Федеральное агентство по образованию
Государственное образовательное учреждение
высшего профессионального образования
"ТВЕРСКОЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ"

В.В. ГРИГОРЬЕВА, В.Г. ШЕРЕТОВ, Ю.В. ШЕРЕТОВ

ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Учебное пособие

ТВЕРЬ 2009

УДК: 519.2 (075.8)
ББК: В171в172я73
Г 49

Рецензенты:
Кандидат физико-математических наук, доцент
Э.А. Сергеев
Кандидат физико-математических наук, доцент
С.Ю. Щербакова

Григорьева В.В., Шеретов В.Г., Шеретов Ю.В.

Г 49 Элементы теории вероятностей и математической статистики: Учеб. пособие. – Тверь: Твер. гос. ун-т, 2009. – 88 с.

Ил. 3. Библиогр.: 14 назв.

ISBN 978-5-7609-0462-1

В доступной форме излагаются основные понятия теории вероятностей и математической статистики как раздела курса высшей математики, изучаемого в технических университетах. Знакомство с этим разделом необходимо не только инженерно-техническим, управленческим и медицинским работникам, биологам, генетикам, но и юристам, экономистам, социологам, психологам, банковским служащим, специалистам по рекламе и маркетингу. Большинство суждений носит эвристический, не вполне строгий характер. Разобрано значительное число примеров. В конце глав даны списки задач для самостоятельного решения. Их количество достаточно для полноценного усвоения предмета. Дополнительная учебная литература приведена в конце пособия.

УДК: 519.2 (075.8)
ББК: В171в172я73

Печатается по решению научно-методического совета
Тверского государственного университета
(протокол № 2 от 4 февраля 2009 г.)

ISBN 978-5-7609-0462-1

Теория вероятностей – вот истинная логика этого мира.

Дж.К. Максвелл

Глава I

ВЕРОЯТНОСТИ СЛУЧАЙНЫХ СОБЫТИЙ

§ 1.1. Случайные события

Теория вероятностей изучает закономерности случайных явлений. Если при осуществлении определенного комплекса условий может произойти или не произойти некоторое событие A, то A называется *случайным событием*, а каждое осуществление указанного комплекса условий называется *испытанием*. Для простоты будем пользоваться одним термином *испытание* для таких понятий, как опыт (эксперимент), наблюдение, измерение и т.п. Будем также считать, что испытание можно повторять в принципе неограниченное число раз.

* Испытание – подбрасывание монеты. Случайное событие – выпадение герба (орла).

* Испытание – производство выстрела по мишени. Случайное событие – промах.

* Испытание – взвешивание на аналитических весах эталона в 10 г. Случайное событие – ошибка при взвешивании превосходит 0.05 мг.

* Испытание – бросание игральной кости (кубика, изготовленного из однородного материала, грани которого пронумерованы числами от 1 до 6). Случайное событие – выпадение числа 6 на верхней грани кости (шести очков).

* Испытание – автоматическое перемешивание шаров в лототроне и извлечение одного шара. Случайное событие – выпавший шар имеет номер 36.

* Испытание – вращение барабана (рулетки) в игре "Поле чудес". Случайное событие – выпадение сектора "приз".

Достоверным называется событие, которое происходит при каждом испытании. *Невозможным* называется событие, которое не может произойти ни в одном испытании.

* При однократном бросании игральной кости выпадение положительного числа очков – достоверное событие, выпадение 7 очков – невозможное событие.

Случайные события A и B называются *несовместными*, если их одновременное появление в одном испытании невозможно.

* Пусть A – выпадение четного числа очков, B – выпадение 5 очков при однократном бросании игральной кости. Тогда A и B – несовместные события.

Пусть A – случайное событие. Символом $\overline{A}$ обозначают *противоположное* A событие, состоящее в том, что в данном испытании A не происходит. Очевидно, что любые два противоположных случайных события являются несовместными. В частности, достоверное и невозможное события противоположны.

* Пусть A – выпадение четного числа очков, B – выпадение нечетного числа очков при однократном бросании игральной кости. Тогда A и B – противоположные события.

В случае, когда события A и B тождественны (то есть A происходит, если и только если происходит B), условимся использовать запись $A = B$. Например, $\overline{\overline{A}} = A$, где $\overline{\overline{A}}$ – событие, противоположное $\overline{A}$.

§ 1.2. Статистическое, классическое и геометрическое определения вероятности

В этом пункте двумя способами вводится числовая мера возможности появления того или иного случайного события – его вероятность. Тем самым строятся простейшие математические модели исследования закономерностей случайных явлений. Пусть A – некоторое случайное событие. Если в серии из n испытаний событие A произошло в k испытаниях, то дробь k/n называется *относительной частотой* события A в данной серии испытаний. Условимся обозначать ее символом $\nu(A)$. Относительная частота наступления события $\overline{A}$ в той же серии испытаний, очевидно, равна $(n-k)/n$. Если относительная частота $\nu(A)$ при больших n слабо зависит от n и колеблется около некоторой константы $p(A), 0 \leqslant p(A) \leqslant 1$, то величина $p(A)$ называется *вероятностью* случайного события A. Это – так называемое *статистическое* определение вероятности.

* Станок, штампующий детали, дает в среднем 1.5 процента брака. Тогда вероятность того, что взятая наугад деталь окажется бракованной, равна 0.015.

* Демографическая статистика утверждает, что среди каждых ста новорожденных в среднем имеется 51 мальчик и 49 девочек. Стало быть, вероятность рождения мальчика равна 0.51, а вероятность рождения девочки – 0.49.

Вероятность достоверного события равна 1, вероятность невозможного события равна 0. В любом случае вероятность случайного события заключена между нулем и единицей. Очевидно также, что $p(\overline{A}) = 1 - p(A)$, поскольку подобным же образом связаны относительные частоты $\nu(A)$ и $\nu(\overline{A})$.

Нередко будет встречаться следующая модельная ситуация. Испытание имеет ровно N попарно несовместимых исходов $H_1, H_2, \cdots, H_N$ (каждый исход – некоторое случайное событие), причем все эти исходы равновероятны, то есть вероятность каждого исхода равна $1/N$. Если случайное событие A наступает в каждом из K перечисленных исходов и не наступает в каждом из остальных $N - K$ исходов, то вероятность события A вычисляется по формуле $p(A) = K/N$. Это – так называемое *классическое* определение вероятности.

* В закрытой урне (лототроне) имеется N одинаковых на ощупь шаров, помеченных номерами от 1 до N, причем K шаров белые, остальные – черные. Испытание состоит в тщательном перемешивании шаров, извлечении из урны наудачу одного шара и фиксировании номера шара. Это испытание имеет ровно N попарно несовместных и равновероятных исходов H_k, $k = 1, \cdots, N$, (H_k – появление шара с номером k). Вероятность появления белого шара при одном испытании равна K/N, так как этому событию "благоприятствуют K элементарных событий".

* Бросается игральная кость. Это испытание имеет 6 попарно несовместных исходов H_k – выпадение k очков на верхней грани, $k = 1, 2, \cdots, 6$. Тогда вероятность события A – выпадения не менее 5 очков – равна $1/3$, поскольку его наступлению благоприятствуют два элементарных исхода H_5 и H_6 из шести возможных.

С помощью следующей задачи о встрече познакомимся еще с так называемым *геометрическим определением вероятности*.

* Два лица условились встретиться в определенном месте в интервале между девятью и десятью часами. Время прибытия на место встречи каждого из них случайно и может быть любым моментом из условлен-

ного интервала от 9 до 10 ч. Каждый ждет 20 мин. и покидает место встречи. Какова вероятность встречи?

На координатной плоскости $x\,o\,y$ рассмотрим квадрат Q, проекциями которого на оси абсцисс и ординат будут отрезки [0,1]. На отрезке оси абсцисс будем отмечать случайные моменты прихода и ухода первого лица, на таком же отрезке оси ординат – времена прихода и ухода второго лица. Начало отсчета этих моментов – с 9 ч. Встреча состоится, если разность между временем x прихода первого лица и временем y прихода второго лица окажется не более $1/3$ ч., то есть $|x-y| \leqslant 1/3$. Множество точек квадрата Q, координаты $x\,,y$ которых удовлетворяют этому неравенству, представляет собой шестиугольник P, являющийся пересечением Q с полосой, лежащей между прямыми $x - y = 1/3$ и $x - y = -1/3$. Очевидно, что площадь P равна разности между площадью Q и суммой площадей двух прямоугольных треугольников с катетами $2/3$, дополняющими P до квадрата Q. Значит, $P = 1 - (2/3)^2 = 5/9$. Согласно геометрическому определению, искомая вероятность встречи равна отношению площади фигуры P, образуемой точками, "благоприятствующими" встрече, к площади квадрата Q, символизирующего все пары моментов возможного прихода лиц к месту встречи. Итак, вероятность встречи равна $5/9$.

§ 1.3. Формулы сложения вероятностей

Объединением или суммой событий A и B называется событие, состоящее в осуществлении хотя бы одного из этих событий. Сумма событий A и B обозначается символом $A + B$. Аналогично определяется сумма нескольких событий. Символ $\sum_{k=1}^{n} H_k$ обозначает сумму событий $H_1, H_2, \cdots, H_n$.

* Бросается игральная кость. Как и ранее, пусть $H_1, H_2, \cdots, H_n$ – события, означающие выпадение $1, 2, \cdots, 6$ очков, соответственно. Тогда событие A, состоящее в выпадении четного числа очков, равно сумме $H_2 + H_4 + H_6$. Событие $B = H_3 + H_6$ состоит в выпадении числа очков, кратного 3.

Заметим, что для любых трех событий A, B, C верны равенства:

$$A + A = A,\ A + B = B + A,\ (A + B) + C = A + (B + C) = A + B + C.$$

Предположим, что события A и B несовместны. Пусть в серии из n испытаний событие A наступило k_1 раз, событие B наступило k_2 раз. Тогда в силу несовместности A и B событие $A + B$ в этой серии испытаний произойдет $k_1 + k_2$ раз. Относительные частоты $\nu(A+B), \nu(A)$ и $\nu(B)$ связаны соотношением

$$\frac{k_1 + k_2}{n} = \frac{k_1}{n} + \frac{k_2}{n}.$$

Отсюда следует, что верна следующая *формула сложения вероятностей двух несовместных событий*:

$$P(A+B) = P(A) + P(B). \tag{1}$$

Формула сложения вероятностей для любого конечного числа попарно несовместных событий $A_1, A_2, \cdots, A_n$ имеет вид

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n). \tag{2}$$

* Пусть в урне имеется N шаров, среди них k_1 белых, k_2 черных, остальные синие. Пусть A – появление белого шара, B – появление черного шара, C – появление синего шара при извлечении из урны после перемешивания. Эти события попарно несовместны. Событие $\overline{C}$ – появление белого либо черного шара очевидно равно $A + B$. Стало быть, $p(\overline{C}) = P(A) + P(B) = (k_1 + k_2)/n$.

Если события A и B совместны, то формула сложения вероятностей в выписанной форме не верна.

* В предыдущем примере с игральной костью имеем $P(A) = 1/2$, $p(B) = 1/3$, $P(A+B) = 2/3$, $P(A)+P(B) = 5/6$, значит, $P(A)+P(B) \neq P(A+B)$. Здесь A и B – совместные события.

§ 1.4. Произведение событий. Теорема о вероятности суммы совместных событий

Произведением или совмещением событий A и B называется событие AB, состоящее в наступлении обоих событий A и B в одном испытании. Аналогично определяется произведение любого числа событий.

* Бросается игральная кость. A – выпадение четного числа очков, B – выпадение числа очков, кратного 3. Тогда AB – выпадение 6 очков.

Пусть A и B – какие-либо события. Если событие A произошло в данном испытании, то одновременно с ним наступит событие B или $\overline{B}$. Символически $A = A(B+\overline{B}) = A\,B + A\,\overline{B}$. Аналогично $B = B\,A + B\,\overline{A}$. Событие $A+B$ можно представить как $A\,B + A\,\overline{B} + B\,\overline{A}$, так как $A\,B + A\,B = A\,B$. Учитывая попарную несовместность событий $A\,B$, $A\,\overline{B}$, $\overline{A}\,B$ и пользуясь формулой сложения вероятностей таких событий, получим

$$P(A+B) = P(A\,B + A\,\overline{B} + B\,\overline{A}) = P(A\,B) + P(A\,\overline{B}) + P(B\,\overline{A}).$$

Стало быть, *формула сложения вероятностей для двух событий* имеет вид
$$P(A+B)=P(A)+P(B)-P(A\,B). \tag{3}$$
В случае, когда события A и B несовместны (не могут произойти в одном испытании), $P(A\,B)=0$ и получается уже известная нам формула (1) из § 3.

* В рассмотренном выше примере с игральной костью имеем
$$P(A+B)=1/2+\ 1/3-\ 1/6=2/3.$$
Для суммы большего числа событий общая формула сложения вероятностей имеет более сложный вид. Например, для трех событий:
$$P(A+B+C)=P(A)+P(B)+P(C)-$$
$$-P(A\,B)-P(B\,C)-P(A\,C)+P(A\,B\,C).$$

§ 1.5. Условные вероятности. Формула умножения вероятностей

Пусть A, B – два события. Наступление в данном испытании события A может повлиять на вероятность наступления в том же испытании события B.

* Бросается игральная кость. A – выпадение четного числа очков, B – выпадение четырех очков. Очевидно, что $P(B)=1/6$. Но если оказалось, что в испытании произошло событие A, то вероятность B станет равной 1/3. Если же в испытании произошло событие $\bar{A}$, то вероятность события B станет равной нулю.

Вероятность события B при условии наступления события A называется его *условной вероятностью и обозначается символом* $P(B/A)$. В предыдущем примере $P(B/A)=1/3$, $\ P(B/\bar{A})=0$.

Чтобы получить формулу для вероятности $P(A\,B)$ произведения двух событий, рассмотрим следующий пример.

* В урне N шаров: M белых, остальные черные. Некоторые шары имеют трещины, таких шаров K, из них L белых. Из урны извлекается наудачу один шар. Пусть A – появление белого шара, B – появление шара с трещиной. Тогда $P(A)=M/N$, $\ P(B)=K/N$, $\ P(A\,B)=L/N$.

Предположим, что извлеченный шар оказался белым и вычислим вероятность $P(B/A)$ того, что он имеет трещины. Так как всего белых шаров M, а трещины имеют L из них, то

$$P(B/A) = L/M.$$

Так как

$$L/N = M/N \;\; L/M, \quad L/N = K/N \;\; L/K,$$

то получим

$$P(A\,B) = P(A)\,P(B/A) \quad \text{или} \;\; P(A\,B) = P(B)\;P(A/B). \tag{4}$$

Формулы умножения вероятностей (4) справедливы для любых двух событий A и B, хотя здесь они были получены на примере.

Для трех событий формула умножения имеет вид

$$P(A\,B\,C) = P(A)\,P(B/A)\,P(C/A\,B). \tag{5}$$

Ее можно получить следующим образом. Обозначим событие $A\,B$ через D. Вычисления дают: $P(A\,B\,C) = P(D\,C) = P(D)\,P(C/D) = P(A\,B)\cdot$ $\cdot P(C/A\,B)$. Подставляя в последнее выражение вместо $P(A\,B)$ правую часть первой формулы (4), получим формулу (5).

Обобщение формулы умножения на конечное число событий имеет вид

$$P(A_1\,A_2\cdots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1\,A_2)\cdots P(A_n/A_1\cdots A_{n-1}).$$

* Имеется 10 однотипных деталей, среди них 2 дефектные. Последовательно наудачу и без возвращения берутся 4 детали. Найти вероятность того, что все 4 детали не имеют дефектов.

Пусть A_k – событие, состоящее в том, что взятая при k–м выборе деталь окажется бездефектной. Рассматривая выбор четырех деталей подряд как одно испытание, запишем интересующее нас событие как $A_1\,A_2\,A_3\,A_4$. Применим формулу умножения

$$P(A_1\,A_2\,A_3\,A_4) = P(A_1)P(A_2/A_1)P(A_3/A_1\,A_2)\,P(A_4/A_1\,A_2\,A_3).$$

Вычисления дают:

$$P(A_1) = 8/10, \;\; P(A_2/A_1) = 7/9, \;\; P(A_3/A_1\,A_2) = 6/8,$$

$$P(A_4/A_1\,A_2\,A_3) = 5/7.$$

Стало быть,

$$P(A_1\,A_2\,A_3\,A_4) = (8/10)\cdot(7/9)\cdot(6/8)\cdot(5/7) = 1/3.$$

* Предыдущую задачу можно решить иначе. Способов выбора четырех деталей из десяти имеется всего C_{10}^4 (число сочетаний из 10 по 4). Из них имеется C_8^4 способов выбора четырех бездефектных деталей. Поскольку все выборы равновероятны, то согласно классическому определению искомая вероятность равна

$$P(A_1\,A_2\,A_3\,A_4) = \frac{C_8^4}{C_{10}^4} = \frac{8\cdot 7\cdot 6\cdot 5}{4!} : \frac{10\cdot 9\cdot 8\cdot 7}{4!} = 1/3.$$

§ **1.6**. **Независимые события**

Из формул (4) предыдущего параграфа находим

$$P(A)P(B/A) = P(B)P(A/B).$$

Рассмотрим случай, когда $P(B/A) = P(B)$. При $P(B) \neq 0$ получим $P(A/B) = P(A)$. Это можно интерпретировать так, что вероятность события B не зависит от наступления события A, а вероятность события A не зависит от наступления события B. Для таких событий A, B формулы (4) принимают вид

$$P(A\,B) = P(A)P(B). \tag{6}$$

События A, B называются *независимыми*, если для них справедлива формула (6).

Если вероятность каждого из событий $A_1,\,A_2,\,\cdots,A_n$ не зависит от реализации любой комбинации остальных событий, то формула умножения для любых из этих событий принимает вид

$$P(A_i\,A_j\,\cdots A_s) = P(A_i)P(A_j)\cdots P(A_s). \tag{7}$$

События $A_1,\;A_2,\;\cdots,A_n$ называются *независимыми*, если справедливы все формулы вида (7) для произведений любой выборки из этих n событий.

* Бросается игральная кость. Пусть A – выпадение четного числа очков, B – выпадения числа очков, делящегося на 3. Тогда $A\,B$ – выпадение шестерки. Вычисления дают:

$$P(A) = 1/2,\;\; P(B) = 1/3,\;\; P(A\,B) = 1/6 = P(A)P(B).$$

Стало быть, в согласии с определением, события A и B независимы.

* Два стрелка производят по одному выстрелу по мишени. Поражение мишени с одного выстрела первым стрелком (событие H_1) имеет вероятность 0.9 и соответственно вторым стрелком (событие H_2) – 0.8. Считаем, что стрелки никак не влияют друг на друга и качество стрельбы одного не зависит от качества стрельбы другого. Это обстоятельство означает независимость событий H_1 и H_2. Вероятность того, что оба стрелка поразят мишень с первого раза, равна $P(H_1\,H_2) = P(H_1)P(H_2) = 0.9 \cdot 0.8 = 0.72$.

Если одно из событий A, B невозможно либо достоверно, то A и B – независимые события и в этих условиях формула (7) справедлива.

Если события A, B независимы, то независимыми будут также пары событий A, $\bar{B}$, $\bar{A}$, B, $\bar{A}$, $\bar{B}$. Покажем, например, независимость событий A, $\bar{B}$. Имеем $A = A\,B + A\,\bar{B}$, поэтому

$$P(A) = P(A\,B) + P(A\,\bar{B}) = P(A)P(B) + P(A\,\bar{B}),$$

следовательно,

$$P(A\,\bar{B}) = P(A)(1 - P(B)) = P(A)P(\bar{B}).$$

Таким образом, для рассматриваемой пары событий A, $\bar{B}$ имеет место формула (7), что означает независимость этих событий. Рекомендуется аналогичным путем проверить независимость пар событий $\bar{A}$, B и $\bar{A}$, $\bar{B}$.

* Продолжим анализ примера с двумя стрелками. Вероятность того, что у обоих первые выстрелы дадут промахи, равна

$$P(\bar{H}_1\,\bar{H}_2) = P(\bar{H}_1)P(\bar{H}_2) = 0.1 \cdot 0.2 = 0.02.$$

Вероятность того, что один из стрелков промахнется (неважно какой), а другой поразит цель, равна

$$P(H_1\,\bar{H}_2 + \bar{H}_1\,H_2) = P(H_1\,\bar{H}_2) + P(\bar{H}_1\,H_2) =$$

$$P(H_1)P(\bar{H}_2) + P(\bar{H}_1)P(H_2) = 0.9 \cdot 0.2 + 0.1 \cdot 0.8 = 0.26.$$

События $A + B$ и $\bar{A}\,\bar{B}$ являются противоположными, поэтому

$$P(A + B) = 1 - P(\bar{A}\,\bar{B}).$$

Если события A и B независимы, то формула примет вид

$$P(A + B) = 1 - P(\bar{A})\,P(\bar{B}).$$

Эта формула по индукции обобщается на любое конечное число независимых событий $A_1,\ A_2,\ \cdots, A_n$:

$$P(A_1 + A_2 + \cdots + A_n) = 1 - P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n). \qquad (8)$$

* Вернемся к рассматриваемому примеру. Вероятность поражения мишени хотя бы одним стрелком равна

$$P(H_1 + H_2) = 1 - P(\bar{H}_1)P(\bar{H}_2) = 1 - 0.1 \cdot 0.2 = 0.98.$$

Вероятность того, что хотя бы один стрелок промахнется, равна

$$P(\bar{H}_1 + \bar{H}_2) = 1 - P(H_1)P(H_2) = 1 - 0.9 \cdot 0.8 = 0.28.$$

§ 1.7. Формула полной вероятности. Формулы Байеса

События $H_1,\ H_2,\ \cdots, H_n$ образуют *полную систему событий*, если они попарно несовместны, а их сумма есть достоверное событие, то есть при каждом испытании происходит хотя бы одно из этих событий. Два события образуют полную систему тогда и только тогда, когда они взаимно противоположны.

