c, \dots есть функция самих этих величин. Поэтому можно поставить вопрос об определении тех значений a, b, c, \dots , при которых получится максимум вероятности. Так как a, b, c, \dots входят только в показатель, то наиболее вероятной будет гипотеза об a, b, c, \dots при таких их значениях, при которых $\sum_{k=1}^n \delta_k^2$ будет наименьшей. А это и есть принцип

Лежандра. Таким образом, мы приходим к следующему заключению:

если дана система равноточных условных уравнений, то наиболее вероятные значения неизвестных получатся из условия минимума суммы квадратов невязок.

§ 80. Обобщение принципа Лежандра на неравноточные условные уравнения. Приведение неравноточных уравнений к равноточным

Пусть результаты измерений по каким-либо причинам нельзя считать равноточными. Предположим, что для них известны средние квадратичные ошибки $\sigma_1, \sigma_2, \dots, \sigma_n$ и найдены веса p_1, p_2, \dots, p_n . Мы можем опять вычислить вероятность получить в k -м измерении приблизительно число l_k при определенной гипотезе о значениях a, b, c, \dots :

$$P_k(\delta_k < \delta < \delta_k + \Delta\delta) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{\delta_k^2}{2\sigma_k^2}} \Delta\delta, \quad (16.12)$$

где согласно (16.6)

$$\delta_k = f_k(a, b, c, \dots) - l_k.$$

Введем веса

$$p_k = \frac{\sigma_0^2}{\sigma_k^2},$$

где σ_0 — средняя квадратичная ошибка единицы веса. Под знаком вероятности в (16.12) вместо δ_k напомним l_k , так как δ_k есть функция l_k при заданных a, b, c, \dots . Применяя сокращенную запись неравенств, стоящих под знаком вероятности в (16.12), получим

$$P(\sim l_k | a, b, c, \dots) = \frac{\sqrt{p_k}}{\sigma_0 \sqrt{2\pi}} \exp\left(-\frac{p_k \delta_k^2}{2\sigma_0^2}\right) \Delta\delta.$$

Как и в предыдущем параграфе, можно написать вероятность совокупности результатов измерений как функцию

$$P(\sim l_1, l_2, \dots, l_n | a, b, c, \dots) = \left(\frac{\Delta\delta}{\sigma_0 \sqrt{2\pi}}\right)^n \left(\prod_{k=1}^n \sqrt{p_k}\right) \exp\left[-\frac{\sum_{k=1}^n p_k \delta_k^2}{2\sigma_0^2}\right].$$

Опять используя следствие из формулы Байеса в предположении равной вероятности возможных совокупностей (a, b, c) , получим

$$P(a, b, c, \dots | l_1, l_2, \dots, l_n) = M \exp\left[-\frac{\sum_{k=1}^n p_k \delta_k^2}{2\sigma_0^2}\right]. \quad (16.13)$$

где M включает и все множители, не зависящие от a, b, c , входившие в предыдущее равенство, и коэффициент пропорциональности в следствии из формулы Бейеса.

Отсюда вытекает, что наиболее вероятную совокупность значений a, b, c получим, если потребуем, чтобы сумма

$$S_p = \sum_{k=1}^n p_k \delta_k^2$$

имела минимум. Таким образом, мы приходим к обобщенному принципу Лежандра, который можно сформулировать так:

если условные уравнения неравноточны, то следует искать неизвестные так, чтобы имела минимум сумма произведений квадратов невязок на веса (весовая сумма квадратов невязок).

Если уравнения равноточные, то все веса можно считать равными единице, и мы получим принцип Лежандра в его простом виде.

Из обобщенного принципа Лежандра легко получается правило приведения неравноточных условных уравнений к равноточным. Представим весовую сумму квадратов невязок так:

$$S_p = \sum_{k=1}^n (\delta_k \sqrt{p_k})^2.$$

Введем обозначение: $\varepsilon_k = \delta_k \sqrt{p_k}$. Тогда

$$S_p = \sum_{k=1}^n \varepsilon_k^2.$$

Это значит, что весовая сумма квадратов невязок является суммой квадратов невязок для таких равноточных условных уравнений, которые имеют невязки ε_k . Заданные уравнения невязок

$$\delta_k = f_k(a, b, c) - l_k$$

легко сделать уравнениями с невязками ε_k , если учесть связь между δ_k и ε_k . Таким образом, получается следующее правило:

чтобы привести неравноточные условные уравнения к равноточным, нужно умножить каждое условное уравнение на корень квадратный из его веса.

Нормальные уравнения в случае неравноточных условных уравнений получаются из условия минимума функции нескольких аргументов:

$$\frac{\partial S_p}{\partial a} = \frac{\partial S_p}{\partial b} = \frac{\partial S_p}{\partial c} = \dots = 0,$$

§ 81. Приведение нелинейных условных уравнений к линейному виду

Составление нормальных уравнений возможно при любом виде условных уравнений, т. е. при любых функциях $f_k(a, b, c)$. Решение же нормальных уравнений в случае нелинейных условных уравнений может представить большие трудности. Еще важнее то, что в этом случае корни нормальных уравнений не будут линейными функциями случайных чисел l_k . Важность этого обстоятельства вытекает из следующих соображений: наиболее вероятные значения корней нормальных уравнений, очевидно, являются функциями случайных величин l_k , следовательно, сами тоже случайны. Для определения возможных колебаний наиболее вероятных значений неизвестных надо знать по крайней мере дисперсии неизвестных. Отметим, что когда мы говорим о дисперсиях неизвестных, то имеется в виду не только та выборочная совокупность чисел l_k , которая получена в изучаемой серии условных уравнений. Мы представляем себе наряду с полученной серией чисел l_k всевозможные серии значений l_k , какие могли бы получиться при других комбинациях причин, вызывающих случайные ошибки. Каждой серии будут соответствовать свои наиболее вероятные значения $\bar{a}, \bar{b}, \bar{c}, \dots$. Совокупность всевозможных $\bar{a}, \bar{b}, \bar{c}, \dots$ образует область возможных значений $\bar{a}, \bar{b}, \bar{c}, \dots$; можно положить, что $E(\bar{a}) = a$, $E(\bar{b}) = b$, $E(\bar{c}) = c, \dots$ и определять дисперсии случайных величин $\bar{a}, \bar{b}, \bar{c}, \dots$, которые и дадут понятие о надежности вычисленных $\bar{a}, \bar{b}, \bar{c}, \dots$ из данной серии измерений.

Именно необходимость определить дисперсии $\bar{a}, \bar{b}, \bar{c}, \dots$ и заставляет заботиться о линейности условных уравнений. В этом случае не только упрощается решение нормальных уравнений, но и становится простой задача об определении дисперсий неизвестных. Действительно, линейные условные уравнения дадут линейные нормальные уравнения. Их корни будут линейными функциями чисел l_k , т. е. линейными функциями случайных ошибок. Если можно считать эти ошибки взаимно независимыми, то будем иметь простую задачу об определении дисперсии линейной функции независимых величин. Вот почему аппарат способа наименьших квадратов развит только для линейных условных уравнений.

В случае задания нелинейных условных уравнений (например, при построении эмпирической формулы, нелинейной относительно параметров), нужно преобразовать их к линейному виду*). Можно указать два метода такого приведения к линейному виду.

*) Эту операцию называют иногда *линеаризацией* уравнений,

Первый, весьма простой метод, заключается в такой замене неизвестных, чтобы после надлежащих выкладок условные уравнения были линейными относительно новых неизвестных.

Пример 1. Если условные уравнения имеют вид

$$\alpha_k \sin(a+b) + \beta_k \sin(a-b) + \gamma_k e^{-2c} = l_k, \quad k = 1, 2, \dots, n,$$

где $\alpha_k, \beta_k, \gamma_k, l_k$ — заданные числа, a, b, c — неизвестные, то подстановкой

$$\sin(a+b) = x, \quad \sin(a-b) = y, \quad e^{-2c} = z$$

получим линейные уравнения

$$\alpha_k x + \beta_k y + \gamma_k z - l_k = 0.$$

После решения системы получим наиболее вероятные значения неизвестных $\bar{x}, \bar{y}, \bar{z}$. Эти случайные значения суть линейные функции случайных величин l_k , и в силу постулата о нормальном законе для случайных ошибок можно утверждать, что $\bar{x}, \bar{y}, \bar{z}$ распределены также нормально. Из формул, связывающих a, b, c и x, y, z , получим

$$\bar{a} = \frac{1}{2} (\arcsin \bar{x} + \arcsin \bar{y}),$$

$$\bar{b} = \frac{1}{2} (\arcsin \bar{x} - \arcsin \bar{y}),$$

$$\bar{c} = -\frac{1}{2} \ln \bar{z}.$$

Поскольку законы распределения $\bar{x}, \bar{y}, \bar{z}$ известны (нормальные), то можно найти законы распределения $\bar{a}, \bar{b}, \bar{c}$ и, следовательно, их дисперсии, что полностью решает задачу в вероятностном смысле.

Этот метод, однако, применим далеко не всегда.

Общий метод, пригодный для всех условных уравнений, опирается на предположение, что невязки условных уравнений малы, для чего нужно положить, что коэффициенты условных уравнений точны, а случайные ошибки чисел l_k малы (по модулю). Иными словами, несовместность системы условных уравнений «слаба».

Найдем теперь любым способом предварительные приближенные значения неизвестных, обозначим их a_0, b_0, c_0 . Можно указать два употребительных способа выбора исходного приближения. Первый способ заключается в том, что отбирают из заданной системы столько условных уравнений, сколько неизвестных, заботясь о том, чтобы отобранные приблизительно равномерно были разбросаны по системе. Решение отобранных уравнений и дает числа a_0, b_0, c_0 . Второй метод применим к тем задачам, которые неоднократно решаются по мере увеличения наблюдательного материала. В астрономической практике это, например, задачи по определению фундаментальных постоянных (параллакс Солнца, постоянная прецессии

и т. п.). Предыдущая работа такого типа дает значения неизвестных, которые и могут быть приняты за исходные в новой работе.

Положим теперь, что

$$a = a_0 + x_1, \quad b = b_0 + y_1, \quad c = c_0 + z_1, \dots; \quad (16.14)$$

подставим эти выражения в условные уравнения, разложим функции f_k в ряды по степеням x_1, y_1, z_1, \dots и ограничимся в разложениях первыми степенями этих поправок.

Тогда получим линейные условные уравнения вида

$$f_k(a_0, b_0, c_0, \dots) - l_k + \left(\frac{\partial f_k}{\partial a}\right)_0 x_1 + \left(\frac{\partial f_k}{\partial b}\right)_0 y_1 + \left(\frac{\partial f_k}{\partial c}\right)_0 z_1 + \dots = 0, \quad (16.15)$$

где нуль у частных производных показывает, что после дифференцирования нужно подставить a_0, b_0, c_0, \dots . Решение системы линейных условных уравнений способом, который будет указан ниже, даст вероятные значения $\bar{x}_1, \bar{y}_1, \bar{z}_1$ и их квадратичные ошибки $\sigma_{\bar{x}_1}, \sigma_{\bar{y}_1}, \sigma_{\bar{z}_1}$. После этого строим приближение

$$a_1 = a_0 + \bar{x}_1, \quad b_1 = b_0 + \bar{y}_1, \quad c_1 = c_0 + \bar{z}_1, \dots$$

с указанными квадратичными ошибками, так как a_0, b_0, c_0, \dots — определенные (неслучайные) числа. С этим приближением поступаем, как с исходным, т. е. составляем линейную систему вида

$$f_k(a_1, b_1, c_1) - l_k + \left(\frac{\partial f}{\partial a}\right)_1 x_2 + \left(\frac{\partial f}{\partial b}\right)_1 y_2 + \left(\frac{\partial f}{\partial c}\right)_1 z_2 + \dots = 0,$$

где единица у частных производных означает, что после дифференцирования подставляется a_1, b_1, c_1 . Решение даст $\bar{x}_2, \bar{y}_2, \bar{z}_2, \dots$ и средние квадратичные ошибки $\sigma_{\bar{x}_2}, \sigma_{\bar{y}_2}, \sigma_{\bar{z}_2}, \dots$. Тогда

$$a_2 = a_1 + \bar{x}_2, \quad b_2 = b_1 + \bar{y}_2, \quad c_2 = c_1 + \bar{z}_2, \dots$$

или

$$a_2 = a_0 + \bar{x}_1 + \bar{x}_2, \quad b_2 = b_0 + \bar{y}_1 + \bar{y}_2, \quad c_2 = c_0 + \bar{z}_1 + \bar{z}_2, \dots$$

Средние квадратичные ошибки в этом случае определяются по теореме о дисперсии суммы:

$$\sigma_{a_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2}$$

(и подобные формулы для b_2, c_2, \dots).

Вопрос о сходимости этого процесса последовательных приближений должен ставиться для каждой данной системы. Так как исследование сходимости достаточно сложно, то обычно ограничиваются выяснением практической сходимости. Все вычисления ведутся с ограниченным числом знаков после запятой (или числом значащих цифр). Если два раза подряд с принятой точностью

получились одинаковые значения, то считают, что процесс последовательных приближений закончен, и полученные значения a , b , c считаются окончательными.

Описанный только что способ линеаризации довольно громоздок, поэтому применяется редко, в наиболее важных задачах. Если в задаче нельзя применить способ подстановки, то предпочитают заменить приближенно условные уравнения такими, к которым можно применить первый способ.

Пример 2. Дана таблица значений приблизительно периодической функции (см. табличку справа). Требуется определить параметры a , b , c , P приближающего ее закона гармонического колебания

$$w = a \sin\left(\frac{2\pi t}{P} + b\right) + c,$$

считая, что в значениях w содержатся случайные ошибки и что измерения являются равноточными.

Подставляя в принятый закон табличные значения t_k и w_k , получим систему 13 нелинейных условных уравнений:

$$a \sin\left(\frac{2\pi t_k}{P} + b\right) + c = w_k, \quad k = 0, 1, 2, \dots, 12.$$

По таблице и составленному на ее основании графику можно принять следующие исходные приближенные значения для неизвестных:

$P_0 = 6,30$ на том основании, что при $t = 6,24$ еще не получено исходное значение (2,02);

$c_0 = 1,08$, $a_0 = 0,90$, так как максимальное значение близко к 2 и среди значений есть близкое к нулю, следовательно a и c одного порядка (за значение c_0 взято среднее арифметическое из всех w);

$b_0 = 1,50$, что можно получить грубым сравнением графика с синусоидой и определением сдвига.

Введем обозначения:

$$a = a_0 + x_1, \quad b = b_0 + y_1, \quad c = c_0 + z_1, \quad P = P_0 + u_1.$$

Проделав операцию линеаризации, получим систему линейных условных уравнений вида

$$\sin \tau_k x_1 + a_0 \cos \tau_k y_1 - a_0 \cos \tau_k \frac{2\pi t_k}{P_0^2} u_1 +$$

$$+ z_1 + (a_0 \sin \tau_k + c_0 - w_k) = 0 \quad (k = 0, 1, 2, \dots, 12),$$

где

$$\tau_k = \frac{2\pi t_k}{P_0} + b_0.$$

Эту систему перепишем в таком виде:

$$a_k x_1 + b_k y_1 + c_k z_1 + d_k u_1 + l_k = 0 \quad (k = 1, 2, \dots, 12),$$

где

$$a_k = \sin \tau_k, \quad b_k = a_0 \cos \tau_k,$$

$$c_k = 1, \quad d_k = -a_0 \cos \tau_k \frac{2\pi t_k}{P_0^2},$$

$$l_k = a_0 \sin \tau_k + c_0 - w_k$$

t	w
0,00	2,02
0,52	1,86
1,04	1,49
1,56	1,02
2,08	0,51
2,60	+ 0,14
3,12	— 0,03
3,64	+ 0,15
4,16	0,53
4,68	0,97
5,20	1,46
5,72	1,90
6,24	1,98
	14,00

В следующих параграфах будет рассмотрен способ решения такой системы условных уравнений, основанный на принципе Лежандра, и определены значения x_1 , y_1 , z_1 , u_1 в рассматриваемом примере. Здесь мы лишь отметим результаты определения первого приближения:

$$\bar{x}_1 = +0,0973; \quad \bar{y}_1 = +0,0875; \quad \bar{z}_1 = -0,0749; \quad \bar{u}_1 = -0,0354.$$

Значения исходных значений неизвестных и результата первого приближения запишем в первых двух столбцах следующей таблицы:

$a_0 = 0,90$	$a_1 = 0,9973$	$(a) = 1$
$b_0 = 1,50$	$b_1 = 1,5875$	$(b) = \frac{\pi}{2} = 1,5708$
$c_0 = 1,08$	$c_1 = 1,0051$	$(c) = 1$
$P_0 = 6,30$	$P_1 = 6,2646$	$(P) = 6,2832 = 2\pi$

Отметим, что для построения этого примера был применен следующий искусственный прием: при значениях параметров (a) , (b) , (c) , (P) , содержащихся в третьем столбце приведенной таблицы, были вычислены точные значения w при значениях t с шагом $\frac{\pi}{6}$; эти значения случайным образом изменены на одну, две, три сотых со знаками плюс и минус, значения t взяты в радианах с точностью до половины сотой доли. Получились значения табличной функции $w = w(t)$, для которой и предлагалось построить приближенную функцию. Результаты вычислений показывают, что уже первое приближение дает значения параметров, близкие к исходным.

§ 82. Линейные условные и нормальные уравнения

Пусть дана система линейных условных уравнений*)

$$a_k x + b_k y + c_k z + d_k u + l_k = 0 \quad (k = 1, 2, \dots, n), \quad (16.16)$$

где x , y , z , u — неизвестные, a_k , b_k , c_k , d_k , l_k — числа, изменяющиеся от уравнения к уравнению, n — число условных уравнений. Следует помнить, что содержащими случайные ошибки мы считаем только числа l_k ; числа a_k , b_k , c_k , d_k могут быть приближенными, но, по предположению, не содержат случайных ошибок. Дальше эти числа будем называть *свободными членами* (l_k) и *коэффициентами* (a_k , b_k , c_k , d_k).

Невязки δ_k уравнений определяются равенствами

$$\delta_k = a_k x + b_k y + c_k z + d_k u + l_k \quad (k = 1, 2, \dots, n), \quad (16.17)$$

в которых под x , y , z , u можно подразумевать пока любые числа. Между системами уравнений (16.16) и (16.17) как будто имеется противоречие. В связи с этим следует вспомнить, что система заданных условных уравнений несовместна, и потому

*) Во всех дальнейших выкладках в этой главе мы для сокращения записи ограничимся случаем четырех неизвестных.

система (16.16) имеет только формальное значение. Сумма квадратов невязок (16.17) имеет вид

$$S = \sum_{k=1}^n \delta_k^2 = \sum_{k=1}^n (a_k x + b_k y + c_k z + d_k u + l_k)^2. \quad (16.18)$$

Считая условные уравнения равноточными, напомним по принципу Лежандра необходимые условия минимума:

$$\left. \begin{aligned} \frac{\partial S}{\partial x} &= 2 \sum_{k=1}^n (a_k x + b_k y + c_k z + d_k u + l_k) a_k = 0, \\ \frac{\partial S}{\partial y} &= 2 \sum_{k=1}^n (a_k x + b_k y + c_k z + d_k u + l_k) b_k = 0, \\ \frac{\partial S}{\partial z} &= 2 \sum_{k=1}^n (a_k x + b_k y + c_k z + d_k u + l_k) c_k = 0, \\ \frac{\partial S}{\partial u} &= 2 \sum_{k=1}^n (a_k x + b_k y + c_k z + d_k u + l_k) d_k = 0. \end{aligned} \right\} \quad (16.19)$$

Эти условия приводят к нормальным уравнениям:

$$\left. \begin{aligned} x \sum a_k^2 + y \sum a_k b_k + z \sum a_k c_k + u \sum a_k d_k + \sum a_k l_k &= 0, \\ x \sum b_k a_k + y \sum b_k^2 + z \sum b_k c_k + u \sum b_k d_k + \sum b_k l_k &= 0, \\ x \sum c_k a_k + y \sum c_k b_k + z \sum c_k^2 + u \sum c_k d_k + \sum c_k l_k &= 0, \\ x \sum d_k a_k + y \sum d_k b_k + z \sum d_k c_k + u \sum d_k^2 + \sum d_k l_k &= 0 \end{aligned} \right\} \quad (16.20)$$

(суммирование всюду ведется от $k=1$ до $k=n$).

Аналогично составляются нормальные уравнения при любом числе неизвестных.

Вопрос о достаточности этих условий минимума функции нескольких аргументов мог бы быть исследован известными средствами анализа, но в этом нет надобности. Сумма квадратов невязок S есть квадратичная форма аргументов x, y, z, u , поэтому она может иметь только один экстремум. Эта форма существенно положительна, поэтому минимум у нее должен быть, и корни нормальных уравнений определяют именно минимум.

При записи нормальных уравнений в астрономии и геодезии почти всегда пользуются обозначениями Гаусса:

$$\sum_{k=1}^n a_k^2 = [aa], \quad \sum_{k=1}^n a_k b_k = [ab], \quad \dots \quad \sum_{k=1}^n a_k l_k = [al] \quad (16.21)$$

и так далее. В обозначениях Гаусса система нормальных уравнений с четырьмя неизвестными имеет вид:

$$\left. \begin{aligned} [aa]x + [ab]y + [ac]z + [ad]u + [al] &= 0, \\ [ba]x + [bb]y + [bc]z + [bd]u + [bl] &= 0, \\ [ca]x + [cb]y + [cc]z + [cd]u + [cl] &= 0, \\ [da]x + [db]y + [dc]z + [dd]u + [dl] &= 0. \end{aligned} \right\} \quad (16.22)$$

При этом имеем очевидные равенства

$$[ba] = [ab], \quad [ca] = [ac] \quad \text{и т. д.}$$

Если заданы неравноточные условные линейные уравнения и их веса p_k , то следует применить общее правило приведения неравноточных уравнений к равноточным: каждое условное уравнение нужно умножить на корень квадратный из его веса. Преобразованная система условных уравнений будет иметь вид

$$a_k \sqrt{p_k} x + b_k \sqrt{p_k} y + c_k \sqrt{p_k} z + d_k \sqrt{p_k} u + l_k \sqrt{p_k} = 0.$$

Применяя к этой системе принцип Лежандра и составляя нормальные уравнения, получим:

$$\left. \begin{aligned} [paa]x + [pab]y + [pac]z + [pad]u + [pal] &= 0, \\ [pba]x + [pbb]y + [pbc]z + [pbd]u + [pbl] &= 0, \\ [pca]x + [pcb]y + [pcc]z + [pcd]u + [pcl] &= 0, \\ [pda]x + [pdb]y + [pdc]z + [pdd]u + [pdl] &= 0, \end{aligned} \right\} \quad (16.23)$$

где

$$[paa] = \sum_{k=1}^n p_k a_k^2; \quad [pab] = \sum_{k=1}^n p_k a_k b_k \quad \text{и т. п.} \quad (16.24)$$

Матрица коэффициентов при неизвестных системы нормальных уравнений обладает следующими двумя свойствами:

1) она симметрична относительно главной диагонали;

2) все элементы главной диагонали — положительные числа. Второе свойство вытекает из того, что эти элементы представляют суммы квадратов коэффициентов условных уравнений при последовательных неизвестных или суммы произведений квадратов коэффициентов на веса, т. е. на положительные числа. Такая сумма может обратиться в нуль только в том случае, когда каждое слагаемое равно нулю. Но это значит, что в условных уравнениях вовсе нет членов с соответствующим неизвестным, а потому оно не подлежит рассмотрению.

Если в нормальных уравнениях переставляют члены или сами уравнения, то допускаются только перестановки, сохраняющие указанные два свойства матрицы.

Мы не будем рассматривать в общем виде вопрос о существовании решения системы нормальных уравнений. Отметим лишь, что в исключительных случаях решения может и не существовать. Чтобы убедиться в этом, рассмотрим систему условных уравнений с двумя неизвестными:

$$a_k x + b_k y + l_k = 0.$$

Определитель системы нормальных уравнений при помощи тождества Лагранжа преобразуется так:

$$\sum_{k=1}^n a_k^2 \sum_{k=1}^n b_k^2 - \left(\sum_{k=1}^n a_k b_k \right)^2 = \sum_{k=1}^n \sum_{m=1}^n (a_k b_m - a_m b_k)^2.$$

Отсюда видно, что определитель обратится в нуль, если $\frac{a_k}{b_k} = \frac{a_m}{b_m}$, т. е. отношение коэффициентов постоянно (при всех k и m). Если и $\frac{a_k}{l_k}$ равно тому же постоянному, то задача неопределенна. Если числа l_k произвольны, то удовлетворить условию Лежандра конечными значениями неизвестных невозможно, ибо мы имеем систему уравнений с двумя неизвестными, в которых члены с неизвестными одинаковы, а свободные члены разные.

В общей задаче при произвольном числе неизвестных нужны более сложные соотношения между коэффициентами при неизвестных, чтобы определитель системы нормальных уравнений обратился в нуль. В задачах обычного типа, когда эти коэффициенты произвольны, определитель в нуль не обращается, и задача имеет единственное решение.

Отметим (без доказательства), что необходимым и достаточным условием неравенства нулю определителя системы нормальных уравнений является максимальность ранга матрицы системы условных уравнений, т. е. равенство этого ранга числу неизвестных. Заметим, впрочем, что на практике легче вычислить определитель нормальной системы, чем проверить это условие.

Составление линейных нормальных уравнений представляет громоздкую работу. Если число условных уравнений равно n , а число неизвестных равно m , то надо выполнить $\frac{nm(m+3)}{2}$ умножений по два сомножителя в каждом и $\frac{m(m+3)}{2}$ сложений по n слагаемых в каждой сумме. Например, при $m=4$ и $n=10$ имеем 140 умножений и 14 сложений. Хотя работа эта и элементарна, но вследствие громоздкости утомительна, вычислитель может ослабить внимание и сделать ошибку. Если n велико, то имеет смысл механизировать вычисления. Простейшая механизация — вычисление коэффициентов нормальных уравнений на арифмометре по способу накопления (первое произведение двух сомножителей не снимается с результатного счетчика арифмометра, к нему прибавляется следующее произведение и т. д.).

§ 83. Контроль составления нормальных уравнений

И при обычных и при механизированных вычислениях необходимо контролировать составление нормальных уравнений. Контроль обычно выполняется следующим простым способом. В каждом условном уравнении образуем сумму s_k коэффициентов при всех неизвестных и свободного члена; иными словами, s_k есть сумма элементов k -й строки матрицы коэффициентов и свободных членов. Определение чисел s_k нужно контролировать. Для этого вычисляем по столбцам полной матрицы суммы всех элементов, т. е. находим суммы коэффициентов при всех неизвестных и сумму свободных членов. Одновременно вычислим сумму всех чисел s_k . Обозначим по Гауссу последнюю сумму через $[s]$, а суммы элементов матрицы по столбцам через $[a]$, $[b]$, $[c]$, $[d]$, $[l]$.

Имеем очевидное контрольное равенство

$$([s]) = [a] + [b] + [c] + [d] + [l],$$

где скобка обозначает контрольное число. Если указанные суммирования выполняются формально точно, то контрольное равенство должно выполняться точно. Такой контроль легко выполнить и на табуляторе, если число условных уравнений велико. После вычисления и проверки чисел s_k образуем контрольные суммы произведений

$$[as], [bs], [cs], [ds];$$

их число равно числу неизвестных. Легко видеть, что контроль нормальных уравнений осуществляется следующими равенствами (в задаче с четырьмя неизвестными):

$$\left. \begin{aligned} [aa] + [ab] + [ac] + [ad] + [al] &= [as], \\ [ba] + [bb] + [bc] + [bd] + [bl] &= [bs], \\ [ca] + [cb] + [cc] + [cd] + [cl] &= [cs], \\ [da] + [db] + [dc] + [dd] + [dl] &= [ds]. \end{aligned} \right\} \quad (16.25)$$

Левые части представляют суммы коэффициентов при неизвестных и свободных членов последовательных нормальных уравнений.

Условимся о нумерации неизвестных в порядке их записи в условных уравнениях и о такой же нумерации последовательных нормальных уравнений. Тогда *правило контроля* можно изложить так:

Чтобы проверить m -е условное уравнение, нужно образовать сумму произведений чисел s_k на коэффициенты со знаками k при m -м неизвестном; эта сумма должна равняться сумме коэффициентов при неизвестных и свободного члена проверяемого условного уравнения.

Для обеспечения надежного контроля удобно при умножениях сначала не учитывать то, что коэффициенты суть приближенные числа и получать коэффициенты нормальных уравнений со всеми

знаками, какие формально получаются. Тогда контрольные равенства должны выполняться точно. Удобно вести контроль параллельно для всех нормальных уравнений; это облегчает поиски ошибок в случае расхождения в контроле. Если, например, при четырех неизвестных контроль расходится только в нормальном уравнении III, то в первую очередь надо проверить коэффициенты $[cc]$, $[cl]$ и контрольную сумму. Причина очевидна: остальные коэффициенты этого уравнения входят в другие уравнения, в которых контроль не сигнализирует об ошибке, поэтому их можно считать надежными.

Для контроля потребуется дополнительно: n сложений по $m+1$ слагаемому в каждой сумме (числа s_k), $m+2$ сложения по n слагаемых (для контроля s_k), nm умножений по два сомножителя, m сложений по n слагаемых в каждой сумме и n сложений по $m+1$ слагаемому. При $m=4$ и $n=10$ имеем следующий объем дополнительной работы:

- 10 сложений по 5 слагаемых,
- 6 сложений по 10 слагаемых,
- 40 умножений по 2 сомножителя,
- 4 сложения этих произведений,
- 10 сложений по 5 слагаемых для
- контроля нормальных уравнений

Пример. В примере 2, рассмотренном в § 81, при помощи линеаризаций была получена в буквенном виде системе 13 линейных условных уравнений с четырьмя неизвестными x_1, y_1, z_1, u_1 . Подставляя в коэффициенты a_k, b_k, c_k, d_k, l_k принятые значения величин $a_0, b_0, c_0, P_0, t_1, t_2, \dots, t_{13}$ запишем эту систему в следующем схематическом виде:

x	y	z	u	L	S
0,998	0,061	1	0,000	- 0,042	2,017
0,897	- 0,398	1	0,033	0,027	1,559
0,566	- 0,742	1	0,122	0,099	1,045
0,081	- 0,897	1	0,222	0,133	0,539
- 0,416	- 0,819	1	0,269	0,196	0,230
- 0,812	- 0,525	1	0,216	0,209	0,088
- 0,995	- 0,093	1	0,046	0,215	0,173
- 0,915	0,365	1	- 0,210	0,107	0,347
- 0,591	0,725	1	- 0,478	0,018	0,674
- 0,113	0,895	1	- 0,663	0,008	1,127
0,386	0,831	1	- 0,683	- 0,033	1,501
0,793	0,547	1	- 0,495	- 0,106	1,739
0,991	0,120	1	- 0,118	- 0,008	1,985
0,870	+ 0,070	+ 13	- 1,739	+ 0,823	13,024
					(13,024)

В заголовках столбцов указаны те неизвестные, коэффициенты при которых даны в столбцах; в столбце, озаглавленном L , даны свободные члены. Суммированием чисел по строкам получается столбец S контрольных сумм. Они проверяются последней строчкой, в которой даны суммы коэффициентов по столбцам. В последнем столбце в скобке дана контрольная сумма 13,024, полученная сложением сумм первых пяти столбцов. Так как она точно равна $[s]$, то можно считать, что в столбце S нет ошибок (при условии, что в этом столбце не были сделаны ошибки, взаимно компенсирующие при сложении).