* Рассмотрим испытание – бросание игральной кости. Возможные исходы испытания – выпадение одного, двух $\cdots$ шести очков – обозначим символами $H_1, H_2,\ \cdots,\ H_6$ соответственно. Система событий $H_1,\ H_2,\ \cdots,$ H_6 является полной. Здесь можно выделить и другие полные системы событий, например, $H_1 + H_2,\ H_3 + H_4,\ H_5 + H_6$.

Пусть A – некоторое событие, а $H_1,\ H_2,\ \cdots, H_n$ – полная система событий. Тогда

$$A = A\,H_1 + A\,H_2 + \cdots + A\,H_n,$$

так как вместе с A должно произойти одно из событий $H_1,\ H_2,\ \cdots, H_n$. Слагаемые в правой части формулы – несовместные события, поэтому применение правил сложения и умножения вероятностей приводит к следующей *формуле полной вероятности*:

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i). \qquad (9)$$

* Однотипные детали штампуются на трех станках и складываются в одну кучу. Первый станок дает 40 % всей продукции, второй – 35 %, третий – 25 %. На первом станке брак составляет 8 %, на втором – 4 %,

на третьем – 10 %. Из кучи наудачу извлекается одна деталь. Найдем вероятность того, что она годная.

Символами H_1, H_2, H_3 обозначим соответственно события, состоящие в том, что деталь изготовлена первым, вторым и третьим станком. Имеем полную систему событий. Применяя формулу полной вероятности, получим

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3) =$$

$$= 0.4 \cdot 0.92 + 0.35 \cdot 0.96 + 0.25 \cdot 0.9 = 0.929.$$

Пусть A – случайное событие с отличной от нуля вероятностью и H_1, H_2, $\cdots, H_n$ – полная система событий. Из формулы умножения вытекают равенства

$$P(H_k)P(A/H_k) = P(A)P(H_k/A), \;\; k = 1, \cdots, n.$$

Следовательно,

$$P(H_k/A) = \frac{P(H_k)P(A/H_k)}{P(A)}, \;\; k = 1, \cdots, n.$$

С учетом формулы полной вероятности приходим к *формулам Байеса*

$$P(H_k/A) = \frac{P(H_k)P(A/H_k)}{\sum_{i=1}^{n} \; P(H_i)P(A/H_i)}, \;\; k = 1, \cdots, n. \qquad (10)$$

Гипотезами (Hypothese) принято называть события H_k из полной системы, а их вероятности $P(H_k)$, вычисленные до испытания, повлекшего наступление события A, – *априорными вероятностями гипотез.* Стоящие в левых частях формул (10) величины называют тогда *вероятностями гипотез после испытания или апостериорными вероятностями гипотез.* Термины происходят от латинских выражений a priori – до опыта, a posteriori – после опыта. Формулы (10) служат для вычисления апостериорных вероятностей гипотез.

* Из кучи белых и черных шаров некто положил в пустую урну 2 шара. О содержимом урны можно высказать три взаимно исключающие друг друга гипотезы:

H_1 – 0 белых, 2 черных шара;

H_2 – 1 белый, 1 черный шар;

H_2 – 2 белых, 0 черных шаров.

Испытание заключается в извлечении наудачу одного шара из урны. До этого испытания вероятности гипотез можно считать одинаковыми, равными $1/3$, поскольку нам не известны соображения, которыми руководствовался человек, положивший шары в урну.

Пусть в результате испытания извлеченный из урны шар оказался белым (наступило событие A). Вычисления дают:

$$P(H_1)P(A/H_1) = \frac{1}{3} \cdot 0 = 0,$$

$$P(H_2)P(A/H_2) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6},$$

$$P(H_3)P(A/H_3) = \frac{1}{3} \cdot 1 = \frac{1}{3}.$$

По формуле полной вероятности находим $P(A) = 0 + 1/6 + 1/3 = 1/2$. Применив формулы Байеса, будем иметь

$$P(H_1/A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0}{1/2} = 0,$$

$$P(H_2/A) = \frac{P(H_2)P(A/H_2)}{P(A)} = \frac{1/6}{1/2} = \frac{1}{3},$$

$$P(H_3/A) = \frac{P(H_3)P(A/H_3)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

* В условиях рассмотренной в начале параграфа задачи о станках, штампующих детали, вероятности гипотез о том, что детали изготовлены на первом, втором и третьем станках, таковы:

$$P(H_1) = 0.4, \ \ P(H_2) = 0.35, \ \ P(H_3) = 0.25.$$

Из кучи выбрали случайным образом одну деталь, которая оказалась дефектной (произошло событие $\bar{A}$). Подсчитаем апостериорные вероятности гипотез, то есть вероятности $P(H_1/A)$,
$P(H_2/A)$, $P(H_3/A)$ того, что выбранная деталь была изготовлена на первом, втором, третьем станке. Вычисления дают:

$$P(\bar{A}) = 1 - P(A) = 0.071,$$

$$P(H_1/\bar{A}) = \frac{P(H_1)P(\bar{A}/H_1)}{P(\bar{A})} = \frac{0.4 \cdot 0.08}{0.071} = 0.451,$$

$$P(H_2/\bar{A}) = \frac{P(H_2)P(\bar{A}/H_2)}{P(\bar{A})} = \frac{0.35 \cdot 0.04}{0.071} = 0.197,$$

$$P(H_3/\bar{A}) = \frac{P(H_3)P(\bar{A}/H_3)}{P(\bar{A})} = \frac{0.25 \cdot 0.1}{0.071} = 0.352.$$

Современная медицина располагает обширными банками данных о болезнях, статистике их распространенности, динамике, об индивидуальных больных. Эти обстоятельства делают популярными применения формул Байеса в задачах диагностики, динамики эпидемий и т.д.

§ **1.8. Энтропия и информация**

Пусть имеется полная система S событий $A_1,\ A_2,\ \cdots,A_n$ и известны (априорные) вероятности $P(A_k) = p_k$ этих событий, $k = 1,\ 2, \cdots, n$. Важной числовой характеристикой этой системы событий, мерой неопределенности описываемой ею ситуации является *энтропия* $H(S) = - \sum_{k=1}^{n} p_k \log_2 p_k$, где $\log_2 p$ – логарифм числа p по основанию 2. Это понятие ввел американский математик К. Шеннон.

* Пусть O – выпадение орла, $\bar{O}$ – выпадение решки при однократном бросании монеты, причем вероятности этих событий равны $1/2$. Энтропия системы $S = \{O, \bar{O}\}$ определяется по формуле

$$H(S) = -(1/2)\log_2(1/2) - (1/2)\log_2(1/2) = \log_2 2 = 1.$$

* В задаче о станках из предыдущего параграфа вычислим энтропию полной системы $S = \{H_1,\ H_2,\ H_3\}$. Получим

$$H(S) = -0.4\ \log_2 0.4 - 0.35\ \log_2 0.35 - 0.25\ \log_2 0.25 \approx 1.414.$$

С понятием теоретико-вероятностной энтропии тесно связано понятие количества *информации*, которое несет сообщение о наступлении того или иного случайного события, совместимого с какими-либо событиями полной системы.

Пусть снова имеется полная система S событий $A_1,\ A_2,\ \cdots,\ A_n$ и известны (априорные) вероятности $P(A_k) = p_k$ этих событий, $k = 1,\ 2, \cdots n$. Далее, поступило сообщение о появлении события B, причем априорные вероятности $P(B\,/A_k) = q_k$, $k = 1,\ 2,\ \cdots n$, также были известны. По формулам Байеса найдем апостериорные вероятности $P(A_k/B) = q_k$, $k = 1,\ 2, \cdots n$, а затем вычислим энтропию полной системы A_1/B, A_2/B, $\cdots$, A_n/B по формуле

$$H(S/B) = -\sum_{i=1}^{n} q_k \log_2 q_k.$$

В описанной модельной ситуации разность $H(S) - H(S/B)$ обозначается символом $I(B)$ и называется *количеством информации* (по Шеннону), которое несет сообщение о наступлении события B. При вычислении двоичного логарифма числа p рекомендуется пользоваться формулой $\log_2 p = \ln p/\ln 2$ – отношению натуральных логарифмов, где $\ln 2 \approx 0.693$. Заметим также, что функция $h(p) = -p\log_2 p$ при $p = 1$ равна нулю. При $p = 0$ ее полагают равной нулю. Эти факты полезно иметь в виду в рассматриваемых вопросах.

* Перед бросанием монеты энтропия полной системы $S = \{A_1, A_2\}$, где $A_1 = O$ – выпадение орла, $A_2 = \bar{O}$ – выпадение решки, как было показано выше, равна 1. Пусть результатом испытания стало появление орла. Ситуация вполне определена: $P(A_1/O) = 1,\ \ P(A_2/O) = 0$, и апостериорная энтропия $H(S/O)$ оказалась равной нулю в силу замечания. Итак, количество информации, которое несет сообщение в выпадении орла, равно $H(S) - H(S/O) = 1 - 0 = 1$.

Единица энтропии и информации безразмерна и носит название 1 бит.

* Перед бросанием игральной кости имеем полную систему событий $S = \{A_1, A_2, \cdots, A_6\}$, где A_m – выпадение m очков при однократном бросании кости, $m = 1, 2, \cdots, 6$. Так как все вероятности $P(A_m) = 1/6$, то энтропия системы $H(S)$ равна $6\cdot(-1/6)\cdot\log_2(1/6) \approx 2.585$. После однократного бросания кости поступило сообщение B, что число выпавших очков оказалось четным. Теперь условные вероятности $P(A_m/B)$ равны нулю при нечетных m и $1/3$ при четных m, так как есть три варианта появления четного числа очков и ни одному из них нельзя отдать предпочтение. Апостериорная энтропия $H(S/B) = 3\cdot\ (-(1/3)\cdot\log_2(1/3) \approx 1.585$. Количество информации, даваемое сообщением о наступлении события B, равно $H(S) - H(S/B) = 1$.

При построении вероятностных моделей реальных случайных событий важно выделить практически достоверные и практически невозможные события. Вероятности первых столь близки к единице, что в модели их объявляют достоверными, а вероятности вторых столь близки к нулю, что в модели их объявляют невозможными. Наступление практически достоверного события несет несущественное количество информации. Напротив, наступление практически невозможного события несет большое количество информации. Например, катастрофа самолета данной серии, не исчерпавшего свой ресурс, – событие с очень малой вероятностью, и информация о такой катастрофе даже с технической и социальной точек зрения крайне важна. Информация о наступлении события, никак не связанного с S, не зависимого от событий этой системы, не несет информации в рамках вероятностной модели, основанной на использовании полной системы S случайных событий.

При построении и совершенствовании вероятностных моделей реальных явлений широко применяются методы математической статистики, о которых подробнее будет сказано в главе 4. Здесь также важны опыт и интуиция инженера, исследователя.

Понятия энтропии и информации играют существенную роль в статистической механике и во многих важных задачах кодирования и теории связи.

§ 1.9. Схемы Бернулли

Пусть A – случайное событие, $P(A) = p$ – его вероятность. Рассмотрим сложное испытание, состоящее из последовательности (серии) n элементарных независимых испытаний, в каждом из которых вероятность наступления события A постоянна и равна p. Исход такого сложного испытания, называемого *серией испытаний Бернулли*, можно представить в виде произведения $B_n = A\,A\,\bar{A}\,A\,\bar{A}\cdots A$, в котором n сомножителей. Некоторые сомножители равны A, остальные равны $\bar{A}$. Наличие на k-й позиции этого произведения символа A означает наступление события A в k-м элементарном испытании. В противном случае на этой позиции должен стоять символ $\bar{A}$. Предположим, что в записи события B_n символ A встречается m раз, а символ $\bar{A}$ – $n - m$ раз. Тогда по теореме умножения вероятность $P(B_n)$ равна произведению

$$p \cdot p \cdot (1-p) \cdot (1-p) \cdots p = p^m (1-p)^{n-m}.$$

Из комбинаторики известно, что число таких исходов серии испытаний Бернулли равно числу способов разместить m символов A на n позициях, то есть числу сочетаний $C_n^m = n!/(m! \cdot (n-m)!)$. Стало быть, вероятность $P_n(m)$ события, заключающегося в m-кратном наступлении события A в серии из n испытаний Бернулли равна сумме вероятностей всех C_n^m образующих его исходов, то есть

$$P_n(m) = C_n^m \cdot p^m (1-p)^{n-m}. \tag{11}$$

Формула (11) получена швейцарским математиком Я. Бернулли – одним из основоположников теории вероятностей.

Заметим, что появление события A в испытаниях Бернулли часто называют успехом, а появление противоположного события – неудачей (промахом). Таким образом, вероятность появления m успехов в серии из n испытаний Бернулли вычисляется по формуле (11).

* Вероятность того, что стрелок с одного выстрела попадет в мишень, равна 0.9. Стрелок производит 5 выстрелов. Вероятность того, что мишень будет поражена 3 раза, равна

$$P_5(3) = C_5^3 \cdot 0.9^3 \cdot 0.1^2 = 0.0729.$$

Далее, вероятность поражения мишени не менее трех раз определяется следующим образом:

$$P_5(m \geqslant 3) = P_5(3) + P_5(4) + P_5(5) =$$

$$= 0.0729 + 0.32805 + 0.59049 = 0.99144.$$

* Рассмотрим отбор четырех деталей в задаче § 1.6. Вероятность того, что среди отобранных деталей окажутся 3 годные, такова:

$$P_4(3) = C_4^3\, 0.8^3 \cdot 0.2 = 0.4096.$$

Биномиальная формула (11) часто встречается в биостатистике и генетике.

* Некоторая тропическая рыба может быть серого или золотого цвета. Цвет передается генами, находящимися в хромосомах. В хромосомах мужской и женской особей имеется по одному гену: С (серый цвет) или З (золотой цвет). Таким образом, каждое оплодотворенное яйцо имеет два гена: СС, СЗ, ЗЗ. Золотой цвет рыба будет иметь только при комбинации ЗЗ. Наличие же в комбинации гена С означает, что рыба будет иметь серый цвет. В такой ситуации ген С называется доминирующим.

Потомство родительской пары с генами СС и ЗЗ наследует лишь комбинацию СЗ и будет иметь серый цвет. Если же родительская пара имеет наборы ген ЗС и ЗЗ, то потомство может унаследовать одну из комбинаций ЗЗ или ЗС, то есть в половине случаев будет иметь золотой цвет. Наконец, если оба родителя имеют набор ген ЗС, то у потомства возможны комбинации СС, СЗ, ЗС, ЗЗ. Комбинации СЗ и ЗС равносильны. Вероятность появления золотой рыбки у такой пары равна 1/4, вероятность появления серой рыбки равна 3/4. В среднем половина потомства будет иметь генный набор ЗС, такой же, как у родителей.

Пусть имеется 10 оплодотворенных икринок от родительской пары с генным набором ЗС у каждого. Назовем успехом рождение рыбки золотого цвета. Тогда вероятность появления в данном поколении трех рыбок золотого цвета равна $C_{10}^3\,(1/4)^3\,(3/4)^7 \approx 0.2503$.

§ 1.10. Задачи для самостоятельного решения

1. Пусть A, B, C – три произвольных события. Найти выражения для событий, состоящих в том, что из A, B, C :

а) произошло только A;
б) произошли A и B, а C не произошло;
в) все три события произошли;
г) произошло по крайней мере одно из событий;
д) ни одно из событий не произошло.

2. Игральная кость брошена два раза. Найти вероятность:

а) того, что в обоих случаях выпадет по одному очку;
б) что сумма выпавших очков будет равна 11;
в) что сумма выпавших очков будет не менее 10;
г) что выпавшие очки будут кратны 3.

3. Разрыв электрической цепи может произойти вследствие выхода из строя элемента K или двух элементов K_1, K_2, которые выходят из строя независимо друг от друга с вероятностями 0.3, 0.2 и 0.2 соответственно. Найти вероятность разрыва цепи.

4. В некотором семействе 5 детей. Вероятность рождения мальчика – 0.51, а вероятность рождения девочки – 0.49. Найти вероятность того, что в семье :

а) три девочки; б) не менее трех мальчиков.

5. Деталь с вероятностью 0.01 имеет дефект A, с вероятностью 0.02 имеет дефект B и с вероятностью 0.005 имеет оба дефекта. Найти вероятность того, что деталь имеет хотя бы один дефект.

6. В партии изделий 90 исправных и 10 бракованных изделий. Найти вероятность того, что среди 10 проданных изделий:

а) ровно одно бракованное;
б) нет бракованных.

7. Вероятность появления в потомстве черного кролика равна 1/4. Найти вероятность появления не менее двух черных кроликов в потомстве из четырех кроликов.

8. Найти вероятность того, что дни рождения 12 человек придутся на разные месяцы года.

9. Определить вероятность того, что в партии из 1000 электрических лампочек нет ни одной неисправной, если из взятых наудачу 100 лампочек все оказались исправными. Предполагается, что число неисправных лампочек из 1000 равновозможно от 0 до 5. Указание: применить формулу Байеса.

10. Известно, что 96 % выпускаемой продукции удовлетворяет стандарту. Упрощенная схема контроля признает пригодной стандартную продукцию с вероятностью 0.98, а нестандартную – с вероятностью 0.05. Найти вероятность того, что изделие, прошедшее упрощенный контроль, удовлетворяет стандарту.

11. Рассмотрим вместо игральной кости тетраэдр Т – правильную треугольную пирамиду с равными гранями, причем три грани окрашены в красный, зеленый и синий цвета соответственно, а на четвертой грани присутствуют все три цвета и только они. Пусть тетраэдр изготовлен из однородного материала, так что при его подбрасывании Т с равной вероятностью $p = 1/4$ падает на любую из граней. Символами К, З, С обозначим случайные события – Т падает на грань, содержащей красный, зеленый, синий цвет соответственно.

а) Вычислить вероятности событий $К \cdot С$, $К \cdot З$, $З \cdot С$, $К \cdot З \cdot С$.

б) Показать, что события К, З, С попарно независимы, но зависимы в совокупности.

12. В некотором эксперименте в каждый из 320 сосудов было помещено по 5 насекомых. После кратковременного их окуривания было подсчитано число насекомых, оставшихся в живых. Оказалось, что в 81 сосуде погибли все насекомые, в 122 сосудах оставалось по одному живому, в 88 – по два, в 21 – по три, в 6 – по четыре насекомых и лишь в двух сосудах выжили все. Какова вероятность того, что в результате такого эксперимента насекомое выживет?

13. В некотором эксперименте на определение всхожести семена были высеяны по 4 в ряд, и были получены следующие результаты.

Число проросших семян в ряду	0	1	2	3	4
Число рядов	0	3	27	43	27

Какова вероятность прорастания (всхожесть) семян?

14. Некоторое изделие может поступить для обработки в случайном порядке на один из трех станков с вероятностями p_1, p_2 и p_3, соответственно равными 0.2, 0.3 и 0.5. При обработке на первом станке вероятность брака равна 0.02, на втором станке – 0.03 и на третьем станке – 0.05. Найти вероятность того, что поступившее в цех изделие после обработки окажется удовлетворяющим техническим условиям.

15. Вычислить количество информации, которое несет сообщение о появлении белого шара в задаче о шарах из § 1.7.

16. Что вероятнее выиграть у равносильного противника:

а) три партии из четырех или пять из восьми,

б) не менее трех партий из четырех или не менее пяти партий из восьми?

Под равносильным противником понимается такой, вероятность выигрыша у которого в каждой партии равна 1/2. Ничьи исключаются.

17. Среди 64 клеток шахматной доски выбирают наудачу две клетки и на них ставят две ладьи – белую и черную. Какова вероятность того, что ладьи не будут бить друг друга?

18. Два стрелка стреляют по одной мишени, делая по одному выстрелу. Вероятность попадания в мишень у первого – 0.8, у второго – 0.4. После стрельбы в мишени обнаружена одна пробоина. Какова вероятность того, что в мишень попал первый стрелок?

19. Грегор Мендель – основатель генетики – обнаружил, что семена сладкого горошка имеют разный цвет – желтый и зеленый, причем желтый цвет доминирует. Пусть Ж и З – гены, ответственные за желтый и зеленый цвет у горошины. Горошина с генным набором ЖЗ проросла, зацвела и самоопылилась. Какова ожидаемая доля желтых горошин в урожае? Найти вероятность того, что все 5 горошин в стручке из этого урожая зеленые.

20. Мендель также обнаружил, что гладкость или морщинистость поверхности горошины определяется соответствующими генами. Гладкие горошины преобладали. Таким образом, пыльца растения характеризуется наличием гена Г (гладкость поверхности горошины) или гена М (морщинистость поверхности горошины). Символически можно выписать четыре комбинации генов в пыльце: ЖГ, ЖМ, ЗГ, ЗМ. Желтая гладкая горошина с генной структурой ЖЗМГ проросла, зацвела и была опылена пыльцой типа ЖМ. Собранный урожай помещен в урну. Найти вероятность извлечения из урны желтой морщинистой горошины.

Глава II

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 2.1. Распределение вероятностей дискретной случайной величины

Величина, которая в каждом испытании определенного типа принимает какое-нибудь действительное или комплексное значение (вообще говоря, разное в разных испытаниях), называется *случайной величиной (с.в.)*. Случайные величины обозначают обычно греческими буквами $\xi, \eta, \zeta, \varphi, \psi, \omega$ и т.д., а принимаемые ими значения – соответствующими латинскими буквами, возможно, снабженными индексами.

* Количество очков, выпавших на игральной кости, есть с.в. ξ, принимающая значения $x_k = k, \quad k = 1,\ 2,\ 3,\ 4,\ 5,\ 6$.

* Число появлений события A в одном испытании есть с.в., которая принимает два значения 0, 1. Она называется *индикатором с.в. A* и будет впредь обозначаться символом η_A.

* Относительная частота наступления события A в серии из n испытаний Бернулли есть с.в. θ, принимающая значения $t_k = k/n, \quad k = 0,\ 1,\ \cdots, n$.

* Суточное количество автомобильных аварий в г. Москве есть с.в. ζ, принимающая неотрицательные целые значения.

* Количество километров, пройденных данным автомобилем до капитального ремонта, есть с.в. ω, принимающая любые положительные значения в довольно широком интервале на числовой прямой.