Суммы произведений коэффициентов при неизвестных и свободных членов условных уравнений, дающие коэффициенты нормальных уравнений, легко вычисляются на арифмометре при ручном счете или на мультиметре при средней механизации без записи отдельных произведений. Это значительно сокращает объем вычислительной работы по сравнению со старыми схемами, когда каждое произведение вычислялось с помощью таблицы логарифмов. Одновременно с коэффициентами нормальных уравнений вычисляются контрольные суммы для всех нормальных уравнений. При этом для полного контроля выписываются все знаки, какие формально получаются при умножениях. Результаты записываются в виде матрицы с заголовками столбцов, обозначающими неизвестные, свободные члены l и контрольные суммы.

x_1	y_1	z_1	u_1	L	S
6,909156	0,080654	0,870000	— 0,439975	— 0,630230	6,789605
0,080654	4,936378	0,070000	— 2,509864	— 0,523411	2,053757
0,870000	0,070000	13,000000	— 1,739000	+ 0,823000	13,024000
— 0,439975	— 2,509864	— 1,739000	1,623981	+ 0,189828	— 2,875030

Последний столбец контрольный; он служит для того, чтобы проверить коэффициенты нормальных уравнений сравнением числа последнего столбца с суммой остальных чисел той же строки. Так как умножения выполнены формально точно, то всякое s_j ($j = 1, 2, 3, 4$) должно точно равняться указанной сумме. Несовпадение означает наличие ошибки либо в коэффициентах нормальных уравнений, либо в контрольной сумме. Следует отметить что последние ошибки встречаются нередко, что объясняют утомлением вычислителя при контроле в конце работы и пониженным вниманием к контролю.

Почти во всех задачах коэффициенты условных уравнений и их свободные члены — это приближенные числа, в которых верны только выписанные знаки, и то не всегда: нередко последний знак ненадежен в том смысле, что он может на несколько единиц отличаться от того, который получился бы при более точных вычислениях. Поэтому далеко не все знаки, выписанные в последней матрице, верны. Как известно (ч. I), в произведении двух чисел число верных знаков не может быть больше наименьшего числа верных знаков в сомножителях. Когда парные произведения складываются, то предельные абсолютные ошибки слагаемых складываются, и при этом на погрешность результата влияет больше не число верных знаков, а число знаков после запятой в слагаемых. Вопрос об

оценке погрешности коэффициентов нормальных уравнений поэтому решается довольно громоздко. С другой стороны, известно, что ошибки округлений, с которыми здесь приходится иметь дело, могут быть положительными и отрицательными и при большом числе операций могут частью взаимно компенсироваться. Поэтому оценки ошибок «по максимуму» почти всегда преувеличены, и в коэффициентах нормальных уравнений имеет смысл оставлять больше знаков, чем это следует из расчета «по максимуму».

В рассмотренном примере элементы матрицы условных уравнений даны с тремя знаками после запятой. Привести к подобному виду можно всякую систему очевидной заменой неизвестных. Если, например, в системе

$$a_k x + b_k y + l_k = 0, \quad k = 1, 2, \dots, n$$

числа a_k даны с тремя знаками после запятой, а b_k с двумя, то достаточно ввести новое неизвестное $\eta = 10y$, чтобы коэффициенты при η имели тоже три знака после запятой. В подобных случаях часто считают, что коэффициенты нормальных уравнений не могут иметь больше верных знаков после запятой, чем исходные числа, и соответственно округляют полученные коэффициенты нормальных уравнений.

В нашем примере после округления полученных чисел до трех знаков после запятой имеем следующую схематически записанную систему нормальных уравнений:

x_1	y_1	z_1	u_1	L	S
6,909	0,081	0,870	— 0,440	— 0,630	6,790
0,081	4,936	0,070	— 2,510	— 0,523	2,054
0,870	0,070	13,000	— 1,7939	0,823	13,024
— 0,440	— 2,510	— 1,739	1,124	0,190	— 2,875

Столбец S также выписывается, так как эти числа в дальнейшем используются для контроля решения системы нормальных уравнений. Результат округления числа S может на 1—2 единицы последнего знака отличаться от суммы округленных коэффициентов нормальных уравнений. Действительно, модуль ошибки в S от округления не превышает половины единицы третьего знака после запятой в нашем примере; это же верно и для каждого коэффициента. Так как как у нас их 5, то предельная погрешность суммы коэффициентов равна $2,5 \cdot 10^{-3}$, предельная погрешность S равна $0,5 \cdot 10^{-3}$, предельная погрешность разности равна $3 \cdot 10^{-3}$. Этот расчет и подтверждает отмеченную возможность расхождения между S и суммой коэффициентов. Использование S для контроля решения нормальных уравнений возможно только при условии точного равенства S и суммы коэффициентов. Поэтому нужно обеспечить это равенство изменением либо числа в столбце S , либо коэффициентов нормальных уравнений. Последнее правильное, так как ошибка S меньше, как видно из сделанного расчета, однако этот путь более неопределен, ибо неясно, какие (или какой) коэффициенты нужно изменить; поэтому обычно изменяют S .

§ 84. Решение системы линейных нормальных уравнений

Если требуется найти только значения неизвестных, то способ решения системы нормальных уравнений может быть любым из известных способов решения системы линейных уравнений. В частности, можно пользоваться способом итерации, если число неизвестных велико.

В § 78 было, однако, отмечено, что в задачах естествознания надо применять способ наименьших квадратов по той причине, что числа l_k содержат случайные ошибки. Корни нормальных уравнений, которые мы будем считать приближенными значениями неизвестных, представляют функции случайных величин l_k и, значит, тоже случайны. Чтобы определить их возможное рассеяние, надо найти средние квадратичные ошибки неизвестных. Когда ставится и такая задача, то обычно применяют только два способа: 1) способ определителей и 2) способ последовательного исключения неизвестных, обычно называемый способом Гаусса *).

Способ решения системы линейных уравнений с помощью определителей общеизвестен, поэтому мы на нем не будем задерживаться. Можно только заметить, что прямое вычисление определителей выше 4-го порядка является весьма трудоемкой работой. Поэтому с определителями решают обычно системы не выше 4-го порядка.

Способ последовательного исключения неизвестных часто применяют и при 3—4 неизвестных и всегда при числе неизвестных больше четырех.

Этот способ весьма прост. Из первого нормального уравнения напишем выражение первого неизвестного через остальные и назовем это равенство *первым уравнением исключения* **). В применении к системе с четырьмя неизвестными оно имеет вид

$$x = -\frac{[ab]}{[aa]} y - \frac{[ac]}{[aa]} z - \frac{[ad]}{[aa]} u - \frac{[al]}{[aa]}. \quad (16.26)$$

Это равенство схематически записывается так: в столбце, озаглавленном «у», под системой нормальных уравнений записывается число $-\frac{[ab]}{[aa]}$, в следующем столбце число $-\frac{[ac]}{[aa]}$ и так до конца.

*) Заметим, что это название мало оправдано. Такой способ собственно решения системы линейных уравнений, по-видимому, был известен еще арабам. Заслуга Гаусса не в указании способа, а в том, что он показал, как при использовании этого способа можно одновременно определить и веса неизвестных.

**) Порядок уравнений и неизвестных часто рекомендуется принять таким, при котором коэффициент при первом неизвестном наибольший из диагональных элементов.

Отметим (пока без пояснений), что с числами столбца S нужно делать те же операции, что и с числами столбца L . В нашем случае под столбцом S должно быть записано число $-\frac{[as]}{[aa]}$; здесь $[as]$ есть верхнее число столбца S . Подставляя найденное выражение для x во все остальные уравнения, получим промежуточную систему с числом неизвестных на единицу меньше. Коэффициенты промежуточной системы записываем в виде следующей матрицы:

y	z	u	L	S
$[bb1]$	$[bc1]$	$[bd1]$	$[bl1]$	$[bs1]$
$[cb1]$	$[cc1]$	$[cd1]$	$[cl1]$	$[cs1]$
$[db1]$	$[dc1]$	$[dd1]$	$[dl1]$	$[ds1]$

Если выполнить фактически подстановку первого уравнения исключения, то после приведения подобных членов получим следующие выражения для коэффициентов:

$$\left. \begin{aligned} [bb1] &= [bb] + \left\{ -\frac{[ab]}{[aa]} [ba] \right\}, \\ [bc1] &= [bc] + \left\{ -\frac{[ac]}{[aa]} [ba] \right\}, \\ [bd1] &= [bd] + \left\{ -\frac{[ad]}{[aa]} [ba] \right\}, \\ [bl1] &= [bl] + \left\{ -\frac{[al]}{[aa]} [ba] \right\}, \\ [cb1] &= [cb] + \left\{ -\frac{[ab]}{[aa]} [ca] \right\} \end{aligned} \right\} \quad (16.27)$$

и т. д.

Можно отметить мнемоническое правило для проверки всех этих формул: первый член справа содержит те же буквы, что и член слева после отбрасывания значка 1; второй член справа должен быть таким, чтобы после условного (формального!) сокращения на одинаковые множители получился такой же член, как и первый; в целом правая часть после формальных сокращений и приведения «подобных» членов должна дать нуль.

Изобразим матрицу коэффициентов нормальных уравнений с контрольным столбцом S точками, под этой матрицей выпишем

строку исключения, обозначим ее элементы кружками:

$$\begin{array}{cccccc}
 x & y & z & u & L & S \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \boxed{\cdot} & \boxed{\cdot} & \boxed{\cdot} & \cdot & \cdot & \cdot \\
 \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
 \circ & \circ & \circ & \circ & \circ & \circ \\
 * & * & * & * & * & * \\
 * & \boxed{*} & * & * & * & * \\
 * & * & * & * & * & *
 \end{array}$$

Матрица промежуточной системы с тремя неизвестными обозначена звездочками. Можно указать простое схематическое правило образования чисел, изображенных звездочками. В основной матрице отбросим первую строку, она использована для получения строки с кружками (строки исключения — «элиминационной строки»); чтобы получить m -ю строку из звездочек, нужно к каждому числу (кроме первого) m -й из сохраненных строк прибавить произведение числа строки исключения, стоящего в том же столбце, на первое число взятой точечной строки. На схеме исходные и полученные числа взяты в рамку. Такую операцию нужно сделать во всех столбцах (кроме первой основной матрицы) и во всех строках (кроме первой в основной матрице). Это правило иногда называют *правилом прямоугольного треугольника*, что оправдывается чертежом на нашей условной схеме: число в вершине прямого угла треугольника нужно сложить с произведением чисел, стоящих в концах гипотенузы.

Полученная после исключения первого неизвестного система обладает указанными выше двумя свойствами нормальной системы: матрица коэффициентов при неизвестных симметрична относительно главной диагонали, на главной диагонали положительные числа. Однако есть смысл вычислять и взаимно симметричные элементы ($[bc1]$ и $[cb1]$ и т. п.): поскольку они получаются не совсем одинаковыми операциями, их неравенство означает наличие ошибки, если разница больше 2—3 единиц последнего знака. Вследствие округлений возможна разница в одну-две единицы последнего знака; в таком случае необходимо изменить одно из чисел (или оба) так, чтобы иметь точное равенство.

Правильность составления строки исключения контролируется в столбце S , если в нем производить те же действия, что и в остальных столбцах: число в столбце S должно равняться сумме всех

остальных чисел этой же строки исключения без единицы. Если, вследствие различия в округлениях будет разница в одну-две единицы последнего знака (иногда и больше, если много неизвестных), то надо изменить число строки исключения в столбце S , чтобы было точное равенство в контроле. Это обеспечит надежность контроля в последующих вычислениях.

Первая промежуточная система, полученная после исключения первого неизвестного, также контролируется числами столбца S . Легко показать, что каждое из чисел этого столбца должно равняться сумме остальных чисел, стоящих в той же строке. И здесь возможно расхождение в единицах последнего знака, которое надо устранить изменением числа в столбце S .

После исключения первого неизвестного получается система нормальных уравнений с числом неизвестных на единицу меньше, чем в начальной системе. Из нее исключается второе неизвестное тем же способом, каким из начальной системы исключалось первое неизвестное, и с теми же контрольными вычислениями. Уравнение исключения имеет вид *):

$$y = -\frac{[bc1]}{[bb1]}z - \frac{[bd1]}{[bb1]}u - \frac{[bl1]}{[bb1]}. \quad (16.28)$$

В схеме записываются в соответствующих столбцах только коэффициенты. Подстановка y во все уравнения первой промежуточной системы, кроме первого, даст систему с числом неизвестных на две единицы меньше. Из системы с четырьмя неизвестными получим систему вида

z	u	L	S
$[cc2]$	$[cd2]$	$[cl2]$	$[cs2]$
$[dc2]$	$[dd2]$	$[dl2]$	$[ds2]$

Коэффициенты вычисляются по формулам, подобным формулам (16.27) для системы с тремя неизвестными:

$$\left. \begin{aligned} [cc2] &= [cc1] + \left\{ -\frac{[bc1]}{[bb1]} [cb1] \right\}, \\ [cd2] &= [cd1] + \left\{ -\frac{[bd1]}{[bb1]} [cb1] \right\}, \\ [cl2] &= [cl1] + \left\{ -\frac{[bl1]}{[bb1]} [cb1] \right\}, \\ [dc2] &= [dc1] + \left\{ -\frac{[bc1]}{[bb1]} [db1] \right\} \end{aligned} \right\} \quad (16.29)$$

*) Иногда рекомендуют перед исключением второго неизвестного изменить порядок уравнений и неизвестных так, чтобы коэффициент при первом неизвестном в промежуточной системе был наибольшим на главной диагонали.

и т. д., причем и здесь вычисления ведутся по схеме прямоугольного треугольника. Вычислять нужно и $[cs2]$, $[ds2]$ для контроля новой промежуточной системы; необходимо проверить также и строку исключения (сразу же после ее составления) по формуле

$$-\frac{[bs1]}{[bb1]} = -1 + \left(-\frac{[bc1]}{[bb1]}\right) + \left(-\frac{[bd1]}{[bb1]}\right) + \left(-\frac{[bl1]}{[bb1]}\right), \quad (16.30)$$

т. е. сумма чисел строки исключения (без S) должна быть на единицу меньше того числа, которое получится в строке исключения в столбце S .

Уравнения второй промежуточной системы контролируются совершенно так же, как и уравнения первой промежуточной и как уравнения основной, с помощью чисел столбца S .

Затем из первого уравнения второй промежуточной системы находится следующее уравнение исключения:

$$z = -\frac{[cd2]}{[cc2]} u - \frac{[cl2]}{[cc2]}; \quad (16.31)$$

в столбце « u » записывается коэффициент при u , в столбце L свободный член, кроме того, в столбце S записывается число $-\frac{[cs2]}{[cc2]}$. Выражение для z подставляется в следующее уравнение и получается одно уравнение вида

$$[dd3]u + [dl3] = 0 \quad (16.32)$$

с одним неизвестным u , которое в схеме записывается в столбцах u , L :

$$\begin{array}{cc} u & L \\ [dd3] & [dl3]; \end{array}$$

в столбце S записывается контрольное число $[ds3]$. Коэффициенты вычисляются по формулам

$$\left. \begin{aligned} [dd3] &= [dd2] + \left\{ -\frac{[cd2]}{[cc2]} [dc2] \right\}, \\ [dl3] &= [dl2] + \left\{ -\frac{[cl2]}{[cc2]} [dc2] \right\}, \\ [ds3] &= [ds2] + \left\{ -\frac{[cs2]}{[cc2]} [dc2] \right\}. \end{aligned} \right\} \quad (16.33)$$

Контроль уравнения (16.32) выполняется по формуле:

$$[dd3] + [dl3] = [ds3].$$

Из уравнения (16.32) для u получаем последнее уравнение исключения

$$\bar{u} = -\frac{[dl3]}{[dd3]}. \quad (16.34)$$

Здесь \bar{u} — обозначение корня системы нормальных уравнений, т. е. наиболее вероятное значение неизвестного u . Подставляя найденное значение \bar{u} в уравнение исключения (16.31), получим

$$\bar{z} = -\frac{[cd2]}{[cc2]} \bar{u} - \frac{[cf2]}{[cc2]}. \quad (16.35)$$

Подстановка \bar{u} и \bar{z} в уравнение исключения (16.28) даст

$$\bar{y} = -\frac{[bc1]}{[bb1]} \bar{z} - \frac{[bd1]}{[bb1]} \bar{u} - \frac{[bf1]}{[bb1]}. \quad (16.36)$$

Наконец, первое уравнение исключения (16.26) дает

$$\bar{x} = -\frac{[ab]}{[aa]} \bar{y} - \frac{[ac]}{[aa]} \bar{z} - \frac{[ad]}{[aa]} \bar{u} - \frac{[af]}{[aa]}. \quad (16.37)$$

Значения неизвестных контролируются на основании следующего простого соображения. Пусть в нормальных уравнениях сделана замена неизвестных:

$$\begin{aligned} x &= \xi + 1, \\ y &= \eta + 1, \\ z &= \zeta + 1, \\ u &= \nu + 1. \end{aligned}$$

Первое нормальное уравнение примет вид:

$$[aa]\xi + [ab]\eta + [ac]\zeta + [ad]\nu + \\ + \{[aa] + [ab] + [ac] + [ad] + [af]\} = 0.$$

В преобразованном уравнении член без неизвестного равен $[af]$ согласно контрольной формуле для первого нормального уравнения. Аналогично преобразуются и остальные уравнения. Отсюда следует: если в нормальных уравнениях заменить столбец L на S , то получится система уравнений для определения ξ , η , ζ , ν , т. е. неизвестных, которые точно на единицу меньше неизвестных x , y , z , u . Когда в столбце S производятся те же операции, что и в столбце L , то фактически решается система с неизвестными ξ , η , ζ , ν . Поэтому при решении системы нормальных уравнений одновременно производятся со столбцом S те же операции, что и со столбцом L , дающим x , y , z , u . Столбец S даст значения $\bar{\xi}$, $\bar{\eta}$, $\bar{\zeta}$, $\bar{\nu}$, которые и используются для контроля: при отсутствии ошибок они должны быть точно на единицу меньше, чем \bar{x} , \bar{y} , \bar{z} , \bar{u} . Как обычно, вследствие ошибок округления разница между \bar{x} и $\bar{\xi}$ и т. д. может на единицы последнего знака отличаться от числа 1. Величина допустимого отличия от единицы растет при увеличении числа неизвестных.

Отметим следующее обстоятельство, важное для вычисления средних ошибок неизвестных; при решении системы нормальных уравнений по способу Гаусса разрешается производить только те

арифметические операции, которые необходимы для этого способа; нельзя умножать или делить основные и промежуточные уравнения на какие-нибудь числа.

Помимо рассмотренной схемы решения системы нормальных уравнений методом последовательного исключения неизвестных применяются и другие схемы. Так, в геодезии широко употребляется так называемая *схема Гаусса — Дулитля*, в которой число записей значительно меньше, чем в рассмотренной схеме; выписываются только строки исключения и по одному (первому) столбцу в каждой промежуточной системе (см., например, Романовский, Математическая статистика). Удобная схема расположения записей для этого случая дана П. Т. Резниковским. С. Г. Маковер разработал матричный способ решения системы нормальных уравнений, который приводит к вычислительной схеме, аналогичной схеме Гаусса — Дулитля.

Кроме метода последовательного исключения и метода определителей, к системе нормальных уравнений можно применять и другие методы, подробное изложение которых читатель сможет найти в книге В. Н. Фаддеевой. Однако не все из этих методов следует применять, если требуется вычислить не только значения неизвестных, но и их веса.

§ 85. Вычисление весов неизвестных

Как уже упоминалось в § 81, наиболее вероятные значения неизвестных, полученные из системы линейных нормальных уравнений, являются линейными функциями случайных величин l_1, l_2, \dots, l_n . Для того чтобы получить эти функции, выпишем решение системы нормальных уравнений с помощью определителей (в случае системы четвертого порядка):

$$\bar{x} = \frac{D_x}{D}, \quad \bar{y} = \frac{D_y}{D}; \quad \bar{z} = \frac{D_z}{D}, \quad \bar{u} = \frac{D_u}{D}, \quad (16.38)$$

где D — определитель системы

$$D = \begin{vmatrix} [aa] & [ab] & [ac] & [ad] \\ [ba] & [bb] & [bc] & [bd] \\ [ca] & [cb] & [cc] & [cd] \\ [da] & [db] & [dc] & [dd] \end{vmatrix}, \quad (16.39)$$

D_x — определитель вида

$$D_x = - \begin{vmatrix} [al] & [ab] & [ac] & [ad] \\ [bl] & [bb] & [bc] & [bd] \\ [cl] & [cb] & [cc] & [cd] \\ [dl] & [db] & [dc] & [dd] \end{vmatrix}, \quad (16.40)$$

а D_y, D_z, D_u имеют аналогичные выражения с той разницей, что столбец свободных членов нормальных уравнений становится последовательно на второе, третье, четвертое места,

Подставляя в D_{ω} вместо элементов первого столбца их развернутые выражения, получим

$$D_{\omega} = - \sum_{k=1}^n l_k \begin{vmatrix} a_k & [ab] & [ac] & [ad] \\ b_k & [bb] & [bc] & [bd] \\ c_k & [cb] & [cc] & [cd] \\ d_k & [db] & [dc] & [dd] \end{vmatrix}. \quad (16.41)$$

Из выражений (16.38) и (16.41) для \bar{x} и D_{ω} и получается искомое линейное выражение \bar{x} через числа l_k :

$$\bar{x} = - \sum_{k=1}^n \frac{\Delta_k}{D} l_k, \quad (16.42)$$

где

$$\Delta_k = \begin{vmatrix} a_k & [ab] & [ac] & [ad] \\ b_k & [bb] & [bc] & [bd] \\ c_k & [cb] & [cc] & [cd] \\ d_k & [db] & [dc] & [dd] \end{vmatrix}. \quad (16.43)$$

Перейдем теперь к определению дисперсии неизвестной величины \bar{x} , считая числа l_k случайными. По теореме о дисперсии линейной функции взаимно независимых случайных величин l_k получим

$$\sigma_{\bar{x}}^2 = \sum_{k=1}^n \frac{\Delta_k^2}{D^2} \sigma_k^2, \quad (16.44)$$

где σ_k^2 — дисперсии последовательных условных уравнений. Мы будем считать условные уравнения равноточными (или приведенными к равноточным); поэтому

$$\sigma_k^2 = \sigma_0^2, \quad k = 1, 2, \dots, n, \quad (16.45)$$

где σ_0 — средняя квадратичная ошибка одного уравнения с весом единица. Таким образом, для равноточных условных уравнений

$$\sigma_{\bar{x}}^2 = \frac{\sigma_0^2}{D^2} \Delta, \quad (16.46)$$

где $\Delta = \sum_{k=1}^n \Delta_k^2$, а Δ_k определяется формулой (16.43).

Для того чтобы получить окончательное выражение для $\sigma_{\bar{x}}$, следует найти величину Δ . Преобразуем эту величину следующим

способом. Учитывая (16.43), имеем:

$$\Delta = \sum_{k=1}^n \Delta_k \Delta_k = \sum_{k=1}^n \Delta_k \begin{vmatrix} a_k & [ab] & [ac] & [ad] \\ b_k & [bb] & [bc] & [bd] \\ c_k & [cb] & [cc] & [cd] \\ d_k & [db] & [dc] & [dd] \end{vmatrix}.$$

Используя правило умножения определителя на число и правило сложения определителей, получим

$$\Delta = \begin{vmatrix} \sum_{k=1}^n a_k \Delta_k & [ab] & [ac] & [ad] \\ \sum_{k=1}^n b_k \Delta_k & [bb] & [bc] & [bd] \\ \sum_{k=1}^n c_k \Delta_k & [cb] & [cc] & [cd] \\ \sum_{k=1}^n d_k \Delta_k & [db] & [db] & [dd] \end{vmatrix}. \quad (16.47)$$

Из линейной алгебры известно, что определитель с двумя одинаковыми столбцами равен нулю. Поэтому

$$\sum_{k=1}^n b_k \Delta_k = \begin{vmatrix} \sum_{k=1}^n b_k a_k & [ab] & [ac] & [ad] \\ \sum_{k=1}^n b_k^2 & [bb] & [bc] & [bd] \\ \sum_{k=1}^n b_k c_k & [cb] & [cc] & [cd] \\ \sum_{k=1}^n b_k d_k & [db] & [dc] & [dd] \end{vmatrix} = 0$$

аналогично

$$\sum_{k=1}^n c_k \Delta_k = 0,$$

$$\sum_{k=1}^n d_k \Delta_k = 0.$$

Далее имеем:

$$\sum_{k=1}^n a_k \Delta_k = \begin{vmatrix} \sum_{k=1}^n a_k^2 & [ab] & [ab] & [ad] \\ \sum_{k=1}^n a_k b_k & [bb] & [bc] & [bd] \\ \sum_{k=1}^n a_k c_k & [cb] & [cc] & [cd] \\ \sum_{k=1}^n a_k d_k & [db] & [dc] & [dd] \end{vmatrix} = D.$$

Подставляя найденные суммы в определитель Δ и разлагая его по элементам первого столбца, получим

$$\Delta = D \cdot D_{11}, \quad (16.48)$$

где D_{11} — алгебраическое дополнение первого диагонального элемента определителя D системы нормальных уравнений.

Подставляя (16.48) в формулу (16.46) для σ_x^2 , получим окончательно

$$\sigma_x^2 = \frac{D_{11}}{D} \sigma_0^2. \quad (16.49)$$

По определению весов имеем

$$p_{\bar{x}} = \frac{\sigma_0^2}{\sigma_{\bar{x}}^2} = \frac{D}{D_{11}}; \quad \left. \begin{array}{l} \text{аналогично можно получить} \\ p_{\bar{y}} = \frac{D}{D_{22}}; \quad p_{\bar{z}} = \frac{D}{D_{33}}; \quad p_{\bar{u}} = \frac{D}{D_{44}}, \end{array} \right\} \quad (16.50)$$

где D_{22} , D_{33} , D_{44} — алгебраические дополнения последовательных элементов главной диагонали определителя системы нормальных уравнений.

Таким образом, можно сформулировать следующее правило нахождения весов способом определителей:

Вес каждого неизвестного равен определителю системы нормальных уравнений, деленному на алгебраическое дополнение того диагонального элемента этого определителя, который является коэффициентом при рассматриваемом неизвестном.

Это правило имеет смысл применять, если сама система нормальных уравнений решается способом определителей, что, как указывалось, целесообразно лишь для систем 2—3-го порядка.

Предположим теперь, что система нормальных уравнений решается способом Гаусса. Докажем, что в этом случае указанное правило

вычисления весов применимо не только к исходной нормальной системе, но и к любой из промежуточных систем *).

Произведем над определителем D системы нормальных уравнений преобразования, аналогичные тем, которые проводились при образовании первой промежуточной системы. Разделим первую строку этого определителя на $[aa]$ и вычтем затем эту строку из II, III и IV строк, умножая ее соответственно на $[ba]$, $[ca]$, $[da]$. Тогда получим:

$$D = [aa] \begin{vmatrix} 1; & \frac{[ab]}{[aa]}; & \frac{[ac]}{[aa]}; & \frac{[ad]}{[aa]} \\ 0; & [bb1] & [bc1] & [bd1] \\ 0; & [cb1] & [cc1] & [cd1] \\ 0; & [db1] & [dc1] & [dd1] \end{vmatrix}. \quad (16.51)$$

т. е.

$$D = [aa] \cdot D^{(1)}, \quad (16.52)$$

где через $D^{(1)}$ обозначен определитель первой промежуточной системы.

Аналогичные преобразования, произведенные над алгебраическими дополнениями диагональных элементов определителя D , дают

$$D_{22} = [aa] D_{22}^{(1)}, \quad D_{33} = [aa] D_{33}^{(1)}, \quad D_{44} = [aa] D_{44}^{(1)},$$

где $D_{22}^{(1)}$, $D_{33}^{(1)}$, $D_{44}^{(1)}$ — алгебраические дополнения элементов $[bb1]$, $[cc1]$, $[dd1]$ определителя $D^{(1)}$. Из последних формул и (16.50) находим

$$p_{\bar{y}} = \frac{D^{(1)}}{D_{22}^{(1)}}, \quad p_{\bar{z}} = \frac{D^{(1)}}{D_{33}^{(1)}}, \quad p_{\bar{u}} = \frac{D^{(1)}}{D_{44}^{(1)}}. \quad (16.53)$$

Таким образом, сформулированное утверждение доказано для первой промежуточной системы. Точно так же оно доказывается и для следующих промежуточных систем. Основываясь на этом утверждении, докажем следующую теорему Гаусса:

При решении системы нормальных уравнений по способу Гаусса коэффициент при последнем неизвестном в последнем промежуточном уравнении (с одним этим неизвестным) равняется его весу, если производить только те операции, какие для этого способа нужны.

Для доказательства образуем при помощи сформулированного правила веса предпоследнего и последнего неизвестного, входящих в предпоследнюю промежуточную систему:

$$p_{\bar{z}} = \frac{\begin{vmatrix} [cc2] & [cd2] \\ [dc2] & [dd2] \end{vmatrix}}{[dd2]}, \quad p_{\bar{u}} = \frac{\begin{vmatrix} [cc2] & [cd2] \\ [dc2] & [dd2] \end{vmatrix}}{[cc2]}.$$

*) Доказательство этого факта, а также нижеследующее доказательство теоремы Гаусса сообщено автору П. Т. Резниковским.

Из формулы для коэффициента $[dd3]$ последнего промежуточного уравнения, приведенной в предыдущем параграфе, следует, что

$$\left| \begin{array}{cc} [cc2] & [cd2] \\ [dc2] & [dd2] \end{array} \right| = [dd3][cc2].$$

Поэтому

$$\left. \begin{array}{l} p_{\bar{x}} = \frac{[dd3][cc2]}{[dd2]}, \\ p_{\bar{u}} = [dd3]. \end{array} \right\} \quad (16.54)$$

Последняя формула и доказывает теорему Гаусса.

Таким образом, при решении системы нормальных уравнений по способу Гаусса вес последнего неизвестного получается как бы автоматически: он равен коэффициенту, который все равно должен быть вычислен.

Вес предпоследнего неизвестного получается легко по первой формуле (16.54). Он равен произведению веса последнего неизвестного на отношение элементов главной диагонали предпоследней системы Гаусса (с двумя неизвестными). Обычно определяют веса последнего и предпоследнего неизвестных, затем изменяют в уравнениях порядок неизвестных так, чтобы последними стали следующие два неизвестных; считая с конца. При этом надо переставить и уравнения так, чтобы получилась симметричная матрица коэффициентов при неизвестных с положительными элементами на главной диагонали. Иными словами, после перестановки неизвестных и уравнений должна получиться система со свойствами системы нормальных уравнений.

Решение преобразованной системы дает веса еще двух неизвестных. При этом контролируется еще раз решение, так как вторично можно получить все неизвестные. Если есть уверенность в результатах первого решения, то можно ограничиться только составлением промежуточных систем, ибо этого достаточно для вычисления весов. При 3 или 4 неизвестных придется два раза решать систему, при 5 или 6 неизвестных — три раза и т. д.

Можно показать, что вес третьего с конца неизвестного определяется формулой *)

$$p_{\bar{y}} = [dd3][cc2] \frac{[bb_1]}{\left| \begin{array}{cc} [cc1] & [cd_1] \\ [dc1] & [dd1] \end{array} \right|}. \quad (16.55)$$

Если, наряду с формулами (16.54), применять и эту формулу, то систему нормальных уравнений 3-го порядка придется решать один раз, системы 4-го, 5-го и 6-го порядка два раза и т. д.

*) Можно было бы выписать также аналогичные формулы для весов четвертого, пятого и т. д. с конца неизвестных, но эти формулы содержат, соответственно, определители 3-го, 4-го и т. д. порядков и поэтому неудобны для употребления.

Изложенный способ определения весов существенно опирается на схему Гаусса решения системы нормальных уравнений, в которой выписываются все промежуточные системы. В геодезической литературе этот способ называют *способом Энке*. Если число неизвестных в системе нормальных уравнений больше шести, то при определении весов способом Энке систему нормальных уравнений приходится решать не менее трех раз (или даже не менее четырех раз, если применять только формулы (16.54)). Поэтому для систем высокого порядка следует обращаться к другим способам вычисления весов.

В способе, принадлежащем Н. И. Идельсону, наряду с данной системой нормальных уравнений m -го порядка, решается (параллельно) еще m систем с тем же определителем и следующими последовательностями свободных членов: $(1, 0, 0, \dots)$, $(0, -1, 0, \dots)$, $(0, 0, -1, \dots)$, ... Легко убедиться в том, что в каждой из таких систем одно из неизвестных равно обратной величине одного из весов неизвестных системы нормальных уравнений. Объем дополнительных вычислений для определения весов в способе Н. И. Идельсона примерно равен объему основных вычислений при решении системы нормальных уравнений способом последовательного исключения неизвестных. Поэтому этот способ безусловно выгоднее способа Энке, если порядок системы больше шести.

Из (16.53) следует, что обратные величины весов равны диагональным элементам матрицы, обратной по отношению к матрице системы нормальных уравнений. Поэтому применение к нормальным уравнениям любого из способов решения системы линейных алгебраических уравнений, основанных на нахождении обратной матрицы, дает не только значения неизвестных, но и их веса. На этом факте основано матричное изложение способа Н. И. Идельсона, данное С. Г. Маковым.

§ 86. Приближенное значение средней квадратичной ошибки на единицу веса. Средние квадратичные ошибки неизвестных

После решения нормальных уравнений получаются наиболее вероятные значения неизвестных \bar{x} , \bar{y} , \bar{z} , \bar{u} . Подстановка их в условные уравнения даст невязки, удовлетворяющие условию минимума суммы квадратов. Эти невязки часто называют *остающимися погрешностями*, сохраняя общее название «невязки» за результатом подстановки произвольных чисел x , y , z , u . Чтобы выяснить, как найденные \bar{x} , \bar{y} , \bar{z} , \bar{u} удовлетворяют условным уравнениям, необходимо вычислить остающуюся погрешность каждого условного уравнения. Это вычисление не представляет никаких затруднений при работе на арифмометре, ибо требуется найти сумму произведений, что делается без записи отдельных слагаемых*). Обозначим остающуюся погрешность через ϵ_k^* ($k = 1, 2, \dots, n$) и будем называть их *остатками*. Определим теперь сумму квадратов остатков, т. е. минимум

*) Знание всех ϵ_k необходимо, так как среди условных уравнений бывают содержащие грубые ошибки. Такие уравнения необходимо отбросить и заново произвести вычисления. Условные уравнения с грубыми ошибками обнаруживаются по слишком большим остающимся погрешностям сравнительно со значительным большинством ϵ_k . Обычно применяется правило трех сигм.

суммы квадратов невязок. Обозначим эту минимальную величину буквой \bar{s} . В случае четырех неизвестных

$$\bar{s} = \sum_{k=1}^n \epsilon_k^2 = \sum_{k=1}^n (a_k \bar{x} + b_k \bar{y} + c_k \bar{z} + d_k \bar{u} + l_k)^2. \quad (16.56)$$

Величину \bar{s} можно легко вычислить непосредственно по ϵ_k , так как все равно отдельные ϵ_k должны быть вычислены. Для вычисления \bar{s} может быть построена довольно простая формула. Несложное вычисление дает:

$$\begin{aligned} \bar{s} = & \sum_{k=1}^n \bar{x} (a_k^2 \bar{x} + a_k b_k \bar{y} + a_k c_k \bar{z} + a_k d_k \bar{u} + a_k l_k) + \\ & + \sum_{k=1}^n \bar{y} (b_k a_k \bar{x} + b_k^2 \bar{y} + b_k c_k \bar{z} + b_k d_k \bar{u} + b_k l_k) + \\ & + \sum_{k=1}^n \bar{z} (c_k a_k \bar{x} + c_k b_k \bar{y} + c_k^2 \bar{z} + c_k d_k \bar{u} + c_k l_k) + \\ & + \sum_{k=1}^n \bar{u} (d_k a_k \bar{x} + d_k b_k \bar{y} + d_k c_k \bar{z} + d_k^2 \bar{u} + d_k l_k) + \\ & + \sum_{k=1}^n l_k^2 + \bar{x} \sum_{k=1}^n a_k l_k + \bar{y} \sum_{k=1}^n b_k l_k + \bar{z} \sum_{k=1}^n c_k l_k + \bar{u} \sum_{k=1}^n d_k l_k. \end{aligned}$$

Каждое слагаемое возведено в квадрат, все члены с удвоенными произведениями выписаны в разных строчках в виде суммы двух одинаковых слагаемых, и сделано разложение на множители. Полученное равенство можно переписать в обозначениях Гаусса:

$$\begin{aligned} \bar{s} = & \bar{x} ([aa] \bar{x} + [ab] \bar{y} + [ac] \bar{z} + [ad] \bar{u} + [al]) + \\ & + \bar{y} ([ba] \bar{x} + [bb] \bar{y} + [bc] \bar{z} + [bd] \bar{u} + [bl]) + \\ & + \bar{z} ([ca] \bar{x} + [cb] \bar{y} + [cc] \bar{z} + [cd] \bar{u} + [cl]) + \\ & + \bar{u} ([da] \bar{x} + [db] \bar{y} + [dc] \bar{z} + [dd] \bar{u} + [dl]) + \\ & + [ll] + \bar{x} [al] + \bar{y} [bl] + \bar{z} [cl] + \bar{u} [dl]. \end{aligned}$$

Так как \bar{x} , \bar{y} , \bar{z} , \bar{u} — корни нормальных уравнений, то все суммы в круглых скобках дадут нули, и остается простое выражение

$$\bar{s} = \bar{x} [al] + \bar{y} [bl] + \bar{z} [cl] + \bar{u} [dl] + [ll], \quad (16.57)$$

для которого потребуется только одно дополнительное вычисление суммы квадратов чисел l_k .

При вычислении \bar{s} по этой формуле ошибки округлений будут значительно меньше, чем при прямом вычислении суммы квадратов остатков. Поэтому ее следует считать основной, а прямую

сумму квадратов остатков — контрольной. При сравнении этих двух значений \bar{s} может быть заметное расхождение, если n — не малое число. Для сравнения придется оценить предельную погрешность прямой суммы квадратов остающихся погрешностей. Пусть, например, все остатки даны с тремя надежными знаками после запятой, и число условных уравнений 10. Если остатки порядка 0,05, то предельная погрешность квадрата одного остатка равна

$$2 \cdot 0,05 \cdot 0,5 \cdot 10^{-3} = 0,5 \cdot 10^{-4};$$

так как их 10, то предельная погрешность прямой суммы квадратов невязок равна $0,5 \cdot 10^{-3}$. Если к этому прибавить предельную погрешность суммы по формуле (16.57), то получим верхнюю границу модуля разности между двумя значениями \bar{s} . Превышение этой границы наверняка обозначает ошибку в одном из чисел (иногда и в обоих).

После вычисления суммы квадратов остатков можно вычислить наиболее вероятное значение средней квадратичной ошибки на единицу веса σ_0 по формуле

$$\sigma_0^2 = \frac{\bar{s}}{n-m}, \quad (16.58)$$

где m — число неизвестных.

Эту формулу мы даем без вывода, так как он довольно громоздок. Вывод подобен тому, который был дан при решении аналогичной задачи для равноточных и неравноточных измерений одной величины. Эти задачи можно считать частным случаем задачи этой главы.

После вычисления σ_0 средние квадратичные ошибки неизвестных определяются по известным формулам, поскольку веса неизвестных уже найдены:

$$\sigma_{\bar{x}} = \frac{\sigma_0}{\sqrt{p_{\bar{x}}}}, \quad \sigma_{\bar{y}} = \frac{\sigma_0}{\sqrt{p_{\bar{y}}}} \quad \text{и т. д.} \quad (16.59)$$

Результат принято записывать в таком виде:

$$x = \bar{x} \pm \sigma_{\bar{x}}, \quad y = \bar{y} \pm \sigma_{\bar{y}} \quad \text{и т. д.}$$

Если принять положение о нормальном распределении случайных ошибок чисел l_k , то \bar{x} , \bar{y} , ..., как линейные функции l_k , также распределены нормально. Поэтому можно вычислять вероятности вида

$$P(|x - \bar{x}| < \sigma_{\bar{x}}) = 0,68, \quad P(|x - \bar{x}| < 3\sigma_{\bar{x}}) = 0,9973$$

(и аналогично для других неизвестных), т. е. находить границы, в которых с заданной вероятностью может заключаться истинное значение, или решать обратную задачу. Если нельзя пользоваться

нормальным законом, то аналогичные вероятности можно найти по неравенствам Чебышева.

Получаемые таким путем оценки правильны, если число наблюдений не мало (примерно 20 или больше). При малом числе наблюдений результаты нужно оценивать, пользуясь распределением Стьюдента (см. книгу В. И. Романовского «Основные задачи теории ошибок»).

§ 87. Пример и схема решения системы линейных условных уравнений

Дана система 13 условных уравнений с четырьмя неизвестными:

x	y	z	u	l	s	s	s^2
+ 0,998	+ 0,061	+ 1,000	+ 0,000	— 0,042	+ 2,017	— 0,015	0,000225
+ 0,897	— 0,398	+ 1	+ 0,033	+ 0,027	+ 1,559	+ 0,003	9
+ 0,566	— 0,742	+ 1	+ 0,122	+ 0,099	+ 1,045	+ 0,010	100
+ 0,081	— 0,897	+ 1	+ 0,222	+ 0,133	+ 0,539	— 0,020	400
— 0,416	— 0,819	+ 1	+ 0,269	+ 0,196	+ 0,230	+ 0,000	0
— 0,812	— 0,525	+ 1	+ 0,216	+ 0,209	+ 0,088	+ 0,001	1
— 0,995	— 0,093	+ 1	+ 0,046	+ 0,215	+ 0,173	+ 0,033	1089
— 0,915	+ 0,365	+ 1	— 0,210	+ 0,107	+ 0,347	— 0,018	324
— 0,591	+ 0,725	+ 1	— 0,478	+ 0,018	+ 0,674	— 0,034	1156
— 0,113	+ 0,895	+ 1	— 0,663	+ 0,008	+ 1,127	+ 0,024	576
+ 0,386	+ 0,831	+ 1	— 0,683	— 0,033	+ 1,501	+ 0,027	729
+ 0,793	+ 0,547	+ 1	— 0,495	— 0,106	+ 1,739	— 0,038	1444
+ 0,991	+ 0,120	+ 1,000	— 0,118	— 0,008	+ 1,985	+ 0,028	0,000784
Суммы							
+ 0,870	+ 0,070	+ 13,000	— 1,739	+ 0,823	+ 13,024	— 0,001	0,006837

Составляем нормальные уравнения, вычисляя их коэффициенты способом накопления, и записываем их по той же схеме, что и условные уравнения.

x	y	z	u	L	S
+ 6,909156	+ 0,080654	+ 0,870000	— 0,439975	— 0,630230	+ 6,789605
+ 0,080654	+ 4,936378	+ 0,070000	— 2,509864	— 0,523411	+ 2,053757
+ 0,870000	+ 0,070000	+ 13,000000	— 1,739000	+ 0,823000	+ 13,024000
— 0,439975	— 2,509864	— 1,739000	+ 1,623981	+ 0,189828	— 2,875030

Сохранение всех формально получаемых знаков и точное выполнение контрольных равенств позволяет считать, что в составлении нормальных уравнений нет ошибок.

Теперь отбросим в нормальных уравнениях те знаки, которые наверно ненадежны, сохраняя во всех числах только три знака после запятой. Следует заметить, что по правилам оценки ошибок предельными погрешностями

по крайней мере третий знак после запятой может быть неверен, но он сохраняется, так как операций не мало, и можно считать с вероятностью, близкой к единице, что действительные ошибки коэффициентов заметно меньше предельных погрешностей.

Переписываем систему нормальных уравнений, сохраняя три знака после запятой, и выпишем полную схему решения:

x	y	z	u	L	S	
6,909	0,081	0,870	-0,440	-0,630	6,790	
0,081	4,936	0,070	-2,510	-0,523	2,054	
<u>0,870</u>	0,070	13,000	<u>-1,739</u>	0,823	13,024	
-0,440	-2,510	-1,739	1,624	0,190	-2,875	
$x = +0,0973$ ($\xi = -0,9027$)	-0,0117	-0,1259	<u>+0,0637</u>	+0,0912	-0,9827	(1)
	4,935	0,060	-2,505	-0,516	1,974	I
	0,060	12,890	<u>-1,684</u>	0,902	12,168	
	-2,505	-1,684	1,596	0,150	-2,443	
	$y = +0,0875$ ($\eta = -0,9125$)	-0,0122	+0,5076	+0,1046	-0,4000	(2)
		12,889	-1,654	0,908	12,143	II
		-1,654	0,325	-0,112	-1,441	
		$z = -0,0749$ ($\zeta = -1,0749$)	0,1283	-0,0704	-0,9421	(3)
			0,113	0,004	0,117	III
			$u = -0,0354$ ($v = -1,0354$)			(4)

Пояснения к схеме вычислений

(1) — строка исключения неизвестной x ; ее подробная запись была бы

$$x = -0,0117y - 0,1259z + 0,637u + 0,0912.$$

Коэффициенты получены делением коэффициентов первого нормального уравнения на 6,909 и переменной знака.

I — система трех уравнений с неизвестными y , z , u . Коэффициенты этой системы получаются из коэффициентов заданной системы по правилу прямоугольного треугольника. Чтобы получить, например, число $-1,684$, нужно к числу $-1,739$ в вершине прямого угла прибавить произведение чисел 0,870 и 0,0637 на концах гипотенузы (числа, образующие треугольник, взяты на схеме в рамки). Во всякой строке левый конец гипотенузы постоянный — первое число строки.

(2) — Строка исключения неизвестного y из системы I.

II — система с неизвестными z и u ; получается из I и (2) по правилу прямоугольного треугольника.

(3) — строка исключения неизвестного z из системы II.

III — уравнение с одним неизвестным u ; получено из II и (3) по той же схеме.

Первые числа в каждой строке исключения — значения неизвестных, они получаются «обратным ходом» с помощью строк исключения. Числа под ними в скобках — контрольные; они получаются тем же «обратным ходом», если в строках исключения брать числа из столбца S вместо столбца L . В двух местах в столбце S последняя цифра уменьшена на единицу. Это было сделано для согласования контроля (устранение допустимого различия).

Последнее в столбце число « u », выделенное курсивом, есть вес последнего неизвестного по теореме Гаусса:

$$p_u = 0,113.$$

Вес \bar{z} получим, если вес \bar{u} умножим на отношение элементов главной диагонали системы II:

$$p_z = 0,113 \cdot \frac{12,889}{0,325} = 4,480.$$

Чтобы получить веса остальных неизвестных, вторично решаем систему нормальных уравнений, изменив порядок неизвестных и порядок уравнений на обратный.

u	z	y	x	L	S
1,624	-1,739	-2,510	-0,440	0,190	-2,875 (0)
-1,739	13,000	0,070	0,870	0,823	13,024 (1)
-2,510	0,070	4,936	0,081	-0,523	2,054 (0)
-0,440	0,870	0,081	6,909	-0,630	6,790 (0)
(\bar{u}) = -0,0424	+1,0708	+1,5456	+0,2709	-0,1170	+1,7703
	11,138	-2,618	0,399	1,026	9,945
	-2,618	1,057	-0,599	-0,229	-2,389
	0,399	-0,599	6,790	-0,579	6,011
	(\bar{z}) = -0,0759	+0,2351	-0,0358	-0,0921	-0,8928
		-0,442	-0,505	+0,012	-0,051
		-0,505	6,776	-0,616	5,655
		(\bar{y}) = +0,0838	+1,1425	-0,0271	+0,1154
			6,199	-0,602	5,597
			(\bar{x}) = +0,0971		

Вычисления выполнены так же, как и в первый раз. Для проверки снова вычислены неизвестные, хотя в этом нет действительной необходимости. Так как все корни малы, то даже заметные отличия корней при втором вычислении приходится считать допустимыми. В качестве окончательных значений можно принять средние из двух вычислений. Проверка первой и второй совокупности корней по нормальным уравнениям показывает, что обе совокупности могут быть одинаково приняты, ибо корни той и другой совокупности достаточно хорошо удовлетворяют нормальным уравнениям (максимальное отклонение 0,0005).

Находим веса двух «последних» неизвестных:

$$p_x = 6,199, \quad p_y = \frac{0,442}{6,776} \cdot 6,199 = 0,404.$$

Напомним еще, что

$$p_z = 4,480, \quad p_u = 0,113.$$

Принятые значения неизвестных (средние из двух решений)

$$\bar{x} = +0,0972; \quad \bar{y} = +0,0856; \quad \bar{z} = -0,0754; \quad \bar{u} = -0,0390$$

подставим во все условные уравнения, чтобы получить остающиеся погрешности. Результаты записаны в столбце ε рядом с матрицей условных уравнений. После этого непосредственно получаем

$$\bar{s} = \sum_{k=1}^n \varepsilon_k^2 = 0,006837.$$

Подсчитывая сумму квадратов остающихся погрешностей по выведенной в предыдущем параграфе формуле (16.57), получим

$$\bar{s}' = 0,007081.$$

Здесь $[U] = 0,182531$. Если принять во внимание накопление погрешностей от округлений при вычислении непосредственным, а также при вычислении по формуле (16.57), то можно считать разницу между \bar{s} и \bar{s}' допустимой. Выписанные четыре знака формально получаются, но не все верны. Мы примем

$$\bar{s} = 0,0070,$$

т. е. число, близкое к \bar{s}' на основании замечания о большей точности \bar{s}' по сравнению с \bar{s} . По формуле (16.58) наиболее вероятное значение средней квадратичной ошибки на единицу веса определится так:

$$\sigma_0^2 = \frac{0,0070}{13-4} = 0,000778,$$

$$\sigma_0 = 0,028.$$

Так как условные уравнения равноточны, то это же число есть средняя квадратичная ошибка одного условного уравнения. По формулам (16.59) найдем дисперсии неизвестных:

$$\sigma_{\bar{x}}^2 = \frac{0,00078}{6,2} = 0,00013; \quad \sigma_{\bar{y}}^2 = \frac{0,00078}{0,40} = 0,0020;$$

$$\sigma_{\bar{z}}^2 = \frac{0,00078}{4,5} = 0,00017; \quad \sigma_{\bar{u}}^2 = \frac{0,00078}{0,113} = 0,0069.$$

Отсюда

$$\sigma_{\bar{x}} = 0,011, \quad \sigma_{\bar{y}} = 0,045, \quad \sigma_{\bar{z}} = 0,013, \quad \sigma_{\bar{u}} = 0,083.$$

Результаты запишем в виде

$$x = 0,097 \pm 0,011, \quad y = 0,086 \pm 0,045,$$

$$z = -0,075 \pm 0,013, \quad u = -0,039 \pm 0,083.$$

Результаты показывают, что из заданной системы условных уравнений оказалось возможным достаточно уверенно определить только неизвестные \bar{x} и \bar{z} , значение \bar{y} определяется неуверенно, значение \bar{u} совершенно ненадежно. Действительно, учитывая нормальный закон распределения, можно написать по правилу сигмы:

$$P(-0,122 < u < +0,044) = 0,68.$$

Следовательно, даже знак u определен ненадежно, ибо с умеренной вероятностью 0,68 можно ожидать, что u примет положительное значение.

ГЛАВА 17

ЭМПИРИЧЕСКИЕ ФОРМУЛЫ

§ 88. Постановка задачи

Важнейшей задачей астрономии и вообще естествознания является разыскание закономерностей в явлениях природы. Средством для этого является накопление наблюдательного материала и извлечение из него тех или иных сведений о законах, которым подчиняется явление. Решение такой задачи исключительно важно, так как установление законов явлений позволяет судить о ходе явлений в будущем. Для иллюстрации достаточно напомнить только один пример из астрономии. Открытие законов движения небесных тел (законы Кеплера и законы Ньютона) позволяет предсказывать солнечные и лунные затмения с точностью, вполне достаточной для практики. В древности, когда затмения считались катастрофическими событиями, служители культов, умевшие чисто эмпирически предвидеть затмения, использовали предсказания их наступлений для усиления своей власти.

В дальнейшем мы будем говорить о закономерностях числовых и функциональных.

Рассмотрим такую (сравнительно) простую задачу: наблюдаются в некотором явлении значения величин t и x , т. е. из n наблюдений получены числа t_k и x_k ($k = 1, 2, \dots, n$). Предполагается, что x есть некоторая неизвестная функция аргумента t , и требуется найти функцию, которая приближенно представляла бы связь между t и x .

Задание наблюдений в такой задаче есть в известном смысле задание табличной функции, которую требуется аппроксимировать некоторой функцией.

К сказанному следует сделать важное замечание. В природе нет таких простых связей только между двумя величинами. Обычно каждая величина зависит от целого ряда величин. Задача о разыскании связи двух величин может ставиться либо при условии, что влиянием на величину x остальных аргументов можно пренебречь, либо при условии, что все остальные аргументы сохраняют постоянные (хотя бы приблизительно) значения во время всех наблюдений.

Первое условие означает, что при той точности, с которой производятся наблюдения, ошибки вследствие пренебрежения другими аргументами мало влияют на величину функции. В силу же второго условия остальные аргументы входят в функциональную зависимость в качестве неизменных параметров.

Кроме основной задачи отыскания закономерностей, при изучении наблюдений приходится иметь дело и с более скромной задачей. Наблюдения не могут дать значения функции при любых значениях аргумента. Поэтому, если требуется найти значение функции для значений аргумента, не данных в таблице, то должен быть использован математический аппарат для вычисления приближенных значений функции при значениях аргумента, отсутствующих в таблице. Это — задача интерполяции (иногда и экстраполяции), с которой мы встречались в части II; там же был указан один из способов решения задачи интерполяции.

В подобных случаях иногда ограничиваются проведением плавной кривой между точками графика, по которой и определяют значения функции при любых значениях аргумента. Если требуются только грубые значения, то этот способ вполне пригоден, ибо он не требует никаких вычислений. Его очевидный недостаток — излишняя зависимость от исследователя, ибо он проводит «плавную» кривую на глаз, и в какой-то мере субъективно, поскольку и понятие о плавности, и глаза у разных исследователей бывают различными. Поэтому результаты разных исследователей затруднительно сравнивать или сводить их в одно, что часто приходится делать в астрономии.

Всякую функцию, приближающую табличную функцию, полученную из наблюдений, называют *эмпирической функцией* или формулой.

Задача о построении эмпирической функции требует введения двух условий, без которых эта задача, так же, как и общая задача о приближении функции, становится неопределенной.

1. Нужно выбрать аналитическое строение функции, которая должна приблизить табличную эмпирическую функцию.

2. Нужно условиться, как удовлетворить естественному требованию, чтобы эмпирическая формула по возможности хорошо представляла наблюдения.

§ 89. Выбор типа формулы

Выбор типа формулы является наиболее неопределенной и затруднительной частью работы, которую иногда приходится повторять не один раз. Работу начинают с построения графика эмпирической таблицы. Уже сравнение этого точечного графика с разными кривыми, уравнения которых известны, дает в ряде случаев указания на возможный тип формул. Например, если точки графика

располагаются так, как на рис. 9, т. е. располагаются с небольшими отклонениями вдоль прямой, то формулу следует искать в виде

$$x = a + bt,$$

где a и b — коэффициенты, x подлежащие определению по материалу наблюдений.

Иногда бывают известны весьма общие свойства связи между величинами, например величины примерно обратно пропорциональны друг другу; одна величина есть периодическая функция другой с приближенно известным периодом, и т. п. Если, например, точки располагаются как на графике (рис. 10), т. е. как будто расположены вдоль ветви гиперболы с асимптотами, параллельными осям координат, то можно испытывать обобщенное уравнение равнобедренной гиперболы вида

$$x = \frac{a + bt}{c + dt},$$

т. е. дробно-линейную функцию (один из коэффициентов можно положить равным единице). Такую же функцию можно взять, когда известно, что величины приблизительно обратно пропорциональны.

В случае периодичности можно искать формулу вида

$$x = a \sin \left(\frac{2\pi b}{P} t + c \right)$$

или тригонометрический полином порядка выше первого.

Наконец, может быть случай, когда можно построить грубую теорию явления. Тогда функция, определяемая этой теорией, может дать представление о возможном виде эмпирической формулы,

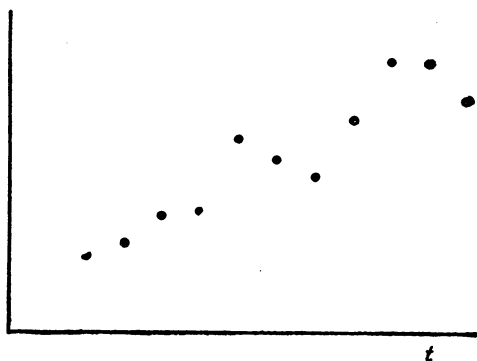


Рис. 9. Пример эмпирического распределения, близкого к прямой пропорциональности.

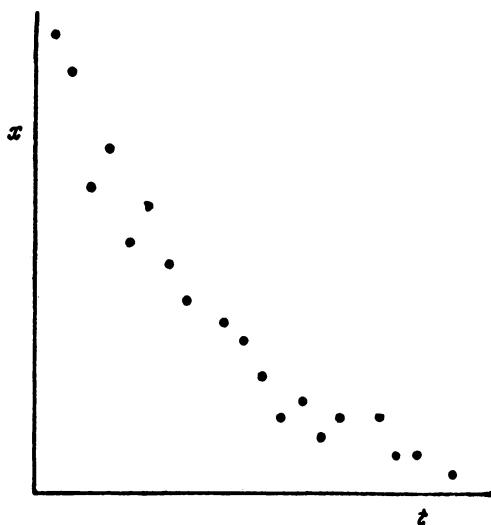


Рис. 10. Пример эмпирического распределения, близкого к обратной пропорциональности.

Пример 1. Строится закон масса — светимость для звезд. По общим соображениям можно грубо считать, что светимость звезд определенного спектрального класса пропорциональна величине поверхности. Считая звезду шаром, получаем, что светимость пропорциональна квадрату радиуса. При том же упрощающем предположении масса пропорциональна кубу радиуса как величина, пропорциональная объему. Отсюда исключением радиуса получается формула вида

$$L = am^{\frac{2}{3}},$$

где L — светимость, m — масса, a — некоторый параметр. В таком виде формула, конечно, непригодна, ибо упрощения были чрезмерно велики. Поэтому следует взять формулу вида

$$L = am^b + c,$$

где a , b , c — параметры, подлежащие определению по наблюдательному материалу.

Пример 2. Строится эмпирический закон растворимости некоторого вещества в определенной жидкости. Нужно получить формулу, выражающую количество растворившегося вещества в функции времени. За основу можно принять простой дифференциальный закон: количество вещества, растворившегося за малый промежуток времени, пропорционально этому промежутку и количеству нерастворенного вещества. Обозначим через M количество вещества в начальный момент, через x — количество растворившегося вещества к текущему моменту t ; тогда по указанному закону

$$dx = k(M - x) dt;$$

$k > 0$ — коэффициент пропорциональности. Решение этого простого дифференциального уравнения с разделяющимися переменными дает связь

$$x = M - be^{-kt},$$

где b — произвольная постоянная. Учитывая, что $x = 0$ при $t = 0$, имеем $b = M$, откуда

$$x = M - Me^{-kt}.$$

Исходя из этой приближенной зависимости, можно пытаться искать формулу вида

$$x = M - ae^{-bt},$$

где a и b — параметры.

Каков бы ни был вид эмпирической формулы, общее у них всех — это наличие нескольких буквенных параметров, которые должны быть определены в соответствии с принятым критерием близости формулы к таблице. Наличие нескольких параметров придает формуле большую гибкость.

Обычно стремятся выполнить еще одно условие: желательно, чтобы формула либо была линейной относительно параметров, либо простыми подстановками приводилась к линейному виду. По этой причине часто пользуются алгебраическими или тригонометрическими полиномами.