* Процент углерода в стали определенной марки есть с.в.ψ, принимающая любые значения в некотором интервале, предусмотренном стандартом.

Случайная величина называется *дискретной*, если ее значения образуют конечную или бесконечную последовательность

$$x_1,\ x_2,\ \cdots, x_n,\ \cdots . \tag{1}$$

В приведенных выше примерах первые четыре с.в. являются дискретными, а остальные две с.в. не дискретны.

Пусть ξ – дискретная с.в. с множеством значений (1). Символом A_k обозначим событие, заключающееся в том, что ξ принимает значение x_k:

$$A_k = \{\xi = x_k\}, \quad k = 1,\, 2, \cdots .$$

Вероятность события A_k условимся обозначать символом p_k. События $A_1,\, A_2, \cdots$ образуют полную систему событий. Стало быть, сумма их вероятностей равна 1. В случае бесконечного числа событий A_k под суммой их вероятностей понимается сумма числового ряда $\sum_1^\infty p_k$. Таким образом, дискретная с.в. с бесконечным множеством значений такова, что $\sum_1^\infty p_k = 1$.

Если последовательность (1) конечная, то можно составить следующую *таблицу распределения вероятностей* или *закон распределения* дискретной с.в. ξ:

$$\begin{array}{cccccc} \xi & : & x_1 & x_2 & \cdots & x_n, \\ p & : & p_1 & p_2 & \cdots & p_n. \end{array}$$

* Пусть ξ – с.в. из примера с игральной костью. Таблица распределения вероятностей ξ имеет вид

$$\begin{array}{cccccccc} \xi & : & 1 & 2 & 3 & 4 & 5 & 6, \\ p & : & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6}. \end{array}$$

* Пусть η_A – индикатор события A. Таблица распределения вероятностей η_A имеет вид

$$\begin{array}{cccc} \eta_A : & 0 & 1, \\ P : & 1-p & p, \end{array}$$

где p=P(A) – вероятность появления события A.

* Из теоремы Я. Бернулли следует, что закон распределения с.в. θ из приведенного выше третьего примера представляется в виде

$$\begin{array}{ccccccc} \theta : & 0 & 1/n & 2/n & \cdots & \cdots & 1, \\ P : & q^n & C_n^1 p\, q^{n-1} & C_n^2 p^2 q^{n-2} & \cdots & & p^n, \end{array}$$

где $p = P(A)$ – вероятность успеха, $q = 1 - p$. Такая с.в. называется *биномиальной.*

Постоянную величину C (константу) можно считать дискретной с.в., принимающей в каждом испытании значение C с вероятностью p=1.

В одном испытании могут появляться значения нескольких с.в.

* При анализе пробы водопроводной воды определяется процентное содержание железа, кальция, хлора, фтора и т.д. Все эти величины – различные с.в. в данном испытании.

Случайные величины $\xi, \eta, \cdots, \zeta$, принимающие свои значения в одном испытании, называются *независимыми*, если вероятности любых значений каждой из этих величин не зависят от того, какие значения принимают остальные с.в. в том же испытании.

* Бросаются три игральные кости, ξ, η, ζ – значения очков, выпавших на первой, второй и третьей кости. Очевидно, с.в. ξ, η, ζ – независимые.

* Пусть η_k – индикатор успеха в k-м испытании Бернулли. Тогда с.в. $\eta_1, \eta_2, \cdots, \eta_n$ независимы.

В дальнейшем нам понадобится понятие функции от с.в. Пусть $h(x)$ – заданная функция, область определения D_h которой содержит все значения с.в. ξ. Символом $\eta = h(\xi)$ обозначается случайная величина, принимающая значение $h(x)$ всякий раз, когда ξ принимает значение x.

* Пусть ξ – с.в., значения которой равны числу выпавших очков при бросании игральной кости, и $h(x) = x^2$. Тогда с.в. $\eta = \xi^2$ принимает каждое из значений 1, 4, 9, 16, 25, 36 с вероятностью $1/6$.

Если данная функция нескольких переменных $g(x, y, \cdots, z)$ определена в области D_g евклидова пространства $\mathbb{R}^n$, содержащей все точки вида $(x_i, y_j, \cdots, z_s)$, где $\{x_i\}$ – значения с.в. ξ, $\{y_j\}$ – значения с.в. η, $\cdots$, $\{z_s\}$ – значения с.в. ζ, то с.в. $\mu = g(\xi, \eta, \cdots, \zeta)$ принимает значение $g(x_i, y_j, \cdots, z_s)$ всякий раз, когда $\xi = x_i, \eta = y_j, \cdots, \zeta = z_s$.

* Пусть $\xi_k, \;\; k = 1, 2, \cdots, n,$ – индикаторы успехов в схеме Бернулли. Тогда

$$\tau = \frac{1}{n}\sum_{1}^{n} \xi_k \;-$$

с.в., равная относительной частоте успеха в серии из n испытаний Бернулли.

§ 2.2. Функции распределения случайных величин. Непрерывные случайные величины

Пусть с.в. ξ принимает действительные значения. *Функция распределения $F(x)$ с.в. ξ* определяется формулой $F(x) = P\{\xi < x\}$, где $x \in \mathbb{R}$. Иными словами, значение этой функции в точке $x \in \mathbb{R}$ равно вероятности события $\{\xi < x\}$. Так как вероятность есть число из промежутка [0, 1], то

$$0 \leqslant F(x) \leqslant 1 \tag{2}$$

для любого $x \in \mathbb{R}$ и график функции распределения любой с.в. ξ лежит в горизонтальной полосе $\{0 \leqslant x \leqslant 1\}$. Кроме того, для $F(x)$ выполняются предельные соотношения

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, \;\; F(+\infty) = \lim_{x \to +\infty} F(x) = 1, \tag{3}$$

так как первый предел дает вероятность того, что с.в. ξ не примет никаких значений (это событие невозможно и имеет нулевую вероятность), а второй предел равен вероятности того, что ξ примет какие угодно значения, что является достоверным событием и имеет вероятность 1.

Пусть $x_1 < x_2$. События $A = \{\xi < x_1\}$ и $B = \{x_1 \leqslant \xi < x_2\}$ несовместны и их сумма $A + B = \{\xi < x_2\}$. Поэтому в силу теоремы сложения вероятностей $P(A + B) = P(A) + P(B)$. Стало быть,

$$P(\xi < x_2) = P(\xi < x_1) + P(x_1 \leqslant \xi < x_2)$$

или

$$P(x_1 \leqslant \xi < x_2) = P(\xi < x_2) - P(\xi < x_1) = F(x_2) - F(x_1). \tag{4}$$

Из (4) следует, что $\Delta F = F(x_2) - F(x_1) \geqslant 0$ при $x_2 > x_1$, то есть $F(x)$ является *неубывающей* функцией на числовой прямой $\mathbb{R}$.

В случае дискретной с.в. имеем $\lim_{x_2 \to x_1} \Delta F = P(\xi = x_1)$. Если окажется, что x_1 принадлежит множеству значений с.в. ξ, то $P(\xi = x_1)$ – положительное число и функция распределения имеет в точке $x = x_1$ *разрыв первого рода со скачком* $P(\xi = x_1)$. Если же x_1 не является значением дискретной с.в. ξ, то функция $F(x)$ непрерывна в точке x_1. Нетрудно еще показать, что $F(x)$ непрерывна справа в каждой точке $x \in \mathbb{R}$. Более того, это свойство справедливо для функции распределения произвольной с.в.

* График функции распределения с.в. η, равной числу очков, выпадающему в результате однократного бросания игральной кости, имеет вид, изображенный на рис. 1.

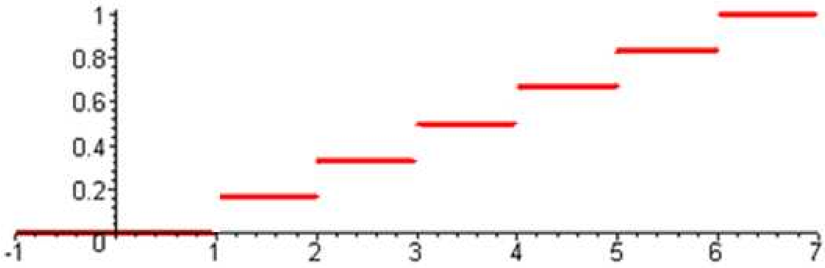

Рис. 1

Случайная величина ξ называется *непрерывной*, если ее функция распределения $F(x)$ непрерывна и

$$F(x) = \int_{-\infty}^{x} F'(t)\,dt. \tag{5}$$

Для непрерывной с.в. функция $f(x) = F'(x)$ называется *дифференциальной функцией распределения вероятностей*. Эта функция существует в точках дифференцируемости функции $F(x)$. По отношению к $f(x)$ функция $F(x)$ называется *интегральной функцией распределения с.в.* ξ.

* Непрерывная с.в. ξ имеет *равномерное распределение вероятностей на промежутке* [a, b], если на этом промежутке $F(x) = (x-a)/(b-a)$ и равна 0 или 1 при $x < a$ и при $x > b$ соответственно. Дифференциальная функция $f(x)$ равна $(b-a)^{-1}$ на интервале (a, b), а вне промежутка [a, b] обращается в ноль. В точках $x = a$ и $x = b$ функция $f(x)$ не существует.

Для непрерывной с.в. ξ вероятность того, что ξ примет какое-нибудь фиксированное значение x, равна нулю:

$$P\{\xi = x\} = \lim_{\Delta x \to 0} P\{x \leqslant \xi < x + \Delta x\} = \lim_{\Delta x \to 0} [F(x + \Delta x) - F(x)] = 0.$$

Последний предел равен нулю в силу непрерывности функции $F(x)$. Поэтому для непрерывной с.в.

$$P\{x_1 \leqslant \xi < x_2\} = P\{x_1 < \xi < x_2\} = P\{x_1 \leqslant \xi \leqslant x_2\} =$$

$$= P\{x_1 < \xi \leqslant x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} F'(t)\,dt = \int_{x_1}^{x_2} f(t)\,dt. \tag{6}$$

Поскольку $F(x)$ – неубывающая функция, то $F'(x) = f(x) \geqslant 0$ во всех точках, где $F(x)$ дифференцируема. Из (5) и (3) имеем

$$\int_{-\infty}^{+\infty} f(x)\,d\,x = \int_{-\infty}^{+\infty} F'(x)\,d\,x = F(+\infty) = 1. \tag{7}$$

Отношение

$$\frac{P\{x \leqslant \xi < x + \Delta\,x\}}{\Delta\,x}$$

называется *средней плотностью распределения вероятности* на промежутке $[x,\, x + \Delta\,x)$. Ее предел при $\Delta\,x \to 0$ называется *плотностью распределения вероятности* в точке x. В тех точках, где этот предел существует, имеем

$$\lim_{\Delta\,x \to 0} \frac{P\{x \leqslant \xi < x + \Delta\,x\}}{\Delta\,x} = \lim_{\Delta\,x \to 0} \frac{F(x + \Delta\,x) - F(x)}{\Delta\,x} = F'(x).$$

Отсюда следует, что для непрерывной с.в. дифференциальная функция распределения вероятностей $f(x)$ равна плотности распределения вероятностей.

§ **2.3**. Математическое ожидание случайной величины

Рассмотрим дискретную с.в. ξ с таблицей распределения вероятностей

$$\begin{aligned} &\xi\ :\ \ x_1\ \ x_2\ \ \cdots\ \ \ x_m,\\ &p\ :\ \ p_1\ \ p_2\ \ \cdots\ \ \ p_m. \end{aligned} \tag{8}$$

Пусть в некоторой серии из n независимых испытаний с.в. ξ приняла

k_1 раз значение x_1,

k_2 раз значение x_2,

....

k_m раз значение x_m.

Сумма всех принятых значений равна $\sum_{i=1}^{m} k_i\,x_i$, а среднее арифметическое, приходящееся на одно испытание, равно

$$\frac{1}{n} \cdot \sum_{i=1}^{m} k_i\,x_i = \sum_{i=1}^{m} x_i\,\frac{k_i}{n}.$$

Если n велико, то

$$\frac{k_i}{n} \approx p_i, \quad i = 1,\, 2,\, \cdots, m,$$

и указанное среднее значение будет близко к величине $\sum_{i=1}^m x_i\, p_i$. Последнее выражение называется *математическим ожиданием* или *средним значением* с.в. ξ. Оно обозначается символом $M\xi$:

$$M\xi = \sum_{i=1}^{m} x_i\, p_i. \tag{9}$$

Для непрерывной с.в. с плотностью распределения вероятностей $f(x)$ математическое ожидание определяется как интеграл

$$M\xi = \int_{-\infty}^{+\infty} x\, f(x)\, d\,x. \tag{10}$$

Заметим, что физическая размерность $M\xi$ совпадает с физической размерностью ξ. Можно дать механическую интерпретацию математического ожидания. Если на числовой прямой даны материальные точки $x_1,\, x_2,\, \cdots, x_m$ с массами $p_1,\, p_2,\, \cdots, p_m$, так, что сумма всех масс равна 1, то $\sum_{i=1}^m i\, p_i$ определяет положение центра этих масс. Тот же механический смысл имеет интеграл $\int_{-\infty}^{+\infty} x\, f(x)\, d\,x$ в случае, когда единичная масса непрерывно распределена на числовой прямой с плотностью $f(x)$, такой, что $\int_{-\infty}^{+\infty} f(x)\, d\,x$.

* Разыгрывается лотерея из 1000 билетов, имеющая следующие выигрыши: один в 1000 рублей, четыре по 500 рублей, восемь по 300 рублей, двадцать по 50 рублей. Стоимость выигрыша на один билет – с.в. ξ со следующей таблицей распределения вероятностей:

ξ :	0	50	300	500	1000,
P :	0.967	0.02	0.008	0.004	0.001.

Средний выигрыш на один билет равен

$$M\xi = 0 \cdot 0.967 + 50 \cdot 0.02 + 300 \cdot 0.008 +$$

$$+ 500 \cdot 0.004 + 1000 \cdot 0.001 = 6.4 \text{ руб.}$$

* Для числа очков ξ, выпадающих на игральной кости, имеем $M\xi = (1+2+3+4+5+6) \cdot (1/6) = 7/2$.

* Для с.в. ξ, равномерно распределенной на $[0,\,1]$, имеем $M\xi = \int_0^1 x \cdot 1\,d\,x = 1/2$.

Рассмотрим функцию $\eta = h(\xi)$ от с.в. ξ, где ξ – дискретная с.в. с таблицей распределения вероятностей (8). Если значения $h(x_1)$, $h(x_2)$, $\cdots, h(x_m)$ различны, то таблица распределения вероятностей с.в. η примет вид

$$\begin{array}{cccccc} \eta: & h(x_1) & h(x_2) & \cdots & h(x_m), \\ p: & p_1 & p_2 & \cdots\cdots & p_m. \end{array} \tag{11}$$

Если же среди перечисленных значений с.в. η встречаются одинаковые, то в таблице их следует объединить, а соответствующие им вероятности во второй строке сложить. В обоих случаях для математического ожидания $M\eta$ получается одно и то же выражение

$$M\eta = \sum_{i=1}^{m} h(x_i)\,p_i.$$

Если ξ – непрерывная с.в. с плотностью распределения вероятностей $f(x)$, то можно показать, что

$$M\eta = M(h(\xi)) = \int_{-\infty}^{+\infty} h(x) f(x)\,d\,x. \tag{12}$$

Сформулируем *основные свойства математического ожидания*:

1. Если $\xi = C$ – постоянная с.в., то $M\,C = C$.

2. Если С – постоянная величина и ξ – с.в., то $M\,(C\xi) = C\,M\xi$.

3. Для любых с.в. ξ, η, появляющихся в одном испытании, справедлива формула сложения $M(\xi + \eta) = M\,\xi + M\,\eta$.

4. Для любых независимых с.в. ξ, η, появляющихся в одном испытании, справедлива формула умножения $M(\xi \cdot \eta) = M\,\xi \cdot M\,\eta$.

5. Если с.в. ξ принимает только неотрицательные значения, то $M\,\xi \geqslant 0$.

6. Если для с.в. ξ и η в каждом испытании $\xi \geqslant \eta$, то $M\,\xi \geqslant M\,\eta$.

Проверим эти свойства в случае дискретных с.в. с конечными множествами значений. Первое свойство очевидно: из таблицы распределения вероятностей $\xi = C$ следует $M\,C = 1 \cdot C = C$.

Далее, из таблиц распределений с.в. ξ и $C\,\xi$ получаем, что $M\,C\xi = \sum_{i=1}^{m} C\,x_i\,p_i = C\,\sum_{i=1}^{m} x_i\,p_i = C\,M\,\xi$.

Переходя к проверке свойства 3, рассмотрим с.в. ξ с законом распределения вероятностей (8) и с.в. η с законом распределения

$$\begin{array}{llllll} \eta: & y_1 & y_2 & \cdots & y_n, \\ p: & q_1 & q_2 & \cdots & q_n. \end{array}$$

Рассмотрим события

$$A_i = \{\xi = x_i\}, \;\; i = 1, \cdots, m,$$

$$B_j = \{\eta = y_j\}, \;\; j = 1, \cdots, n.$$

Положим $r_{i\,j} = P(A_i\,B_j)$. Тогда математическое ожидание суммы $\xi + \eta$ можно преобразовать к виду

$$M(\xi + \eta) = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_i + y_j)\, r_{i\,j} =$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i\, r_{i\,j} + \sum_{i=1}^{m} \sum_{j=1}^{n} y_j\, r_{i\,j} =$$

$$= \sum_{i=1}^{m} x_i \Big(\sum_{j=1}^{n} r_{i\,j} \Big) + \sum_{j=1}^{n} y_j \Big(\sum_{i=1}^{m} r_{i\,j} \Big). \tag{13}$$

По формуле полной вероятности

$$p_i = P(A_i) = \sum_{j=1}^{n} P(A_i\,B_j) = \sum_{j=1}^{n} r_{i\,j}, \tag{14}$$

$$q_j = P(B_j) = \sum_{i=1}^{m} P(A_i\,B_j) = \sum_{i=1}^{m} r_{i\,j}. \tag{15}$$

Из (13)–(15) следует требуемый результат:

$$M(\xi + \eta) = \sum_{i=1}^{m} x_i\, p_i + \sum_{j=1}^{n} y_j\, q_j = M\,\xi + M\,\eta.$$

Методом математической индукции свойство 3 обобщается на сумму конечного числа с.в. Именно, если ξ_1, ξ_2, $\cdots, \xi_m$ – с.в., наступление которых возможно в одном испытании, то

$$M \sum_{i=1}^{m} \xi_i = \sum_{i=1}^{m} M\,\xi_i.$$

Свойство 4 проверяется следующим образом. В силу независимости ξ и η имеем $r_{i\,j} = P(A_i\,B_j) = P(A_i)\,P(B_j) = p_i\,q_j$. Следовательно,

$$M(\xi\,\eta) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i\,y_j\,r_{i\,j} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i\,y_j\,p_i\,q_j =$$

$$= \Big(\sum_{i=1}^{m} x_i\,p_i\Big)\Big(\sum_{j=1}^{n} y_j\,q_j\Big) = M\,\xi\cdot M\,\eta.$$

Свойство 5 очевидно, так как если все x_i неотрицательны, то

$$M\,\xi = \sum_{i=1}^{m} x_i\,p_i \geqslant 0.$$

Наконец, свойство 6 вытекает из 3 и 5, поскольку из $M(\xi-\eta) \geqslant 0$ следует, что $M\,\xi - M\,\eta \geqslant 0$.

* Рассмотрим дискретную с.в. μ, которая принимает неотрицательные целые значения $m = 0,\,1,\,2,\cdots$ с вероятностями $P\{\mu = m\} = \lambda^m\,e^{-\lambda}/m!$, где λ – заданное положительное число (параметр с.в.). Вычислим ее математическое ожидание:

$$M\,\mu = \sum_{m=1}^{\infty} m\cdot\frac{\lambda^m\,e^{-\lambda}}{m!} =$$

$$= \lambda\,e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^{m-1}}{(m-1)!} = \lambda,$$

поскольку согласно формуле Тейлора сумма последнего ряда равна e^{λ}.

Случайная величина μ с законом распределения

$$P\{\mu = m\} = \lambda^m\,e^{-\lambda}/m!,\ m = 0,\,1,\,2,\,\cdots,$$

называется *пуассоновской* с.в. или с.в., *распределенной по закону Пуассона.* Итак, $M\,\mu = \lambda$. В этом состоит вероятностный смысл параметра λ в законе Пуассона.

* Биномиальная с.в. β – число успехов в серии из n испытаний Бернулли –nможет быть представлена как сумма с.в. I_k-индикаторов успехов в k-том испытании, $k = 1,\,2,\cdots,n$. Так как $M\,I_k = 1\cdot p + 0\cdot(1-p) = p$ для каждого индикатора, то $M\,\beta = M\,\sum_{k=1}^{n} I_k = \sum_{k=1}^{n} M\,I_k = \sum_{k=1}^{n} p = n\,p)$.

Стало быть, *математическое ожидание биномиальной с.в.* β *равно* $n\,p$.

§ 2.4. Дисперсия и средне-квадратичное отклонение случайной величины

Пусть ξ – с.в., $M\,\xi = a$ – ее математическое ожидание. *Отклонением с.в.* ξ *от* a *называется с.в.* $\xi - a$. Очевидно, что $M\,(\xi - a) = M\,\xi - M\,a = a - a = 0$, то есть $M\,(\xi - a) = 0$.

Важной числовой характеристикой с.в. является ее *дисперсия* $D\,\xi$, определяемая формулой $D\,\xi = M\,(\xi - a)^2$. Квадратный корень из дисперсии называется *средним квадратичным отклонением* ξ. Оно обозначается символом $\sigma\,\xi$. Дисперсия $D\,\xi$ и среднее квадратичное отклонение $\sigma\,\xi = \sqrt{D\,\xi}$ являются оценками рассеяния значений с.в. ξ относительно ее математического ожидания a или оценками среднего отклонения ξ от a.