§ 90. Определение значений параметров по принципу Лежандра

Способ вычисления параметров эмпирических формул определяется соглашением о критерии близости значений, даваемых формулой, к табличным. В задаче точечной интерполяции (часть II) мы подбирали алгебраический (или тригонометрический) полином так, чтобы он в узлах давал точно табличные значения функции. Это соглашение, как упоминалось, имеет смысл, если таблица точная, т. е. если можно ручаться за все знаки чисел, приведенных в таблице. В нашей задаче, где и значения функции и значения аргумента получаются из наблюдений, их нельзя считать точными во всех формально полученных знаках. Если, например, наблюдения угла даются с точностью до $0''$, 1, то случайные и неучтенные систематические ошибки в совокупности могут достигать $1 - 2'$, а иногда и больше (например, при наблюдениях комет). Поэтому нет смысла требовать от эмпирической формулы, чтобы она совершенно точно представляла табличную функцию *).

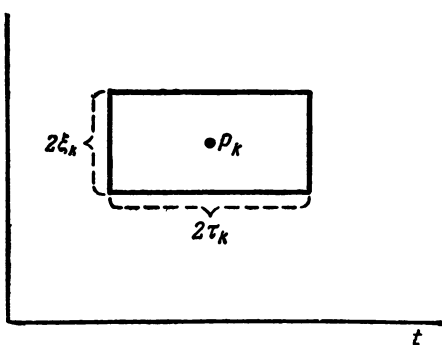


Рис. 11. «Прямоугольник возможных отклонений» точки эмпирического распределения.

Поясним это графически (рис. 11).

Положим, что известны предельные погрешности наблюдений или средние квадратичные их ошибки. Предположим, что мы знаем предельные погрешности t_k и x_k ; обозначим их τ_k и ξ_k . Точку графика с координатами (t_k, x_k) обозначим P_k . Наблюдения дают последовательность точек $P(t_k, x_k)$. Если ошибки t и x можно считать независимыми, то абсцисса t при других комбинациях факторов, вызывающих ошибки, могла бы дать любое значение от $t_k - \tau_k$ до $t_k + \tau_k$. Аналогично, вместо x_k могло бы получиться любое число от $x_k - \xi_k$ до $x_k + \xi_k$. Таким образом, вследствие ошибок точка P_k может попасть в любое место прямоугольника со сторонами $2\tau_k$ и $2\xi_k$ и центром в точке (t_k, x_k) . Такие прямоугольники следует построить около всякой точки, полученной из наблюдений. Было бы,

*) Это было бы трудно и чисто технически, ибо желательно брать побольше наблюдений, а тогда пришлось бы пользоваться полиномами высоких порядков (по условию точечной интерполяции степень полинома на единицу меньше числа узлов).

в сущности, достаточно, чтобы график эмпирической формулы прошел через все прямоугольники.

Если вместо предельных погрешностей даны средние квадратичные $\sigma_{t,k}$ и $\sigma_{x,k}$, то можно свести их к предельным с вероятностью, близкой к единице (2σ с вероятностью 0,95 или 3σ с вероятностью 0,9973).

Однако в такой постановке задача построения эмпирических формул выше не ставилась, ибо она была бы слишком сложной. Поэтому на практике был выбран другой путь, а именно, использование уже знакомого нам принципа Лежандра:

Параметры выбранной эмпирической формулы определяются так, чтобы сумма квадратов отклонений от табличных значений функции была наименьшей.

Это условие приближения эмпирической табличной функции является простым соглашением, которому нельзя дать вероятностного объяснения. Можно только мотивировать допустимость этого соглашения следующими соображениями. Число значений эмпирической функции в обычных задачах не очень мало, поэтому условие малости отклонений для отдельных точек осуществить невозможно. Можно ввести только некоторое общее условие, относящееся ко всей совокупности точек. Нет смысла ставить условие минимума с учетом знаков отклонений, так как обычно отклонения бывают как положительными, так и отрицательными; такие отклонения, даже будучи большими по абсолютной величине, могут взаимно компенсироваться. Было бы довольно естественным условие минимума суммы модулей отклонений. Однако оно привело бы к более громоздким вычислениям. Можно бы, конечно, потребовать, чтобы сумма четвертых или шестых или еще более высоких степеней имела минимум, но это было бы формальным упражнением в обобщениях. Вот почему в настоящее время общепринято определять параметры по способу наименьших квадратов, что приводит к довольно простым вычислениям.

Определение параметров эмпирической формулы на основе принципа Лежандра производится следующим образом. Последовательная подстановка всех табличных значений аргумента и функции в выбранную формулу с буквенными параметрами приводит к системе условных уравнений для определения численных значений параметров по наблюдениям. Эти уравнения в общем случае несовместны, причем несовместность объясняется не только случайными ошибками измерений, но и тем, что выбранная формула есть только приближение к неизвестной точной формуле и что фактически мы пренебрегаем зависимостью функции от других аргументов. По условным уравнениям составляются нормальные уравнения, которые и решают одним из указанных выше способов.

Так как вычисления удобно выполнять в случае линейных условных уравнений, то формулы и следует выбирать так, чтобы они

были линейными относительно параметров или надлежащими подстановками легко приводились бы к линейным.

Вычисление средних квадратичных ошибок неизвестных здесь утрачивает вероятностный смысл; их правильнее называть *средними квадратичными отклонениями*. Однако эти величины все-таки следует вычислять, так как они дают представление о надежности вычисленных значений параметров, в частности показывают, сколько знаков в параметрах следует сохранить.

§ 91. Проверка эмпирической формулы

Так как, в отличие от точечной интерполяции, эмпирическая формула не представляет точно табличных значений функции, то естественный способ первичной проверки заключается в вычислении по выведенной формуле значений функции при табличных значениях аргументов и сравнении результатов с наблюдаемыми значениями функции. Иными словами, вычисляются *остающиеся погрешности* условных уравнений. Подобная таблица и может служить основанием для проверки пригодности эмпирической формулы.

Если все остатки по модулю не превышают предельных ошибок измерений функции, то можно считать выведенную эмпирическую формулу хорошей. Если все остатки заметно превышают по модулю предельные погрешности, то формула явно непригодна. Чаще всего бывает промежуточный случай — есть и большие и малые остатки. Тогда можно использовать еще знаки остатков; если знаки чередуются, то формула может быть принята; если же положительные и отрицательные остатки идут группами, то формула вызывает сомнения, так как имеются систематические отклонения.

Наконец, для оценки пригодности формулы можно использовать среднюю квадратичную ошибку одного уравнения. Если эта величина имеет порядок около половины предельной погрешности измерений функции, то формулу можно принять. Если она превышает предельную погрешность, то формула непригодна.

Если после обследования формула совсем непригодна или вызывает сомнения, то рекомендуется работу повторить с несколько измененной формулой; из полученных нескольких формул лучшей следует считать ту, у которой средняя квадратичная ошибка на единицу веса меньше. Выбор лучшей из формул удобно начинать с построения графика эмпирической формулы и сравнения его с графиком табличных значений.

При оценке пригодности формулы иногда можно использовать физические сведения об определяемой функции. Например, если эмпирическая формула дает максимум при некотором значении аргумента, а по общим сведениям о явлении функция должна быть монотонной в окрестности этого значения аргумента, то построенная эмпирическая формула не может считаться удовлетворительной,

§ 92. Пример вывода эмпирической формулы

Наблюдения дали табличную функцию:

t	0,0	0,1	0,2	0,3	0,4	0,5
x	1,3	1,0	0,8	0,6	0,4	0,1

Построение графика в малом масштабе дает приблизительно прямую. Поэтому пробуем построить линейную формулу вида

$$x = a + bt,$$

где a и b — неизвестные коэффициенты.

Составляем условные уравнения по схеме, в которую включены также контрольные числа s , т. е. суммы коэффициентов и свободных членов, перенесенных влево. В схему включаются также столбцы остающихся погрешностей e и их квадратов (эти два столбца, разумеется, заполняются после решения системы нормальных уравнений). Последний столбец можно и не за-

полнять, если $\bar{s} = \sum_{k=1}^6 e_k^2$ вычисляется на арифмометре по способу накопления.

a	b	$-x$	s	e	$e^2 \cdot 10^4$
1	0,0	-1,3	-0,3	-0,03	9
1	0,1	-1,0	+0,1	+0,04	16
1	0,2	-0,8	+0,4	+0,01	1
1	0,3	-0,6	+0,7	-0,02	4
1	0,4	-0,4	+1,0	-0,05	25
1	0,5	-0,1	+1,4	+0,03	9
6	1,5	-4,2	+3,3	-0,02	0,0064

Нормальные уравнения имеют вид:

a	b	L	S
6,0	1,5	-4,2	3,3
1,5	0,55	-0,65	1,4
7,5	2,05	-4,85	4,7

Контроль составления нормальных уравнений осуществляется числами S , которые получены в первом уравнении сложением всех чисел s , так как коэффициенты при a все равны единице, во втором уравнении S есть сумма произведений чисел s на коэффициенты при неизвестном b в условных уравнениях. Кроме того, дополнительный контроль получается сложением коэффициентов и свободных членов нормальных уравнений и сравнением этой суммы с суммой всех чисел S . Этот контроль записан в третьей строке.

Решаем систему нормальных уравнений с помощью определителей. Чтобы избежать накопления погрешностей от приближенных вычислений, производим вычисления так, как если бы коэффициенты нормальных уравнений были точными числами. В нашей задаче это не доставит затруднений, поскольку все коэффициенты имеют мало значащих цифр. Вычислим

определители:

$$D = \begin{vmatrix} 6 & 1,5 \\ 1,5 & 0,55 \end{vmatrix}, \quad D_a = \begin{vmatrix} 4,2 & 1,5 \\ 0,65 & 0,55 \end{vmatrix}, \quad D_\alpha = \begin{vmatrix} -3,3 & 1,5 \\ -1,4 & 0,55 \end{vmatrix},$$

$$D_b = \begin{vmatrix} 6 & 4,2 \\ 1,5 & 0,65 \end{vmatrix}, \quad D_\beta = \begin{vmatrix} 6 & -3,3 \\ 1,5 & -1,4 \end{vmatrix}.$$

Определители D_a и D_β получаются заменой столбца L на столбец S , чтобы иметь контроль решения нормальных уравнений, описанный при решении системы нормальных уравнений: неизвестные a и β должны быть на единицу меньше, чем a и b . Вычисление определителей дает числа:

$$D = 1,05, \quad D_a = 1,335, \quad D_b = -2,4, \quad D_\alpha = 0,285, \quad D_\beta = -3,45.$$

Вычисляем основные и контрольные неизвестные:

$$a = \frac{1,335}{1,05} = 1,271, \quad b = -\frac{2,4}{1,05} = -2,286,$$

$$\alpha = \frac{0,285}{1,05} = 0,271, \quad \beta = -\frac{3,45}{1,05} = -3,286.$$

Вычисляем среднюю квадратичную ошибку одного условного уравнения, для чего подставляем в условные уравнения найденные значения коэффициентов и вычисляем остатки, обозначенные в схеме условных уравнений буквой ϵ :

$$\sigma_0^2 = \frac{0,0064}{6-2} = 0,0016, \quad \sigma_0 = 0,04.$$

Определяем веса неизвестных с помощью определителей:

$$p_a = \frac{1,05}{0,55} = 1,91, \quad p_b = \frac{1,05}{0,18} = 0,18.$$

Определяем средние квадратичные ошибки (точнее, средние квадратичные отклонения) неизвестных:

$$\sigma_a^2 = \frac{0,0016}{1,91} = 0,00084, \quad \sigma_b^2 = \frac{0,0016}{0,18} = 0,0089,$$

$$\sigma_a = 0,029, \quad \sigma_b = 0,094.$$

Результат определения коэффициентов можно записать так:

$$a = 1,271 \pm 0,029, \quad b = -2,286 \pm 0,094.$$

Учитывая величины средних квадратичных ошибок, результат запишем в виде

$$a = 1,27 \pm 0,03, \quad b = -2,29 \pm 0,09;$$

эмпирическая формула имеет вид

$$x = 1,27 - 2,29 t.$$

Рассмотрение остающихся погрешностей (остатков) показывает, что формула вполне удовлетворительно представляет наблюдения, если считать, что значения x даны с предельной погрешностью 0,05; модуль только одного остатка равен 0,05, а модули остальных остатков меньше.

ЧАСТЬ V

ОБРАБОТКА СТАТИСТИЧЕСКОГО МАТЕРИАЛА

ГЛАВА 18

ОБРАБОТКА ОДНОМЕРНОЙ СТАТИСТИЧЕСКОЙ СОВОКУПНОСТИ

§ 93. Статистические совокупности

Одна из задач естествознания заключается в изучении предметов или явлений, которые объединяются совместно по некоторым признакам (отдельные объекты совокупности отличаются друг от друга по каким-либо другим признакам). В астрономии и в других областях знаний такая задача встречается достаточно часто. Звезды любого каталога, каталог элементов орбит малых планет и т. п. представляют собой подобные совокупности. Например, в каталоге элементов орбит малых планет известные астероиды включены в один каталог, так как это малые по массам и размерам небесные тела, движение которых определяется главным образом притяжением к Солнцу. Можно сказать, что объекты каталога объединены в совокупность по некоторому качественному признаку.

Но каждый объект совокупности имеет собственные признаки, меняющиеся при переходе от одной малой планеты к другой. Если мы выделим из различных признаков малых планет те, которые определяют их движение, т. е. элементы оскулирующих эллипсов какой-нибудь эпохи, то получим *статистическую совокупность*, в которой каждому члену ее соответствуют числовые значения шести элементов орбиты, описывающих движение. Выделив какой-нибудь элемент орбит, например, наклонность, мы получим *одномерную статистическую совокупность*.

Мы ограничимся рассмотрением одномерных и двумерных совокупностей, хотя иногда приходится иметь дело с совокупностями большей размерности. В таких случаях следует обращаться к более полным пособиям — например, книге В. И. Романовского.

Первая задача при изучении статистических совокупностей заключается в нахождении приближенного выражения для функции распределения или плотности вероятности по эмпирическому материалу и в выводе способа получения нескольких

рационально выбранных числовых характеристик всей совокупности, которые дали бы общее представление о всей совокупности.

С этой задачей тесно связана вторая: найти вероятность того, что случайно выбранный объект совокупности имеет значение величины в заданных границах.

Этими задачами можно было бы ограничиться, если бы эмпирическая совокупность содержала все объекты исследуемого типа; такие полные совокупности называют *генеральными*. Простейший пример генеральной совокупности — материалы переписи населения, в которых имеются сведения о всех гражданах. В совокупностях, исследуемых в астрономии (совокупность малых планет, совокупность переменных звезд определенного типа и т. п.), генеральная совокупность почти всегда неизвестна, поскольку ежегодно открываются новые объекты каждой совокупности. Исключением можно считать совокупность звезд до определенной видимой величины, доступной обычным инструментам. Такие звезды открыты действительно все, но совокупность одних видимых величин звезд не представляет для науки большого интереса.

В обычных задачах известна совокупность, представляющая *выборку* из существующей, вообще говоря, генеральной совокупности, только часть которой удалось обнаружить наблюдениями. Если одни объекты генеральной совокупности чисто случайно попадают в наблюдаемый материал, а другие не попадают, то мы имеем *случайную выборку* (выборочную совокупность) из генеральной совокупности. Но в целом ряде задач объекты генеральной совокупности не все попадают в статистический материал не по случайным причинам, а потому, что они либо не могут быть открыты при современных средствах наблюдений, либо потому, что по разным причинам нет возможности выделить у них тот признак, по которому подобрана совокупность. В таких случаях говорят, что совокупность есть результат *селекции*.

Примеры первого вида селекции (по ограниченности наблюдательных средств) очевидны: современными инструментами нельзя пока обнаружить звезды, имеющие видимую величину $+24^m$ и больше; очень малые астероиды обнаруживаются только при сближении с Землей и т. п. Несколько менее очевидны случаи селекции вследствие свойств небесных тел. Можно указать следующие два примера: 1) обнаружить затменную двойную звезду возможно только при условии, что наблюдатель находится в некоторой ограниченной области пространства, образованной конусом, касающимся поверхностей обеих звезд; 2) в силу ограниченности разрешающей способности спектральных приборов нельзя найти лучевые скорости звезд, если они не превышают определенную величину. Если для выборочной совокупности полностью решены первые две задачи, то возникает третья задача: выяснить, в какой мере полученные числовые

параметры выборочной совокупности описывают генеральную совокупность.

В этой части книги мы будем рассматривать главным образом первые две задачи. Мы ограничимся в основном вопросом, указанным выше: разработать способы получения нескольких чисел, которые можно было бы считать достаточно полно описывающими данную статистическую совокупность. Чтобы подойти к решению такой задачи, значение того числового признака (или двух и т. д.), который изучается, будем считать случайной величиной. Если возможно найти приближенно закон распределения этой случайной величины, то параметры его могут считаться достаточными характеристиками всей совокупности, так как по этим параметрам можно вычислить вероятность того, что значение изучаемой величины заключено между любыми заданными границами. Знание вероятности позволит определить с той же вероятностью количество соответствующих объектов умножением общего числа объектов на значение вероятности.

Способ нахождения закона распределения по эмпирическому материалу является таким же, как и вообще способ построения эмпирических формул (см. предыдущую главу). Разница в том, что количество известных теоретических законов распределения ограничено. Довольно часто сначала проверяют пригодность нормального закона распределения.

Выбранный теоретический закон должен быть проверен по наблюдательному материалу.

§ 94. Дискретное эмпирическое распределение и его числовые характеристики

В астрономических работах редко, но все же приходится иметь дело с изучением дискретного эмпирического распределения (например, совокупности кратных звезд). Дискретное распределение следует рассмотреть еще и потому, что техника обработки непрерывного эмпирического распределения почти такая же, как и дискретного.

Материал, полученный из наблюдений, представляет собой каталог, в котором для каждого объекта совокупности (для каждого наблюдения) указано значение дискретной случайной величины X . Пусть величина X может принимать значения

$$x_1, x_2, \dots, x_s,$$

которые выписаны в порядке возрастания. Количество наблюдений, давшее каждое из возможных значений x_1, x_2, \dots, x_s величины X , называются *численностями* и обозначаются через n_1, n_2, \dots, n_s ; отношение численности n_k к общему объему n совокупности назовем *эмпирической вероятностью* того, что величина X примет значение x_k .

Первичная обработка материала, содержащего n объектов, начинается с подсчета численностей и оформляется в виде следующей таблицы:

$x_1, x_2, \dots, x_{s-1}, x_s$	$(x_{s+1} > x_s)$	$\sum_{k=1}^s n_k = n,$
$n_1, n_2, \dots, n_{s-1}, n_s$	(0)	
$p_1, p_2, \dots, p_{s-1}, p_s$	(0)	$\sum_{k=1}^s p_k = 1, p_k = \frac{n_k}{n},$
$0 = N_1, N_2, \dots, N_{s-1}, N_s$	$N_{s+1} = n$	$N_k = \sum_{l=1}^{k-1} n_l,$
$0 = P_1, P_2, \dots, P_{s-1}, P_s$	$P_{s+1} = 1$	$P_k = \frac{N_k}{n}.$

В этой таблице числа n_1, n_2, \dots, n_s представляют численности значений величины. Первые две строки назовем *таблицей эмпирического распределения* случайной величины. Вместо чисел n_k следовало бы ввести эмпирические вероятности $p_k = \frac{n_k}{n}$, помещенные в третьей строке. Однако на практике удобнее иметь дело с целыми числами, а деление на n для перехода к вероятностям отнести к концу вычислений; поэтому числа p_k редко включаются в таблицу. В четвертой строке даны числа N_k объектов, у которых значения случайной величины меньше x_k . Очевидно, имеем

$$N_1 = 0, \quad N_2 = n_1, \\ N_3 = n_1 + n_2, \dots, N_s = n_1 + n_2 + \dots + n_{s-1} = n - n_s.$$

Добавим в первой строке еще одно значение x_{s+1} , как угодно близкое к x_s и несколько большее его; так как наблюдения не дают такого значения, то $n_{s+1} = 0$; тогда по определению чисел N_k будет $N_{s+1} = n$. Если все числа $N_k (k=1, 2, \dots, s, s+1)$ разделить на общее число наблюдений, то получим числа P_k записанные в пятой строке. Их можно считать эмпирическими значениями функции распределения дискретной случайной величины X . В случае дискретной величины функция распределения имеет разрыв при каждом из значений x_k .

Основная задача — определение вероятности случайной величине принять какое-нибудь из своих значений между заданными границами α и β — с помощью построенной таблицы решается очень просто. Если не ставится вопрос о совпадении X с одной из заданных границ (или с обеими), то всегда

$$P(\alpha < X < \beta) = \sum_{\substack{x < \beta \\ x > \alpha}} p_k.$$

Суммирование производится по всем k , при которых выполняется неравенство под знаком вероятности.

Если требуется вычислить

$$P(\alpha \leq X \leq \beta),$$

то способ решения зависит от совпадения или несовпадения одной или обеих границ с табличными значениями. Если, например, $\beta = x_r$, где $r \leq s$, а α не равно никакому из чисел x_k , то имеем

$$P(\alpha \leq X \leq \beta) = \sum_{x > \alpha}^{k=r} p_k;$$

суммирование начинается с члена, соответствующего значению X , ближайшему справа к числу α .

Рассмотренная таблица дает полные сведения о дискретной случайной величине, если таблицу возможно построить и она невелика. Такую таблицу можно использовать для описания совокупности и решения основной задачи, если речь идет об одной совокупности определенного типа. Если нужно сравнивать разные совокупности одного типа (например, совокупности кратных звезд в разных областях Галактики), то сравнение полных таблиц далеко не всегда легко провести. Удобно иметь определенные однообразно построенные немногие числовые характеристики совокупности, чтобы сравнивать эти характеристики, вместо сравнения полных таблиц. Рассмотрим некоторые характеристики.

1. Среднее значение совокупности

Простейшей суммарной числовой характеристикой совокупности считают среднее значение случайной величины, определяемое, как и математическое ожидание дискретной величины, по формуле

$$\bar{x} = \sum_{k=1}^s x_k p_k. \quad (18.1)$$

Однако, в отличие от определения математического ожидания, p_k суть эмпирические вероятности.

Среднее значение, как и математическое ожидание, есть число размерное, поэтому при сравнении однотипных совокупностей нужно, чтобы X измерялись одинаковыми единицами. Если измеряются совокупности разных типов, то сравнивать средние значения нельзя, так как это разнородные величины. Если \bar{x} не равняется ни одному из возможных значений величины, то можно считать \bar{x} формальной характеристикой и брать в качестве приближения ближайшее возможное значение.

2. Медиана

Подобно медиане теоретического распределения непрерывной случайной величины, медиана \bar{x}_m эмпирического распределения дискретной величины определяется условием: эмпирическая вероятность получить значение меньше медианы должна равняться половине (говорят еще, что медиана «делит совокупность пополам»). Медиану дискретной величины удобнее всего определить по таблице чисел N_k . В ней числа, как было указано, монотонно растут (только иногда, в отдельных точках, могут повторяться). Если среди чисел N_k найдется такое N_m , которое равно точно половине общего числа наблюдений, то за медиану \bar{x}_m можно принять любое число между x_{m-1} и x_m , ибо, по определению,

$$N_m = \sum_{k=1}^{m-1} n_k = nP(x \leq x_{m-1}), \quad P(x \leq x_{m-1}) = \frac{1}{2}. \quad (18.2)$$

Если нет числа N_k , которое точно равнялось бы $\frac{n}{2}$, то находим число $N_r < \frac{n}{2}$ так, чтобы было $N_{r+1} > \frac{n}{2}$; соответствующие значения величины X будут x_r и x_{r+1} . В этом случае можно бы сказать, что медиана \bar{x}_m заключена между числами x_r и x_{r+1} . Можно принять за \bar{x}_m либо x_r , если $\frac{n}{2} - N_r < N_{r+1} - \frac{n}{2}$, либо x_{r+1} , если знак неравенства обратный. Так как рассматривается дискретная величина, не имеющая значений между x_r и x_{r+1} , то ответ может быть дан только приближенный в указанной форме (либо x_r , либо x_{r+1}).

Можно также определить медиану \bar{x}_m линейной интерполяцией по значениям N_r и N_{r+1} , x_r и x_{r+1} и считать ее формальной характеристикой совокупности.

3. Среднее квадратичное отклонение

Согласно определению среднее квадратичное отклонение есть корень квадратный из дисперсии. Заменяя в определении дисперсии для дискретной случайной величины вероятности эмпирическими вероятностями, получим формулу для определения эмпирического среднего квадратичного отклонения

$$\sigma_x^2 = \sum_{k=1}^s p_k (x_k - \bar{x})^2. \quad (18.3)$$

Во многих задачах числа x_k достаточно просты — в том смысле, что они имеют мало знаков. Если, например, рассматривается совокупность кратных звезд, то числа x_k имеют значения 1, 2, 3, 4 ...

(конец таблицы пока нельзя считать достаточно определенным). В общем же случае число \bar{x} не является в указанном смысле простым, т. е. имеет несколько значащих цифр, и вычисление по написанной формуле может осложниться. Поэтому лучше воспользоваться формулой для вычисления дисперсии (см. § 44, часть III)

$$\sigma_w^2 = \sum_{k=1}^s p_k x_k^2 - \bar{x}^2. \quad (18.4)$$

Число σ_w именованное, той же размерности, что и числа x_k . Поэтому значение σ_w зависит от выбора единиц измерения. При сравнении однотипных совокупностей, т. е. совокупностей одних и тех же величин, подобранных по разным признакам, нужно, чтобы величины x элементов разных совокупностей измерялись одинаковыми единицами.

Если требуется сравнить разные совокупности, то удобно ввести безразмерные числовые характеристики. За одну из таких безразмерных характеристик рассеяния можно принять величину

$$V_w = \frac{\sigma_w}{\bar{x}}, \quad (18.5)$$

которую называют *коэффициентом изменчивости*. Она неудобна, если \bar{x} близко к нулю.

Указанных двух числовых характеристик (\bar{x} и σ_w) достаточно чтобы по неравенству Чебышева

$$P(|X - \bar{x}| < t\sigma_w) > 1 - \frac{1}{t^2}$$

оценить вероятность для X принять значение в заданных границах. Границы, однако не вполне произвольны: они должны быть симметричны относительно среднего значения \bar{x} .

Оценка, которую дает неравенство Чебышева, довольно груба; использование полной таблицы, если оно возможно, дает точную вероятность.

4. Моменты распределения

В части III были определены моменты распределения для непрерывных случайных величин, как математические ожидания степеней случайной величины (начальные моменты) и математические ожидания степеней отклонения случайной величины от центра распределения (центральные моменты).

Воспользуемся этим же определением для дискретной случайной величины, заменяя в математических ожиданиях теоретические

вероятности эмпирическими. Тогда для начальных моментов порядка r получим формулу

$$\nu_r = \sum_{k=1}^s p_k x_k^r, \quad (18.6)$$

где r — целые положительные числа.

Так как начальные моменты зависят от начала отсчета, то более показательны для распределения центральные моменты, определяемые формулами:

$$\mu_r = \sum_{k=1}^s p_k (x_k - \bar{x})^r; \quad (18.7)$$

они не зависят от начала отсчета, так как при смещении начала отсчета на столько же изменяется \bar{x} .

При вычислениях удобно по большей части найти сначала начальные моменты, а затем центральные по формулам, которые легко получаются разложением по биному Ньютона и приведением двух последних членов:

$$\begin{aligned} \mu_r = \nu_r - r\bar{x}\nu_{r-1} + \frac{r(r-1)}{2!} \bar{x}^2 \nu_{r-2} + \dots \\ \dots + (-1)^r \frac{r(r-1)}{2} \bar{x}^{r-2} \nu_2 + (-1)^{r+1} (r-1) \bar{x}^r. \end{aligned}$$

Например,

$$\left. \begin{aligned} \mu_2 &= \nu_2 - \bar{x}^2, & \mu_3 &= \nu_3 - 3\bar{x}\nu_2 + 2\bar{x}^3, \\ \mu_4 &= \nu_4 - 4\bar{x}\nu_3 + 6\bar{x}^2\nu_2 - 3\bar{x}^4. \end{aligned} \right\} \quad (18.8)$$

За безразмерные числовые характеристики можно принять числа

$$M_r = \frac{\mu_r}{\sigma^r}$$

или

$$m_r = \frac{\sqrt[r]{\mu_r}}{\sigma} \quad (r = 1, 2, \dots). \quad (18.9)$$

При помощи этих безразмерных характеристик можно сравнивать не только разные распределения одной величины, но и распределения различных величин.

§ 95. Непрерывное эмпирическое распределение

Почти во всех задачах статистической обработки материала приходится иметь дело с непрерывными величинами.

Материал наблюдений находится в разных каталогах или в специально сделанных наблюдениях. «Сырой» материал есть список элементов совокупности с указанием числовых значений величины,

полученных из наблюдений. Если объектов совокупности мало (например, один-два десятка), то обрабатывают непосредственно материал наблюдений. Если материала не очень мало, то работа непосредственно с материалом была бы громоздкой, и эта громоздкость не оправдывалась бы улучшением результатов обработки, так как материал почти всегда выборочный.

Сырой материал подвергают *первичной обработке*, состоящей из следующих операций.

Находим просмотром каталога наименьшее и наибольшее значения величины. Первое несколько уменьшаем (или оставляем без изменения), второе — либо оставляем без изменения, либо увеличиваем. Все это делается для того, чтобы границы принятой области значений X имели мало значащих цифр. Некоторое расширение табличной области значений X обычно допустимо, так как почти всегда материал выборочный, и действительная область значений X может быть шире полученной наблюдениями. Установленную область делим на s равных частей (интервалов), заботясь при этом (для упрощения вычислений), чтобы границы интервалов были числами с малым количеством значащих цифр. Вопрос о числе интервалов пока не имеет теоретического обоснования и решается пробами. Принято брать от 10 до 20 интервалов, но это лишь ориентировочные числа.

Выбрав некоторое число интервалов, обозначим границы интервалов $x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(s)}$. Теперь подсчитываем число элементов совокупности, у которых значения X попадают в каждый из интервалов. Может случиться, что отдельные значения X , полученные из наблюдений, равны значениям границ некоторых интервалов. В таких случаях нужно условиться, к какому из интервалов отнести такие элементы совокупности, к правому или левому. Довольно часто приписывают по половине случая к каждому из смежных интервалов. Чтобы не иметь дела с дробями, в этом случае все количество элементов умножают на два. Это не вносит ошибки, так как и общее количество (наблюдений) умножается на два; следовательно, эмпирическая вероятность попасть в каждый из интервалов не изменится. Подсчет нужно делать по каждому из интервалов в отдельности, затем контролировать суммой по всем интервалам, и она должна равняться общему числу наблюдений.