Для дискретной с.в. ξ с таблицей распределения (8) имеем

$$D\,\xi = M\,(\xi - a)^2 = \sum_{i=1}^{m} (x_i - a)^2\,p_i. \tag{16}$$

Соответствующая формула для дисперсии непрерывной с.в ξ с плотностью вероятности $f(x)$ –

$$D\,\xi = \int_{-\infty}^{+\infty} (x - a)^2\,f(x)\,d\,x. \tag{17}$$

Как и в предыдущем параграфе, рассмотрим механическую интерпретацию распределения вероятностей как распределение единичной массы на числовой прямой. Тогда выражения (16) и (17) будут равны моменту инерции относительно центра масс $x = a$ соответственно для дискретно и непрерывно распределенных масс.

Преобразуем выражение для дисперсии, пользуясь свойствами математического ожидания. Будем иметь

$$D\,\xi = M\,(\xi - a)^2 = M\,(\xi^2 - 2a\,\xi + a^2) =$$
$$= M\,\xi^2 - 2a\,M\,\xi + a^2 = M\,(\xi^2) - [M\,\xi]^2. \tag{18}$$

Для дискретной с.в. формула (18) представляется в виде

$$D\,\xi = \sum_{i=1}^{m} x_i^2\,p_i - a^2,$$

а для непрерывной с.в. – соответственно интегралом:

$$D\,\xi = \int_{-\infty}^{+\infty} x^2\, f(x)\, d\,x - a^2.$$

Сформулируем *основные свойства дисперсии.*

1. Если C – постоянная, то $D\,C = 0$.

2. Для постоянной C и любой с.в. ξ верно равенство $D\,(C\xi) = C^2\,D\,\xi$.

3. Для любых независимых с.в. ξ и η, появляющихся в одном испытании, $D(\xi + \eta) = D\,\xi + D\,\eta$.

4. Для постоянной C и любой с.в. ξ верно равенство $D\,(C + \xi) = D\,\xi$.

Методом математической индукции третье свойство распространяется на сумму конечного числа независимых с.в. ξ_i, $i = 1, 2, \cdots, m$: $D\sum_{i=1}^{m}\xi_i = \sum_{i=1}^{m} D\,\xi_i$.

Доказательства этих свойств проводятся с помощью свойств математического ожидания и формулы (18).

В самом деле, $D(C) = M(C^2) - (M\,C)^2 = C^2 - C^2 = 0$,

$$D(C\xi) = M(C^2\xi^2) - \big(M(C\xi)\big)^2 = C^2\,M(\xi^2) - C^2(M\,\xi)^2 = C^2\,D\,\xi.$$

Проверим третье свойство. Вычисления дают

$$D(\xi + \eta) = M(\xi + \eta)^2 - [M(\xi + \eta)]^2 =$$

$$= M(\xi^2) + 2M(\xi\,\eta) + M(\eta^2) - (M\,\xi)^2 - 2(M\xi)(M\eta) - (M\,\eta)^2 =$$

$$= D\,\xi + D\,\eta + 2M\xi\,\eta - 2M\,\xi \cdot M\,\eta. \tag{19}$$

Но для независимых ξ, η в силу четвертого свойства математического ожидания имеем $M(\xi\,\eta) = M\,\xi \cdot M\,\eta$. Стало быть, последнее выражение в (19) равно $D\,\xi + D\,\eta$, что и требовалось доказать.

Свойство 4 вытекает из свойств 1 и 3, если принять во внимание, что ξ и C независимые с.в.

* Вычислим дисперсию пуассоновской с.в. μ, имеющей математическое ожидание λ. Применим формулу $D\,\mu = M\,(\mu^2) - (M\,\mu)^2$. Получим

$$D\,\mu = \sum_{m=1}^{\infty} m^2 \frac{\lambda^m\, e^{-\lambda}}{m!} - \lambda^2 =$$

$$\lambda \cdot e^{-\lambda} \cdot \sum_{m=1}^{\infty} m \frac{\lambda^{m-1}}{(m-1)!} - \lambda^2 =$$

$$= \lambda \cdot e^{-\lambda} \cdot \Big(\lambda \cdot \sum_{m=2}^{\infty} \frac{\lambda^{m-2}}{(m-2)!} + \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!}\Big) - \lambda^2 =$$
$$= \lambda \cdot e^{-\lambda}\big(\lambda e^{\lambda} + e^{\lambda}\big) - \lambda^2 = \lambda.$$

Итак, *дисперсия пуассоновской с.в.* μ *с математическим ожиданием* λ *равна* λ, $D\,\mu = \lambda$.

* Займемся вычислением дисперсии и среднего квадратичного отклонения биномиальной с.в. $\beta = \sum_{k=1}^{n} I_k$. Сначала найдем дисперсию индикатора I_k по формуле $D\,I_k = M\,(I_k^2) - (M\,I_k)^2$. Будем иметь

$$D\,I_k = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p(1-p).$$

Затем, в силу независимости индикаторов $I_1, I_2, \cdots, I_n$, можно применить правило вычисления дисперсии суммы конечного числа независимых с.в. Получим

$$D\beta = \sum_{k=1}^{n} D\,I_k = \sum_{k=1}^{n} p(1-p) = n\ \cdot p(1-p).$$

Таким образом, *дисперсия биномиальной с.в.* β *равна* $n \cdot p(1-p)$. Соответственно ее *среднее квадратичное отклонение равно* $\sqrt{n \cdot p(1-p)}$.

§ 2.5. Корреляционная связь случайных величин

Пусть (ξ, η) – *векторная с.в.*, ее компоненты ξ, η – случайные величины, которые могут наблюдаться в одном испытании. Полная картина взаимной зависимости компонент этого случайного вектора дается его *функцией распределения вероятностей* $F(x, y)$, определяемой следующим образом. Возьмем два числа x, y и рассмотрим события $A = \{\xi < x\}$, $B = \{\eta < y\}$. Тогда величина $F(x, y)$ полагается равной вероятности произведения этих событий: $F(x, y) = P(A \cdot B) = P\{\xi < x, \eta < y\}$.

Для многих практических целей, однако, желательно иметь более простую и удобную характеристику зависимости между двумя случайными величинами. Возможны следующие три случая.

1. Между ξ и η существует функциональная зависимость $\eta = h(\xi)$, то есть η должна принять значение $h(x)$, если в том же испытании ξ приняла значение x.

2. ξ и η – независимые с.в., то есть распределение вероятностей каждой из этих величин не зависит от того, какое значение принимает вторая величина.

3. Между ξ и η нет функциональной зависимости, но они также не являются независимыми, то есть распределение вероятностей каждой из этих величин зависит от значений, принимаемых другой величиной. Такая связь между ξ и η называется *корреляционной.*

Случаи 1 и 2 можно рассматривать как крайние, предельные для третьего – общего случая корреляционной связи.

Укажем несколько примеров с.в., между которыми может существовать корреляционная связь:

* ξ – освещенность рабочего места, η – производительность труда на этом рабочем месте;

* ξ – количество осадков в летние месяцы в данном регионе, η – собранный в том же регионе урожай;

* ξ и η – процентное содержание двух минералов в образцах руды из одного месторождения;

* ξ – процент хороших, η – процент отличных оценок у студентов в экзаменационную сессию.

Перейдем теперь к более строгим определениям. Пусть с.в. ξ и η имеют математические ожидания a и b и положительные дисперсии $D\,\xi$ и $D\,\eta$ соответственно. Удобной и полезной количественной характеристикой связи между рассматриваемыми с.в. является *коэффициент корреляции*, определяемый формулой

$$\rho(\xi,\,\eta)=\frac{M(\xi-a)\cdot(\eta-b)}{\sqrt{D\,\xi\cdot D\,\eta}}. \tag{20}$$

Нетрудно показать, что возможные значения коэффициента корреляции принадлежат промежутку [-1, 1]. Далее, в случае независимых с.в. ξ, η величины $\xi-a$, $\eta-b$ также независимы и в силу четвертого свойства математического ожидания

$$M(\xi-a)(\eta-b)=M(\xi-a)\cdot M(\eta-b)=(M\,\xi-a)(M\,\eta-b)=0.$$

Стало быть, в этом случае коэффициент корреляции равен нулю, то есть $\rho(\xi,\,\eta)=0$. Если однако $\rho(\xi,\,\eta)=0$, то не факт, что рассматриваемые с.в. независимы.

Если $\rho(\xi,\,\eta)\neq 0$, то принято говорить, что между с.в. ξ, η имеется *корреляционная зависимость или корреляционная связь.* Если же $\rho(\xi,\,\eta)=0$, то с.в. ξ, η называются *некоррелированными.* Таким образом, две независимые с.в. являются также некоррелированными.

Числитель дроби, стоящей в правой части формулы (20), называется *ковариацией с.в. ξ и η* и обозначается символом $cov(\xi,\,\eta)$.

* Найдем коэффициент корреляции между с.в. ξ и $\eta = \xi^2$, если $P(\xi = 0) = 1/3$, $P(\xi = 1) = 1/2$, $P(\xi = -1) = 1/6$. Сначала вычислим математические ожидания

$$a = M\,\xi = 0 \cdot 1/3 + 1 \cdot 1/2 + (-1) \cdot 1/6 = 1/3,$$

$$b = M\,(\xi^2) = 0^1 \cdot 1/3 + 1^2 \cdot 1/2 + (-1)^2 \cdot 1/6 = 2/3.$$

Случайные величины $(\xi - a)$ и $(\eta - b)$ принимают значения $-1/3$, $2/3$, $-4/3$ и $-2/3$, $1/3$ с вероятностями $p_1 = 1/3$, $p_2 = 1/2$, $p_3 = 1/6$ и $q_1 = 1/3$, $q_2 = 2/3$ соответственно. Теперь можно найти дисперсии:

$$D\,\xi = M((\xi - a)^2) = (-1/3)^2 \cdot (1/3) + (2/3)^2 \cdot (1/2) + (-4/3)^2 \cdot (1/6) = 5/9,$$

$$D\,\eta = M((\eta - b)^2) = (-2/3)^2 \cdot (1/3) + (1/3)^2 \cdot (2/3) = 2/9.$$

Квадратный корень из произведения дисперсий, стоящий в знаменателе правой части формулы (20), равен $\sqrt{10}/9$.

Произведение с.в. $\xi - a$ и $\eta - b$ есть с.в. ζ, принимающая значения $2/9$, $-4/9$, $8/9$, $-1/9$, $2/9$, $-4/9$ с вероятностями $1/9$, $1/6$, $1/18$, $2/9$, $1/3$, $1/9$ соответственно. Объединяя общие значения, придем к закону $P(\zeta = -4/9) = 5/18$, $P(\zeta = -1/9) = 2/9$, $P(\zeta = 2/9) = 4/9$, $P(\zeta = 8/9) = 1/18$. Ковариация с.в. ξ, η, равная числителю дроби в правой части формулы (20), может быть вычислена по формуле

$$cov(\xi, \eta) = M\,\zeta = (-4/9) \cdot (5/18) + (-1/9) \cdot (2/9) + (2/9) \cdot (4/9) +$$

$$+ (8/9) \cdot (1/18) = 0.$$

Стало быть, данные с.в. ξ и ξ^2 между собой не коррелируют. Впрочем, все это прямо следует из функциональной зависимости $\eta = \xi^2$.

§ 2.6. Задачи для самостоятельного решения

1. Найти таблицу распределения вероятностей с.в. $\eta = \sin(\pi\xi/3)$, где ξ – число очков, выпадающее при бросании игральной кости.

2. Построить график функции распределения вероятностей с.в. η из предыдущей задачи.

3. Вычислить математическое ожидание и дисперсию с.в. η из задачи 1.

4. Случайная величина ξ имеет равномерное распределение на отрезке $[0, 1]$. Найти математическое ожидание с.в. $\eta = -\xi^2 + 3\xi - 2$.

5. Случайная величина ξ имеет показательное распределение, то есть

$$F_\xi(x) = 1 - e^{-a\,x}, \ \text{ если } \ x \geqslant 0,$$

и $F_\xi(x) = 0$ при $x < 0$.

а) Найти плотность вероятности, математическое ожидание и дисперсию ξ.

б) Найти закон распределения вероятностей с.в. $\eta = a^{-1} \cdot \ln \xi$.

6. Непрерывная случайная величина ξ имеет функцию распределения вероятностей $F_\xi(x)$, равную нулю при $x < -1$, единице при $x > 1$ и $a + b \arcsin x$ при $-1 \leqslant x \leqslant 1$. Найти значения параметров a и b, а также математическое ожидание и дисперсию с.в. ξ. Изобразить эскизы графиков функции распределения и ее плотности.

7. В единичный квадрат $Q = \{0 \leqslant x \leqslant 1, \ 0 \leqslant y \leqslant 1\}$ наудачу брошена точка $A(\xi,\, \eta)$. Вероятность попадания этой точки в какую-либо часть D квадрата равна площади D. Найти вероятность того, что корни квадратного трехчлена $x^2 + \xi x + \eta$ действительны.

8. Привести пример зависимых с.в. ξ, η, для которых $\rho(\xi,\, \eta) = 0$.

Все верят в нормальный закон: математики – потому, что думают , что физики наблюдают его на опыте; физики же – потому, что думают, что математики способны доказать теоретически, что нормальный закон должен выполняться.

А. Пуанкаре

Глава III

КЛАССИЧЕСКИЕ ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

§ 3.1. Неравенства Чебышева

В теории вероятностей важную роль играют некоторые виды неравенств для случайных величин. Здесь будут доказаны два неравенства, открытые еще в XIX веке великим русским математиком П.Л.Чебышевым.

Пусть ξ – какая-либо случайная величина. *Первое неравенство Чебышева* имеет вид

$$P(|\xi| \geqslant \varepsilon) \leqslant \frac{M\,|\xi|}{\varepsilon} \qquad (1)$$

и справедливо для любого наперед заданного положительного числа ε. Левая часть (1) есть вероятность того, что значение с.в. ξ в эксперименте окажется по абсолютной величине большим или равным ε. Для доказательства рассмотрим индикаторы I_A и $I_{\bar{A}}$ двух противоположных событий $A = \{|\xi| \geqslant \varepsilon\}$ и $\bar{A} = \{|\xi| < \varepsilon\}$. Ясно, что $A + \bar{A} = E$ – достоверное событие, поэтому $I_A + I_{\bar{A}} = I_E = 1$. Дальнейшие вычисления дают:

$$|\xi| = |\xi| \cdot I_E = |\xi|(I_A + I_{\bar{A}}) \geqslant |\xi| \cdot I_A \geqslant \varepsilon \cdot I_A,$$

то есть $|\xi| \geqslant \varepsilon \cdot I_A$. Беря математическое ожидание от обеих частей этого неравенства, получим

$$M\,|\xi| \geqslant \varepsilon \cdot P(A) = \varepsilon \cdot P(|\xi| \geqslant \varepsilon).$$

После деления крайних выражений неравенства на положительное ε приходим к первому неравенству Чебышева (1).

Применим первое неравенство Чебышева к с.в. $\eta = (\xi - M\,\xi)^2$. Получим

$$P((\xi - M\,\xi)^2 \geqslant \delta^2) \leqslant \frac{M(\xi - M\,\xi)^2}{\delta^2}.$$

Вместо ε назначили положительное число δ^2. Величину $M\,(\xi - M\,\xi)^2$ в числителе правой части неравенства заменим на дисперсию $D\,\xi$, а в левой части стоящее под знаком P событие $(\xi - M\,\xi)^2 \geqslant \delta^2$ заменим на равносильное событие $|\xi - M\,\xi| \geqslant \delta$. В итоге получим *второе неравенство Чебышева* в форме

$$P(|\xi - M\,\xi| \geqslant \delta) \leqslant \frac{D\,\xi}{\delta^2}. \tag{2}$$

Неравенство (2) верно для любой с.в. ξ с конечной дисперсией $D\,\xi$ и для любого наперед заданного положительного числа δ.

Из (2) легко получается оценка вероятности события $B = \{|\xi - M\,\xi| < < \delta\}$. Имеем

$$P(B) = 1 - P(\bar{B}) = 1 - P(|\xi - M\,\xi| \geqslant \delta) \geqslant 1 - \frac{D\,\xi}{\delta^2},$$

то есть

$$P(|\xi - M\,\xi| < \delta) \geqslant 1 - \frac{D\,\xi}{\delta^2}. \tag{3}$$

Это – другая форма второго неравенства Чебышева, равносильная (2).

* С помощью второго неравенства Чебышева оценим вероятность того, что при 1000 бросаниях монеты число выпадений орла будет заключено между 450 и 550. Так как вероятности появления орла и решки в одном испытании равны $1/2$, число испытаний $n = 100$, а испытания независимы, то мы имеем дело с биномиальной с.в. β. Ее математическое ожидание $M\,\beta$ равно $np = 1000 \cdot (1/2) = 500$, а дисперсия $D\,\beta = n\,p(1-p) = 250$. Очевидно также, что $\delta = 50$. Неравенство (3) дает

$$P(|\beta - 500| < 50) \leqslant \frac{250}{50^2} = 0.1.$$

Искомая вероятность равна 0.1.

§ 3.2. Теоремы Чебышева и Бернулли

Неравенства Чебышева позволяют просто доказать некоторые предельные соотношения, в которых участвуют средние арифметические последовательности $\{\xi_n\}$ независимых с.в. в предположении, что они могут появляться в однотипных испытаниях. Пусть $\xi_1, \xi_2, \cdots$ независимы и одинаково распределены с математическими ожиданиями $M\,\xi_k = a$ и конечными дисперсиями $D\,\xi_k = \sigma^2,\ k = 1, 2, \cdots$. Среднее арифметическое n таких с.в обозначим символом ζ_n. То есть $\zeta_n = (1/n)\sum_{k=1}^{n} \xi_k$. Математическое ожидание $M\,\zeta = (1/n)\sum_{k=1}^{n} M\,\xi_k = a$, дисперсия $D\,\zeta = n^{-2}\sum_{k=1}^{n} D\,\xi_k = \sigma^2/n$. Принимая во внимание эти факты, из неравенства (3) находим:

$$P\Big(\Big|\frac{\xi_1+\xi_2+\cdots+\xi_n}{n}-a\Big|<\delta\Big)=$$

$$=P\Big(|\zeta_n-a|<\delta\Big)\geqslant 1-\frac{D\,\zeta_n}{\delta^2}\geqslant 1-\frac{\sigma^2}{n\delta^2}.$$

Переход к пределу при $n\to\infty$ в этом неравенстве дает соотношение

$$\lim_{n\to\infty} P\Big(\Big|\frac{\xi_1+\xi_2+\cdots+\xi_n}{n}-a\Big|<\delta\Big)=1, \tag{4}$$

которое выполняется при любом $\delta > 0$. Доказана теорема, называемая *законом больших чисел в форме Чебышева.*

Теорема Чебышева. *Если $\xi_1, \xi_2, \cdots$ одинаково распределены и независимы, $M\,\xi_i = a,\ D\,\xi_i = \sigma^2 < \infty$ для всех $i = 1, 2, \cdots$, то средние арифметические*

$$\zeta_n=\frac{\xi_1+\xi_2+\cdots+\xi_n}{n}$$

стремятся по вероятности к $M\,\zeta_n = a$ при $n\to\infty$, то есть при любом $\delta > 0$ выполняется предельное соотношение (4). Исторически первой теоремой, выражающей *закон больших чисел для испытаний Бернулли*, является следующая теорема, которая получается как следствие теоремы Чебышева.

Теорема Бернулли. *Пусть β_n – число успехов в n испытаниях Бернулли с вероятностью успеха p, $0 < p < 1$, в каждом испытании. Тогда при любом $\delta > 0$*

$$\lim_{n\to\infty} P\Big\{\Big|\frac{\beta_n}{n}-p\Big|\leqslant\delta\Big\}=1. \tag{5}$$

Действительно, мы можем представить β_n как сумму индикаторов I_k успехов в k-м испытании Бернулли. Так как $M\,I_k = p$ и $D\,I_k = p(1-p)$

для каждого натурального числа k, то к с.в.

$$\zeta_n = \frac{\beta_n}{n} = \frac{1}{n}\sum_{k=1}^{n} I_k$$

применима теорема Чебышева. Теорема Бернулли доказана. Соотношение (5) показывает, что при больших n разность между относительной частотой β_n/n и вероятностью p успеха в отдельном испытании Бернулли мала с вероятностью, близкой к 1. Иными словами, указанная относительная частота успехов в n испытаниях сходится по вероятности к p. Это полезно сравнить со статистическим определением вероятности, данным в первой главе.

В условиях, когда справедливо свойство устойчивости относительных частот, можно применять следующий принцип: *при единичном испытании маловероятное событие практически невозможно.* Считая серию из n испытаний Бернулли за единичное испытание и выбирая $\delta > 0$ таким образом, чтобы

$$\frac{D\,\beta_n}{n^2\delta^2} = \frac{p(1-p)}{n\,\delta^2}$$

было мало, мы можем утверждать, что неравенство $|\beta_n/n - p| > \delta$ практически невозможно.

Как было отмечено выше, вопрос о том, какие вероятности считать малыми, зависит от конкретной прикладной задачи и не может быть решен средствами одной лишь теории вероятностей.

§ 3.3. Теорема Пуассона о редких событиях

Биномиальная с.в. β_n с вероятностью p успеха в каждом из n испытаний имеет закон распределения вероятностей $P(\beta_n = m) = C_n^m \cdot p^m\, q^{n-m}$ для $m = 0,\, 1,\, 2,\, \cdots,n$, где $q = 1 - p$.

* Игральная кость подбрасывается 12000 раз. Какова вероятность P того, что число шестерок будет лежать в промежутке (1800, 2100]?

Искомая вероятность равна

$$P = \sum_{1800 < m \leqslant 2100} C_{12000}^m \left(\frac{1}{6}\right)^m \left(\frac{5}{6}\right)^{n-m}.$$

Понятно, что точное вычисление этой суммы представляет весьма трудоемкую задачу даже с использованием персонального компьютера. Дело в том, что в число сочетаний C_n^m входят факториалы $m!$, $n!$, $(n-m)!$, так что значения C_n^m очень велики, в то время как величины $p^m\,q^{n-m}$ очень малы. При вычислении тех и других величин, а также их произведений и сумм произведений нужно следить, чтобы не было переполнений при работе ПК. Нет смысла составлять таблицы значений подобных вероятностей: они громоздки и неудобны для пользователя.

Поскольку схема независимых испытаний Бернулли служит вероятностной моделью многих реальных случайных явлений, представляет значительный интерес задача о нахождении асимптотических формул, позволяющих с хорошей точностью приближенно вычислять величины вида $C_n^m\,p^m\,q^{n-m}$ и их суммы. Одна из первых теорем в этом направлении доказана французским математиком С.Д. Пуассоном.

Содержанием ее служит приближенное выражение для вероятности $P(\beta_n = m)$ биномиальной с.в. β_n в том случае, когда число испытаний n велико, а *вероятность p наступления события A – успеха, – напротив, мала*, так что произведение $n\,p$ оказывается равным некоторому числу λ, не зависящему от n. События такого типа принято называть *редкими событиями.* Ситуацию предпочтительно описывать следующим образом. Вместо бесконечной последовательности независимых испытаний с постоянной вероятностью осуществления некоторого события A рассмотрим последовательность удлиняющихся серий испытаний, таких, что в каждой из серий испытания независимы и вероятность появления наблюдаемого события в пределах каждой серии сохраняет одно и то же значение $p = \lambda/n$, где n – число испытаний в этой серии.

Теорема Пуассона. *Пусть дана описанная выше последовательность серий независимых испытаний. Если в серии из n испытаний вероятность $P(A) = p$ успеха для каждого отдельного испытания постоянна и равна λ/n, а число успехов в пределах такой серии равно β_n, то*

$$P(\beta_n = m) \to \frac{\lambda^m \cdot e^{-\lambda}}{m!} \text{ при } n \to \infty, \tag{6}$$

где $\lambda > 0$ – данное, не зависящее от n число. Более строго, речь идет о существовании предела по вероятности (а не предела в смысле классического математического анализа). Стало быть, теорема Пуассона утверждает, что в случае редких событий асимптотикой биномиальной с.в. β_n при $n \to \infty$ является пуассоновская с.в. μ с законом распределения вероятностей

$$P(\mu = m) = \frac{\lambda^m \cdot e^{-\lambda}}{m!},\ \ m = 0,\ 1,\ 2,\ \cdots .$$

Предельное соотношение (6) вытекает из возможности записать $P(\beta_n = m)$ в виде

$$C_n^m\, p^m \cdot q^{n-m} = \frac{(n\,p)^m}{m!}\Big(1-\frac{1}{n}\Big)\Big(1-\frac{2}{n}\Big)\cdots\Big(1-\frac{m-1}{n}\Big)(1-p)^{n-m}$$

и учесть, что при $n \to \infty$ и $n\,p \to \lambda$ предел $(1-p)^{n-m}$ равен пределу $(1-\lambda/n)^n$, а последний предел равен $e^{-\lambda}$.

* В партии из 1000 изделий имеется 8 бракованных. Из этой партии выбирается наудачу 100 штук. Какова вероятность того, что среди выбранных изделий: а) нет брака; б) точно одно бракованное изделие; в) точно два бракованных изделия.

Вероятность p выбора бракованного изделия из данной партии в результате одного испытания равна 0.008. Значит, мы имеем дело с редким событием. Испытания будем считать независимыми. Тогда можно воспользоваться приближением Пуассона с параметром $\lambda = 100 \cdot 0.008$, то есть $\lambda = 0.8$. В случае а) получаем $P(\mu = 0) = e^{-0.8}/0! \approx 0.45$; в случае б) – $P(\mu = 1) = 0.8 \cdot e^{-0.8}/1! \approx 0.36$; в случае в) – $P(\mu = 2) = (0.8)^2 \cdot e^{-0.8}/2! \approx 0.144$.

§ 3.4. Нормальные случайные величины. Теоремы Муавра – Лапласа

Если успех не является редким событием, то вместо теоремы Пуассона используют локальную или интегральную теоремы Муавра – Лапласа. Пусть β_n – биномиальная с.в. Тогда *нормированная биномиальная случайная величина* β_n^o определяется как

$$\beta_n^o = \frac{\beta_n - M\,\beta_n}{\sqrt{D\,\beta_n}} = \frac{\beta_n - n\cdot p}{\sqrt{n\,p\,q}}.$$

Вычисления показывают, что $M\,\beta_n^o = 0,\;\; D\,\beta_n^o = 1$. Эти свойства справедливы для любой нормированной с.в. $\mu^o = (\mu - M\,\mu)/\sqrt{D\,\mu}$: $M\,\mu^o = 0,\;\; D\,\mu^o = 1$.

В теории вероятностей и математической статистике исключительную роль играет открытый великим немецким математиком К.Ф. Гауссом класс $N(a,\,\sigma)$ непрерывных случайных величин ξ_n с плотностями вероятностей вида

$$p(x) = \frac{1}{\sigma\sqrt{2\,\pi}} \cdot e^{-(x-a)^2/2\sigma^2}. \tag{7}$$

Такие с.в. ξ_n называются *нормальными или гауссовыми случайными величинами с параметрами* a *и* σ.

Заметим, что цитированное выше высказывание Пуанкаре относится к началу XX века. Прошедшие сто лет принесли много новых результатов, которые, пожалуй, только укрепили веру в нормальный закон, но вместе с тем более четко обозначили границы его применимости.

Параметры нормальной с.в. ξ_n имеют вполне определенный вероятностный смысл: $M\,\xi_n = a$, $D\,\xi_n = \sigma^2$.

* Прежде чем доказать это, вычислим два вспомогательных интеграла

$$J_1 = \int_{-\infty}^{+\infty} e^{-t^2/2}\,d\,t, \quad J_2 = \int_{-\infty}^{+\infty} t^2\,e^{-t^2/2}\,d\,t.$$

Будем иметь

$$J_1^2 = \int_{-\infty}^{+\infty} e^{-x^2/2}\,d\,x \cdot \int_{-\infty}^{+\infty} e^{-y^2/2}\,d\,y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2}\,d\,x\,d\,y.$$

Переходя к полярным координатам, получим

$$J_1^2 = \int_0^{2\,\pi} d\,\varphi \int_0^{+\infty} e^{-\rho^2/2}\,\rho\,d\,\rho.$$

Внутренний интеграл $\int_0^{+\infty} e^{-\rho^2/2}\,\rho\,d\,\rho$ равен 1. Следовательно, $J_1^2 = = \int_0^{2\,\pi} d\,\varphi = 2\,\pi$ и $J_1 = \sqrt{2\,\pi}$. Теперь найдем интеграл J_2. Применяя формулу интегрирования по частям при $u = t$, $v = -e^{-t^2/2}$, получим

$$J_2 = \left(-t\,e^{-t^2/2}\right)\Big|_{-\infty}^{+\infty} + J_1.$$

Внеинтегральный член в правой части определяется по правилу Лопиталя. Он равен нулю, поэтому $J_2 = \sqrt{2\,\pi}$. Вернемся к вычислению $M\,\xi_n$ и $D\,\xi_n$. В выражении для математического ожидания

$$M\,\xi_n = \int_{-\infty}^{+\infty} x\,p(x)\,d\,x = \frac{1}{\sigma\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} x\,e^{-(x-a)^2/2\sigma^2}\,d\,x$$

сделаем замену переменной интегрирования $t = (x-a)/\sigma$. Будем иметь

$$M\,\xi_n = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} (\sigma\,t + a)\,e^{-t^2/2}\,d\,t =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-t^2/2}dt + \frac{a}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{-t^2/2}dt = a.$$

Предпоследний интеграл равен нулю, так как подынтегральная функция является нечетной. Последний же слагаемое равно a, поскольку $J_1 = \sqrt{2\,\pi}$. Итак, $M\,\xi_n = a$.

В выражении для дисперсии

$$D\,\xi_n = \int_{-\infty}^{+\infty} (x-a)^2\, p(x)\, d\,x = \frac{1}{\sigma\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} (x-a)^2\, e^{-(x-a)^2/2\sigma^2}\, d\,x$$

сделаем ту же замену переменной и получим

$$D\,\xi_n = \frac{\sigma^2}{\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} t^2\, e^{-t^2/2}.$$

Принимая во внимание значение интеграла J_2, приходим к окончательному результату $D\,\xi_n = \sigma^2$. *Нормированная гауссова с.в.* $\xi_n^o = (\xi_n - a)/\sigma$ принадлежит классу $N(0,\,1)$ и имеет плотность вероятности

$$p(x) = \frac{1}{\sqrt{2\,\pi}}\, e^{-x^2/2}. \tag{8}$$

Соответствующая интегральная функция распределения $F_{\xi_n^o}(x)$, называемая *интегралом вероятностей*, обозначается ниже символом $\Psi(x)$ и представляется в виде

$$\Psi(x) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, d\,u. \tag{9}$$

Заметим, что интеграл вероятностей $\Psi(x)$, как и связанная с ним *функция Лапласа*

$$\Phi(x) = \frac{1}{\sqrt{2\,\pi}} \int_{0}^{x} e^{-u^2/2}\, d\,u = \Psi(x) - 0.5,$$

не относятся к классу элементарных функций. Таблицы значений функции Лапласа либо интеграла вероятностей можно найти во многих учебных пособиях и справочниках [3–13]. Рекомендуется также использовать специальные пакеты программ для ЭВМ.

Полезно знать, как выглядят графики функций (8), (9). Первый график – кривая Гаусса – симметричен относительно оси ординат, расположен над осью абсцисс, имеет единственный максимум в начале координат $x = 0$, равный $1/\sqrt{2\,\pi} \approx 0.4$. Кроме того, ось абсцисс является асимптотой кривой Гаусса при $x \to \pm\infty$ (см. рис. 2).

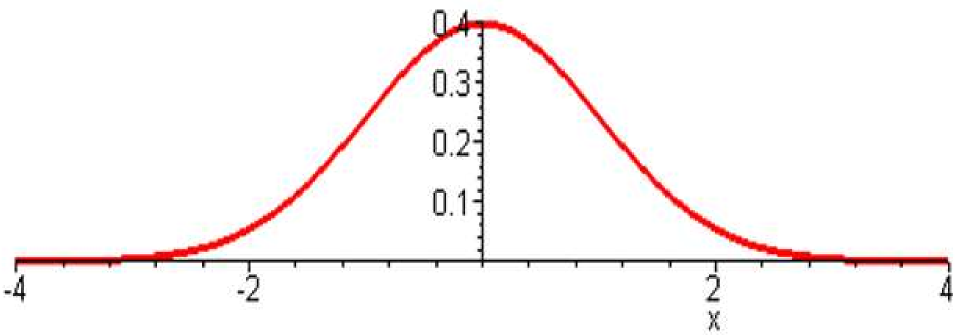

Рис. 2

График функции $\Psi(x)$ расположен между осью абсцисс и горизонтальной прямой $y = 1$ и имеет две горизонтальные асимптоты. Функция $\Psi(x)$ возрастает на всей числовой прямой от предельного значения $y = 0$ при $x \to -\infty$ до предельного значения $y = 1$ при $x \to +\infty$ (см. рис. 3).

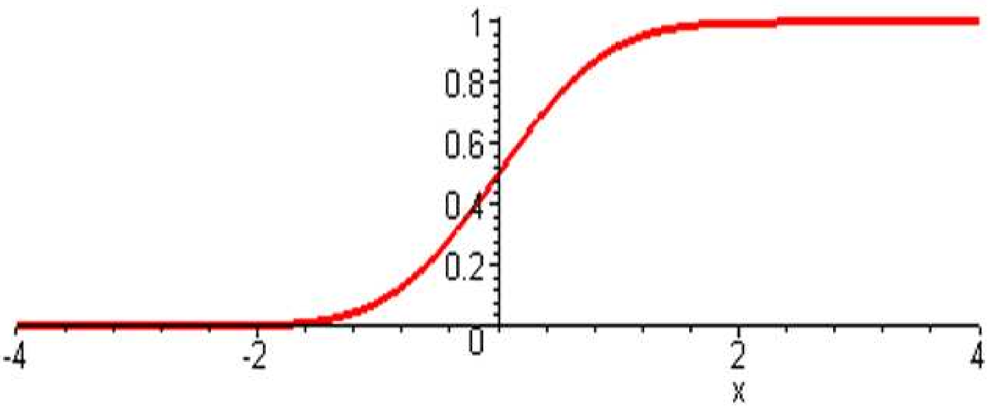

Рис. 3

На промежутке $(-\infty, 3]$ значения интеграла вероятностей положительны и не превосходят 0.00135, а на промежутке $[3, \infty)$ они меньше 1 и превосходят 0.99865. Практически достаточно иметь таблицу значений $\Psi(x)$ (или функции Лапласа $\Phi(x)$) на сегменте $[\text{-}3, 3]$. Теперь мы подготовлены к тому, чтобы дать формулировку *локальной предельной* *теоремы Муавра – Лапласа* об асимптотике значения нормированной биномиальной с.в. β_n^o.

Теорема. *Пусть в схеме Бернулли испытаний с постоянной вероятностью p успеха среднее квадратичное отклонение $\sigma = \sqrt{npq}$ стремится к бесконечности при $n \to \infty$. Тогда для любой наперед назначенной постоянной $C > 0$ для каждого значения $\beta_n^o = (m - n\,p)/\sigma$, такого, что $|(m - n\,p)/\sigma| \leqslant C$, для вероятности $P(\beta_n = m) = P(\beta_n^o = (m - n\,p)/\sigma)$ справедлива асимптотическая формула*

$$P(\beta_n = m) \approx \frac{1}{\sqrt{2\,\pi\,n\,p\,q}}\, e^{-x^2/2}, \tag{10}$$

где $x = x_{m,n} = (m - n\,p)/\sigma$. Иными словами, разность между левой и правой частями формулы (10) стремится к нулю при $n \to \infty$.

Доказательство можно найти в более подробных курсах теории вероятностей и математической статистики. Здесь ограничимся указанием о том, что в основе доказательства лежит известная асимптотическая при $n \to \infty$ формула Стирлинга

$$n! = n^n \cdot e^{-n} \cdot \sqrt{2\,\pi\,n}(1 + \varepsilon_n),$$

где $0 < \varepsilon_n < 1/(12n)$.

* Рассмотрим серию из $n = 1000$ подбрасываний монеты. Требуется найти вероятность того, что орел выпадет ровно 500 раз.

В данном случае $p = q = 1/2$,

$$x_{m\,n} = \frac{500 - 1000 \cdot (1/2)}{\sqrt{1000 \cdot (1/2) \cdot (1/2)}} = 0.$$

Стало быть, приближенное значение искомой вероятности в силу локальной теоремы Муавра – Лапласа равно

$$P(\beta_{1000} = 500) \approx \frac{1}{\sqrt{2\,\pi\, \cdot 1000 \cdot (1/2) \cdot (1/2)}} \approx 0.02523.$$

Найденное значение вероятности оказалось небольшим, что объясняется требованием точного совпадения биномиальной с.в. с назначенным числом $m = 500$. Однако если видоизменить задачу так, чтобы требовалось найти вероятность $P(m_1 \leqslant \beta_n \leqslant m_2)$, где m_1, m_2 – заданные границы, то ее величина уже не будет малой, сколь бы большим ни было число испытаний n. При этом границы предполагаются определенным образом связанными с n, а именно:

$$-C \leqslant x_{m_1,\,n} < x_{m_2,\,n} \leqslant C,$$

где C – наперед заданное положительное число.

Вероятности $P(m_1 \leqslant \beta_n \leqslant m_2)$ и $P(x_{m_1,\,n} \leqslant \beta_n^o \leqslant x_{m_2,\,n})$ совпадают, и по теореме сложения вероятностей несовместных событий можно записать

$$P(m_1 \leqslant \beta_n \leqslant m_2) = \sum_{m_1 \leqslant m \leqslant m_2} P(\beta_n = m), \tag{11}$$

где суммирование идет по всем значениям m, принадлежащим отрезку $[m_1,\, m_2]$. Отвечающие им значения $x_{m,\,n} = (m - n\,p)/\sqrt{n\,p\,q}$ принадлежат отрезку $[x_{m_1,\,n},\, x_{m_2,\,n}]$. Поэтому каждое $x_{m,\,n}$ из этого промежутка подчинено ограничению $|x_{m,\,n}| \leqslant C$, и мы вправе применить для приближенного вычисления вероятности $P(\beta_n = m)$ локальную теорему Муавра – Лапласа. Тогда (11) представляется в виде

$$P(m_1 \leqslant m \leqslant m_2) = \frac{1}{\sqrt{2\,\pi}} \sum_{x_{m,\,n} \in [x_{m_1,\,n},\, x_{m_2,\,n}]} e^{-x_{m,\,n}^2/2} \cdot \frac{1}{\sqrt{n\,p\,q}} +$$
$$+ \sum \alpha_{m,\,n}, \tag{12}$$

где $\alpha_{m,n}$ –разность между левой и правой частями в (9) и последняя сумма распространяется на все m, такие, что $m_1 \leqslant m \leqslant m_2$. Принимая во внимание, что

$$\Delta x_{m,\,n} = x_{m+1,\,n} - x_{m,\,n} =$$
$$= \frac{m + 1 - n\,p}{\sqrt{n\,p\,q}} - \frac{m - n\,p}{\sqrt{n\,p\,q}} = \frac{1}{\sqrt{n\,p\,q}},$$

перепишем (12) как

$$P(m_1 \leqslant m \leqslant m_2) = \frac{1}{\sqrt{2\,\pi}} \sum_{x_{m,\,n} \in [x_{m_1,\,n},\, x_{m_2,\,n}]} e^{-x_{m,\,n}^2/2} \cdot \Delta x_{m,\,n} + \sum \alpha_{m,\,n}. \tag{13}$$

Первая сумма в правой части (13) представляет собой интегральную сумму для интеграла

$$\int_{x_{m_1,\,n}}^{x_{m_2,\,n}} e^{-x^2/2}\,d\,x.$$

Далее, строгое доказательство локальной теоремы Муавра – Лапласа содержит в качестве одного из пунктов обоснование утверждения о том, что для любого $\varepsilon > 0$ начиная с некоторого номера n выполняется неравенство $|\alpha_{m,n}| < \varepsilon$ для каждого m, такого, что $|x_{m,n}| \leqslant C$. Значит, последняя сумма в (13) становится произвольно малой при достаточно больших номерах n. Выполнив предельный переход в формуле (13), получим

$$\lim_{n\to\infty} P(m_1 \leqslant m \leqslant m_2) = \frac{1}{\sqrt{2\,\pi}} \int_{x_{m_1,\,n}}^{x_{m_2,\,n}} e^{-x^2/2}\,d\,x.$$

Итак, доказана *интегральная теорема Муавра – Лапласа.*

Теорема. *Пусть x_1, x_2 – заданные числа, $x_1 < x_2$. Вероятность $P(x_1 \leqslant (\beta_n - n\,p)/\sqrt{n\,p\,q} \leqslant x_2)$ того, что число успехов β_n в серии n испытаний по схеме Бернулли окажется равным какому-либо из чисел m, удовлетворяющих неравенствам $x_1 \leqslant (m - n\,p)/\sqrt{n\,p\,q} \leqslant x_2$, имеет при $n \to \infty$ предел*

$$\lim_{n\to\infty} P(x_1 \leqslant \beta_n^o \leqslant x_2) = \frac{1}{\sqrt{2\,\pi}} \int_{x_1}^{x_2} e^{-x^2/2}\,d\,x. \tag{14}$$

На практике формулу (14) используют как точную, пренебрегая возможной малой ошибкой, уже в случаях, когда $n\,p\,q$ достигает нескольких десятков, хотя она имеет предельный характер. Интеграл в (14) удобно вычислять по формуле Ньютона-Лейбница:

$$\frac{1}{\sqrt{2\,\pi}} \int_{x_1}^{x_2} e^{-x^2/2}\,d\,x = \Phi(x_2) - \Phi(x_1), \tag{15}$$

где $\Phi(x)$ – функция Лапласа. Таким образом, в условиях теоремы Муавра-Лапласа при больших n справедлива приближенная формула

$$P(m_1 \leqslant \beta_n \leqslant m_2) \approx \Phi(x_2) - \Phi(x_1). \tag{16}$$

В Приложении приводится краткая таблица значений функции Лапласа (табл. I). Следует иметь в виду, что если в данной задаче для оценки вероятности можно применить как теорему Чебышева, так и интегральную теорему Муавра – Лапласа, то предпочтение нужно отдать более точной теореме Муавра – Лапласа.

* Монета подбрасывается $n = 400$ раз. Найти вероятность того, что число m выпадений орла будет лежать в пределах от $m_1 = 194$ до $m_2 = 208$.

Здесь $p = q = 1/2$, $n\,p = 200$, $n\,p\,q = 100$, $x_1 = \frac{m_1 - n\,p}{\sqrt{n\,p\,q}} = \frac{194-200}{10} = -0.6$, $x_2 = \frac{m_2 - n\,p}{\sqrt{n\,p\,q}} = \frac{208-200}{10} = 0.8$.

В силу интегральной теоремы Муавра – Лапласа

$$P(194 \leqslant m \leqslant 208) \approx \frac{1}{\sqrt{2\,\pi}} \int_{-0.6}^{0.8} e^{-u^2/2}\,d\,u = \Phi(0.8) - \Phi(-0.6) \approx 0.5139.$$

Замечание. Более точный математический анализ показывает, что ошибка при приближенном вычислении $P(m_1 \leqslant \beta_n \leqslant m_2)$ будет существенно меньшей, если вместо формулы (16) использовать уточненную формулу

$$P(m - 1 \leqslant \beta_n \leqslant m_2) \approx \Phi(\hat{x}_2) - \Phi(\hat{x}_1), \tag{17}$$

где

$$\hat{x}_1 = \frac{m_1 - \frac{1}{2} - n\,p}{\sqrt{n\,p\,q}}, \;\; \hat{x}_2 = \frac{m_2 + \frac{1}{2} - n\,p}{\sqrt{n\,p\,q}}.$$

При n порядка нескольких сотен формула (17) дает точность до трех – четырех знаков после запятой. Обычно применяемая формула (16) такой точности не дает. Последовательность $\{\xi_k\}$, $k = 1, 2, \cdots$, одинаково распределенных независимых с.в. ($M\,\xi_k = a$, $D\,\xi_k = \sigma^2$) называется *асимптотически нормальной*, если для соответствующей последовательности нормированных с.в. $\xi_k^o = (\xi_k - a)/\sigma$ выполняется предельное соотношение

$$\lim{}_{k\to\infty} P(x_1 \leqslant \xi_k^o \leqslant x_2) = \Phi(x_2) - \Phi(x_1),$$

каковы бы ни были $x_1, x_2 \in \mathbb{R}$, $x_1 < x_2$.