Если наблюдений много (больше сотни), то подсчет по каждому интервалу в отдельности становится утомительным, ибо нужно перелистывать каталог (список) столько раз, сколько взято интервалов. Для упрощения подсчета можно применить следующий простой способ. Прямоугольник произвольного размера, начерченный на миллиметровке или на бумаге в клетку, делится на одинаковые прямоугольники параллельными прямыми. Число этих прямоугольников равно принятому числу интервалов, на границах прямоугольников у разделительных линий пишутся значения $x^{(k)}$, соответствующие

границам интервалов. Читая каталог (один раз), ставим точку в малых прямоугольниках в зависимости от величины x каждого элемента совокупности. Когда весь каталог прочитан, в каждом из малых прямоугольников будет стоять ряд точек, которые остается подсчитать. Если какое-нибудь каталожное значение равно граничному значению, то на разделительной линии лучше поставить крестик, а не точку.

После подсчета числа элементов в каждом интервале получим приведенную на этой странице таблицу. В первом столбце выписаны границы интервалов, во втором столбце значения X , соответствующие серединам интервалов, в четвертом столбце число наблюдений в каждом интервале.

Границы интервалов	Средины интервалов	Условные единицы	Численность
$x^{(0)}$			
$x^{(1)}$	x_1	$-(r-1)$	n_1
$x^{(2)}$	x_2	$-(r-2)$	n_2
...
...	x_r	0	n_r
$x^{(s-1)}$
$x^{(s)}$	x_s	$s-r$	n_s

После подсчета по таблице значений n_1, n_2, \dots, n_s эти численности относят к серединам интервалов, и получается как бы дискретная величина, принимающая значения x_1, x_2, \dots, x_s с эмпирическими вероятностями

$$\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_s}{n}.$$

Вычисление основных характеристик распределения (среднее значение, среднее квадратичное отклонение, моменты) может быть выполнено так же, как для дискретной величины по формулам § 94.

Для облегчения вычислений полезно по таблице численностей прикинуть, в каком интервале должно находиться среднее значение величины. Чаще всего это будет тот интервал, в котором n_k наибольшее. Если считать n_k массами материальных точек, помещенных в середины интервалов, то среднее значение \bar{x} соответствует центру масс. По этому признаку можно уточнить интервал, в котором может быть среднее значение, если распределение заметно асимметрично *).

Пусть x_r — середина интервала, в котором, по предположению находится среднее значение. Припишем этому интервалу условный номер «нуль»; интервалы, следующие за «нулевым», пронумеруем подряд целыми положительными числами от 1 до $(s-r)$, а предшествующие «нулевому» целыми отрицательными числами от -1

*) Отметим, однако, что прикидка места среднего значения не имеет принципиального значения: если интервал указан не совсем удачно, то лишь немного осложнятся вычисления.

до $-(r-1)$. Эти условные номера или *условные единицы* приведены в третьем столбце таблицы на стр. 295.

Введем теперь вместо x новый аргумент совокупности t , значения которого равны указанным условным единицам. Величины x и t связаны очевидной линейной зависимостью $x_{r+t} = x_r + ht$, где h — шаг таблицы по x .

Введение аргумента t вместо x существенно облегчает вычисления, поэтому все параметры сначала вычисляются для аргумента t . Среднее значение и среднее квадратичное отклонение выражают затем также и в единицах той размерности, в каких обычно измеряют изучаемую величину, что делается с помощью формул

$$\bar{x} = x_r + ht, \quad \sigma_x = h\sigma_t, \quad (18.10)$$

Если распределение отличается от нормального, то для описания совокупности и для вероятностных расчетов параметров \bar{x} и σ_x (или \bar{t} и σ_t) недостаточно. В таких случаях принято вычислять еще асимметрию и эксцесс по формулам:

$$A = \frac{\mu_3(t)}{\sigma_t^3}, \quad E = \frac{\mu_4(t)}{\sigma_t^4} - 3, \quad (18.11)$$

где $\mu_3(t)$ и $\mu_4(t)$ — центральные моменты 3-го и 4-го порядков для аргумента t . Эти два параметра необходимы, если предполагается приближать исследуемое эмпирическое распределение функцией Шарлье или, как чаще говорят, «сравнивать эмпирическое распределение с соответствующей кривой Шарлье».

Выясним смысл асимметрии и эксцесса.

Если распределение симметрично относительно центра распределения, то центральные моменты нечетного порядка обращаются в нуль, и асимметрия равна нулю. Чем более несимметрично распределение, тем более A отличается от нуля; следовательно, A растет с увеличением асимметрии распределения.

Если относительный максимум распределения находится справа от центра распределения, то $\mu_3 > 0$, так как в распределении тогда больше случаев с положительными отклонениями от центра, чем с отрицательными. Величина σ положительна по определению. Поэтому мы получим по формуле положительную асимметрию. Если же максимум расположен слева от центра распределения, то асимметрия окажется отрицательной. Таким образом, знак асимметрии указывает направление несимметричности кривой распределения.

Значение эксцесса E можно выяснить, если вспомнить, что для нормального распределения мы имеем $\mu_4 = 3\sigma^4$. Поэтому, согласно (18.11), эксцесс нормального распределения равен нулю.

Положим для простоты, что эмпирическое распределение симметрично. Предположим также, что построено нормальное распре-

деление, соответствующее эмпирическому, т. е. имеющее центр и дисперсию эмпирического.

Если вершина графика эмпирического распределения выше, чем вершина нормального, то $E > 0$; если вершина эмпирического распределения ниже вершины нормального распределения, то $E < 0$. Таким образом, положительный эксцесс указывает на то, что эмпирическая кривая распределения у центра выше нормальной (начиная же с некоторого расстояния от центра ниже нормальной, так как площадь, ограниченная каждой кривой, равна единице); если эксцесс отрицателен, то у центра кривая ниже нормальной (с некоторого расстояния выше).

Если распределение асимметрично, вычисляют еще медиану и моду.

Медиана есть (как и раньше отмечалось) значение величины, меньше которого в эмпирическом распределении половина случаев. Она может быть определена с помощью чисел x_r и x_{r+1} , определяемых так же, как для дискретного распределения. Здесь, однако, не ограничиваются определением ближайшего к медиане дискретного значения, поскольку величина непрерывна. Поэтому поступают так: после разыскания двух значений середин интервалов x_r и x_{r+1} , между которыми находится медиана, определяют медиану линейной интерполяцией по числам x_r и x_{r+1} и по соответствующим числам N_r и N_{r+1} при помощи очевидной формулы:

$$\bar{x}_m = x_r + \frac{x_{r+1} - x_r}{N_{r+1} - N_r} \left(\frac{n}{2} - N_r \right). \quad (18.12)$$

Модой теоретического распределения называется значение величины, при котором плотность вероятности имеет относительный максимум. В эмпирическом распределении можно принять, что мода равна значению середины интервала, в котором численность наибольшая. Если желательно уточнение, то подбирают теоретическую плотность, достаточно удовлетворительно приближающую эмпирическую плотность, и за моду принимают значение величины, соответствующее максимуму принятого закона распределения.

Иногда в результате смешения разнородного материала таблица на стр. 296 имеет больше одного относительного максимума. Если они устойчивы, т. е. не исчезают при изменении интервалов, то можно ставить задачу о разделении материала и определении параметров составляющих распределений. Такая задача решена для двух составляющих нормальных распределений.

В приводимом ниже примере дана схема вычисления моментов до четвертого порядка включительно и столбец значений N_k эмпирической функции распределения (см. табл. А *).

*) Таблица Б будет рассмотрена в следующем параграфе.

Т а б л и ц а А

Пр и м е р. Обработка распределения абсолютных величин звезд спектральных классов В5 — В9. Вычисление моментов

I	II	III	IV	V	VI	VII	VIII
M	$n_k^{(0)}$	t	$n_k t_k$	$n_k t_k^2$	$n_k t_k^3$	$n_k t_k^4$	N_k
— 2,0	1	— 4	— 4	16	— 64	256	0
— 1,5	3	— 3	— 9	27	— 81	243	1
— 1,0	22	— 2	— 44	88	— 176	352	4
— 0,5	60	— 1	— 60	60	— 60	60	26
0,0	78	0	0	0	0	0	86
+ 0,5	20	+ 1	+ 20	20	+ 20	20	164
+ 1,0	3	+ 2	+ 6	12	+ 24	48	184
+ 1,5	2	+ 3	+ 6	18	+ 54	162	187
+ 2,0							189
Суммы	189		— 85	241	— 283	1141	

Пояснения к схеме

В столбце I даны границы интервалов абсолютной величины M . В этом же столбце можно было бы вписать значения M , соответствующие серединам интервалов, которые мы обозначим M_k , но из них на практике потребуется только одно, и его легко определить в уме, поэтому в записи нет надобности.

Столбец II содержит численности во всех интервалах, указанных в столбце I (значок «0» в скобках обозначает, что численности n_k — наблюдаемые).

В столбце III поставлены значения абсолютных величин в условных единицах от условного нуля, соответствующие серединам интервалов; эти числа обозначены t_k . Условный нуль взят в середине того интервала, которому соответствует наибольшая численность (у нас 78).

Числа столбца IV получаются перемножением чисел II и III столбцов; сумма чисел IV столбца дает после деления на n начальный момент первого порядка, т. е. среднее значение в единицах t .

Числа столбца V получаются перемножением чисел IV и III столбцов, сумма чисел, деленная на n , дает начальный момент второго порядка.

Аналогично получаются столбцы VI и VII умножением предыдущего столбца на числа t_k ; сложением и делением на n получим начальные моменты 3-го и 4-го порядков.

В столбце VIII дано эмпирическое распределение (суммарное) рассматриваемой величины. Числа N_k относятся к границам интервалов; статистический смысл этих чисел следующий: ни одного значения меньше — 2,0, одно значение меньше — 1,5, четыре случая, когда величина меньше — 1,0 и т. д.

Таблица Б

Сравнение эмпирического распределения с нормальным

IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII
$t^{(k)}$	$t^{(k)} - \bar{t}$	$(t^{(k)} - \bar{t}) : \sigma_t$	$\Phi(XI)$	$p_k^{(c)}$	$n_k^{(c)}$	$(0-c)$	$(0-c)^2$	$\frac{(0-c)^2}{c}$	$N_k^{(c)}$
-4,5	-4,050	-3,909	-0,4999	0,0016	0,3				0,0
-3,5	-3,050	-2,944	-0,4983	0,0222	4,2	-0,5	0,25	0,06	0,3
-2,5	-2,050	-1,979	-0,4761	0,1322	25,0	-3,0	9,00	0,36	4,5
-1,5	-1,050	-1,014	-0,3439	0,3248	61,4	-1,4	1,96	0,03	29,5
-0,5	-0,050	-0,048	-0,0191	0,3395	64,2	+13,8	190,44	2,97	90,9
+0,5	+0,950	+0,917	+0,3204	0,1496	28,3	-8,3	68,89	2,43	155,1
+1,5	+1,950	+1,882	+0,4700	0,0278	5,3				183,4
+2,5	+2,950	+2,847	+0,4978	0,0021	0,4	-0,7	0,49	0,09	188,7
+3,5	+3,950	+3,813	+0,4999						189,1
Суммы	—	—	—	0,9998	189,1	-0,1		5,94	

Вычисления вне схемы

а) Начальные моменты в условных единицах:

$$\begin{aligned}
 \bar{t} = \nu_1 &= -85 : 189 = -0,450 & \nu_1^2 &= +0,202 \\
 \nu_2 &= +241 : 189 = +1,275 & \nu_1^3 &= -0,0911 \\
 \nu_3 &= -283 : 189 = -1,497 & \nu_1^4 &= +0,0410 \\
 \nu_4 &= +1141 : 189 = +6,037
 \end{aligned}$$

б) Центральные моменты в условных единицах:

$$\begin{aligned}
 \mu_2 &= +1,275 - 0,202 = 1,073, \quad \sigma_t = \sqrt{1,073} = 1,036 \\
 \mu_3 &= -1,497 - 3 \cdot 1,275 (-0,450) + 2 \cdot (-0,0911) = +0,042 \\
 \sigma_t^3 &= 1,112 \\
 \mu_4 &= 6,037 - 4 \cdot (-1,497) \cdot (-0,450) + 6 \cdot (1,275) \cdot 0,202 - 3 \cdot 0,410 = 4,765 \\
 \sigma_t^4 &= 1,151.
 \end{aligned}$$

Параметры (числовые характеристики) эмпирического распределения

Среднее значение. Здесь среднее значение должно быть вычислено в первоначальных единицах, т. е. в тех, в каких были данные каталога. Условный нуль был взят в середине интервала от 0,0 до 0,5, значит, условному нулю соответствует значение $M = 0,25$; шаг в действительных единицах равен 0,5, поэтому

$$\bar{M} = 0,25 + (-0,450) \cdot 0,5 = +0,025.$$

Медиана. В таблице чисел N_k (столбец VIII) имеем: при $M = 0,0$ $N_k = 86$, при $M = 0,5$ $N_{k+1} = 164$. Линейной интерполяцией нужно найти такое M , при котором $N_k = 94,5$ (половина полного числа наблюдений). По формуле (18.12) получим

$$M_m = 0,0 + \frac{94,5 - 86}{164 - 86} \cdot 0,5 = 0,054.$$

Среднее квадратичное отклонение

$$\sigma_M = 1,036 \cdot 0,5 = 0,513.$$

Асимметрия и эксцесс [по формулам (18.11)]

$$A = \frac{+0,042}{1,112} = +0,038, \quad E = \frac{4,765}{1,151} - 3 = +1,14.$$

Мода. Как было сказано, за величину моды эмпирического распределения можно принять значение в середине того интервала, в котором численность наибольшая. В нашем примере получается по этому способу значение моды 0,25 (в начальных единицах, т. е. звездных величинах).

Мода определяется более уверенно, если найден в аналитическом виде закон распределения, удовлетворительно представляющий эмпирическое распределение. Если считать, что наш материал достаточно удовлетворительно представляется нормальным законом распределения с параметром $\bar{M} = +0,025$, то надо принять, что мода равна $+0,025$. Вычисленные параметры эмпирического распределения дают описание совокупности с помощью нескольких чисел. По ним можно получить некоторые общие сведения об исследуемой случайной величине, не делая сравнения с теоретическим распределением.

Величина среднего квадратичного отклонения показывает, что рассеяние значений M не велико. Действительно, в интервале от $\bar{M} - \sigma_M$ до $\bar{M} + \sigma_M$ содержится больше 70% всего материала наблюдений. Небольшая разница между средним значением и медианой показывает, что распределение приблизительно симметрично относительно центра распределения. Это же подтверждается малым значением асимметрии. Знак асимметрии показывает, что мода должна быть несколько больше среднего значения. Величина эксцесса не может считаться малой. Это означает, что около центра есть некоторое преобладание наблюдений по сравнению с тем, что было бы в случае нормального распределения. Это согласуется с замечанием о доле наблюдений в интервале длиной $2\sigma_M$ около центра распределения.

§ 96. Сравнение эмпирического распределения с теоретическим

Теоретическая функция распределения (или плотность вероятности) должна содержать в общем случае буквенные параметры. Сравнение полученного из наблюдений распределения с некоторым теоретическим распределением заключается в том, что сначала

вычисляют параметры теоретического распределения по данным наблюдений, а после этого строят теоретическое распределение с найденными числовыми значениями параметров и сравнивают его с эмпирическим распределением.

Параметры теоретического распределения нужно вычислить так, чтобы оно в каком-то смысле возможно лучше представило наблюдения. Для этого Пирсон предложил метод моментов.

По этому методу нужно вычислить центральные моменты 2-го и более высоких порядков и нецентральные моменты 1-го порядка теоретического распределения и приравнять их одноименным моментам эмпирического распределения. Моменты теоретического распределения суть функции буквенных параметров распределения; моменты наблюдаемого распределения — числа, полученные обработкой статистического материала. Приравняв эти числа упомянутым функциям параметров (т. е. моментам теоретического распределения), получим систему уравнений для определения параметров.

Если, например, выражение для плотности вероятности включает три параметра, то приравняются теоретические моменты до 3-го порядка включительно и получаются три уравнения с тремя неизвестными. Отметим, что должны быть равны и моменты нулевого порядка, но это условие будет автоматически выполнено, если плотность вероятности нормирована (т. е. интеграл от нее равен единице). Равенство нецентральных моментов первого порядка означает, что должны быть равны между собой среднее теоретическое и среднее эмпирическое. Равенство центральных моментов второго порядка означает равенство дисперсий или средних квадратичных отклонений.

Особенно часто эмпирическое распределение сравнивают с нормальным. В плотность вероятности нормального закона входят два параметра. Поэтому для сравнения наблюдаемого распределения с нормальным достаточно приравнять моменты до 2-го порядка включительно. Сравнение моментов 1-го и 2-го порядков дает

$$\bar{t} = \bar{v}_1, \quad \mu_2 = \sigma^2 = \bar{\mu}_2,$$

где через \bar{t} и σ обозначены параметры нормального закона, а через \bar{v}_1 и $\bar{\mu}_2$ — вычисленные моменты эмпирического распределения (в условных единицах).

После получения параметров теоретического (например, нормального) закона нужно построить теоретическое распределение, а затем сравнить его с эмпирическим. При таком сравнении можно воспользоваться различными критериями близости теоретического распределения к эмпирическому (*критериями согласия*). Мы укажем здесь два критерия.

1. Критерий А. Н. Колмогорова. При помощи параметров теоретического распределения вычисляются вероятности

того, что случайная величина принимает значения, меньшие границ интервалов аргумента распределения. Умножая эти вероятности на общее число наблюдений n , получим числа $N_k^{(c)}$, являющиеся значениями теоретической функции распределения (значок (c) обозначает принадлежность к теоретическому распределению).

Находим максимум модуля разности чисел $N_k^{(c)}$ и значений функции эмпирического распределения N_k . Обозначая эту величину через D ,

$$D = \max |N_k^{(c)} - N_k|, \quad (18.13)$$

получаем аргумент λ критерия А. Н. Колмогорова:

$$\lambda = \frac{D}{\sqrt{n}}. \quad (18.14)$$

По аргументу λ при помощи прилагаемой таблицы находится величина $P(\lambda)$, равная вероятности того, что разность D превысит полученное значение.

Таблица для критерия А. Н. Колмогорова:

λ	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
$P(\lambda)$	1,000	0,997	0,964	0,864	0,711	0,544	0,393	0,270

Чем ближе $P(\lambda)$ к единице, тем лучше принятое теоретическое распределение представляет эмпирическое. Если $P(\lambda)$ мало, то теоретическое распределение неудовлетворительно представляет эмпирический материал.

2. Коэффициент точности. При помощи теоретического закона вычисляются вероятности того, что изучаемая случайная величина принимает значения, заключенные в интервалах эмпирического распределения, для которых даны числа случаев. Умножая полученные вероятности на общее число наблюдений, получаем теоретические численности c . Сравнение этих чисел с численностями 0 , полученными из наблюдений, может показать, насколько наблюдаемое распределение близко к теоретическому.

При помощи теоретического распределения два-три крайних интервала объединяют в один (см. пример). Коэффициент точности H вычисляется по формуле

$$H = \frac{1}{s' - 1} \sum_{k=1}^{s'} \left[\frac{(0 - c)^2}{c} \right]_k, \quad (18.15)$$

где s' обозначает окончательное число интервалов эмпирического распределения в отличие от начального числа интервалов. Можно показать, что математическое ожидание H равно единице. Поэтому построенное теоретическое распределение близко к эмпирическому, если H близко к единице.

Пример. В таблице Б на стр. 299 приведена схема вычислений, необходимых для сравнения построенного в предыдущем параграфе эмпирического распределения с нормальным.

Столбец IX таблицы Б содержит границы интервалов в условных единицах $t^{(k)}$; так как шаг в условных единицах равен единице, то к числам столбца таблицы А, относящимся к серединам интервалов, нужно прибавлять по 0,5 и отнимать по 0,5, чтобы получить границы интервалов.

Чтобы иметь возможность вычислить вероятность случайной величине принять значения в каждом из интервалов, нужно согласно формуле (11,31) части III, иметь значения интеграла вероятностей в начале и конце каждого интервала, при значениях аргумента интеграла, представляющих отклонения от среднего значения, деленные на среднее квадратичное отклонение нормального закона. Поэтому в столбце X подготовлены значения отклонений границ интервалов от среднего значения, т. е. числа $t^{(k)} - \bar{t}$ ($\bar{t} = -0,450$), а в столбце XI даны частные от деления этих чисел на σ_t ($\sigma_t = 1,036$); эти частные и будут значениями аргументов для интеграла вероятностей.

Столбец XII. По каждому из чисел

$$\frac{t^{(k)} - \bar{t}}{\sigma_t}$$

столбца XI определяем по таблице интеграла вероятностей (таблица III в конце книги) числа

$$\Phi\left(\frac{t^{(k)} - \bar{t}}{\sigma_t}\right);$$

если аргумент функции Φ отрицателен, то Φ берется для модуля аргумента и к полученному числу приписывается знак «—». Столбец XIII содержит вероятности $p_k^{(e)}$ попасть в каждый из интервалов. Если границы какого-нибудь из интервалов α и β в условных единицах, то вероятность попасть в этот интервал вычисляется по формуле

$$p(\alpha < t < \beta) = p_k^{(e)} = \Phi\left(\frac{\beta - \bar{t}}{\sigma_t}\right) - \Phi\left(\frac{\alpha - \bar{t}}{\sigma_t}\right).$$

Поэтому, чтобы получить числа $p_k^{(e)}$, нужно из каждого числа столбца XII вычесть соседнее предыдущее число того же столбца (вычислить разности первого порядка чисел столбца XII). При этом нужно обратить внимание на то место этого столбца, где происходит смена знака (минус на плюс), так как в этом интервале нужно складывать модули значений функции Φ . Полученные таким образом вероятности попасть в интервалы имеют верхние значки (с), чтобы отметить, что это вероятности вычисленные (в предположении, что распределение удовлетворительно приближается нормальным законом распределения). Контроль — сумма чисел столбца приблизительно равна единице.

Столбец XIV. Чтобы не иметь дела с дробями, числа столбца XIII умножаем на общее число наблюдений, т. е. считаем, что вероятности определяют доли и вычисляем теоретические численности в интервалах, обозначенные через $n_k(c)$. Контроль — сумма чисел $n_k(c)$ — должна мало

отличаться от общего числа наблюдений n . Возможно отличие в несколько единиц, так как в случае нормального теоретического распределения случайная величина формально может принимать значения от $-\infty$ до $+\infty$, а мы ограничиваемся конечной областью; в нашем примере сумма отличается от n только на единицу запасного знака (числа $n_k(c)$ считались с десятичными долями, которые и дают запасный знак).

Столбец XV. Из каждого члена столбца II вычитаем соответствующее число столбца XIV. Полученные разности обозначены $0-s$ (наблюденные минус вычисленные). Сумма полученных чисел должна равняться разности между суммами чисел столбцов II и XIV. При образовании чисел столбца XV были соединены первые два и последние два интервала таблицы, так как теоретические численности оказались малыми, что затрудняет в дальнейшем определение коэффициента точности*). Рассматриваемый столбец может быть использован для качественного сравнения наблюдаемого распределения с соответствующим нормальным распределением.

В столбцах XVI и XVII выполняется подготовка к вычислению коэффициента точности — одного из критериев пригодности теоретического распределения, в нашей задаче — нормального. Для его вычисления нужна сумма чисел столбца XVII.

В столбце XVIII приведены значения $N_k^{(e)}$ теоретической функции распределения. Эти значения получаются сложением соседних чисел столбца XIV (таким образом, столбец XIV состоит из первых разностей чисел столбца XVIII).

Сравнение эмпирического распределения с нормальным

Использование асимметрии и эксцесса. Это — простейший способ, не требующий дополнительных вычислений, кроме тех, которые выполняются для получения параметров эмпирического распределения. У нормального распределения и асимметрия и эксцесс равны нулю. В нашем примере асимметрия имеет достаточно малое числовое значение. Эксцесс нельзя считать достаточно малым. Таким образом, распределение приблизительно симметрично относительно своего центра, как нормальное, но вблизи центра плотность вероятности больше, чем у нормального распределения.

Использование таблицы $0-s$ (столбец XV). Здесь даны разности численностей наблюдаемых и вычисленных в предположении справедливости нормального закона. Рассмотрение этих чисел подтверждает заключение, полученное по эксцессу: крупное положительное отклонение около центра и малые отрицательные отклонения во всех остальных интервалах, кроме смежного с центром, где получилось $-8,3$. Можно, следовательно, отметить систематический до некоторой степени ход чисел $0-s$.

Вычисление коэффициента точности. По формуле (18,15), где s' — число интервалов, в которых даны числа $0-s$, с помощью суммы чисел столбца XVII получаем

$$H = \frac{5,94}{6-1} = 1,2.$$

Так как H немного отличается от единицы, то нормальное распределение может считаться допустимым приближением для рассматриваемого эмпирического распределения.

*) Первые два интервала сверху дадут $n(0) = 4$, $n(c) = 4,5$, поэтому $(0-s)$ будет $-0,5$; аналогично в последних двух интервалах

Критерий А. Н. Колмогорова. Сравнивая числа столбца XVIII с числами столбца VIII таблицы на стр. 298, находим максимум D модуля разности $N_k - N_k^{(c)}$: $D = 8,9$.

Вычисляем аргумент

$$\lambda = \frac{D}{\sqrt{n}} = \frac{8,9}{\sqrt{189}} = 0,65.$$

По таблице на стр. 303 находим вероятность того, что отклонение D превысит полученное число 8,9: $P(\lambda) = 0,78$. Так как вероятность получить еще большее отклонение, чем в рассматриваемом случае, заметно больше половины, то приближение нормальным законом можно считать удовлетворительным.

§ 97. Доверительные вероятности и доверительные границы

Рассмотренные числовые характеристики эмпирического распределения (среднее значение, среднее квадратичное отклонение, коэффициент изменчивости, мода, медиана, асимметрия, эксцесс) достаточны для общего описания различных распределений, а некоторые из них — и для сравнения распределений разных величин. Таким образом, их назначение — заменить сырой материал (т. е. список результатов наблюдений) несколькими числами, которые бы по возможности полно характеризовали всю совокупность наблюдений. Однако сами по себе эти характеристики не дают возможности некоторого «предвидения» возможных результатов будущих наблюдений случайной величины. Как указывалось раньше, это «предвидение» может заключаться в вычислении вероятности того, что значение величины будет заключаться в заданных границах. На практике считают, что такое указание границ можно считать практически достоверным, если вероятность близка к единице (например, 0,99 или 0,999). В связи с этим введем понятие о доверительных вероятностях и доверительных границах.

Определение. *Доверительной вероятностью* того, что случайная величина примет какое-нибудь значение в некоторых границах, называется такое значение этой вероятности, которое по соглашению считается достаточно близким к единице. Соответствующие границы называются *доверительными*.

Близость доверительной вероятности к единице означает, что событие (попадание в границы) практически достоверно, т. е. только иногда не будет происходить. Слово «иногда» имеет следующий смысл. Если, например, вероятность равна 0,99, то по закону больших чисел (теорема Я. Бернулли) при большом числе наблюдений доля неоявлений события близка к 0,01 или $\sim 1\%$.

Чтобы иметь возможность вычислять доверительные вероятности и доверительные границы, надо построить теоретический закон распределения, решив задачу о приближении эмпирического распределения выбранным по каким-нибудь основаниям теоретическим законом. Параметры этого закона определяются указанным в § 96

способом. Построенный закон распределения (плотность вероятности) должен быть проверен по наблюдениям. При помощи принятой теоретической функции распределения можно составить таблицы для вычисления доверительной вероятности по заданным доверительным границам и для решения обратной задачи.

Для нормального распределения при помощи табл. II (в конце книги) легко получить таблицу доверительных границ для некоторых доверительных вероятностей (\bar{x} — среднее значение, σ — среднее квадратичное отклонение) (табл. А). Приводим также таблицу доверительных вероятностей для разных доверительных границ (табл. Б).

Таблица А

Доверительные вероятности	Нижняя граница	Верхняя граница
0,9	$\bar{x} - 1,645\sigma$	$\bar{x} + 1,645\sigma$
0,99	$\bar{x} - 2,577\sigma$	$\bar{x} + 2,577\sigma$
0,999	$\bar{x} - 3,291\sigma$	$\bar{x} + 3,291\sigma$

Таблица Б

Доверительные границы	Доверительные вероятности
$\bar{x} \pm 1\sigma$	0,6827
$\bar{x} \pm 2\sigma$	0,9545
$\bar{x} \pm 3\sigma$	0,9973
$\bar{x} \pm 4\sigma$	0,9999

Подобные таблицы составлены также для распределений Шарлье и Пирсона.

§ 98. Графическое представление эмпирической совокупности

Наряду с численной обработкой эмпирической совокупности применяется графическое представление. Полученные графики затем сравниваются с графиками принятого теоретического распределения.

Рассмотрим сначала графики дискретного распределения.

График таблицы эмпирического распределения имеет простой вид (рис. 12). По оси абсцисс откладываются дискретные значения x_k , по оси ординат — соответствующие им значения эмпирических вероятностей p_k . График состоит из изолированных точек. Их можно соединить пунктирными прямыми, чтобы несколько более наглядно изобразить изменение p_k при изменении x_k . В промежутках между точками, изображающими x_k , а также вне области (x_1, x_s) значения p равны нулю.

График функции распределения имеет следующий вид: от $x = -\infty$ до $x = x_1$ график совпадает с осью абсцисс; при $x = x_1$ происходит разрыв непрерывности — слева $P = 0$, справа $P = P_1 = p_1$ до значения $x = x_2$, при котором опять разрыв непрерывности — слева p_1 , справа $p_1 + p_2$ до $x = x_3$ и т. д. Таким образом, график состоит из полуоси абсцисс, из отрезка, параллельного оси абсцисс на расстоянии p_1 от оси абсцисс между значениями x_1 и x_2 аргу-

мента, далее отрезка, параллельного оси абсцисс на расстоянии $p_1 + p_2$ между $x = x_2$ и $x = x_3$ и т. д. В конце будет: между $x = x_{s-1}$ и $x = x_s$ отрезок, параллельный оси абсцисс на расстоянии $1 - p_s$ от нее, затем луч, параллельный оси абсцисс на расстоянии единицы от $x = x_s$ до $+\infty$. Такой график нередко называют *ступенчатым*.

В случае непрерывного распределения на оси абсцисс откладывают значения случайной величины. Для построения графика функции распределения в тех точках оси абсцисс, которые изображают верхние границы интервалов, строятся ординаты длиной N_k или N_k/n . На левой границе первого интервала ставится точка на оси абсцисс. Если на ординатах откладывается N_k , то длина последней ординаты равна всему числу случаев, если же откладываются эмпирические вероятности, то ее длина равна единице.

В отличие от графика для дискретной величины, здесь точки должны быть соединены, поскольку возможны любые промежуточные значения нашей величины; обычно нанесенные точки соединяют отрезками прямых. График представляет ломаную, начинающуюся на оси абсцисс и заканчивающуюся в точке, ордината которой равна n или 1; всякая

последующая ордината ломаной не может быть ниже предыдущей. График этого вида носит название *огива* (рис. 13). Огивой пользуются для определения медианы. Через середину крайней правой ординаты проводят прямую, параллельную оси абсцисс, до пересечения с огивой, и из точки пересечения опускают перпендикуляр на ось абсцисс. Основание перпендикуляра приближенно определяет медиану.