Имеется в виду сходимость по вероятности. В описанной ситуации употребительны записи $\xi_k \sim N(a, \sigma^2)$ и $\xi_k^o \sim N(0, 1)$ при $k \to \infty$. Утверждения о существовании предельных соотношений такого сорта носят название *центральных предельных теорем.* Примером является теорема Муавра – Лапласа. Существует много ее обобщений, наиболее известна центральная предельная теорема А.М. Ляпунова.

Пусть с.в. ξ_k, $k = 1, 2, 3, \cdots$, независимы и имеют не обязательно одно и то же распределение. Обозначим

$$M\,\xi_k = a_k, \;\; D\,\xi_k = \sigma_k^2, \;\; M\,|\xi_k - a_k|^3 = c_k^3,$$

$$A_n = \sum_{k=1}^{n} a_k, \;\; B_n^2 = \sum_{k=1}^{n} \sigma_k^2, \;\; C_n^3 = \sum_{k=1}^{n} c_k^3.$$

Теорема Ляпунова. *Если $\xi_1, \xi_2, \cdots$ независимы, A_n, B_n, C_n конечны и $C_n/B_n \to 0$ при $n \to \infty$, то для любого $x \in \mathbb{R}$ выполняется предельное соотношение*

$$\lim_{n\to\infty} P\Big\{\frac{\xi_1 + \cdots + \xi_n - A_n}{B_n} \leqslant x\Big\} = \Psi(x),$$

где $\Psi(x)$ – интеграл вероятностей. Доказательство можно найти в учебной литературе [1; 2].

§ 3.5. Задачи для самостоятельного решения

1. Пусть с.в. η_n равна сумме всех очков, появившихся при $n = 100$ подбрасываниях игральной кости. Пользуясь неравенством Чебышева, оценить снизу вероятность $P(|\eta_{100}/100 - 3.5| < 2)$.

2. Устройство состоит из 10 независимо работающих элементов. Вероятность отказа (выхода из строя) каждого в течение полугода равна 0.05. Используя неравенство Чебышева, оценить вероятность того, что абсолютная величина разности между числом отказавших элементов и средним числом (математическим ожиданием) отказов за те же полгода окажется меньшим 2.

3. Пусть в результате некоторого испытания вероятность p появления события A равна 0.001, то есть в среднем на каждую тысячу испытаний событие A появляется только один раз. Выполняется $n = 2000$ таких взаимно независимых испытаний. Какова вероятность, что число μ появлений события A окажется равным нулю (не произойдет ни разу?). Найти также соответствующие вероятности для $\mu = 1,\ \mu = 2,\ \mu = 3$.

4. В городе N в среднем в течение одного дневного часа поступает один вызов на скорую помощь. С какой вероятностью за три дневных часа поступит более 10 вызовов?

5. В некоторой местности имеется 3% больных малярией. Производится обследование 500 человек. С какой вероятностью среди обследованных окажется $3\pm0.5\%$ больных малярией?

6. Какова вероятность, что в группе, состоящей из 30 студентов, никто не родился в январе? Вычислить эту вероятность по точной формуле и по пуассоновскому приближению.

7. В поселке 2500 жителей. Каждый из них примерно 6 раз в месяц ездит на поезде в город, выбирая дни поездок по случайным мотивам

независимо от остальных. Какой наименьшей вместимостью должен обладать поезд, чтобы он переполнялся не чаще одного раза в сто дней? (Поезд ходит один раз в сутки).

8. Производится $n = 200$ испытаний по схеме Бернулли с вероятностью успеха в отдельном испытании $p = 1/4$. Чему равна вероятность того, что число β_{200} всех успехов будет равно $m = 50$?

9. В условиях предыдущей задачи найти вероятность того, что число успехов будет лежать в пределах от $m_1 = 30$ до $m_2 = 70$.

10. Рассматривается серия из $n = 300$ испытаний по схеме Бернулли с вероятностью успеха в отдельном испытании $p = 1/3$. При каком значении C вероятность осуществления неравенства $|\beta_{300} - 100| \leqslant C$ будет равна 0.98?

11. Рассматривается серия из $n = 169$ испытаний по схеме Бернулли с вероятностью успеха в одном испытании $p = 9/13$. Требуется указать границы возможных уклонений (одинаковых) в обе стороны числа μ наступлений успехов от значения $n\,p = 169 \cdot (9/13) = 117$, внутри которых μ будет заключаться с вероятностью 0.9.

12. Производится серия из n испытаний по схеме Бернулли с вероятностью успеха в отдельном испытании $p = 0.4$.

а) При каком значении n с вероятностью 0.99 можно гарантировать осуществление неравенства $|\beta_n/n - 0.4| < 0.05$?

б) Какова вероятность осуществления неравенства $|\beta_n - 40| < 5$, если число испытаний $n = 100$?

в) Каковы возможные уклонения (в обе стороны) числа успехов β_n от $n\,p = 80$, которые могут быть гарантированы с вероятностью 0.92, если число испытаний $n = 200$?

Глава IV

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

§ 4.1. Оценки параметров генеральной совокупности

При практическом применении теории вероятностей естественно возникают два вопроса: 1. Можно ли вообще говорить о вероятностях рассматриваемых событий, то есть обладают ли их относительные частоты свойством статистической (иногда говорят – стохастической) устойчивости? 2. Как найти эти вероятности из экспериментальных данных в случае, когда ответ на первый вопрос является утвердительным? Пусть, например, исследователь произвел n опытов, каждый из которых состоял в некотором измерении, и получил численные результаты x_1, x_2, $\cdots, x_n$. Спрашивается, надо ли их обрабатывать статистически, и если да, то как именно. Следует иметь в виду, что теория вероятностей не всемогуща и не претендует на математическое моделирование каких-угодно случайных событий. В современной математике существуют другие возможности для описания явлений, которые на первый взляд должны быть полем приложений теории вероятностей. Речь идет о синергетике, фракталах, странных аттракторах, дающих новый взгляд на хаотические, беспорядочные явления и процессы [14].

Эти разделы математики лежат вне теории вероятностей и математической статистики, и мы упомянули о них лишь с той целью, чтобы подчеркнуть необходимость аккуратного построения и применения математических моделей. Формальный подход может приводить к абсурдным результатам, как показывает следующий пример Р. Мизеса.

* Теннисист K может поехать на турнир либо в Москву, либо в Лондон, причем оба турнира проходят одновременно. Вероятность того, что K займет первое место в Москве, равна 0.8 (если, конечно, он туда поедет), а вероятность того, что он займет первое место в Лондоне, равна 0.4. Чему равна вероятность того, что K займет где-либо первое место?

Если следовать математически не вполне строгим определениям, данным в главе 1, то ничто не мешает следующим суждениям: события A="выигрыш в Москве" и B="выигрыш в Лондоне" несовместны, поэтому $P(A+B) = P(A) + P(B) = 0.8 + 0.4 = 1.2$. Но вероятность по определению не может быть больше единицы!

Полученное противоречие со здравым смыслом показывает, что одних лишь эвристических соображений для построения теории вероятностей недостаточно. Продолжим обсуждение поставленных вопросов.

Чрезвычайно важно искоренить заблуждение, нередко встречающееся у недостаточно знакомых с теорией вероятностей инженеров и естествоиспытателей, что результат любого эксперимента можно рассматривать как случайную величину. Сформулированные в первых главах настоящего учебного пособия предпосылки о неизменности условий эксперимента и их воспроизводимости принципиально неограниченное число раз и о независимости друг от друга результатов эксперимента на языке теории вероятностей превращаются в следующую основную предпосылку: *результаты n экспериментов $x_1, x_2, \cdots, x_n$ являются независимыми в совокупности случайными величинами с одной и той же функцией распределения $F(x)$*. Кратко это предположение выражают словами о том, что *числа $x_1, x_2, \cdots, x_n$ образуют выборку с теоретическим законом $F(x)$*, или *выборку из распределения $F(x)$*, или, наконец, *выборку из генеральной совокупности с законом распределения $F(x)$. Число n называется объемом выборки.*

Строго говоря, результаты наблюдений следовало бы обозначать $\xi_1, \xi_2, \cdots, \xi_n$, если придерживаться обозначений из второй и третьей глав. Однако в математической статистике, долгие годы развивавшейся автономно от теории вероятностей, сложилась традиция, в силу которой $x_1, x_2, \cdots, x_n$ обозначают иногда *случайные величины*, а иногда – *неслучайные числа или переменные, например переменные интегрирования*, причем в одних и тех же формулах. По ходу дела приходится учиться различать эти ситуации в обозначениях.

Итак, отнюдь не всякий набор $x_1, x_2, \cdots, x_n$ экспериментальных данных мы вправе называть выборкой и обрабатывать методами математической статистики. Выборочным данным должно быть присуще нечто общее, их организующее, именно – теоретический закон $F(x)$.

Кроме терминов "выборка" и "генеральная совокупность" в математической статистике употребительны и другие термины и понятия, которые будут вводиться по ходу изложения материала.

Сейчас же предположим, что имеется выборка $x_1,\, x_2,\, \cdots, x_n$, и поставим вопрос о нахождении закона распределения $F(x)$. Практически известны лишь результаты опытов $x_1,\, x_2,\, \cdots, x_n$, и по ним нужно найти $F(x)$. Ясно, что этой информации не достаточно, чтобы получить достоверный ответ. Речь можно вести о получении наиболее вероятного ответа. Основным принципом здесь является принцип подбора $F(x)$ из определенного семейства законов распределения, зависящего от нескольких параметров (обычно не более трех или четырех). Например, закон Пуассона зависит от одного параметра λ, нормальный закон типа $N(a,\, \sigma)$ – от двух параметров.

Чаще всего в качестве параметров берутся *теоретические моменты k-го порядка* с.в. ξ с функцией распределения $F(x)$:

$$a_k = \int_{-\infty}^{+\infty} x^k\, F'(x)\, d\, x, \quad k = 1,\, 2,\, \cdots .$$

Очевидно, что $a_1 = M\, \xi,\;\; a_2 - a_1^2 = D\, \xi$.

Что касается параметрического семейства функций, среди которых надлежит выбирать $F(x)$, то его определяют из каких-либо общих соображений, обусловленных практикой статистического моделирования реальных явлений в конкретной области техники, экономики и т.д. Например, решают, что теоретический закон является нормальным, а параметры a и σ подбирают исходя из выборочных значений. В математической статистике вводятся также понятия *эмпирической (или выборочной) функции распределения, выборочного среднего, выборочной дисперсии, выборочных моментов k-го порядка.* Удобным правилом для запоминания этих определений является следующее: образуем фиктивную "случайную величину" $\tilde{\xi}$, принимающую каждое значение $x_1,\, x_2,\, \cdots, x_n$ с вероятностью $1/n$. Тогда выборочные характеристики являются соответствующими характеристиками этой фиктивной с.в.

Например, значение *эмпирической функции распределения* в точке $x \in \mathbb{R}$ определяется формулой

$$F_n(x) = \frac{\text{число } x_i, \text{ таких, что } x_i < x}{n}.$$

Выборочным средним называется величина $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$. *Выборочная дисперсия* есть величина

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \Big(\frac{1}{n} \sum_{i=1}^{n} x_i^2\Big) - (\bar{x})^2.$$

Выборочным k-м моментом α_k называется величина

$$\alpha_k = \frac{1}{n} \sum_{i=1}^{n} (x_i)^k.$$

Важность выборочных характеристик заключается в том, что при больших n они становятся близкими к соответствующим теоретическим характеристикам. Действительно, выборочные характеристики являются функциями случайных величин, а теоретические характеристики – функциями $F_\xi(x)$ и числа, то есть не являются случайными величинами. Поэтому близость должна пониматься в смысле сходимости по вероятности. Если случайные величины $x_1, x_2, \cdots, x_n$, образующие выборку, расположить в неубывающем порядке, то получим так называемый *вариационный ряд* $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$. Здесь $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$. Для построения графика эмпирической функции распределения $F_n(x)$ удобно выписать члены вариационного ряда бесповторно и рассмотреть фиктивную дискретную с.в. ν, принимающую выборочные значения бесповторно с вероятностями $P(\nu = x_{(j)}) = k_j/n,\ j = 1, 2, \cdots, l,\ l \leqslant n$, где k_j – число повторений значения $x_{(j)}$ в исходной выборке или в вариационном ряду. График функции распределения этой фиктивной с.в. совпадает с графиком эмпирической функции распределения $F_n(x)$. Он имеет вид "лестницы с горизонтальными ступенями". Высота "ступени"в точке $x = x_{(j)}$ равна k_j/n, $j = 1, 2, \cdots, l$.

Нередко для наглядного представления выборки строят *гистограмму*. Отрезок $[x_{(1)},\ x_{(l)}]$, в котором заключены все выборочные значения, разбивают на несколько примерно равных частей $\Delta_k = [A_k, B_k],\ k = 1, 2, \cdots, m$ (m – от 5 до 25 и более, в зависимости от объема выборки) так, чтобы выборочные значения не являлись концами для $[A_k, B_k]$, $k = 1, 2, \cdots, m$. Вычисляются величины $h_k = q_k/n$, где q_k – число выборочных значений, принадлежащих $[A_k, B_k]$. Гистограмму образуют прямоугольники с основаниями $[A_k, B_k]$ на оси абсцисс и высотами h_k, $k = 1, 2, \cdots, m$. Центры верхних сторон этих прямоугольников расположены примерно на графике функции плотности вероятности генеральной совокупности, из которой извлечена исходная выборка. Для наглядности гистограммы единицу масштаба по оси ординат выбирают близкой к $x_{(l)} - x_{(1)}$. Произвольная функция $g(x_1, x_2, \cdots, x_n)$ от выборочных данных называется *статистикой*. Распределение вероятностей этой функции от с.в. $x_1, x_2, \cdots, x_n$ называется *выборочным распределением статистики* $g(x_1, x_2, \cdots, x_n)$. Статистика $g(x_1, x_2, \cdots, x_n)$, близкая в том или ином смысле к числу c – параметру генеральной совокупности, называется *точечной оценкой* этого параметра. Оценка $g(x_1,\, x_2,\, \cdots, x_n)$ параметра c называется *состоятельной оценкой* этого параметра, если при $n \to \infty$ имеем

$$g(x_1,\, x_2,\, \cdots, x_n) \to \infty$$

по вероятности. Можно доказать в общем случае, что при $n \to \infty$:

1. $F_n(x) \to F_\xi(x)$ для любого $x \in \mathbb{R}$ по вероятности;
2. $\bar{x} \to a = M\,\xi$ по вероятности;
3. $S^2 \to \sigma^2 = D\,\xi$ по вероятности. Оценка $g(x_1,\, x_2,\, \cdots, x_n)$ называется

несмещенной оценкой параметра c, если при любом n имеем

$$M\, g(x_1,\, x_2,\, \cdots, x_n) = c.$$

Таким образом, в случае несмещенной оценки отклонение $g(x_1,\, x_2,\, \cdots,$ $x_n) - c$ носит несистематический характер. Наличие смещенной оценки для математического ожидания с.в., генерируемой каким-либо измерительным прибором, нередко означает, что прибор плохо отрегулирован. Большинство практически применимых оценок являются состоятельными и несмещенными. Например, выборочное среднее $\bar{x}$ является несмещенной оценкой математического ожидания $a = M\,\xi$ теоретического закона (генеральной совокупности), поскольку

$$M\,\bar{x} = \frac{1}{n}\sum_{i=1}^{n} M\,x_i = a.$$

Учтено, что все $M\,x_i = a$.

Отметим, однако, что *выборочная дисперсия S^2 является смещенной оценкой параметра σ^2 генеральной совокупности.*

Действительно,

$$M\,(S^2) = \frac{1}{n}\sum_{i=1}^{n} M\big((x_i - a + a)^2\big) - M\,(\bar{x} - a + a)^2 =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n}\big(M\,(x_i-a)^2+2\,a\,M(x_i-a)+a^2\big)-\big(M\,(\bar{x}-a)^2+2\,a\,M(\bar{x}-a)+a^2\big) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\big(D\,x_i + a^2\big) - (\sigma^2 + a^2) = \frac{n-1}{n}\sigma^2,$$

то есть $M\,S^2 \neq \sigma^2$.

Очевидно, что несколько скорректированная статистика

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

является уже несмещенной оценкой параметра σ^2. Тем не менее статистика $s = \sqrt{s^2}$ не является несмещенной оценкой для среднего квадратичного отклонения σ генеральной совокупности. Предпочтительное использование s вместо S связано не столько со свойством несмещенности s, сколько с тем обстоятельством, что в случае нормального теоретического закона распределения именно s входит во все важнейшие формулы.

Если имеются две оценки $g_1(x_1, x_2, \cdots, x_n)$, $g_2(x_1, x_2, \cdots, x_n)$ параметра a, то более эффективной считают ту, у которой дисперсия меньше. *Несмещенная оценка g параметра с минимальной дисперсией называется эффективной.* На практике при вычислении моментов могут быть полезными следующие замечания. Пусть выборочные значения x_1, $x_2, \cdots, x_n$ группируются около некоторого числа m. Тогда выборочное среднее удобно вычислять по немного видоизмененной формуле

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m) + m,$$

а выборочную дисперсию – по формулам

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 - (\bar{x} - m)^2; \;\; s^2 = \frac{n}{n-1}S^2.$$

* Пусть, например, выборка состоит из 10 значений:

$$799,\ 801,\ 803,\ 778,\ 805,\ 800,\ 802,\ 796,\ 798,\ 805.$$

В качестве m можно взять число 800. Тогда разности $x_i - 800$ равны

$$-1,\ 1,\ 3,\ -2,\ 5,\ 0,\ 2,\ -4,\ -2,\ 4$$

соответственно. Сумма этих значений равна 6, поэтому $\bar{x} = \frac{1}{10}6 + 800 = 800.6$.

Сумма квадратов чисел $x_i - 800$ равна 76, значит $S^2 = 7.6 - (0.6)^2 = 7.24$.

§ 4.2. Оценки максимального правдоподобия

Одним из распространенных методов получения точечных оценок параметров генеральной совокупности, характеризуемой законом $F(x,\, \theta_1,\, \cdots,\, \theta_k)$, зависящим от параметров $\theta_1,\, \cdots,\, \theta_k$, путем обработки выборочных данных $x_1,\, x_2,\, \cdots,\, x_n$ является *метод максимального правдоподобия.* Для некоторого упрощения рассмотрим сначала случай, когда F зависит только от одного параметра, то есть $F = F(x,\, \theta)$. Тогда метод сводится к следующему алгоритму. На первом шаге строится *функция правдоподобия* $\Lambda(\theta,\, x_1,\, x_2,\, \cdots,x_n)$. Например, если дана плотность вероятности $p(x,\, \theta) = F'_x\,(x,\, \theta)$, то полагают

$$\Lambda(\theta,\, x_1,\, x_2,\, \cdots,x_n) = \prod_{j=1}^{n} p(x_j,\, \theta),$$

то есть Λ определяют как произведение всех функций вида $p(x_j,\, \theta)$, в которых первый аргумент совпадает с соответствующим выборочным значением. Эта функция есть не что иное, как *плотность совместного распределения с.в.* $x_1,\, x_2,\, \cdots,x_n$.

На втором шаге находят величину $\hat{\theta}_n$ как точку максимума функции $\Lambda(\theta,\, x_1,\, x_2,\, \cdots,x_n)$. Согласно математическому анализу, сначала находят стационарные точки, решая уравнение $\Lambda'_\theta = 0$, а затем среди них отбирают точку максимума. Величину $\hat{\theta}$ и принимают за искомую. При достаточно общих условиях такие оценки максимального правдоподобия оказываются состоятельными, то есть $\lim_{n\to\infty} P\{|\hat{\theta} - \theta| < \delta\} \to 1$ для любого наперед заданного $\delta > 0$, а также асимптотически нормальными.

В случае дискретных с.в. функция правдоподобия определяется несколько иначе:

$$\Lambda(\theta,\, x_1,\, x_2,\, \cdots,x_n) = \prod_{j=1}^{n} P(\theta,\, \xi = x_j).$$

Здесь вероятности событий $\{\xi = x_j\}$ предполагаются дифференцируемыми функциями одного параметра θ.

Рассмотрим два примера.

* Пусть генеральная совокупность, из которой извлечена выборка $x_1,\, x_2,\, \cdots,x_n$, распределена нормально с плотностью вероятности

$$p(\theta,\, x) = \frac{1}{\sqrt{2\,\pi}}\, e^{-(x-\theta)^2/2},$$

зависящей от параметра θ. Строим функцию правдоподобия:

$$\Lambda(\theta,\, x_1,\, x_2,\, \cdots, x_n) = \Big(\frac{1}{\sqrt{2\,\pi}}\Big)^n \prod_{j=1}^{n} e^{-(x_j-\theta)^2/2} =$$

$$= \Big(\frac{1}{\sqrt{2\,\pi}}\Big)^n e^{-(\sum_{j=1}^n (x_j-\theta)^2)/2}.$$

Вычисляя производную, получим

$$\Lambda'_\theta = \Lambda(\theta,\, x_1,\, x_2,\, \cdots, x_n)\cdot\Big(-\sum_{j=1}^{n}(x_j-\theta)\Big).$$

Условие $\Lambda'_\theta = 0$ дает

$$\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} x_j = \bar{x}.$$

Очевидно, что найденная стационарная для функции правдоподобия точка является ее единственной точкой максимума.

Итак, выборочное среднее $\bar{x}$ является оценкой максимального правдоподобия для математического ожидания θ в рассмотренной задаче.