Для построения графика эмпирической плотности вероятности на каждом интервале как на основании, строят прямоугольник, площадь которого равна либо числу случаев, либо относительной частоте. В первом случае площадь всего графика равна числу случаев, во втором равна единице. Из способа построения следует, что высота каждого прямоугольника представляет среднее число

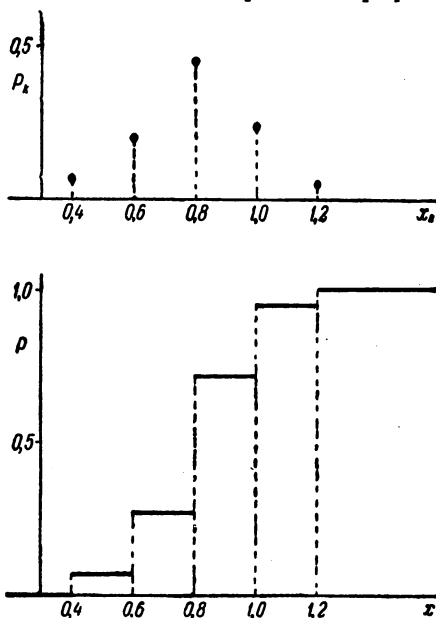


Рис. 12. Графики распределения вероятностей и функции распределения дискретной случайной величины.

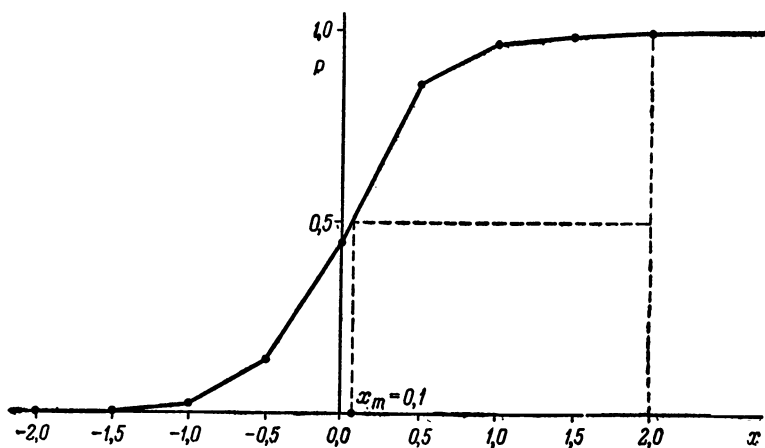


Рис. 13. График эмпирической функции распределения. Пунктиром показано графическое определение медианы (см. пример на стр. 298).

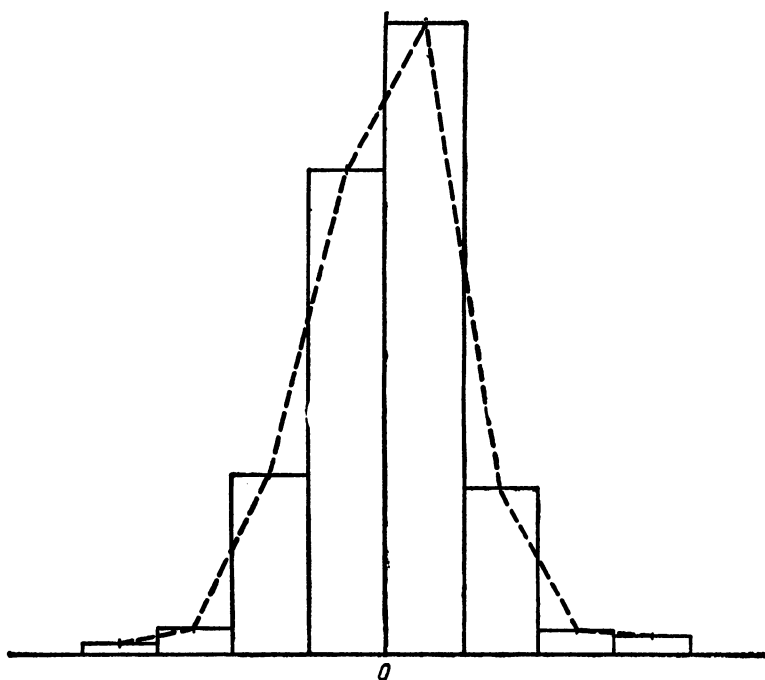


Рис. 14. Гистограмма (сплошные линии) и полигон (пунктир) распределения непрерывной случайной величины. Площади прямоугольников пропорциональны вероятностям значений величины в основаниях прямоугольников.

случаев, приходящихся на единицу соответствующего интервала, или среднюю статистическую вероятность на единицу интервала. Можно сказать, что высота прямоугольника есть средняя плотность эмпирической вероятности того, что значение величины находится в интервале.

Графики подобного вида называют *гистограммами*. Иногда вместо гистограммы строят *полигон*, получающийся из гистограммы, если соединить отрезками прямых середины верхних сторон прямоугольников (рис. 14). Графическое сравнение эмпирического распределения с теоретическим производится при помощи гистограмм этих распределений и полигонов.

Пример. Произведем сравнение эмпирического распределения с теоретическим для примера, рассмотренного в §§ 95—96 (см. табл. А и Б, стр. 298 и 299).

Построим на одном чертеже (рис. 15) гистограмму эмпирического и нормального распределений. Эмпирическая начерчена сплошными линиями по таблице значений n_k (столбец II). Гистограмма нормального распределения, приближающего данное распределение, начерчена пунктиром по числам $n_b(c)$ (столбец XIV). На таком графике можно не следить за тем, чтобы общая площадь равнялась единице или общему числу наблюдений, так как это требование всегда можно выполнить надлежащим выбором масштабов на осях координат. Поэтому высоты прямоугольников можно брать пропорциональными числу случаев (для наблюдаемого — числа столбца II, для теоретического — столбца XIV).

Сравнение гистограмм показано на рис. 15.

На рис. 16 изображены кривая Гаусса и полигон частот, который можно считать графиком плотности вероятности эмпирического распределения.

Полигон частот представляет собой ломаную, построенную по узлам (точкам M_k, n_k). Дополнительные вычисления, необходимые для построения кривой Гаусса в форме, удобной для сравнения, приведены в следующей таблице:

1	2	3	4	5	6	7
M	$M - \bar{M}$	$\frac{M - \bar{M}}{\sigma}$	$\Phi' \left(\frac{M - \bar{M}}{\sigma M} \right)$	$\Phi' : \sigma M$	y_c	y_0
— 1,75	— 1,775	— 3,427	0,0011	0,0021	0,4	2
— 1,25	— 1,275	— 2,461	0,0193	0,0373	7,0	6
— 0,75	— 0,775	— 1,496	0,1303	0,2515	47,5	44
— 0,25	— 0,275	— 0,531	0,3465	0,6689	126,4	120
+ 0,025	0	0	0,3989	0,7701	145,6	—
+ 0,25	+ 0,225	+ 0,434	0,3631	0,7010	132,5	156
+ 0,75	+ 0,725	+ 1,400	0,1497	0,2890	54,6	40
+ 1,25	+ 1,225	+ 2,365	0,0243	0,0469	8,9	6
+ 1,75	+ 1,725	+ 3,330	0,0016	0,0031	0,6	4

В 5-м столбце даны значения плотности вероятности в точках, указанных в 1-м столбце; в 6-м столбце даны произведения y_c этих плотностей на общее число наблюдений. В следующем столбце

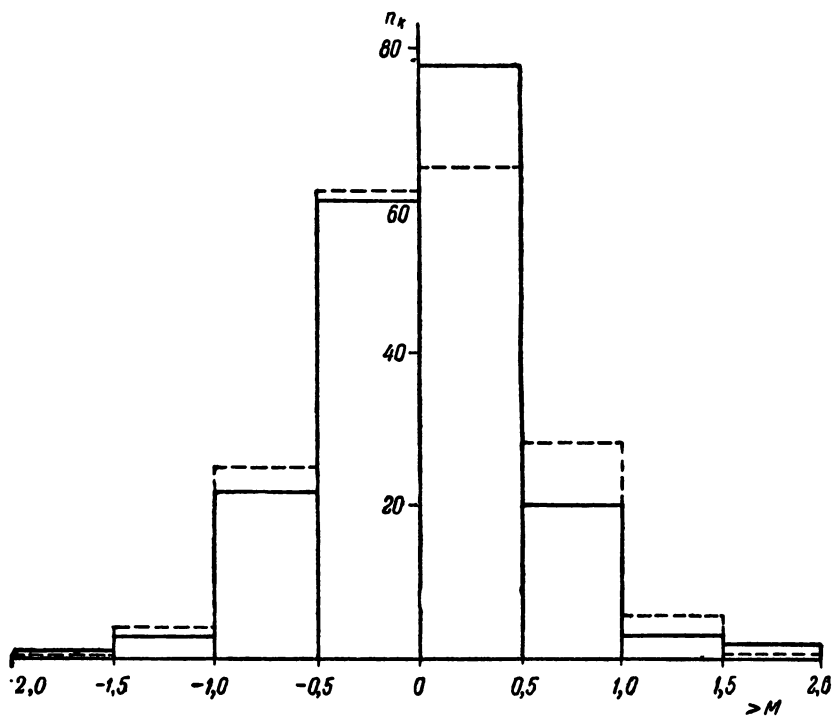


Рис. 15. Сравнение гистограмм эмпирического (сплошная линия) и теоретического (пунктир) распределений.

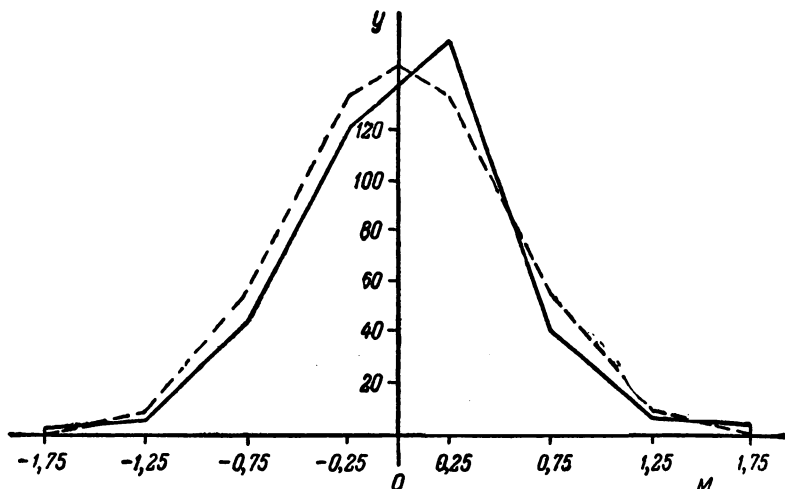


Рис. 16. Сравнение полигон эмпирического (сплошная линия) и нормального (пунктир) распределений.

даны сравниваемые с ними числа u_0 , пропорциональные эмпирической плотности вероятности, полученные делением численностей на величину шага (0,5)*).

Оба рисунка подтверждают заключения об удовлетворительном (в общем) согласии с нормальным распределением.

§ 99. Средние ошибки параметров выборочной совокупности

Как уже указывалось, в большинстве задач эмпирические совокупности являются выборочными. Поэтому выводы, которые делаются по эмпирическим распределениям, нельзя без дополнительного исследования считать достаточно надежными.

В силу закона больших чисел при достаточно большом объеме выборки с вероятностью, близкой к единице, можно ожидать, что параметры выборочной совокупности будут сколь угодно близки к параметрам генеральной совокупности. Поэтому числовые характеристики достаточно большой выборки можно считать приближенными значениями тех вероятностных параметров распределения, которые следовало бы вычислить для генеральной совокупности. По этой же причине и те вероятности, которые вычисляются по эмпирическому распределению, следует считать приближенными значениями вероятностей, которые нужно было бы, если возможно, вычислить по сведениям о генеральной совокупности.

Можно составить некоторое представление о неизвестной генеральной совокупности, если найти, хотя бы приближенно, средние квадратичные ошибки параметров выборочной совокупности, т. е. рассеяние этих параметров относительно параметров генеральной совокупности.

Чтобы оценить указанные средние ошибки, будем рассуждать так. Представим себе, что мы делаем всевозможные выборки из генеральной совокупности. В каждой выборке мы будем получать некоторое значение \bar{x} (для краткости будем говорить только об \bar{x} , но то же относится и к другим параметрам распределений). Совокупность выборочных значений \bar{x} составит некоторое распределение. Если вывести функцию распределения \bar{x} , то можно оценить вероятности различных выборочных \bar{x} . Имея же одно выборочное значение, можно по параметрам распределения оценить, насколько выборочное среднее значение может отличаться от генерального.

Исследование выборочных значений \bar{x} и $\sigma_{\bar{x}}$ из генеральной совокупности, подчиненной нормальному закону, привело к следующим выводам.

*) Строка $M = 0,025$ включена, чтобы отметить центр теоретического распределения; в эмпирическом материале нет соответствующего интервала, поэтому в 7-м столбце поставлена черта.

Закон распределения выборочных средних значений \bar{x} оказывается нормальным, причем выражение для средней квадратичной ошибки величины \bar{x} имеет вид

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}. \quad (18.16)$$

Однако закон распределения выборочных значений дисперсий σ_x оказывается отличным от нормального. Для средней квадратичной ошибки $\sigma(\sigma_x)$ выборочного среднего квадратичного отклонения получается следующее значение:

$$\sigma(\sigma_x) = \frac{\sigma_x}{\sqrt{2n}}. \quad (18.17)$$

Пример. В рассмотренном выше примере

$$\bar{M} = +0,025, \quad \sigma_M = 0,513, \quad n = 189.$$

Поэтому

$$\begin{aligned} \sigma_{\bar{M}} &= \frac{0,513}{\sqrt{189}} \approx 0,037, \\ \sigma(\sigma_M) &= \frac{0,513}{\sqrt{378}} \approx 0,026. \end{aligned}$$

Подводя итог сказанному в этой главе, можно наметить следующий план обработки одномерной статистической совокупности:

1. Первичная обработка.
 2. Нахождение числовых характеристик эмпирического распределения (среднее значение, медиана, среднее квадратичное отклонение, асимметрия, эксцесс, мода, медиана).
 3. Определение параметров принятого теоретического распределения методом моментов.
 4. Построение по найденным параметрам теоретического распределения.
 5. Численное сравнение теоретического распределения с эмпирическим при помощи критериев согласия.
 6. Построение графиков эмпирического и теоретического распределений и их сравнение.
 7. Вычисление средних ошибок параметров эмпирической совокупности (если она выборочная).
-

ГЛАВА 19

ЭЛЕМЕНТАРНАЯ ТЕОРИЯ КОРРЕЛЯЦИИ ДВУХ ВЕЛИЧИН

§ 100. Эмпирическое распределение двух случайных величин

Пусть дана статистическая совокупность, состоящая из объектов (наблюдений), для каждого из которых указаны полученные из наблюдений числовые значения двух случайных величин X и Y (мы ограничимся случаем непрерывных величин). Результаты наблюдений могут быть первоначально записаны в простой таблице, где указаны соответствующие друг другу значения x_k и y_k . Таблицу такого вида будем в дальнейшем обозначать буквой A :

$$\begin{array}{c} X \\ Y \end{array} \left\| \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{array} \right. \quad (A)$$

Таблица A по форме ничем не отличается от той таблицы — списка результатов наблюдений, по которой в предыдущей главе требовалось составить эмпирическую формулу, связывающую X и Y . При этом, однако, в предыдущем параграфе предполагалось, что величины X и Y связаны функциональной зависимостью, неизвестной нам и искаженной ошибками наблюдений. Сейчас же относительно величин X и Y мы не будем делать такого предположения.

Особенность той статистической связи, которая будет рассматриваться в этой главе, заключается в том, что каждому значению одной величины соответствует неопределенное количество значений другой, среди которых могут быть как неравные, так и равные. Эту особенность можно сделать наглядной, если составить еще две таблицы, которые мы условимся обозначать B_1 и B_2 . Для этого выберем разные значения величины X , примем их за значения аргумента и к каждому из таких значений припишем все значения величины Y , какие с ним находятся в таблице A . Получится таблица B_1 вида

$$\begin{array}{ccccccc} x^{(1)} & y_{11} & y_{12} & \dots & y_{1r_1} \\ x^{(2)} & y_{21} & y_{22} & \dots & y_{2r_2} \\ \dots & \dots & \dots & \dots & \dots \\ x^{(k)} & y_{k1} & y_{k2} & \dots & y_{kr_k} \end{array} \quad (B_1)$$

В ней $x^{(1)}, x^{(2)}, \dots, x^{(l)}$ — разные значения X ; числа с двойными знаками представляют соответствующие значения Y .

Аналогичный вид имеет таблица B_2 , отличающаяся от B_1 тем, что значения X и Y поменялись местами

$$\begin{array}{ccccccc} y^{(1)} & x_{11} & x_{12} & \dots & x_{1s} & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ y^{(m)} & x_{m1} & x_{m2} & \dots & x_{ms} & & \end{array} \quad (B_2)$$

Таблица А непосредственно используется для обработки в том случае, когда материала очень мало. Таблицы Б строятся очень редко; нам они потребуются для выяснения некоторых основных понятий.

Если в таблице А количество наблюдений не мало (сто или больше), то составляют таблица В, представляющую двухмерный аналог таблицы распределения вероятностей, которая в предыдущей главе строилась для одной непрерывной величины. Таблица В строится следующим образом. Как и в случае одной величины, разобьем промежуток изменения каждой величины на равные интервалы и подсчитаем число случаев, когда X и Y принимают значения в каждой из пар соответствующих интервалов:

$\begin{array}{c} X \\ Y \end{array}$	x_1	x_2	\dots	x_k	\dots	x_r	Распр. Y
y_1	n_{11}	n_{21}	\dots	n_{k1}	\dots	n_{r1}	n_{01}
y_2	n_{12}	n_{22}	\dots	n_{k2}	\dots	n_{r2}	n_{02}
\vdots							\vdots
y_l	n_{1l}	n_{2l}	\dots	n_{kl}	\dots	n_{rl}	n_{0l}
\vdots							\vdots
y_s	n_{1s}	n_{2s}	\dots	n_{ks}	\dots	n_{rs}	n_{0s}
Распр. X	n_{10}	n_{20}	\dots	n_{k0}	\dots	n_{r0}	$\begin{array}{c} n \\ n \end{array}$

В первой строке таблицы выписаны значения величин X , соответствующие серединам интервалов этой величины (границы интервалов не выписаны); в первом столбце — середины интервалов величины Y . Величины с двумя знаками, не равными нулю, представляют численности частей полной совокупности, имеющих значения

x и y в интервалах, соответствующих той клетке, в которой число написано. Если шаги величин X и Y равны соответственно g и h , то n_{kl} есть число наблюдений, в которых X имеет значения от $x_k - \frac{1}{2}g$ до $x_k + \frac{1}{2}g$, а Y в то же время имеет значения от $y_l - \frac{1}{2}h$ до $y_l + \frac{1}{2}h$. Последний столбец получается сложением всех чисел n_{kl} в каждой строке. Полученный столбец дает распределение одной величины Y (при любых X). Последняя строка получается сложением чисел n_{kl} по столбцам и представляет распределение величины X . Сумма всех чисел n_{kl} ($k \neq 0, l \neq 0$) должна равняться числу наблюдений n .

Такого рода таблицы называются *корреляционными* (иногда употребляется название *сводная таблица*). Составление таких таблиц не рекомендуется, когда число случаев мало, поскольку при обработке считают, что число случаев в каждой клетке относится к серединам интервалов, а это может дать заметные ошибки, когда число наблюдений незначительно.

Эмпирическое распределение двух непрерывных величин можно представить графически. Примем числа x и y в таблице А за прямоугольные координаты на плоскости и построим точки с координатами x_m, y_m ($m=1, 2, \dots, n$). Тогда получим совокупность точек, разбросанных на плоскости; эту совокупность называют *полем корреляции*.

Найдем средние значения \bar{x} и \bar{y} каждой из величин в отдельности. Точка поля с координатами \bar{x} и \bar{y} называется *центром распределения*. Если построено поле корреляции, то легко составить корреляционную таблицу вида В. Для этого достаточно построить координатную сетку через точки, которые определяют границы интервалов. Плоскость разобьется на прямоугольники; подсчитав число точек в каждом прямоугольнике, мы и получим корреляционную таблицу.

На каждом прямоугольнике, принимая его за основание, построим параллелепипеды так, чтобы объем каждого параллелепипеда равнялся отношению числа случаев в прямоугольнике к общему числу случаев. Полученную пространственную ступенчатую фигуру можно назвать *эмпирической поверхностью распределения* или *гистограммой* двумерной задачи. Этот способ представляет естественное обобщение того способа изображения на графике распределения одной величины, который описан выше.

§ 101. Корреляционная зависимость. Задачи теории корреляции

Соотношение между X и Y , определяемое их распределением, обычно заметно отличается от той связи между двумя переменными, которую называют функциональной. В случае взаимно однозначной функциональной зависимости каждому значению x соответствует только одно значение y и наоборот. Из таблиц Б₁ и Б₂ видно, что одному

значению x может соответствовать неопределенное количество значений y , и одному значению y может соответствовать несколько значений x , причем они могут весьма заметно отличаться одно от другого. Чередование числа этих разных значений таково, что говорить о многозначной функциональной зависимости не приходится.

Из рассмотрения корреляционной таблицы В также видно отличие от функциональной зависимости. Когда значения одной из величин заключены в некотором интервале, другая величина может принимать значения, заключенные в довольно широких границах. Эта особенность становится еще более наглядной при рассмотрении поля корреляции.

В случае точной функциональной зависимости между X и Y точки поля должны лежать на некоторой вполне определенной кривой. Если X и Y связаны точной функциональной зависимостью, но числа таблицы А получены из наблюдений, то вследствие случайных ошибок наблюдений точки поля не будут лежать точно на теоретической кривой. Тем не менее, если ошибки малы, точки должны расположиться в некоторой узкой полосе вдоль теоретической кривой. В случае же корреляционной связи точки поля расположены на плоскости довольно беспорядочно.

Для изучения особенностей корреляционной связи дополним таблицы В еще одним столбцом, в котором даны средние значения величины Y при каждом из значений X , выписанных в первом столбце. Первый и последний столбцы расширенной таблицы B_1 определяют зависимость средних значений \bar{y}_x величины Y от соответствующих значений величины X . Аналогично расширим таблицу B_2 и получим эмпирическую зависимость средних значений \bar{x}_y величины X от соответствующих значений величины Y . Говорят, что две таблицы (x, \bar{y}_x) и (y, \bar{x}_y) определяют *эмпирическую регрессию Y по X (или Y на X)* и *эмпирическую регрессию X по Y* .

Рассмотрение подобных таблиц приводит к следующему заключению. Когда величина x растет, то в отдельных наблюдениях соответствующие значения y могут и расти и убывать; одному и тому же значению x могут соответствовать как большие, так и малые значения y . То же самое можно сказать и для y и x . Тем не менее в ряде задач средние значения одной величины, соответствующие значениям другой, обнаруживают известную зависимость от значений второй величины. Можно иногда даже говорить о почти функциональной зависимости средних значений одной величины от соответствующих значений другой.

В тех случаях, когда распределение двух величин обладает указанными особенностями, говорят, что между величинами есть *корреляционная зависимость* или просто *корреляция*. Таким образом, между двумя случайными величинами есть корреляционная зависимость, если каждому значению одной из них соответствует

неопределенное количество значений другой, но средние из этих значений зависят от значений первой величины.

Однозначную функциональную зависимость между величинами X и Y можно считать частным случаем корреляционной зависимости. Если каждому значению x соответствует только одно вполне определенное значение y и наоборот, и все точки поля располагаются при неограниченном увеличении их числа на кривой, то корреляционная связь переходит в функциональную.

Построение полей корреляции для различных пар случайных величин дает различные результаты. Иногда точки поля разбросаны беспорядочно, иногда, наоборот, точки поля почти все расположены вдоль некоторой воображаемой кривой или прямой. Таким образом, корреляционная зависимость может в большей или меньшей степени уклоняться от функциональной зависимости. В связи с этим первой задачей теории корреляции является вывод числового критерия для оценки степени близости корреляционной зависимости к функциональной. Когда из наблюдений выделяются только две величины, то естественно, что между ними не может быть функциональной зависимости, так как эти величины связаны с другими, которые нами не учитываются. Для приложения важно уметь выделить те величины, которые наиболее тесно связаны между собой. Именно для решения таких вопросов и вводится числовой критерий, оценивающий степень связи.

В предыдущем параграфе было указано, что средние значения одной величины, вычисленные для ограниченного интервала другой, обнаруживают зависимость от значений этой другой величины. Поэтому вторая задача, которая рассматривается в теории корреляции, заключается в выводе эмпирических формул для определения средних значений одной величины по значениям другой. Таких формул будет две; они называются *уравнениями регрессии*.

Мы ограничимся только теорией линейной корреляции. Это значит, что мы будем, во-первых, выводить критерий для оценки степени уклонения корреляционной зависимости от линейной функциональной зависимости и, во-вторых, выводить для определения среднего значения каждой величины по значениям другой линейные эмпирические формулы вида

$$\bar{y}_x = a + bx, \quad \bar{x}_y = c + dy. \quad (19.1)$$

§ 102. Вывод линейной эмпирической формулы

Для решения задачи о выводе уравнений регрессии решим сначала задачу о построении линейной эмпирической формулы в таком виде, который будет удобен для получения прямых регрессии.

Пусть из наблюдений получена таблица вида таблицы А значений x_k и соответствующих им значений y_k ($k = 1, 2, \dots, n$).

Требуется построить эмпирическую формулу вида

$$Y = a + bX. \quad (19.2)$$

Согласно § 90 для определения коэффициентов a и b получаем условные уравнения

$$a + bx_k - y_k = 0 \quad (k = 1, 2, \dots, n). \quad (19.3)$$

По ним составляем два нормальных уравнения, коэффициенты которых здесь удобнее записывать не в обозначениях Гаусса, а в виде сумм:

$$na + b \sum_{k=1}^n x_k - \sum_{k=1}^n y_k = 0, \quad (19.4)$$

$$a \sum_{k=1}^n x_k + b \sum_{k=1}^n x_k^2 - \sum_{k=1}^n x_k y_k = 0. \quad (19.5)$$

Предположим теперь, что все x_k и y_k равновероятны. Тогда частное от деления $\sum x_k$ на n дает среднее значение величины x , а частное от деления $\sum y_k$ на n есть среднее значение величины y . Поэтому нормальное уравнение (19.4) после деления на n может быть записано в виде

$$a + b\bar{x} - \bar{y} = 0. \quad (19.6)$$

Для преобразования уравнения (19.5) разделим его тоже на n .

Коэффициент $\left(\sum_{k=1}^n x_k^2\right)$: n можно заменить выражением $\sigma_x^2 + \bar{x}^2$, которое формально получится, если определить дисперсию по формуле: дисперсия равна разности между математическим ожиданием квадрата величины и квадратом ее математического ожидания. Для преобразования последнего члена уравнения (19.5) введем обозначения:

$$\nu_{11} = \frac{1}{n} \sum_{k=1}^n x_k y_k, \quad \mu_{11} = \nu_{11} - \bar{x}\bar{y}.$$

Если обобщить понятие о моментах распределения на совокупность двух случайных величин и считать все пары (x_k, y_k) равновероятными, то можно назвать ν_{11} *начальным моментом порядка один* распределения (X, Y) , а μ_{11} — *центральным моментом* того же порядка рассматриваемого распределения.

После указанных замен второе уравнение примет вид

$$a\bar{x} + b(\sigma_x^2 + \bar{x}^2) - (\bar{x}\bar{y} + \mu_{11}) = 0.$$

Перепишем его так:

$$a\bar{x} + b\bar{x}^2 - \bar{x}\bar{y} + b\sigma_x^2 - \mu_{11} = 0.$$

Первые три члена дадут нуль в силу (19.6) и мы получим систему уравнений:

$$\left. \begin{aligned} a + b\bar{x} &= \bar{y}, \\ b\sigma_x^2 &= \mu_{11}. \end{aligned} \right\} \quad (19.7)$$

Отсюда

$$\left. \begin{aligned} b &= \frac{\mu_{11}}{\sigma_x^2}, \\ a &= \bar{y} - \frac{\mu_{11}\bar{x}}{\sigma_x^2}. \end{aligned} \right\} \quad (19.8)$$

Линейная эмпирическая формула (19.6) по таблице (x_k, y_k) может быть теперь написана в форме

$$Y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2} (X - \bar{x}); \quad (19.9)$$

ее параметры вычисляются по формулам:

$$\left. \begin{aligned} \bar{x} &= \frac{1}{n} \sum_{k=1}^n x_k, & \bar{y} &= \frac{1}{n} \sum_{k=1}^n y_k, \\ \sigma_x^2 &= \frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2, & \mu_{11} &= \frac{1}{n} \sum_{k=1}^n x_k y_k - \bar{x}\bar{y}. \end{aligned} \right\} \quad (19.10)$$

При выводе было предположено, что все пары (x_k, y_k) равновероятны; это эквивалентно предположению, что каждая пара встречается в таблице значений один раз. Если каждая пара чисел (x_k, y_k) встречается в таблице n_k раз, и количество разных пар равно s , то вероятность каждой пары и каждого из чисел, взятого отдельно, надо считать равной $n_k : n$. Написанные формулы поэтому заменятся такими:

$$\left. \begin{aligned} \bar{x} &= \frac{1}{n} \sum_{k=1}^s n_k x_k, & \bar{y} &= \frac{1}{n} \sum_{k=1}^s n_k y_k, \\ \sigma_x^2 &= \frac{1}{n} \sum_{k=1}^s n_k x_k^2 - \bar{x}^2, & \mu_{11} &= \frac{1}{n} \sum_{k=1}^s n_k x_k y_k - \bar{x}\bar{y}. \end{aligned} \right\} \quad (19.11)$$

Эти формулы можно назвать *весовыми*, так как числа n_k играют при вычислениях роль весов.

Мы строили эмпирическую линейную формулу, выражающую Y через X . Ясно, что X и Y могут быть неравноправными с точки зрения физической, но с точки зрения вычислительной они совершенно эквивалентны. Поэтому, кроме эмпирической формулы

построенного вида, можно построить еще одну эмпирическую формулу, выражающую X через Y . Очевидно, что она будет следствием первой формулы только при условии, что X и Y связаны точной линейной зависимостью. Если такой связи нет, то вторая формула не будет следствием первой, а имеет самостоятельное значение. Мы не будем выводить второй формулы (это рекомендуется читателям сделать самостоятельно) и дадим только ее окончательный вид для случая равной вероятности (x_k, y_k) .

$$X - \bar{x} = \frac{\mu_{11}}{\sigma_y^2} (Y - \bar{y}), \quad (19.12)$$

где

$$\sigma_y^2 = \frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2; \quad (19.13)$$

остальные параметры имеют также те же выражения, как и в предыдущем случае. При этом величина μ_{11} не изменяется, так как она симметрична относительно X и Y .

Если пары чисел (x_k, y_k) встречаются n_k раз, то вид формулы не изменится, а параметры вычисляются по весовым формулам, причем формула для σ_y^2 подлежит очевидному изменению.

§ 103. Вывод линейных уравнений регрессии

В предыдущем параграфе были выведены формулы для определения коэффициентов линейных эмпирических формул:

$$Y = a + bX \quad \text{и} \quad X = c + dY.$$

Мы получили уравнения прямых:

$$Y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2} (X - \bar{x}) \quad \text{и} \quad X - \bar{x} = \frac{\mu_{11}}{\sigma_y^2} (Y - \bar{y}); \quad (19.14)$$

из этих уравнений видно, что обе прямые проходят через центр распределения.

Необходимо подчеркнуть, что мы вывели эти формулы, пользуясь таблицей А так, как если бы нужно было получить эмпирические формулы для определения Y по X и X по Y . Между тем, согласно определению, уравнения регрессии должны быть эмпирическими формулами для определения среднего значения \bar{y} по значению x и среднего значения \bar{x} по величине y . Покажем, исходя из определений уравнений регрессии, что эти уравнения имеют тот же вид, что и эмпирические формулы (19.14). Составим таблицы вида Б₁ и Б₂; в каждой строке вычислим средние значения Y и X , которые мы обозначим \bar{y}_x и \bar{x}_y .

Для этих средних и соответствующих им значений X и Y следует составить уравнения регрессии, пользуясь формулами предыдущего параграфа. При этом нужно, однако, учесть, что если некоторому значению x_m соответствует, например, три значения y , то полученная после определения среднего y пара чисел $(x_m, \bar{y}_{\omega, m})$ должна считаться эквивалентной трем парам чисел и средние значения следует определять по весовым формулам. Пусть таблица B_1 (расширенная) имеет вид

$$\begin{array}{c|cccc} x^{(1)} & y_{11} & y_{12} & \dots & y_{1r_1} & \bar{y}_{\omega 1} \\ x^{(2)} & y_{21} & y_{22} & \dots & y_{2r_2} & \bar{y}_{\omega 2} \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ x^{(l)} & y_{l1} & y_{l2} & \dots & y_{lr_l} & \bar{y}_{\omega l} \end{array} \quad (B_1)$$

В этой таблице l — число различных значений x , встречающихся в таблице A , r_1, r_2, \dots, r_l — числа значений y , соответствующих величинам $x^{(1)}, x^{(2)}, \dots, x^{(l)}$. Очевидно,

$$r_1 + r_2 + \dots + r_l = n, \quad l \leq n.$$

Таким образом, нужно построить эмпирическую формулу для таблицы

X	Y	Вес
$x^{(1)}$	$\bar{y}_{\omega 1}$	r_1
$x^{(2)}$	$\bar{y}_{\omega 2}$	r_2
\vdots	\vdots	\vdots
$x^{(l)}$	$\bar{y}_{\omega l}$	r_l

Числа $\bar{y}_{\omega 1}, \bar{y}_{\omega 2}, \dots$ этой таблицы получены по формулам:

$$\bar{y}_{\omega 1} = \frac{y_{11} + y_{12} + \dots + y_{1r_1}}{r_1}, \quad \bar{y}_{\omega 2} = \frac{y_{21} + y_{22} + \dots + y_{2r_2}}{r_2}, \dots$$

Весовое среднее значение \bar{y} будет:

$$\begin{aligned} \bar{y} &= \frac{r_1 \bar{y}_{\omega 1} + r_2 \bar{y}_{\omega 2} + \dots + r_l \bar{y}_{\omega l}}{r_1 + r_2 + \dots + r_l} = \\ &= \frac{y_{11} + y_{12} + \dots + y_{1r_1} + y_{21} + \dots + y_{2r_2} + \dots + y_{lr_l}}{n}. \end{aligned}$$

Так как в числителе последней дроби просто имеем сумму всевозможных значений y , то весовое среднее равно простому среднему, которое получается из таблицы A . Аналогичное обстоятельство имеет место и для x .

Далее, по весовой формуле

$$\sigma_x^2 = \frac{r_1 (x^{(1)})^2 + r_2 (x^{(2)})^2 + \dots + r_l (x^{(l)})^2}{r_1 + r_2 + \dots + r_l} - \bar{x}^2.$$

Но по такой же формуле пришлось бы вычислять простое среднее квадратичное отклонение \bar{X} по таблице А, так как в ней число значений $x^{(1)}$ есть r_1 и т. д. Для определения μ_{11} по таблице Б₁ нужно было бы написать

$$\mu_{11} = \frac{r_1 x^{(1)} \bar{y}_{w1} + r_2 x^{(2)} \bar{y}_{w2} + \dots + r_l x^{(l)} \bar{y}_{wl}}{r_1 + r_2 + \dots + r_l} - \bar{x} \bar{y};$$

если учесть выражение для \bar{y}_{w1} , \bar{y}_{w2} , ..., то

$$\begin{aligned} \mu_{11} &= \\ &= \frac{x^{(1)} y_{11} + x^{(1)} y_{12} + \dots + x^{(1)} y_{1r_1} + x^{(2)} y_{21} + \dots + x^{(2)} y_{2r_2} + \dots + x^{(l)} y_{lr_l}}{n} - \bar{x} \bar{y}. \end{aligned}$$

В числителе опять стоит сумма произведений всех x (в том числе и повторных) на соответствующие значения y , т. е. μ_{11} пришлось бы вычислять так же, как и для таблицы А.

Таким образом, составляя эмпирическую формулу для определения Y по X , мы получим такой же результат, как и при составлении уравнения регрессии \bar{y}_x по x .

Итак, уравнения регрессии имеют вид

$$\bar{y}_x - \bar{y} = \frac{\mu_{11}}{\sigma_x^2} (X - \bar{x}), \quad \bar{x}_y - \bar{x} = \frac{\mu_{11}}{\sigma_y^2} (Y - \bar{y}). \quad (19.15)$$

Часто для упрощения \bar{y}_x и \bar{x}_y заменяют на y и x ; в этом, как было показано только что, не будет ошибки. Числа

$$\frac{\mu_{11}}{\sigma_x^2} = \rho_{yx}, \quad \frac{\mu_{11}}{\sigma_y^2} = \rho_{xy} \quad (19.16)$$

называются *коэффициентами регрессии*.

§ 104. Коэффициент корреляции

Для упрощения дальнейших выводов предположим, что определен центр распределения (\bar{x}, \bar{y}) и все числа x и y заменены их отклонениями от средних: эти отклонения обозначим u и v . Числа u и v связаны с x и y очевидными соотношениями:

$$u = x - \bar{x}, \quad v = y - \bar{y}. \quad (19.17)$$

Если от x и y перейти к u и v , то уравнения регрессии (19.15) примут вид

$$\bar{u}_v = \frac{\mu_{11}}{\sigma_u^2} u, \quad \bar{u}_v = \frac{\mu_{11}}{\sigma_v^2} v. \quad (19.18)$$

В левых частях можно для сокращения писать v и u соответственно.

Так как из (19.17)

$$x = u + \bar{x}, \quad y = v + \bar{y}$$

и \bar{x} , \bar{y} — определенные числа, то

$$\bar{u} = \bar{v} = 0, \quad \sigma_u^2 = \sigma_x^2, \quad \sigma_v^2 = \sigma_y^2. \quad (19.19)$$

Число μ_{11} и дисперсии σ_u , σ_v могут быть вычислены по формулам,

$$\mu_{11} = \sum_{k=1}^n \frac{u_k v_k}{n}, \quad (19.20)$$

$$\left. \begin{aligned} \sigma_u^2 &= \frac{\sum_{k=1}^n u_k^2}{n}, \\ \sigma_v^2 &= \frac{\sum_{k=1}^n v_k^2}{n}. \end{aligned} \right\} \quad (19.21)$$

По таким формулам удобно проводить вычисления, пользуясь непосредственно таблицей (если n мало). Если же n не мало, лучше использовать (19.11).

Для простоты будем считать заданными отклонения от средних, т. е. u_k , v_k .

Мы получили две прямые регрессии, которые, как указано, проходят через центр распределения. Их направления определяются коэффициентами регрессии. Первый из них есть тангенс угла, образованного прямой регрессии v по u с осью u ; второй — тангенс угла между прямой регрессии u по v и осью v так как в этом уравнении u и v обменялись местами. Обозначим упомянутые углы через α и β (рис. 17).

Коэффициенты регрессии могут быть оба положительными ($\mu_{11} > 0$) или оба отрицательными ($\mu_{11} < 0$). В общем случае две прямые регрессии не совпадают. Если корреляционная зависимость перейдет в точную линейную функциональную, то обе прямые регрессии должны совпасть, ибо в этом случае все равно, выражено ли v через u или наоборот. Для совпадающих прямых регрессии

$$\operatorname{tg} \alpha \operatorname{tg} \beta = 1,$$

поскольку тогда

$$\alpha + \beta = \frac{\pi}{2} \quad (\mu_{11} > 0).$$

Если в среднем связи между u и \bar{v} нет, то \bar{v}_u мало изменяется при изменении u и наоборот. В этом случае α и β близки к нулю и

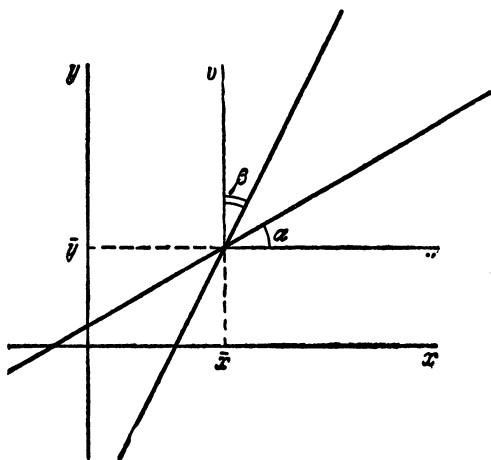


Рис. 17. Прямые регрессии.

в пределе $\operatorname{tg} \alpha \operatorname{tg} \beta = 0$. Корень квадратный из числа

$$\operatorname{tg} \alpha \operatorname{tg} \beta = \frac{\mu_{11}^2}{\sigma_u^2 \sigma_v^2} \quad (19.22)$$

принимают за критерий степени близости корреляционной связи к линейной функциональной зависимости.

Определение. Коэффициентом линейной корреляции двух случайных величин x и y называют число r , определяемое формулой

$$r = \frac{\mu_{11}}{\sigma_u \sigma_v}, \quad (19.23)$$

где u и v — отклонения x и y от их средних значений, а

$$\mu_{11} = \frac{\sum_{k=1}^n u_k v_k}{n}. \quad (19.24)$$

Из формулы (19.23) видно, что r — безразмерная величина, т. е. не зависящая от единиц, которыми измеряются исследуемые величины. Она не зависит и от начал отсчета, так как входят только центральные моменты.

Из (19.16) и (19.23) видно, что коэффициент корреляции равен корню квадратному из произведения коэффициентов регрессии. В уравнения регрессии (19.18) можно подставить из (19.23)

$$\mu_{11} = r \sigma_u \sigma_v;$$

тогда они примут вид

$$\bar{v}_u = \frac{\sigma_v}{\sigma_u} r u, \quad \bar{u}_v = \frac{\sigma_u}{\sigma_v} r v. \quad (19.25)$$

Как мы видели, корреляционная зависимость переходит в функциональную линейную, если $\operatorname{tg} \alpha \operatorname{tg} \beta = 1$, т. е. $r = \pm 1$ (в зависимости от знака μ_{11}).

С другой стороны, если представить себе, например, что точки поля корреляции попарно симметричны относительно осей координат, то прямые регрессии совпадут с осями координат, и r обратится в нуль. Но в этом случае можно считать, что u и v независимы.

Все градации перехода от полной независимости к линейной функциональной зависимости мы получим при таком изменении углов α и β , когда $\alpha + \beta$ растет от 0 до $\frac{\pi}{2}$, если $\mu_{11} > 0$. При этом r монотонно изменяется от 0 до ± 1 (формально возможный случай $\alpha \rightarrow 0, \beta \rightarrow \frac{\pi}{2}$ нами исключается, так как он не имеет значения в практике).

Если $\mu_{11} > 0$, то v растет в среднем при увеличении u и u в среднем растет с увеличением v . Выбор формулы для r дает в этом случае $r > 0$, что представляет некоторое удобство.

Число μ_{11} , входящее в формулу для r , можно вычислять разными способами. Согласно (19.24)

$$\mu_{11} = \frac{1}{n} \sum_{k=1}^n u_k v_k;$$

Здесь u_k и v_k суть отклонения от средних. Пользоваться отклонениями от средних неудобно в большинстве случаев. Поэтому обычно применяют формулы, содержащие x_k и y_k :

$$\mu_{11} = \frac{1}{n} \sum_{k=1}^n x_k y_k - \bar{x} \bar{y}, \quad (19.26)$$

если обрабатывается основная таблица А. Однако это следует делать только тогда, когда n мало.

Если обрабатывается корреляционная таблица В, то, как и в случае распределения одной величины, условимся относить все случаи, указанные в каждой клетке, к серединам интервалов x и y .

Напомним обозначения, уже использованные нами: x_k и y_k — значения x и y , соответствующие серединам интервалов величин x и y ; r — число интервалов величины x ; s — число интервалов величины y ; n_{kl} — число случаев в клетке, соответствующей x_k и y_l . Тогда для вычисления μ_{11} можно написать формулу

$$\mu_{11} = \frac{1}{n} \sum_{k=1}^r \sum_{l=1}^s n_{kl} x_k y_l - \bar{x} \bar{y}. \quad (19.27)$$

Числа \bar{x} , \bar{y} , σ_x , σ_y вычисляются по правилам определений характеристик распределения одной величины:

$$\left. \begin{aligned} \bar{x} &= \frac{1}{n} \sum_{k=1}^r n_{k0} x_k, & \bar{y} &= \frac{1}{n} \sum_{l=1}^s n_{0l} y_l, \\ \sigma_x^2 &= \frac{1}{n} \sum_{k=1}^r n_{k0} x_k^2 - \bar{x}^2, & \sigma_y^2 &= \frac{1}{n} \sum_{l=1}^s n_{0l} y_l^2 - \bar{y}^2. \end{aligned} \right\} \quad (19.28)$$

§ 105. Средние ошибки уравнений регрессии; границы значений коэффициента корреляции

Если воспользоваться уравнением регрессии для определения y по заданному значению x (или наоборот), то вычисленное y_c будет отличаться от каждого из значений y , которые фактически соответствовали числу x при наблюдениях.

Можно сказать, что, подставляя x в уравнение регрессии для вычисления y , мы получим y_c с ошибкой (правильнее,

с отклонением от наблюдений). Естественно поставить вопрос о величине средней «ошибки» уравнения регрессии, чтобы иметь возможность судить, насколько рассеяны точки поля относительно прямой регрессии. При определении средней ошибки (т. е. среднего квадратичного отклонения), удобно иметь дело не со значениями величины, а с их отклонениями от средних. Поэтому уравнение регрессии (19.18) напомним в форме (19.25):

$$v = \frac{\sigma_v}{\sigma_u} r u, \quad u = \frac{\sigma_u}{\sigma_v} r v;$$

мы опускаем черточки и значки, так как речь идет о сравнении σ_v , вычисленного по уравнению регрессии для определенного u , со всеми значениями v , полученными из наблюдений (таблица А).

Обозначим средние квадратичные ошибки уравнений регрессии через s_v и s_u . По определению,

$$n s_v^2 = \sum_{k=1}^n \left(v_k - \frac{\sigma_v}{\sigma_u} r u_k \right)^2 = \sum_{k=1}^n v_k^2 - 2 \frac{\sigma_v}{\sigma_u} r \sum_{k=1}^n u_k v_k + \frac{\sigma_v^2}{\sigma_u^2} r^2 \sum_{k=1}^n u_k^2.$$

Так как u_k, v_k — отклонения от средних, то

$$\left. \begin{aligned} \sum_{k=1}^n v_k^2 &= n \sigma_v^2, \\ \sum_{k=1}^n u_k^2 &= n \sigma_u^2, \\ \sum_{k=1}^n u_k v_k &= n \mu_{11} = n r \sigma_u \sigma_v. \end{aligned} \right\} \quad (19.29)$$

Поэтому

$$s_v^2 = \sigma_v^2 (1 - r^2), \quad s_v = \sigma_v \sqrt{1 - r^2}. \quad (19.30)$$

Аналогичным способом получим

$$s_u = \sigma_u \sqrt{1 - r^2}. \quad (19.31)$$

Из формул (19.30) можно сделать следующее заключение: если воспользоваться уравнением регрессии для определения значения одной величины по значению другой, то средняя ошибка этого определения меньше средней ошибки первой величины, если ее заменить средним значением, полученным из всего распределения.

Формулы (19.30) и (19.31) позволяют точно установить, в каких пределах может изменяться r ; s_v^2 — положительное число; поэтому из формулы для s_v^2 следует, что $r^2 \leq 1$ или

$$-1 \leq r \leq +1. \quad (19.32)$$

Такое же заключение можно получить для s_u^2 .

Коэффициент корреляции r может обращаться в -1 или $+1$ только в том случае, когда $s_v^2 = 0$ или $s_u^2 = 0$; величина ns_v^2 есть сумма квадратов отклонений и может равняться нулю тогда, когда каждое отклонение равно нулю; то же можно сказать и о ns_u^2 . Таким образом, равенство $r = \pm 1$ означает, что уравнения регрессии дают точные значения v по u и наоборот, а это означает, что мы имеем линейную функциональную зависимость.

Отметим особо, что коэффициент корреляции должен указывать степень связи между двумя случайными величинами. Вполне понятно, что одного этого числа недостаточно, чтобы охарактеризовать распределение с этой стороны*). Поэтому нужно осторожно относиться к заключениям, основанным на величине коэффициента корреляции. Если $r \approx 0$, то это еще не означает, что связи нет: связь может быть близка к нелинейной, которая не учитывается величиной r .

§ 106. Средние ошибки выборочных коэффициентов корреляции и регрессии

Понятие о генеральной совокупности можно распространить на случай, когда мы изучаем распределение нескольких случайных величин (в частности, двух). Если нас, например, интересует корреляция между видимой величиной и расстоянием звезд некоторого спектрального класса, то генеральной совокупностью следует считать все те наблюдения, которые теоретически можно было бы сделать (пронаблюдав все звезды). При исследовании связи между ростом призываемых в армию и их весом генеральной совокупностью будет полное собрание сведений о росте и весе всех освидетельствованных. На практике изучение генеральной совокупности либо бывает слишком громоздко, как во втором примере, либо практически неосуществимо, как в первом примере, где понятие генеральной совокупности имеет теоретический характер. Обычно мы имеем дело с выборкой из генеральной совокупности; так, например, четыре срочных наблюдения температуры и влажности, собранные за год, представляют выборку из теоретической генеральной совокупности этих величин.

Однако представляет интерес на основании выборки составить себе представление о характере генеральной совокупности. Для этого необходимо выяснить, в какой мере выборочный коэффициент

*) На практике принято считать, что величины достаточно связаны, если $|r| > 0,7$ или $0,6$. Однако можно сделать заключение о наличии связи и при меньших значениях r , если это заключение можно подкрепить физическими соображениями. С другой стороны, даже когда $|r| = 0,9$, то не следует категорически говорить о зависимости, если это нельзя объяснить физическими причинами.

корреляции и выборочные коэффициенты регрессии представляют такие же величины в генеральной совокупности.

На основании закона больших чисел можно только сказать, что при достаточно большой выборке выборочные параметры могут как угодно мало отличаться от параметров генеральной совокупности, и этого можно ожидать с вероятностью, близкой к единице.

Пирсоном было исследовано распределение выборочных значений r из генеральной совокупности, подчиненной нормальному закону распределения (см. гл. 12). Закон распределения оказался довольно сложным: распределение выборочных коэффициентов корреляции отличается от нормального и только приближается к нему, если выборка велика (но при этом значение коэффициента корреляции не близко к единице). Из функции распределения можно вывести приближенное выражение для средней ошибки выборочного коэффициента корреляции,

$$\sigma_r = \frac{1-r^2}{\sqrt{n}}, \quad (19.33)$$

где n — объем выборки (количество наблюдений).

Согласно сказанному о распределении выборочных r , величине σ_r можно придать такое же значение, как в нормальном распределении, если выборка велика. В противном случае σ_r только грубо оценивает возможные границы для генерального коэффициента (например, по правилу трех сигм).

Для средних ошибок выборочных коэффициентов регрессии получаются следующие приближенные выражения:

$$\left. \begin{aligned} \sigma(\rho_{yx}) &= \frac{\sigma_y}{\sigma_x} \sqrt{\frac{1-r^2}{n}}, \\ \sigma(\rho_{xy}) &= \frac{\sigma_x}{\sigma_y} \sqrt{\frac{1-r^2}{n}} \end{aligned} \right\} \quad (19.34)$$

для регрессии Y по X и X по Y соответственно.

Если выборка очень мала, то вычисление средних ошибок по указанным формулам практически ничего не дает. В таких случаях необходимо оценивать корреляцию по теории малой выборки, пользуясь исследованиями Фишера. Очень ясное изложение этого вопроса можно найти в книге В. И. Романовского «Элементарный курс математической статистики».

§ 107. Вероятностное значение элементарной теории корреляции

В предыдущих параграфах этой главы строились линейные эмпирические формулы, приближенно представляющие зависимость средних значений каждой из величин от соответствующих значений другой величины. Вместе с тем был введен коэффициент корреляции как мера отклонения указанных связей от линейных.

Такая постановка задачи об изучении статистических связей между двумя величинами по наблюдавшимся парам их значений была дана нами без связи с соответствующими разделами теории вероятностей. Следует установить соответствие между гл. 12, содержащей учение о совокупности двух случайных величин, и гл. 19, чтобы выяснить, при каких условиях можно признать развитый в гл. 19 вычислительный аппарат.

Построенный аппарат теории корреляции дает в качестве суммарных характеристик рассмотренной выборочной совокупности пять чисел: средние значения каждой из величин, средние квадратичные отклонения каждой величины в отдельности и коэффициент корреляции. При каких условиях этих пяти чисел достаточно для описания выборочной совокупности и приближенного определения параметров генеральной совокупности, из которой сделана выборка при наблюдениях?

В гл. 12 было рассмотрено нормальное распределение совокупности двух случайных величин. Там было показано, что условное математическое ожидание каждой величины (при заданном значении другой) есть линейная функция другой величины. Это соответствует принятому в гл. 19 условию строить линейные уравнения регрессии.

Введенный в гл. 19 коэффициент корреляции r можно формально определить как частное от деления среднего значения произведения отклонений обеих величин от их средних значений на произведение их средних квадратичных отклонений.

Как было показано в § 64, параметр R двумерного нормального закона равен частному от деления математического ожидания произведения отклонений величин от их математических ожиданий на произведение средних квадратичных отклонений этих величин. Если учесть что математическое ожидание случайной величины есть ее среднее значение, то получаем полное соответствие между числами r и R . Можно сказать, что r есть приближенное значение R , полученное из выборочной совокупности, если считать, что генеральная совокупность подчинена нормальному закону распределения.

Из сказанного об уравнениях регрессии и о коэффициенте корреляции вытекает, что исследование линейной корреляции допустимо, если рассматриваемые величины имеют нормальное распределение. Вычисляемые при этом параметры распределения (x , y , σ_x , σ_y , r) можно считать достаточными для описания изучаемой совокупности двух величин, если ее можно считать нормально распределенной.

Учет средних квадратичных ошибок этих параметров необходим, так как это позволяет оценить надежность вычисленных значений параметров.

Если их можно считать надежными, то можно написать плотность вероятности с приближенными значениями параметров и вычислять вероятность того, что значения величин принадлежат некоторой заданной области интегрированием плотности по этой области.

§ 108. Пример и схема исследования корреляции при большом числе наблюдений

Если материала мало (около 50), то параметры вычисляются по формулам § 103—104.

Рассмотрим вычисление коэффициента корреляции и уравнений регрессии для большого наблюдательного материала.

Вычислим коэффициент корреляции и выведем уравнения регрессии для видимых звездных величин m и логарифмов расстояний в парсеках P звезд спектрального типа B5 — B9.

Вычисления следует располагать в схеме, которая приводится на стр. 330.

I	II	III									IV	V	VI	VII	VIII	IX	X	XI
(1) $\frac{m}{p}$	$\frac{x}{y}$	1,5	2,5	3,5	4,5	5,5	6,5	7,5	Распредел. р (у)									
		-4	-3	-2	-1	0	+1	+2	y_i	n_{0i}	$n_{1i} \cdot y_i$	$\sum n_{ki} x_k$	$y_i \sum n_{ki} x_k$	VIII : V	x_i из уравн. регрессии			
(2)																		
(3)	1,1	-4	2							-4	2	-8	32	-8	+32	-4,00	-3,80	
	1,3	-3	2	4						-3	6	-18	54	-28	+60	-3,33	-3,06	
	1,5	-2		3	3					-2	6	-12	24	-15	+30	-2,50	-2,32	
	1,7	-1			25	7				-1	32	-32	32	-57	+57	-1,78	-1,58	
	1,9	0			7	64	37	1		0	109	0	0	-77	0	-0,71	-0,84	
	2,1	+1			4	16	143	14		+1	177	+177	177	-10	-10	-0,06	-0,09	
(4)	2,3	+2				4	20	16	2	+2	42	+84	168	+16	+32	+0,38	+0,65	
	2,5	+3						1	1	+3	2	+6	18	+3	+9	+1,50	+1,39	
											Сумма	Сумма	Сумма	Сумма	Сумма			
(5)	$\frac{x}{n}$	-4	-3	-2	-1	0	+1	+2		Сумма	376	+197	505	210				
(6)	$n_{k0} \cdot x_k$	-16	-21	-78	-91	0	+32	+6		Сумма	-168							
(7)	$n_{k0} \cdot x_k^2$	64	63	156	91	0	32	12		Сумма	418							
(8)	$\sum n_{k2} y_i$	-14	-18	-27	+17	+183	+49	+7										
(9)	$x_k \cdot \sum n_{k2} y_i$	+56	+54	+54	-17	0	+49	+14		Сумма	210							
(10)	(8) : (5)	-3,50	-2,57	-0,69	+0,19	+0,92	+1,53	+2,33										
(11)	$\bar{y}_{\text{д}}$ из уравн. регрессии	-2,56	-1,70	-0,83	-0,04	+0,91	+1,78	+2,65										

Описание и пояснение схемы (стр. 330)

В столбце I даны значения величины P в серединах ее интервалов; так как шаг равен 0,2, то в первом интервале содержатся звезды со значением P от 1,0 до 1,2, во втором — от 1,2 до 1,4 и т. д. В строке (1) даны значения величины m в серединах ее интервалов; здесь шаг равен единице, поэтому в первый интервал попадут звезды, у которых видимая величина заключена в пределах от 1 до 2, во второй — от 2 до 3 и т. д. Столбец II содержит значения середин интервалов величины P в условных единицах от условного нуля. Строка (2) содержит такие же значения для величины m . Обе эти линии заполняются после вычисления столбца V и строки (5) так же, как в случае одномерного распределения.

Числовая матрица, обозначенная III — (3), представляет корреляционную таблицу типа В. Она получается подсчетом по каталогу количества звезд, у которых видимый блеск и величина P имеют значения в интервалах, соответствующих каждой клетке. Сумма всех элементов матрицы должна равняться общему числу наблюдений. Корреляционная таблица обведена жирными линиями. Справа к ней примыкает столбец IV и за ним столбец V. Столбец IV является повторением столбца II, и, следовательно, заполняется после столбца V. Можно несколько упростить схему, исключая либо столбец II, либо столбец IV; удобнее сохранить столбец IV, рядом со столбцом V. Последний получается сложением всех чисел в каждой строке матрицы III. Столбец V совместно со столбцом I представляет распределение одной величины P , а совместно со столбцом II или IV — распределение той же величины в условных единицах, т. е. величины Y , что и отмечено в надписях на схеме. Все сказанное можно повторить о строках (4) и (5) с той разницей, что нужно говорить о величине m или X .

Подсчет столбца V и строки (5) контролируется сложением всех чисел линии; сумма должна равняться полному числу наблюдений. Поэтому на схеме число 376 написано дважды: под столбцом V и на продолжении строки (5). Необходимость сложения как здесь, так и в других местах схемы указана словом «сумма»; на тех линиях, на которых не нужно суммировать, заключительные клетки перечеркнуты.

После вычисления столбца V нужно выбрать место условного нуля для величины P . Максимум распределения этой величины, т. е. наибольшее значение численности (чисел n_{01}) находится в том интервале, в котором середина интервала имеет значение 2,1. Можно было бы взять условный нуль здесь же, но предшествующее значение n_{01} довольно велико; есть еще несколько предшествующих значений, поэтому было принято место условного нуля в середине того интервала, где n_{01} равно 109*). Для величины m условный нуль взят в середине того интервала, где n_{00} имеет максимум.

Окаймляющие матрицу III столбцы IV и V и строки (4), (5) представляют распределения исследуемых величин Y и X , взятых отдельно, поэтому они ограничены жирными линиями.

В столбце VI производится подготовка к вычислению среднего значения величины P в условных единицах, т. е. величины Y . Поэтому в заголовке столбца написано $n_{01} \cdot y_i$, т. е. каждое число столбца V нужно умножить на стоящее рядом число столбца IV. Полученные числа складываются. Совершенно аналогично операция производится в строке (6).

В столбце VII выполняются подготовительные вычисления для определения начального момента второго порядка величины Y , который необходим для вычисления среднего квадратичного отклонения этой величины. Числа

*) Как показали вычисления, такой выбор не совсем удачен, но это не имеет существенного значения: если условный нуль выбран неудачно, то только незначительно усложнятся вычисления.

столбца VII получаются умножением чисел столбца VI на стоящие в тех же строчках числа IV столбца. Совершенно аналогичные операции производятся в строке (7); они дадут сумму, необходимую для вычисления среднего квадратичного отклонения величины m . Столбцы VI, VII и строки (6), (7) содержат обработку одномерных распределений величин P и m .