* Теперь пусть генеральная совокупность характеризуется однопараметрической плотностью вероятности вида

$$p(\sigma,\, x) = \frac{1}{\sigma\sqrt{2\,\pi}}\, e^{-x^2/2\,\sigma^2},$$

то есть мы снова имеем дело с нормальным законом, но на этот раз типа $N(0,\, \sigma^2)$. При заданной выборке $x_1,\, x_2,\, \cdots, x_n$ функция правдоподобия представляется в виде

$$\Lambda(\sigma,\, x_1,\, x_2,\, \cdots, x_n) = \Big(\frac{1}{\sigma\sqrt{2\,\pi}}\Big)^n e^{-(\sum_{j=1}^n (x_j)^2)/2\sigma^2}.$$

Дифференцируя Λ по параметру, получим

$$\Lambda'_\sigma = -\frac{n}{\sigma(\sigma\sqrt{2\,\pi})^n}\cdot e^{-(\sum_{j=1}^n x_j^2)/2\sigma^2} -$$

$$-\Big(\frac{1}{\sigma\sqrt{2\,\pi}}\Big)^n e^{-(\sum_{j=1}^n (x_j)^2)/2\sigma^2}\cdot\Big(-\sum_{j=1}^{n}\frac{x_j^2}{\sigma^3}\Big).$$

Условие $\Lambda'_\sigma = 0$ дает

$$\sum_{j=1}^{n}\frac{x_j^2}{\sigma^3} - \frac{n}{\sigma} = 0.$$

Находим стационарную точку $\hat{\sigma^2} = \frac{1}{n}\sum_{j=1}^{n} x_j^2$, которая является единственной точкой максимума для функции правдоподобия в данном примере.

Часто удобнее брать в качестве функции правдоподобия не Λ, а $\ln\Lambda$. Поскольку $\Lambda > 0$, то $\Lambda'_\theta = 0$ тогда и только тогда, когда

$$(\ln\,\Lambda)'_\theta = \frac{\Lambda'_\theta}{\Lambda} = 0.$$

Рассмотрим еще два примера.

* Пусть генеральная совокупность задается как дискретная биномиальная случайная величина: $P(\beta_n = k) = C_n^k \cdot p^k\,(1-p)^{n-k}$, $k = 0,\,1,\,2,\,\cdots,n$, причем параметр p, $\;0 < p < 1$, не известен и подлежит оценке с помощью выборки $x_1,\,x_2,\,\cdots,x_n$. Разумеется, выборочные значения целочисленны, $0 \leqslant x_j \leqslant n$ для каждого номера j. Строим функцию правдоподобия:

$$\Lambda(p,\,x_1,\,x_2,\,\cdots,x_n) = \prod_{j=1}^{n} C_n^{x_j} p^{x_j}\,(1-p)^{1-x_j}.$$

Следовательно,

$$\ln\Lambda = \sum_{j=1}^{n}(\ln\,C_n^{x_j}) + \sum_{j=1}^{n} x_j \cdot \ln\,p + \sum_{j=1}^{n}(n - x_j)\cdot\ln(1-p).$$

Логарифмическая производная

$$(\ln\,\Lambda)'_p = \frac{1}{p}\sum_{j=1}^{n} x_j - \frac{1}{1-p}\sum_{j=1}^{n}(n - x_j).$$

Уравнение $(\ln\,\Lambda)'_p = 0$ определяет стационарную точку $\hat{p} = \bar{x}$, которая является искомой оценкой максимального правдоподобия в данном примере.

* Пусть генеральная совокупность распределена по закону Пуассона

$$P(\xi = k) = \frac{\lambda^k\,e^{-\lambda}}{k!},\;\; k = 0,\,1,\,2,\,\cdots$$

и задача состоит в получении оценки максимального правдоподобия для неизвестного параметра $\lambda = M\,\xi = D\,\xi$ по заданным выборочным значениям $x_1,\,x_2,\,\cdots,x_n$. Логарифм функции правдоподобия имеет вид

$$\ln\Lambda = \ln\Big(\prod_{j=1}^{n}\frac{\lambda^{x_j}\,e^{-\lambda}}{x_j!}\Big) =$$

$$= \ln \prod_{j=1}^{n} \frac{1}{x_j!} + \sum_{j=1}^{n} x_j \cdot \ln \lambda - n\lambda.$$

Приравнивая нулю логарифмическую производную

$$(\ln \Lambda)'_\lambda = \frac{1}{\lambda} \sum_{j=1}^{n} x_j - n,$$

получим искомую оценку максимального правдоподобия $\hat{\lambda} = \bar{x}$.

§ 4.3. Доверительные интервалы

Напомним, что согласно теореме Чебышева последовательность средних арифметических $\{\bar{\xi}_n = \frac{1}{n} \sum_{j=1}^{n} \xi_j\}$ одинаково распределенных с.в. с $M\,\xi_k = a$ и $D\,\xi_k = \sigma^2$ для всех номеров k сходится по вероятности к a при $n \to \infty$, причем $M\,\bar{\xi}_n = a$ и $D\,\bar{\xi}_n = \sigma^2/n$ для любого номера n. В частном случае биномиальной с.в. $\beta_n = \bar{\xi}_n$ результат получен Бернулли.

Часто приходится находить вероятность тех или иных возможных отклонений значений выборочной средней $\bar{x}_n$ от математического ожидания a, например, вероятность

$$P(|\bar{x}_n - a| \leqslant \delta), \tag{1}$$

где δ – какое-нибудь положительное число. Неравенство $|\bar{x}_n - a| \leqslant \delta$ или $a - \delta \leqslant \bar{x}_n \leqslant a + \delta$ называется *доверительной оценкой* для $\bar{x}_n$, интервал $(a - \delta,\, a + \delta)$ – *доверительным интервалом*, а вероятность (1) – *надежностью* соответствующей доверительной оценки.

Воспользуемся центральной предельной теоремой для построения доверительных интервалов с заданной надежностью. Имеет место приближенная формула

$$P(|\bar{x}_n - a| \leqslant \delta) \approx \Phi(t_2) - \Phi(t_1). \tag{2}$$

Здесь

$$t_1 = \frac{(a-\delta) - a}{\sigma\sqrt{n}} = -\frac{\delta\sqrt{n}}{\sigma}, \;\; t_2 = \frac{(a+\delta) - a}{\sigma\sqrt{n}} = \frac{\delta\sqrt{n}}{\sigma}.$$

В случае интегральной теоремы Муавра – Лапласа $\sigma = \sqrt{n\,p\,q}$, $a = n\,p$. В силу нечетности функции Лапласа

$$\Phi(t_2) - \Phi(t_1) = \Phi\Big(\frac{\delta\sqrt{n}}{\sigma}\Big) - \Phi\Big(-\frac{\delta\sqrt{n}}{\sigma}\Big) = 2\Phi\Big(\frac{\delta\sqrt{n}}{\sigma}\Big).$$

Стало быть,

$$P(|\bar{x}_n - a| \leqslant \delta) \approx 2\Phi(t), \text{ где } t = \frac{\delta\sqrt{n}}{\sigma}. \tag{3}$$

Пренебрегая погрешностью в интегральной теореме Муавра – Лапласа, для относительной частоты ν_n в серии из n испытаний Бернулли имеем формулу

$$P(|\nu_n - p| \leqslant \delta) = 2\Phi(t), \text{ где } t = \frac{\delta\sqrt{n}}{\sqrt{p\,q}}. \tag{4}$$

Формулы (3) и (4) можно рассматривать как уравнения, связывающие три величины – объем выборки (число испытаний) n, величину предельного отклонения δ и надежность доверительной оценки $\beta = P(|\bar{x}_n - a| \leqslant \delta)$ (соответственно $P(|\nu_n - p| \leqslant \delta)$ в формуле (4)). Зная две из этих величин, с помощью (3) либо (4) можно находить третью величину. То есть можно отвечать на вопросы трех типов:

1. Какова надежность β того, что при данном объеме n выборки отклонение $|\bar{x}_n - a|$ не превзойдет заданного числа δ?

2. Какая величина δ предельного отклонения $\bar{x}_n$ от a гарантируется с заданной надежностью β при данном объеме выборки n?

3. Каков минимальный объем выборки, при котором с данной надежностью β можно гарантировать, что отклонение $|\bar{x}_n - a|$ не превзойдет заданного числа δ?

Рассмотрим числовые примеры для каждого из трех типов задач применительно к формуле (4), считая во всех примерах вероятность успеха $p = 0.2$.

* Найти надежность β при $n = 10000$, $\delta = 0.003$. Имеем $t = 0.003 \cdot \sqrt{10000/0.2 \cdot 0.8} = 0.75$. Поэтому $\beta = 2\Phi(0.75) \approx 2 \cdot 0.2734 = 0.5468$.

* Найти δ при $n = 10000$, $\beta = 0.95$. Имеем $2\Phi(t) = \beta = 0.95$, $\Phi(t) = 0.475$. По таблице значений функции Лапласа находим $t \approx 1.96$. Следовательно,

$$\delta = t \cdot \sqrt{\frac{p(1-p)}{n}} = 1.96 \cdot \sqrt{\frac{0.2 \cdot 0.8}{10000}} = 0.0078.$$

* Найти n при $\delta = 0.02$, $\beta = 0.99$. Имеем $2\Phi(t) = 0.99$, $\Phi(t) = 0.495$. По таблице значений функции Лапласа находим $t = 2.576$. Стало быть,

$$n = p\,q\frac{t^2}{\delta^2} = 0.2 \cdot 0.8 \cdot \frac{2.576^2}{0.02^2} = 2654.$$

Предположим, что дана выборка объема n из генеральной совокупности, характеризуемой нормальным законом типа $N(a,\,\sigma^2)$, причем дисперсия

σ^2 известна, а параметр a подлежит оценке. Можно показать, что выборочное среднее $\bar{x}$ также имеет нормальное распределение с параметрами $M\,\bar{x} = a,\;\; D\,\bar{x} = \sigma^2/n$. Поэтому

$$P(|\bar{x} - a| \leqslant \delta) = 2\Phi(t), \text{ где } t = \delta\frac{\sqrt{n}}{\sigma}. \tag{5}$$

Особенность доверительной оценки $|\bar{x} - a| \leqslant \delta$ в данном случае состоит в том, что параметр a неизвестен, но зато известно выборочное среднее $\bar{x}$. Доверительным интервалом для a служит $(\bar{x} - \delta,\, \bar{x} + \delta)$, а формула (5) дает вероятность (надежность) того, что в результате испытаний получится такое выборочное среднее $\bar{x}$, для которого указанный интервал будет содержать оцениваемый параметр a.

* Пусть $\sigma = 0.01$. Выборка объема 5 из нормальной генеральной совокупности с неизвестным $a = M\,\xi$ имеет вид 4.781, 4.795, 4.769, 4.792, 4.779. Находим выборочное среднее: $\bar{x} = 4.7832$. Пусть доверительная вероятность $\beta = 0.95$. Тогда $\Phi(t_1) = 0.475$ и по таблице значений функции Лапласа $t_1 = 1.96$. Поэтому $\delta_1 = t_1\,\sigma/\sqrt{n} = 0.0088$. Искомый доверительный интервал – (4.774, 4.792).

Если увеличить надежность, взяв $\beta = 0.99$, то получим $\Phi(t_2) = 0.495$, $t_2 = 0,576$, $\delta_2 = 0.0115$. Доверительный интервал станет шире: (4.7717, 4.7947). Ниже нам потребуется знакомство с тремя новыми видами непрерывных случайных величин. Все они тесно связаны с нормальными с.в.

Распределением Пирсона или распределением χ^2 (читается "хи квадрат") с n степенями свободы называется распределение случайной величины

$$\chi_n^2 = \xi_1^2 + \xi_2^2 + \cdots + \xi_n^2,$$

где $\xi_1,\, \xi_2,\, \cdots, \xi_n$ – независимые с.в., каждая из которых имеет нормальное распределение типа $N(0,\, 1)$. *Распределением Стьюдента или t-распределением с n степенями свободы* называется распределение с.в.

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}},$$

где ξ – нормальная с.в. типа $N(0,\, 1)$, а χ_n^2 – распределенная по Пирсону с.в. с n степенями свободы, не зависящая от ξ. *Распределением Фишера (F-распределением, распределением v^2) с (m, n) степенями свободы* называется распределение случайной величины

$$F_{m,\,n} = \frac{\frac{1}{m}\chi_m^2}{\frac{1}{n}\chi_n^2},$$

где в числителе и знаменателе стоят независимые с.в., имеющие распределения Пирсона с m и n степенями свободы соответственно.

При обработке наблюдений нередко приходится иметь дело с таблицами введенных здесь распределений. Наоборот, явные выражения для их плотностей почти не используются, и поэтому мы их не приводим. При необходимости следует обратиться к более полным учебникам математической статистики и справочникам [2; 6; 9; 12; 13] . Сокращенные таблицы распределений Стьюдента и Пирсона приведены в Приложении.

Дадим краткую справку о приближенных выражениях этих распределений при большом числе степеней свободы. Из определения видно, что пирсоновская и фишеровская с.в. принимают неотрицательные значения. Известно, что $M\,\chi_n^2 = n,\;\; D\chi_n^2 = 2\,n$. Нормированная с.в. $\frac{\chi_n^2 - n}{\sqrt{2\,n}}$ при $n \to \infty$ асимптотически нормальна и имеет тип $N(0,\,1)$. Лучшей в данном случае является аппроксимация $\frac{\chi_n^2 - n}{\sqrt{2\,n-1}}$. Отметим еще, что сумма двух независимых с.в. χ_n^2 и χ_m^2 является пирсоновской с.в. типа χ_{m+n}^2.

Далее, в силу закона больших чисел с.в. χ_n^2/n по вероятности стремится к 1. Поэтому пределом с.в. Стьюдента t_n по вероятности является нормальная с.в. типа $N(0,\,1)$. Обычно считают, что при $n > 30$ практически без существенной погрешности вместо t_n можно пользоваться распределением $N(0,\,1)$. Это правило не действует, когда p очень близка к 0 или 1.

Гораздо сложнее обстоит дело с табулированием распределения Фишера уже по той причине, что $F_{m,\,n}(x)$ зависит от трех аргументов $m,\,n,\,x$ и таблица должна иметь три входа. В доступной литературе [6; 9] имеются лишь таблицы процентных точек F-распределения, то есть таблицы точек $x(m,\,n,\,\alpha)$, таких, что $P\{F_{m,\,n} < x(m,\,n,\,\alpha)\} = \alpha$ для нескольких значений α. Обычно этого бывает достаточно.

Вернемся к вопросу о построении доверительных интервалов. Пусть имеется выборка $x_1,\,x_2,\,\cdots,x_n$ из генеральной совокупности, характеризуемой нормальным законом типа $N(a,\,\sigma^2)$, причем a и σ^2 не известны. Поиск доверительных интервалов для этих параметров в этой ситуации осложняется. Воспользуется теоретическим результатом, утверждающим, что статистика

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\bar{x} - x_j)^2$$

такова, что нормированная с.в.

$$\tau_n = \frac{\bar{x} - a}{s/\sqrt{n}}$$

распределена по закону Стьюдента, не зависящему от a и σ.

Пусть $p(x,\, n)$ – плотность вероятности статистики τ_n, $\mathcal{F}(x,\, n)$ – ее интегральная функция распределения,

$$\mathcal{F}_o(x,\, n) = \mathcal{F}(x,\, n) - 0.5 = \int_0^x p(t,\, n)\, dt \;-$$

аналог функции Лапласа для распределения Стьюдента. Тогда

$$P(|\tau_n| \leqslant t) = 2\int_0^t p(x,\, n)\, dx = 2\,\mathcal{F}_o(t,\, n).$$

Оценку $|\tau_n| \leqslant t$ можно записать в виде $|\bar{x} - a| \leqslant \frac{s\,t}{\sqrt{n}}$. Стало быть,

$$P(|\bar{x} - a| \leqslant \delta) = 2\mathcal{F}_o(t,\, n), \;\text{ где }\; t = \frac{\delta\sqrt{n}}{s}.$$

Используя известное значение $\bar{x}$, находим $|\bar{x} - a| \leqslant \delta$, то есть $\bar{x} - \delta \leqslant a \leqslant \bar{x} + \delta$. Таким образом, построен доверительный интервал $(\bar{x} - \delta,\, \bar{x} + \delta)$, с надежностью $\beta = 2\,\mathcal{F}_o(t,\, n)$ накрывающий параметр a, где $t = \frac{\delta\sqrt{n}}{s}$. Для удобства расчетов в Приложении 2 приведена таблица значений t, таких, что

$$2\,\mathcal{F}_o(t,\, n) = \beta,$$

где β – близкая к 1 заданная надежность оценки.

* Рассмотрим выборку объема 5 из предыдущего примера. Кроме найденного ранее значения $\bar{x} = 4.7832$, нам понадобится величина

$$s = \sqrt{\frac{1}{4}\sum_{j=1}^{5}(x_j - \bar{x})^2} = 0.0105.$$

Возьмем $\beta = 0.99$. По табл. II Приложения находим $t = 4.60$. Поэтому $\delta = t\, s/\sqrt{5} = 0.0216$ и доверительным интервалом будет $(4.7616,\ 4.8048)$.

Он является более широким, чем в предыдущем примере, где σ была известна.

Обсудим схему построения доверительного интервала для неизвестного значения дисперсии σ^2 нормально распределенной генеральной совокупности. Воспользуемся известным утверждением о том, что если статистика $n\,s^2/\sigma^2$ распределена по закону Пирсона с $n-1$ степенями свободы, то этот закон не зависит от среднего a генеральной совокупности. Пусть β – назначенная доверительная вероятность, близкая к 1. Выберем числа x_1, x_2 так, чтобы

$$P\Big(\frac{n\,s^2}{\sigma^2} \leqslant x_1\Big) = P\Big(\frac{n\,s^2}{\sigma^2} \geqslant x_2\Big) = \frac{1}{2}(1-\beta).$$

Численные значения x_1, x_2 находятся по табл. III в приложении для χ^2-распределения с числом степеней свободы $(n-1)$ и с входным значением $\alpha/2 = \frac{1}{2}(1-\beta)$. Как только найдены x_1, x_2, можно записать

$$P\Big(x_1 < \frac{n\,s^2}{\sigma^2} < x_2\Big) = \beta.$$

Неравенства $x_1 < \frac{n\,s^2}{\sigma^2} < x_2$ и $\frac{n\,s^2}{x_2} < \sigma^2 < \frac{n\,s^2}{x_1}$ равносильны, поэтому

$$P\Big(\frac{n\,s^2}{x_2} < \sigma^2 < \frac{n\,s^2}{x_1}\Big) = P\Big(s\sqrt{\frac{n}{x_2}} < \sigma < s\sqrt{\frac{n}{x_1}}\Big) = \beta.$$

Это означает, что параметр σ имеет доверительный интервал $(s\sqrt{n/x_2},\ s\sqrt{n/x_1})$.

Отметим, что этот интервал не зависит от a, поскольку распределение статистики $n\,s^2/\sigma^2$ не зависит от a.

* Рассмотрим следующую выборку объема 30 из нормально распределенной генеральной совокупности:

$$-1.90,\ 1.37,\ -0.89,\ -0.13,\ 0.15,\ -0.79,\ -0.96,\ 1.55,\ 0.40,\ 0.69,$$
$$-0.90,\ 0.15,\ 0.90,\ 0.82,\ 1.53,\ -0.34,\ 0.98,\ -1.38,\ 1.48,\ -0.65,$$
$$1.10,\ 0.30,\ -0.13,\ -1.90,\ -0.32,\ -0.42,\ 0.77,\ 0.08,\ 0.17,\ 0.87.$$

Вычисление величин $\bar{x}$ и s^2 по указанным ранее формулам дает следующие результаты: $\bar{x} = 0.087$, $s^2 = 0.949$. Назначим надежность $\beta = 0.9$. Тогда $\gamma = \alpha/2 = 0.05$ и таблица χ^2-распределения с $n-1 = 29$ степенями свободы дает $x_1 = 17.7$, $x_2 = 42.6$. Следовательно, доверительный интервал для σ, отвечающий данной выборке и $\beta = 0.9$, имеет границы:

$$\sqrt{n\,s^2/x_2} \approx 0.80; \quad \sqrt{n\,s^2/x_1} \approx 1.25.$$

§ 4.4. Статистическая проверка гипотез

Одна из наиболее часто встречающихся на практике задач, связанных с применением статистических методов, состоит в принятии решения о том, следует ли на основании данной выборки принять или отвергнуть некоторое предположение (гипотезу) относительно параметров генеральной совокупности или ее теоретического закона. Примеры такого сорта вопросов могут быть приведены из самых разных областей практической или научной деятельности. Допустим, что для лечения некоторой болезни предложено новое лекарство, уже испытанное на некотором числе больных. Спрашивается, можно ли по данным о результатах такого лечения сделать обоснованное заключение о том, что новое лекарство более эффективно, чем применявшиеся ранее методы лечения, или же следует признать, что наблюдаемая разница в результатах лечения при старом и новом методах лежит в допустимых границах и носит чисто случайный характер? Рассматривая данные о результатах лечения как выборку из некоторой генеральной совокупности, нужно по этой выборке принять или отклонить гипотезу о том, что новое лекарство более эффективно, чем старые методы лечения.

Аналогичная задача возникает в агрономической практике, когда по урожаю на опытных участках следует сделать заключение о целесообразности тех или иных нововведений в методы обработки почвы или ухода за посевами. Тот же вопрос возникает перед руководителями промышленного предприятия, когда они должны принять решение об изменении технологического процесса или признать техническую целесообразность (нецелесообразность) замены одного оборудования другим.

Схема постановки задачи статистической проверки гипотез выглядит следующим образом. Относительно некоторой генеральной совокупности выдвигается та или иная гипотеза H. Из этой генеральной совокупности извлекается выборка $x_1,\ x_2,\ \cdots, x_n$. Требуется *указать правило, при помощи которого можно было бы по каждой данной выборке решить вопрос о принятии или отклонении гипотезы* H.