Для вычисления коэффициента корреляции по таблице типа В и коэффициентов регрессии нужно вычислить двойную сумму

$$\sum_k \sum_l n_{kl} x_k y_l = \sum_l \left[y_l \left(\sum_k n_{kl} x_k \right) \right].$$

Ее можно было бы вычислить прямым путем: всякое число матрицы III умножить на соответствующие этому числу значения x_k и y_l и сложить все элементы преобразованной матрицы. Так как при этом надо было бы выписать преобразованную корреляционную таблицу, и не было бы контроля, то предпочитают обычно другой порядок вычислений. Суммирование сначала производится по одному аргументу, затем по другому. Изменяя порядок аргументов, можно получить двойную сумму вторично; так как все операции производятся формально точно (нет приближенных действий), то вторая сумма должна точно равняться первой, что обеспечивает надежный контроль вычислений. Рассмотрим подробно вычисления в столбцах VIII и IX, которые дают двойную сумму. Как указано в заголовке столбца VIII, в каждой строке числа n_{kl} умножаются на соответствующие значения x_k и полученные произведения складываются. Эти операции выполняются либо в уме, либо на арифмометре, если числа n_{kl} велики. В рассматриваемой задаче все эти действия выполнены в уме: первая строка $-2 \times (-4) = -8$; вторая строка $-2 \times (-4) + 4 \times (-3) = -20$, ..., шестая строка $-4 \times (-2) + 16 \times (-1) + 143 \times (0) + 14 \times 1 = -10$. После умножения каждого из чисел VIII столбца на y_l получим столбец IX. Сложение чисел IX столбца дает указанную двойную сумму. Совершенно так же строки (8) и (9) дадут ту же сумму, но сначала выполняется суммирование по y , затем по x . Например, число 49 в строке (8) получено так:

$$1 \times 0 + 14 \times 1 + 16 \times 2 + 1 \times 3 = 49.$$

Умножением на $+1$ получим результат суммирования по y при $x=1$. Сложение всех чисел девятой строки заканчивает вычисление двукратной суммы в порядке сначала по y , затем по x . Должны получиться точно одинаковые числа. Столбцы VIII, IX и строки (8) и (9) вместе с шестыми и седьмыми линиями дают все нужные параметры эмпирического распределения совокупности двух величин (при условии, что достаточно рассматривать линейные уравнения регрессии). На этом можно закончить работу по схеме и перейти к вычислениям коэффициентов регрессии, к выводу уравнений регрессии и к вычислению коэффициента корреляции. Приведенные в схеме линии десятая и одиннадцатая заполняются, если необходимо сравнение эмпирической регрессии с теоретической (линейной), построенной нами. Как указано в заголовке десятых линий, каждое число восьмой линии нужно разделить на число одноименной пятой линии. Так как, например, n_{01} есть число звезд при определенном значении y , то $\frac{n_{kl}}{n_{01}}$ есть эмпирическая вероятность всякого x_k при том же условии, поэтому числа столбца X представляют весовые средние значения \bar{x}_y при последовательных y , а это и значит, что мы получили эмпирическую регрессию X по Y . Аналогично строка (10) дает последовательность средних значений y при заданных x , т. е. эмпирическую регрессию y по x (напомним, что при вычислениях считается: мы имеем n_{kl} звезд, у которых значения величин равны x_k и y_l). На одиннадцатых линиях даны значения средних X при последовательных y .

и средних \bar{Y} при разных x_k , вычисленные по уравнениям регрессии. Сравнение эмпирической регрессии с вычисленной по уравнениям регрессии дано

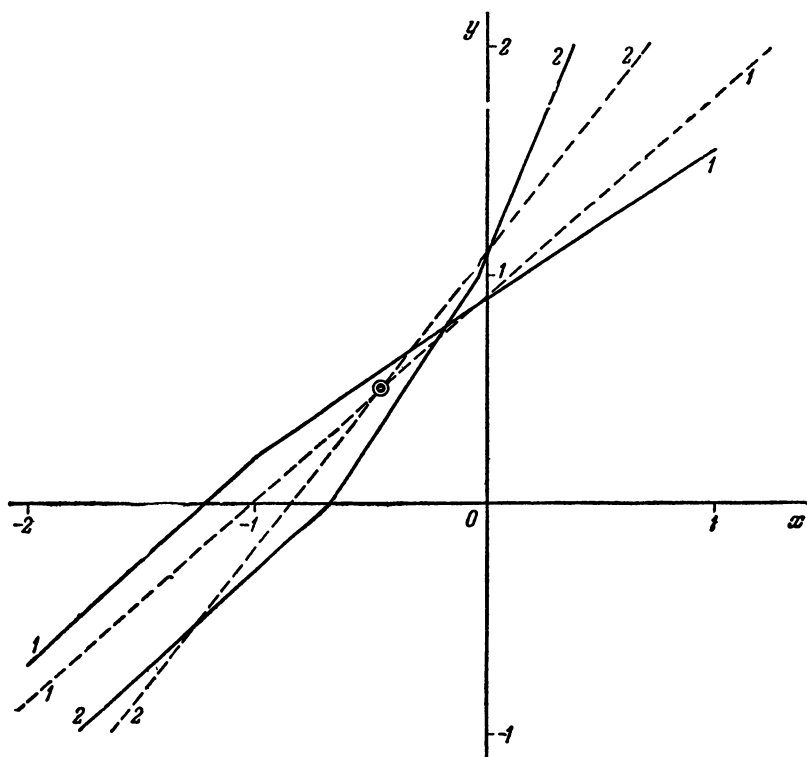


Рис. 18. Эмпирические (сплошные линии) и теоретические (пунктир) линии регрессии вблизи центра распределения. Центр распределения обведен кружком. Цифрами обозначены соответствующие линии регрессии: 1 — y по x ; 2 — x по y .

и графически на рис. 18; эмпирическая связь представлена ломаными линиями, проведенными через точки, построенные по числам десятых линий (столбцов и строк).

Вычисление параметров распределения

а) Средние значения. Деление суммы чисел столбца VI на общее число наблюдений дает среднее значение \bar{Y} :

$$\bar{y} = \frac{+197}{376} = 0,524.$$

Учитывая место условного нуля и величину шага в каталожных единицах, получим $\bar{P} = 1,9 + 0,2 \times 0,524 = 2,005$. Деление суммы чисел шестой строки

дает среднее значение величины X :

$$\bar{x} = \frac{-168}{376} = -0,447;$$

$$\bar{m} = 5,5 + 1 \times (-0,447) = 5,05,$$

так как шаг таблицы по m равен единице, а условный нуль в середине интервала с видимой величиной 5,5.

б) Средние квадратичные отклонения. Сумма чисел седьмого столбца, разделенная на число наблюдений, дает начальный момент порядка нуль — два, т. е. начальный момент второго порядка одной величины Y . Вычитание из этого момента квадрата среднего значения этой величины по формуле дисперсии даст дисперсию величины Y , а корень квадратный из дисперсии даст, по определению, среднее квадратичное отклонение:

$$\sigma_y^2 = \frac{505}{376} - (0,524)^2 = 1,068, \quad \sigma_y = \sqrt{1,068} = 1,033.$$

Аналогично по 7-й строке

$$\sigma_x^2 = \frac{418}{376} - (0,447)^2 = 0,912, \quad \sigma_x = \sqrt{0,912} = 0,955.$$

в) Коэффициент корреляции. Деление суммы чисел IX столбца (или 9-й строки) на полное число наблюдений дает начальный момент порядка один — один данного двухмерного распределения:

$$\gamma_{11} = \frac{210}{376} = 0,559.$$

Чтобы получить центральный момент того же порядка, нужно из начального момента вычесть произведение средних значений исследуемых величин:

$$\mu_{11} = 0,559 - (-0,447) \times (+0,524) = 0,793.$$

Для вычисления коэффициента корреляции заготовим произведение средних квадратичных отклонений: $\sigma_y \sigma_x = 1,033 \cdot 0,955 = 0,987$; после этого получаем:

$$r = \frac{0,793}{0,987} = 0,803.$$

г) Коэффициенты регрессии, уравнения регрессии. Мы их вычислим в условных единицах по формулам, содержащим коэффициент корреляции. Предварительно заготовим отношения средних квадратичных отклонений:

$$\frac{\sigma_y}{\sigma_x} = 1,082, \quad \frac{\sigma_x}{\sigma_y} = 0,924.$$

После этого для коэффициентов регрессии получим значения

$$r_{yx} = 0,803 \times 1,082 = 0,869, \quad r_{xy} = 0,803 \times 0,924 = 0,742.$$

Уравнения регрессии имеют вид:

регрессия у по x

$$\bar{y}_x - 0,524 = 0,869 \times (x + 0,447)$$

или

$$\bar{y}_x = 0,869x + 0,912$$

(обозначение \bar{y}_x можно заменить просто на y , но не следует забывать о смысле и происхождении уравнений регрессии);
регрессия x по y

$$\bar{x}_y + 0,447 = 0,742 \times (y - 0,524)$$

или

$$\bar{x}_y = 0,742y - 0,836.$$

По этим уравнениям и были просчитаны столбец и строка одиннадцатые, для чего в правые их части подставлялись последовательно значения x и y из четвертой строки и четвертого столбца соответственно.

д) Средние квадратичные ошибки параметров

Описанные в этом пункте дополнительные вычисления делаются в тех задачах, в которых статистический материал представляет выборку из бесконечной генеральной совокупности, подчиненной нормальному закону распределения.

В § 99 и 106 были приведены без доказательства формулы, по которым вычисляются средние квадратичные отклонения основных параметров одномерного и двумерного распределений.

Некоторые из приводимых формул известны читателям из теории случайных ошибок.

1. Средние квадратичные ошибки средних арифметических:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}, \quad \sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}};$$

у нас $\sqrt{n} = 19,4$, поэтому

$$\sigma_{\bar{x}} = \frac{0,955}{19,4} = 0,049; \quad \sigma_{\bar{y}} = \frac{1,033}{19,4} = 0,053.$$

2. Средние квадратичные ошибки выборочных средних квадратичных ошибок:

$$\sigma(\sigma_x) = \frac{\sigma_x}{\sqrt{2n}}, \quad \sigma(\sigma_y) = \frac{\sigma_y}{\sqrt{2n}};$$

в нашем примере $\sqrt{2n} = 27,4$, поэтому

$$\sigma(\sigma_x) = \frac{0,955}{27,4} = 0,035, \quad \sigma(\sigma_y) = \frac{1,033}{27,4} = 0,038.$$

Указанные величины характеризуют надежность параметров одномерной задачи. Следующие числа оценивают надежность параметров двумерного распределения.

3. Средние квадратичные ошибки уравнений регрессии s_y и s_x вычисляются по выведенным ранее формулам (19.30). У нас $r^2 = 0,645$, $1 - r^2 = 0,355$, $\sqrt{1 - r^2} = 0,596$, $s_y = 1,033 \cdot 0,596 = 0,616$, $s_x = 0,955 \cdot 0,596 = 0,569$. Числа s_y и s_x представляют средние квадратичные ошибки, которые получатся, если сравнивать значения, вычисленные из уравнений регрессии, с отдельными (но всеми) значениями, полученными из наблюдений.

4. Средняя квадратичная ошибка коэффициента корреляции. По формуле (19.33) для изучаемого распределения получим

$$\sigma_r = \frac{0,355}{19,4} = 0,018.$$

5. Средние квадратичные ошибки коэффициентов регрессии вычисляются по формулам (19.34). Вычисления дают следующие результаты:

$$\frac{\sqrt{1-r^2}}{\sqrt{n}} = \frac{0,596}{19,4} = 0,0307, \quad \sigma(r_{yx}) = 1,082 \cdot 0,0307 = 0,033, \\ \sigma(r_{xy}) = 0,924 \cdot 0,0307 = 0,028.$$

Сводка результатов

Запишем результаты вычисления параметров распределения с учетом средних квадратичных отклонений подобно тому, как это делалось в теории случайных ошибок.

А. Сводка в условных единицах:

$$\begin{aligned} \bar{x} &= -0,447 \pm 0,049, & \bar{y} &= 0,524 \pm 0,053 \\ \sigma_x &= 0,955 \pm 0,035, & \sigma_y &= 1,033 \pm 0,038 \\ r_{yx} &= 0,869 \pm 0,033, & r_{xy} &= 0,742 \pm 0,028 \\ y_x &= 0,869x + 0,912, & \bar{x}_y &= 0,742y - 0,836 \\ r &= +0,803 \pm 0,018. \end{aligned}$$

Б. Сводка в основных единицах.

В обычных задачах нужны параметры не только в условных единицах, но и в тех единицах, в которых общепринято давать исследуемые величины, причем значения отсчитываются от общепринятого начала отсчета (нуль-пункта).

Способ перехода от условных единиц и условного начала отсчета к основным единицам и началу отсчета для одномерного распределения был указан в предыдущей главе. Поэтому достаточно добавить указания об изменении параметров собственно двухмерного распределения. Так как коэффициенты регрессии и коэффициент корреляции зависят от центральных моментов, то изменение начала отсчета не влияет на них, нужно учесть только величины шагов. Коэффициент корреляции вовсе не изменяется, так как это величина безразмерная, как было указано при введении этого параметра. Поэтому рассмотрим только вопрос об изменении коэффициентов регрессии. Из выражений для них

$$r_{yx} = r \frac{\sigma_y}{\sigma_x}, \quad r_{xy} = r \frac{\sigma_x}{\sigma_y}$$

видно, что для перехода к основным единицам нужно в первом из них числитель умножить на шаг второй величины в основных единицах, а знаменатель — на шаг первой величины; во втором коэффициенте — противоположный порядок преобразований. При этой первой величиной считается та, которая обозначена через x в условных единицах. Выполним переход в нашем примере. Так как условный нуль величины x соответствует значению 5,5 видимой величины, а шаг по этой величине равен единице, условный нуль величины y соответствует значению 1,9 величины P , и шаг этой величины равен 0,2, то

$$\begin{aligned} \bar{m} &= 5,5 - 0,447 \times 1 = 5,053, & \bar{P} &= 1,9 + 0,524 \times 0,2 = 2,005 \\ \sigma_m &= 0,955 \times 1 = 0,955, & \sigma_P &= 1,033 \times 0,2 = 0,207 \\ r_{Pm} &= 0,869 \times \frac{0,2}{1} = 0,174, & r_{mP} &= 0,742 \times \frac{1}{0,2} = 3,71. \end{aligned}$$

Уравнения регрессии преобразуем непосредственно, используя полученные коэффициенты регрессии и средние значения. Уравнение регрессии P по m имеет вид

$$\bar{P}_m - 2,005 = 0,174 \times (m - 5,053) \quad \text{или} \quad \bar{P}_m = 0,174m + 1,126.$$

Уравнение регрессии m по P

$$\bar{m}_P - 5.053 = 3.71 \times (P - 2.005)$$

или

$$\bar{m}_P = 3.71P - 2.39.$$

Проведенные вычисления дают возможность сделать следующие выводы о корреляции величин P и m . По значениям средних квадратичных ошибок параметров видно, что все они получены достаточно уверенно, в особенности коэффициент корреляции. Поэтому связь между P и m по исследованному материалу может считаться близкой к линейной в среднем. Это же подтверждается сравнением чисел в столбцах X и XI , а также в строках с такими же номерами. В последних можно, однако, отметить более заметные отклонения теоретических средних от эмпирических на левом краю распределения, что, по-видимому, можно объяснить малым количеством ярких звезд.

Из сказанного следует, что уравнениями регрессии можно пользоваться для вычисления значений одной величины по значениям другой, результаты вычислений дадут в среднем небольшое расхождение с эмпирическими средними.

Рекомендуется читателю самостоятельно провести сравнение эмпирических регрессий с теоретическими в основных единицах. Для этого нужно составить столбец и строку десятые в основных единицах, а линии одиннадцатые просчитать по теоретическим уравнениям регрессии в основных единицах. Если найти отклонения чисел столбца X от чисел столбца XI и вычислить весовое среднее квадратичное отклонение, то получим суммарную оценку качества приближения материала в среднем уравнениями регрессии.

§ 109. Пример исследования корреляции по малому числу наблюдений

Обработка двухмерной статистической совокупности при малом числе наблюдений значительно упрощается. Рассматриваемый ниже пример такого рода взят из книги В. И. Романовского.

Исследуется корреляция между количеством осадков в миллиметрах по январю месяцу последовательных девяти лет в Ташкенте и на опытной сельскохозяйственной станции в 12 км от Ташкента. Вследствие близости станций можно ожидать, что связь тесная, но проверка необходима, так как возможны местные особенности, искажающие связь. Результаты наблюдений в Ташкенте обозначены через x , на станции — y .

Мы приводим схему вычислений с внесенными в нее наблюдениями.

Схема проста, и для ее понимания достаточно надписей над столбцами. Можно отметить, что необходимые для дальнейшего суммы в последней строке можно получить на арифмометре «способом накопления» без записи

	x^2	x	xy	y	y^2
	7 569	87	7 482	86	7 396
	2 209	47	2 632	56	3 136
	5 476	74	6 216	84	7 056
	7 396	86	6 192	72	5 184
	1 444	38	1 786	47	2 209
	225	15	255	17	289
	1 681	41	1 763	43	1 849
	64	8	152	19	361
	6 241	79	6 952	88	7 744
Суммы	32 305	475	33 430	512	35 224

отдельных слагаемых. Здесь вычисления сделаны подробно, чтобы виден был порядок вычислений полностью. По формулам § 103, 104 выполняем следующие вычисления:

$$\begin{aligned}\bar{x} &= \frac{475}{9} = 52,8, & \bar{y} &= \frac{512}{9} = 56,9 \\ v_{20} &= \frac{32305}{9} = 3589,44, & v_{02} &= \frac{35224}{9} = 3913,78, \\ \mu_{20} &= 3589,44 - 52,8^2 = 801,60, & \mu_{02} &= 3913,78 - 56,9^2 = 676,17. \\ \sigma_x &= \sqrt{801,60} = 28,3, & \sigma_y &= \sqrt{676,17} = 26,0, \\ v_{11} &= \frac{33430}{9} = 3710, & \bar{x}\bar{y} &= 52,8 \cdot 56,9 = 3000, \\ \mu_{11} &= 3710 - 3000 = 710, & \sigma_x \sigma_y &= 28,3 \times 26,0 = 736, \\ r &= \frac{710}{736} = 0,965, & r^2 &= 0,931, & 1 - r^2 &= 0,069, \\ \sigma_r &= \frac{0,069}{\sqrt{9}} = 0,023, & r &= 0,965 \pm 0,023.\end{aligned}$$

Коэффициент корреляции близок к единице, и его средняя квадратичная ошибка очень мала, несмотря на крайне малое число наблюдений. Вычисления вначале велись формально точно, т. е. так, как если бы значения x и y были совершенно точными числами. Начиная со средних квадратичных отклонений вычисления трехзначные.

Результат определения коэффициента корреляции показывает, что уравнениями регрессии можно пользоваться для вычисления значений одной величины по значениям другой с небольшой средней квадратичной ошибкой.

Читателю рекомендуется самостоятельно вывести уравнения регрессии и сравнить результаты вычисления по уравнениям регрессии с наблюдениями. Ответ (неполный): $\rho_{yx} = 0,887$, $\rho_{xy} = 1,050$.

ЛИТЕРАТУРА

Часть I

- Безикович Я. С., Приближенные вычисления, изд. 6-е, Гостехиздат, 1949.
- Березин И. С. и Жидков Н. П., Методы вычислений, т. I, Физматгиз, 1959
- Демидович Б. П. и Марон И. А., Основы вычислительной математики, Физматгиз, 1960.
- Крылов А. Н., Лекции о приближенных вычислениях, изд. 6-е, Гостехиздат, 1954; Собр. трудов, т. 3, ч. I, Изд. АН СССР, 1949.
- Яковлев К. П., Математическая обработка результатов измерений, изд. 2-е, Гостехиздат, 1953.

Часть II

- Блажко С. Н., Курс сферической астрономии, изд. 2-е Гостехиздат, 1954, гл. II.
- Гончаров В. Л., Теория интерполирования и приближения функций, изд. 2-е, Гостехиздат, 1954.
- Казаков С. А., Курс сферической астрономии (дополнение об интерполяции), изд. 2-е, Гостехиздат, 1940.
(См. также литературу к части I.)

Часть III

- Бернштейн С. Н., Теория вероятностей, изд. 4-е, Гостехиздат, 1946.
- Гнеденко Б. В., Курс теории вероятностей, изд. 2-е, Гостехиздат, 1954.
- Гнеденко Б. В. и Хинчин А. Я., Элементарное введение в теорию вероятностей, изд. 3-е, Гостехиздат, 1952.
- Гончаров В. Л., Теория вероятностей. Оборонгиз, 1939.
- Колмогоров А. Н., Основные понятия теории вероятностей. ОНТИ, 1936.
- Романовский В. И., Элементарный курс математической статистики Госпланиздат, 1939.

Часть IV

- Идельсон Н. И., Способ наименьших квадратов и теория математической обработки наблюдений, Геодезиздат, 1947.
- Колмогоров А. Н., К обоснованию метода наименьших квадратов, Успехи матем. наук I, вып. I, 1946.
- Лахтин Л. К., Курс теории вероятностей, Гос. изд-во, 1924.
- Линник Ю. В., Метод наименьших квадратов и основы теории обработки наблюдений, Физматгиз, 1958.

- Маковер С. Г., Решение системы нормальных уравнений при помощи матриц, Астрон. ж. 33, вып. 3, 1956.
- Резниковский П. Т., Об одном варианте решения системы нормальных уравнений в методе наименьших квадратов, Сообщ. Гос. астрон. ин-та им. Штернберга, № 54, 1950.
- Романовский В. И., Основные задачи теории ошибок, Гостехиздат, 1947.
- Семендяев К. А., Эмпирические формулы, ОНТИ, 1933.
- Уиттекер Э. и Робинсон Г., Математическая обработка результатов измерений, Пер. с англ., ОНТИ, 1933.
- Фаддеева В. Н., Вычислительные методы линейной алгебры, Гостехиздат, 1950.
- (См. также литературу к части I, а также книги Б. В. Гнеденко и Б. В. Гнеденко и А. Я. Хинчина к части III.)

Часть V

- Крамер Г., Математические методы статистики, Пер. с англ., ИЛ, 1948.
- Романовский В. И., Математическая статистика, ОНТИ, 1938.
- Слуцкий Е. Е., Теория корреляции и элементы учения о кривых распределения, Киев, 1912.
- Смирнов Н. В., Приближение законов распределения случайных величин по эмпирическим данным, Успехи матем. наук, вып. 10, 1944.
- Смирнов Н. В. и Дунин-Барковский И. В., Краткий курс математической статистики для технических приложений, Физматгиз, 1959.
- (См. также книги Б. В. Гнеденко и В. И. Романовского к части III.)
-

ПРИЛОЖЕНИЕ

В *таблице I* даны значения функции Лапласа — Гаусса, которая используется как приближенное выражение вероятности числа повторений в задаче о повторении испытаний. Эта же функция является плотностью вероятности нормального распределения с центром, равным нулю и средним квадратичным отклонением, равным единице. Если n — число испытаний, p — вероятность события в каждом испытании, k — заданное число повторений, вероятность которого нужно вычислить, то вычисление делается по формуле:

$$P_{k,n} \approx \frac{1}{\sqrt{npq}} \Phi' \left(\frac{k - np}{\sqrt{npq}} \right), \quad q = 1 - p,$$

$$\Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

В *таблице II* даны приближенные значения вероятности в задаче о повторении испытаний, когда вероятность события мала *). В таблице k есть число повторений, $a = np$, n — число испытаний, p — вероятность события в каждом испытании,

$$P_{k,n} \approx P_k(a) = \frac{a^k e^{-a}}{k!}.$$

Таблица III содержит значения интеграла вероятностей, используемые для вычисления вероятности случайной величине принять значение в заданном интервале, если она подчиняется нормальному закону распределения. Если \bar{x} — среднее значение величины X (центр распределения) и σ — среднее квадратичное отклонение, то вероятность получить значение между α и β вычисляется по формуле

$$P(\alpha < X < \beta) = \Phi \left(\frac{\beta - \bar{x}}{\sigma} \right) - \Phi \left(\frac{\alpha - \bar{x}}{\sigma} \right),$$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}} dt.$$

При пользовании таблицей нужно учесть, что $\Phi(-z) = -\Phi(z)$.

Если число в табл. I и II заканчивается цифрой 5 и после нее стоит знак плюс, то это приближенное число с недостатком; если такую цифру 5 потребуется отбросить, то предыдущую цифру нужно увеличить на единицу. Если после пяти стоит знак минус, то имеем приближенное значение с избытком; при отбрасывании такой цифры пять предыдущая цифра сохраняется.

*) Использована таблица из книги Б. В. Гнеденко «Курс теории вероятностей».

Таблица I

$$\text{Значения функции } \Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

z	Φ'	z	Φ'	z	Φ'	z	Φ'	z	Φ'	z	Φ'
0,00	0,399	0,45	0,361	0,90	0,266	1,35	0,160	1,80	0,079	2,25	0,032
0,01	0,399	0,46	0,359	0,91	0,264	1,36	0,158	1,81	0,078	2,26	0,031
0,02	0,399	0,47	0,357	0,92	0,261	1,37	0,156	1,82	0,076	2,27	0,030
0,03	0,399	0,48	0,356	0,93	0,259	1,38	0,154	1,83	0,075	2,28	0,030
0,04	0,399	0,49	0,354	0,94	0,256	1,39	0,152	1,84	0,073	2,29	0,029
0,05	0,398	0,50	0,352	0,95	0,254	1,40	0,150	1,85	0,072	2,30	0,028
0,06	0,398	0,51	0,350	0,96	0,252	1,41	0,148	1,86	0,071	2,31	0,028
0,07	0,398	0,52	0,348	0,97	0,249	1,42	0,146	1,87	0,069	2,32	0,027
0,08	0,398	0,53	0,347	0,98	0,247	1,43	0,144	1,88	0,068	2,33	0,026
0,09	0,397	0,54	0,345	0,99	0,244	1,44	0,141	1,89	0,067	2,34	0,026
0,10	0,397	0,55	0,343	1,00	0,242	1,45	0,139	1,90	0,066	2,35	0,025+
0,11	0,396	0,56	0,341	1,01	0,240	1,46	0,137	1,91	0,064	2,36	0,025-
0,12	0,396	0,57	0,339	1,02	0,237	1,47	0,135+	1,92	0,063	2,37	0,024
0,13	0,396	0,58	0,337	1,03	0,235	1,48	0,133	1,93	0,062	2,38	0,023
0,14	0,395	0,59	0,335+	1,04	0,232	1,49	0,131	1,94	0,061	2,39	0,023
0,15	0,394	0,60	0,333	1,05	0,230	1,50	0,130	1,95	0,060	2,40	0,022
0,16	0,394	0,61	0,331	1,06	0,228	1,51	0,128	1,96	0,058	2,41	0,022
0,17	0,393	0,62	0,329	1,07	0,225+	1,52	0,126	1,97	0,057	2,42	0,021
0,18	0,392	0,63	0,327	1,08	0,223	1,53	0,124	1,98	0,056	2,43	0,021
0,19	0,392	0,64	0,325+	1,09	0,220	1,54	0,122	1,99	0,055+	2,44	0,020
0,20	0,391	0,65	0,323	1,10	0,218	1,55	0,120	2,00	0,054	2,45	0,020
0,21	0,390	0,66	0,321	1,11	0,216	1,56	0,118	2,01	0,053	2,46	0,019
0,22	0,389	0,67	0,319	1,12	0,213	1,57	0,116	2,02	0,052	2,47	0,019
0,23	0,388	0,68	0,317	1,13	0,211	1,58	0,114	2,03	0,051	2,48	0,018
0,24	0,388	0,69	0,314	1,14	0,208	1,59	0,113	2,04	0,050	2,49	0,018
0,25	0,387	0,70	0,312	1,15	0,206	1,60	0,111	2,05	0,049	2,50	0,018
0,26	0,386	0,71	0,310	1,16	0,204	1,61	0,109	2,06	0,048	2,51	0,017
0,27	0,385	0,72	0,308	1,17	0,201	1,62	0,107	2,07	0,047	2,52	0,017
0,28	0,384	0,73	0,306	1,18	0,199	1,63	0,106	2,08	0,046	2,53	0,016
0,29	0,382	0,74	0,303	1,19	0,196	1,64	0,104	2,09	0,045	2,54	0,016
0,30	0,381	0,75	0,301	1,20	0,194	1,65	0,102	2,10	0,044	2,55	0,015
0,31	0,380	0,76	0,299	1,21	0,192	1,66	0,101	2,11	0,043	2,56	0,015
0,32	0,379	0,77	0,297	1,22	0,190	1,67	0,099	2,12	0,042	2,57	0,015
0,33	0,378	0,78	0,294	1,23	0,187	1,68	0,097	2,13	0,041	2,58	0,014
0,34	0,376	0,79	0,292	1,24	0,185	1,69	0,096	2,14	0,040	2,59	0,014
0,35	0,375+	0,80	0,290	1,25	0,183	1,70	0,094	2,15	0,040	2,60	0,014
0,36	0,374	0,81	0,287	1,26	0,180	1,71	0,092	2,16	0,039	2,61	0,013
0,37	0,373	0,82	0,285+	1,27	0,178	1,72	0,091	2,17	0,038	2,62	0,013
0,38	0,371	0,83	0,283	1,28	0,176	1,73	0,089	2,18	0,037	2,63	0,013
0,39	0,370	0,84	0,280	1,29	0,174	1,74	0,088	2,19	0,036	2,64	0,012
0,40	0,368	0,85	0,278	1,30	0,171	1,75	0,086	2,20	0,036	2,65	0,012
0,41	0,367	0,86	0,276	1,31	0,169	1,76	0,085	2,21	0,035	2,66	0,012
0,42	0,365+	0,87	0,273	1,32	0,167	1,77	0,083	2,22	0,034	2,67	0,011
0,43	0,364	0,88	0,271	1,33	0,165	1,78	0,082	2,23	0,033	2,68	0,011
0,44	0,362	0,89	0,268	1,34	0,163	1,79	0,080	2,24	0,032	2,69	0,011

Продолжение

[illegible]

Т а б л и ц а II

Распределение Пуассона $P(a) = \frac{a^k e^{-a}}{k!}$

$\begin{matrix} \alpha \\ k \end{matrix}$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,905	0,819	0,741	0,670	0,607	0,549	0,497	0,449	0,407	0,368
1	0,090	0,164	0,222	0,268	0,303	0,329	0,348	0,359	0,366	0,368
2	0,005—	0,016	0,033	0,054	0,076	0,099	0,122	0,144	0,165—	0,184
3	0,000	0,001	0,003	0,007	0,013	0,020	0,028	0,038	0,049	0,061
4			0,000	0,001	0,002	0,003	0,005	0,008	0,011	0,015+
5						0,000	0,001	0,001	0,002	0,003
6								0,000	0,000	0,001