* В теории надежности для многих классов изделий установлен экспоненциальный закон отказов: если τ – время безотказной работы изделия (прибора, автомобиля, самолета, вертолета), то $P(\tau > x) = e^{-x/T}$, где $x \geqslant 0$ и T – среднее время безотказной работы изделия. Стало быть, вероятность Q безотказной работы изделия в течение времени T_0 равна $e^{-T_0/T}$.

Допустим, что техническими условиями эксплуатации изделия предусмотрено обеспечение неравенства $Q \geqslant 1 - \varepsilon$, где ε – малое положительное число. Это неравенство равносильно такому:

$$T \geqslant \frac{T_0}{\ln \frac{1}{1-\varepsilon}} =: T_1.$$

Таким образом, в качестве подлежащей проверке гипотезы можно взять $H = \{F_\tau(x) = 1 - e^{-x/T}, T \geqslant T_1\}$. Альтернативная гипотеза $\bar{H} = \{F_\tau(x) = 1 - e^{-x/T}, T < T_1\}$. Продолжим обсуждение вопроса о статистической проверке гипотез. Пусть при гипотезе H известно распределение вероятностей некоторой статистики $g = g(x_1, x_2, \cdots, x_n)$ и, в частности, пусть для данного значения вероятности $\beta = 1 - \alpha$, близкого к 1, существует такой интервал (t_1, t_2), что

$$P\{g \in (t_1, t_2)\} = P\{t_1 < g(x_1, x_2, \cdots, x_n) < t_2\} = \beta. \tag{6}$$

Предположим α столь малым числом, что мы решаем скорее отказаться от гипотетического равенства (6) и тем самым отвергнуть гипотезу H, чем допустить возможность события $g \notin (t_1, t_2)$ столь малой вероятности:

$$P\{g \notin (t_1, t_2)\} = \alpha. \tag{7}$$

Пусть теперь из рассматриваемой генеральной совокупности извлечена выборка $x_1, x_2, \cdots, x_n$, для которой фактически оказалось, что численное значение статистики $g(x_1, x_2, \cdots, x_n)$ лежит вне интервала (t_1, t_2), то есть пусть произошло практически невозможное событие. В соответствии с изложенным в этом случае принимается решение о том, что *гипотеза H должна быть отвергнута.* Если же численное значение статистики $g(x_1, x_2, \cdots, x_n)$ оказалось лежащим в интервале (t_1, t_2), то выборка $x_1, x_2, \cdots, x_n$ считается *не противоречащей гипотезе H или согласующейся с гипотезой H*. Говорят также, что в этом случае выборка $x_1, x_2, \cdots, x_n$ *подтверждает гипотезу.*

Следует отчетливо представлять себе, что заключения об отклонении или принятии гипотезы H на основании статистического материала $x_1, x_2, \cdots, x_n$ не могут быть достоверными на все 100%. Если руководствоваться описанным выше правилом, то в принципе не исключено, что при справедливости гипотезы H все-таки осуществится событие $g \notin (t_1, t_2)$, которое влечет за собой отклонение гипотезы H. Вероятность этого события равна α. Такого рода ошибка, когда отклоняется верная в действительности гипотеза H, называется *ошибкой первого рода.* Может также случиться, что численное значение статистики g окажется принадлежащим интервалу $(t_1,\, t_2)$, что повлечет за собой принятие гипотезы H, тогда как на самом деле вместо гипотезы H справедлива какая-либо другая гипотеза H^*, не сильно изменяющая равенство (6). Такого типа ошибка, когда принимается ложная в действительности гипотеза H, называется *ошибкой второго рода.* Наконец, возможен и четвертый вариант, когда отклоняется ложная в действительности гипотеза. Понятно, что статистика $g(x_2,\, x_2,\, \cdots, x_n)$ тем с большим основанием может быть использована как инструмент для проверки гипотезы H, чем меньшее значение будут иметь вероятности ошибок обоих родов.

Изложенное правило проверки гипотез называется *критерием значимости*, а применяемая при этом статистика $g(x_1,\, x_2,\, \cdots, x_n)$ часто называется просто *критерием.* Вероятность α ошибки первого рода называется *уровнем значимости данного критерия.* Обычно уровень значимости в зависимости от требуемой точности принимается равным 0.1, 0.05 или 0.01. Нередко, особенно в литературе прикладного характера, уровень значимости выражается в процентах. Только что названные три значения α будут обозначаться соответственно как десяти-, пяти- или однопроцентный уровень значимости. Область значений статистики g, лежащая вне интервала $(t_1,\, t_2)$, называется *критической областью.* Интервал $(t_1,\, t_2)$ называется *областью допустимых значений.*

В рамках данного учебного пособия не может быть изложена теория построения критериев для статистической проверки гипотез. Рассмотрим пример теоретического характера.

* Пусть имеются две генеральные совокупности с законами типа $N(a_1,\, \sigma_1)$ и $N(a_2,\, \sigma_2)$. Необходимо проверить гипотезу о том, что $\sigma_1 = \sigma_2$. Извлекаются две независимые выборки $x_1,\, x_2,\, \cdots, x_n$ и $y_1,\, y_2,\, \cdots, y_m$, по одной из каждой генеральной совокупности. Гипотезу удобно записать в виде $\sigma_1/\sigma_2 = 1$ и применить для ее проверки рапределение Фишера. Рассмотрим отношение

$$\frac{s_x^2}{s_y^2} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}{\frac{1}{m-1}\sum_{j=1}^{m}(y_j-\bar{y})^2}.$$

Числитель имеет распределение $\frac{\sigma_1^2}{n-1}\chi^2_{n-1}$, а знаменатель – распределение $\frac{\sigma_2^2}{m-1}\chi^2_{m-1}$. Если гипотеза верна, то отношение s_x^2/s_y^2 имеет распределение Фишера $F_{n-1,\, m-1}$. Назначают уровень значимости α и с помощью таблиц распределения Фишера находят область $(t_1,\, t_2)$ допустимых значений для этого распределения. Если окажется, что величина s_x^2/s_y^2 попала в эту область, то принимается решение о том, что с надежностью $\beta = 1 - \alpha$ гипотеза $\sigma_1 = \sigma_2$ верна. В противном случае гипотеза отвергается. Во многих случаях закон распределения генеральной совокупности не известен и требуется по выборке $x_1, x_2, \cdots, x_n$ решить вопрос о том, допустимо ли считать это неизвестное распределение совпадающим с тем или иным гипотетическим распределением. Критерии значимости для проверки таких гипотез о распределении генеральной совокупности носят специальное название *критериев согласия.* Остановимся на двух критериях такого рода. Первый из них – это критерий χ^2, тесно связанный с χ^2-распределением Пирсона.

Пусть проверяется гипотеза о том, что исследуемая генеральная совокупность имеет данную функцию распределения $F(x)$. Все множество значений x генеральной совокупности разбивается каким-либо образом на некоторое число промежутков $\Delta_j,\ j = 1, 2, \cdots, k$. Символом p_j обозначим вероятность принадлежности выборочного значения x промежутку Δ_j. При этом все промежутки Δ_j должны быть выбраны таким образом, чтобы все вероятности p_j были положительны и их сумма равнялась 1.

Пусть, далее, выборка $x_1, x_2, \cdots, x_n$ содержит n_j своих значений в промежутке $\Delta_j,\ j = 1, 2, \cdots, k$. Если объем выборки достаточно велик, то по теореме Муавра – Лапласа случайные величины n_j будут приближенно распределены нормально с математическим ожиданием $n\,p_j$ и дисперсией $n\,p_j(1-p_j)$. Нормальным типа $N(0,\,1)$ будет (приближенно) и распределение с.в. $\frac{n_j - n\,p_j}{\sqrt{n\,p_j}}$ для каждого из номеров $j = 1, 2, \cdots, k$. Принимая во внимание данное ранее определение χ^2-распределения, можно ожидать, что сумма

$$\xi = \sum_{j=1}^{k} \frac{(n_j - n\,p_j)^2}{n\,p_j} \tag{8}$$

будет иметь асимптотически при $n \to \infty$ (приближенно при достаточно больших n) χ^2-распределение. Это действительно так, и соответствующая теорема гласит, что предельное распределение с.в. ξ в (8) совпадает с χ^2-распределением при числе степеней свободы $k-1$. Уменьшение на единицу числа степеней свободы по сравнению с числом слагаемых в правой части (8) объясняется тем, что с.в. n_j не являются независимыми, а связаны соотношением $n_1 + n_2 + \cdots + n_k = n$.

На основе сформулированной теоремы строится следующий критерий согласия. По выбранному уровню значимости α находится значение x_α, такое, что $P(\chi^2 > x_\alpha) = \alpha$. Принимается, что величина ξ в формуле (8) имеет χ^2-распределение с $(k-1)$-й степенью свободы. Если окажется, что вычисленное по выборке значение величины ξ из формулы (8) попало в область допустимых значений, то есть $\xi \leqslant x_\alpha$, то выборка считается подтверждающей гипотезу о том, что функцией распределения исследуемой генеральной совокупности является $F(x)$. Если же указанное значение ξ окажется в критической области, то есть $\xi > x_\alpha$, то сформулированная гипотеза отвергается.

По поводу условий применимости критерия χ^2 следует сделать замечания. Во–первых, нужно иметь в виду, что принятие χ^2-распределения для с.в. ξ, определяемой формулой (8), считается допустимым, если для всех значений j выполняются неравенства $n\,p_j \geqslant 10$.

Далее, функция $F(x)$ считается вполне заданной, то есть в ее аналитическом выражении отсутствуют неизвестные параметры.

Если же проверяемая гипотеза формулируется как гипотеза о принадлежности функции распределения генеральной совокупности заданному параметрическому семейству (например, семейству $N(a,\,\sigma^2)$) без фиксации значений параметров, то применение χ^2-критерия согласия требует предварительной подготовки. А именно по имеющейся выборке $x_1,\, x_2,\, \cdots, x_n$ находятся оценки для неизвестных параметров, которые затем используются вместо истинных параметров. Относительно возникающего закона $F^*(x)$ и формулируется гипотеза и применяется χ^2-критерий согласия для ее принятия или отклонения. Естественно, требования к малости уровня значимости α в этой ситуации должны быть повышены.

Наконец, нужно помнить о том, что приведенные сведения первичны и не выделяют ряд тонких моментов, о которых полезно знать тем, кто намерен воспользоваться критериями согласия в своей практической деятельности. Читатель должен тогда обратиться к более полным руководствам по теории вероятностей и математической статистике. Дадим в заключение краткую справку о критерии согласия, построенном А.Н. Колмогоровым. Он позволяет принять либо отклонить гипотезу о том, что функцией распределения исследуемой генеральной совокупности является заданная непрерывная функция $F(x)$ (неизвестных параметров не содержащая). Критерий основан на теореме Колмогорова об асимптотике при $n \to \infty$ статистики

$$D_n = \max |F(x) - F_n(x)|,$$

где $F_n(x)$ – эмпирическая функция распределения, построенная по выборке $x_1, x_2, \cdots, x_n$ данного объема n и максимум (точнее, супремум) берется по всем $x \in \mathbb{R}$.

А.Н. Колмогоров доказал, что для любой функции распределения $F(x)$ непрерывной с.в. справедливо предельное соотношение

$$P\Big\{D_n < \frac{x}{\sqrt{n}}\Big\} \to K(x) \tag{9}$$

при $n \to \infty$, где x – произвольное положительное число, а функция Колмогорова представляется суммой ряда:

$$K(x) = 1 - 2 \cdot \sum_{j=1}^{\infty} (-1)^{j-1} \cdot e^{-2(j\,x)^2}.$$

При $x \leqslant 0$ функция Колмогорова равна нулю.

Функция $K(x)$ табулирована ввиду ее важности для практики. Схема применения теоремы Колмогорова к построению критерия согласия находится в полном соответствии с изложенным выше материалом применительно к критерию Пирсона в непараметрическом случае. Еще раз отметим, что гипотетическая функция распределения $F(x)$ должна быть непрерывной.

§ 4.5. Сглаживание экспериментальных зависимостей методом наименьших квадратов

Проводя прикладные геодезические и астрономические исследования, К.Ф. Гаусс столкнулся со следующей задачей. Пусть требуется найти формулу, выражающую функциональную связь $y = f(x)$ между переменными величинами x и y. С этой целью экспериментально измеряются соответствующие друг другу пары чисел (x_j, y_j), $j = 1, 2, \cdots, n$. Эти пары чисел можно изобразить точками в прямоугольной декартовой системе координат. Произвольная гладкая или кусочно-гладкая кривая, проведенная через эти точки, может служить графиком функциональной зависимости между x и y. Приближениями могут служить и кривые, не обязательно проходящие через все точки, поскольку вследствие неизбежных ошибок в измерениях экспериментальные точки $P_j = (x_j, y_j)$

могут и не принадлежать графику истинной зависимости. Возникает вопрос об оптимальном (наилучшем в каком-либо смысле) и конструктивном способе построения исследуемой зависимости. Можно, конечно, подобрать формулу $y = f^*(x)$ в виде интерполяционного многочлена степени $n-1$, график которого непременно проходит через все точки P_j. Но высказанное соображение об ошибках эксперимента подсказывает нам, что скорее всего истинный график будет проходить вблизи точек P_j.

Несколько идеализируя ситуацию, предположим, что все числа x_j мы умеем находить абсолютно точно. Пусть искомая функциональная зависимость $y = f(x)$ принадлежит параметрическому семейству $y = F(x, a, b, \cdots, c)$ непрерывных функций. Тогда при произвольно заданных значениях параметров $a, b, \cdots, c$ можно вычислить $F(x_j, a, b, \cdots, c)$. Вследствие опытных ошибок при измерении чисел y_j некоторые разности $\delta_j = y_j - F(x_j, a, b, \cdots, c)$ могут быть отличными от нуля. Идея метода наименьших квадратов, разработанного Гауссом, состоит в данном случае в выборе параметров $a, b, \cdots, c$ из условия минимума функции $\Phi(a, b, \cdots, c) = \sum_{j=1}^n \delta_j^2$ (суммы квадратов "невязок" δ_j).

Итак, рассматривается задача минимизации функции

$$\Phi(a, b, \cdots, c) = \sum_{j=1}^{n} \big(y_j - F(x_j, a, b, \cdots, c)\big)^2.$$

Так как Φ принимает неотрицательные значения и непрерывна, то минимум достигается. Стационарные точки, среди которых находятся искомые точки минимума, определяются из условия обращения в нуль всех частных производных функции Φ. Таким путем приходим к системе *нормальных уравнений*:

$$\sum_{j=1}^{n} \big(y_j - F(x_j, a, b, \cdots, c)\big)\frac{\partial F(x_j, a, b, \cdots, c)}{\partial a} = 0,$$

$$\sum_{j=1}^{n} \big(y_j - F(x_j, a, b, \cdots, c)\big)\frac{\partial F(x_j, a, b, \cdots, c)}{\partial b} = 0,$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$\sum_{j=1}^{n} \big(y_j - F(x_j, a, b, \cdots, c)\big)\frac{\partial F(x_j, a, b, \cdots, c)}{\partial c} = 0.$$

Для наиболее часто встречающихся на практике типов параметрических зависимостей $F(x, a, b, \cdots, c)$ эта система имеет единственное решение

$a,\ b,\ \cdots,c$. На нем достигается минимум функции невязок. Рассмотрим нормальные уравнения для некоторых частных случаев.

Для формулы типа $F(x,\ a,\ b) = a\,x+b$ нормальные уравнения примут вид

$$\sum_{j=1}^{n}(y_j - a\,x_j - b)\,x_j = 0,$$

$$\sum_{j=1}^{n}(y_j - a\,x_j - b) \;=\; 0.$$

После приведения подобных членов получим

$$a\sum_{j=1}^{n} x_j + b\,n \;=\; \sum_{j=1}^{n} y_j,$$

$$a\sum_{j=1}^{n} x_j^2 + b\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} x_j\,y_j. \tag{10}$$

* Пусть опытным путем найдены следующие 6 точек: (0.5, 0.31), (1.0, 0.82), (1.5, 1.29), (2.0, 1.85), (2.5, 2.51), (3.0, 3.02). Нетрудно видеть, что все они располагаются примерно на одной прямой линии, поэтому в качестве F выбираем линейную функцию от x.

Нормальная система (10) в данном случае примет вид

$$10.5\,a + 6\,b = 9.8, \quad 22.75\,a + 10.5\,b = 21.945.$$

Ее решением будет $a = 1.096,\ b = -0.285$. Искомая формула $y = 1.096$, x-0.285. Если экспериментальные точки лежат на параболе, то выбирают $F = a\,x^2 + b\,x + c$. Соответствующая система нормальных уравнений

$$\sum_{j=1}^{n}(y_j - a\,x_j^2 - b\,x_j - c) \;=\; 0,$$

$$\sum_{j=1}^{n}(y_j - a\,x_j^2 - b\,x_j - c)x_j \;=\; 0,$$

$$\sum_{j=1}^{n}(y_j - a\,x_j^2 - b\,x_j - c)x_j^2 = 0$$

снова линейна и имеет единственное решение.

Если экспериментальные точки лежат примерно на гиперболе, асимптоты которой параллельны осям координат, то можно выбрать $F = a + b/(x - c)$. Нередко путем подходящей замены переменных удается преобразовать экспериментальную зависимость к виду $F = a\,x + b$. Например, формула типа $y = a\,e^{b\,x}$ равносильна формуле $Y = A\,x + B$, где

$$Y = \lg\, y,\ \ A = b\,\lg\, e,\ \ B = \lg\, a. \tag{11}$$

Постоянные A, B зависимости $Y = A\,x + B$ находятся из уравнений (10), в которых вместо a, b, y_j следует брать соответственно A, B, Y_j. После отыскания значений A и B находят a и b с помощью формул (11) и выписывают искомую зависимость.

* Опытные данные помещены в двух первых строках следующей таблицы:

x :	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2,
y :	30.4	25.0	21.0	17.5	14.5	11.8	9.9	8.5,
Y :	1.48	1.40	1.32	1.24	1.16	1.07	1.00	0.93,

где $Y_j = \lg\, y_j$, $\ j = 1,\, 2,\, \cdots, n$. Несложно проверить, что в координатах x, $Y = \lg y$ точки $(x_j,\, Y_j)$ расположены примерно на одной прямой линии. Поэтому выбираем параметрическую зависимость $F = a\,e^{b\,x}$. Значения A, B находятся из системы вида (10), которая в данном случае выглядит так:

$$12\,A + 8\,B = 9.61,\ \ 19.68\,A + 12\,B = 13.74.$$

Отсюда $A \approx -0.40$, $\ B \approx 1.80$. Из формул (11) получаем $a \approx 63.2$, $\ b \approx -0.92$. Искомая формула –

$$y = 63.2\,e^{-0.92\,x}.$$

Метод наименьших квадратов широко применяется и в других вопросах. Его теоретическое обоснование основано на принципе максимального правдоподобия (см. первые два примера в § 4.2). Если считать, что измерительный прибор генерирует нормальную выборку $y_1,\, y_2,\, \cdots, y_n$, то сумма квадратов невязок является зависящей от нескольких параметров случайной величиной, которой отвечает определенная функция правдоподобия. Условие минимума суммы квадратов невязок равносильно условию максимума функции правдоподобия. Подробности можно найти в литературе. Еще одним примером применения метода наименьших квадратов может служить способ *регуляризации систем алгебраических уравнений по А.Н. Тихонову*. Идея заключается в следующем. Пусть имеется система линейных алгебраических уравнений (СЛАУ)

$$\sum_{i=1}^{n} a_{ij}\, x_i \ = \ b_j, \ \ j = 1,\, 2,\, \cdots, m. \tag{12}$$

Коэффициенты системы (12) считаем полученными опытным путем. Они, конечно, найдены с некоторыми ошибками – нормально распределенными случайными величинами. Анализ средствами классической линейной алгебры вполне может указать на несовместность СЛАУ (12). Но ведь точное решение этой системы нам знать и не нужно, поскольку практический интерес представляет некоторое приближенное ее решение, отвечающее физическому смыслу исходной задачи. Чтобы найти такое приближенное решение, рассмотрим сумму квадратов невязок для системы (12):

$$\Psi(x_1,\, x_2,\, \cdots, x_n) = \sum_{j=1}^{m} \Big(b_j - \sum_{i=1}^{n} a_{ij}\, x_i \ \Big)^2 . \tag{13}$$

Набор координат точки $(x_1^o,\, x_2^o,\, \cdots, x_n^o)$ минимума функции Ψ называется *приближенным решением системы (12) по Тихонову*. Для его нахождения выписывается соответствующая функции Ψ система нормальных уравнений

$$\frac{\partial \Psi}{\partial x_i} = 0, \ \ i = 1,\, 2,\, \cdots, n. \tag{14}$$

Ее принято называть *регуляризованной по Тихонову системой* для СЛАУ(12). Доказывается, что регуляризованная система (14) совместна и имеет единственное решение, которое и будем искомым приближенным решением исходной СЛАУ (12).

* Опытным путем установлено, что неизвестные величины x и y связаны линейными соотношениями

$$2\,x + 3\,y = 7.2, \ \ 3\,x + y = 6.9, \ \ x + 2\,y = 4.0.$$

Функция Ψ имеет вид

$$\Psi(x,\, y) = (2\,x + 3\,y - 7.2)^2 + (3\,x + y - 6.9)^2 + (x + 2\,y - 4.0)^2.$$

Регуляризованной по Тихонову системой будет $14\,x + 11\,y = 39.1, \ \ 11\,x + 14\,y = 36.5$. Искомое приближенное решение: $x \approx 1.95, \ \ y \approx 1.08$.

Наконец, следует иметь в виду, что не любые экспериментальные зависимости имеет смысл сглаживать. Например, кардиограммы, энцефалограммы, сейсмограммы при сглаживании теряют наиболее ценную информацию. Сглаживание записей колебаний курсов валют на биржах также приводит к потере значительной части информации. Современные исследования показывают, что обработку подобных зависимостей (в математической статистике их называют временными рядами) целесообразно проводить не только сглаживанием, но и методами теории фракталов, оценкой динамики фрактальной размерности. Изложение этих методов выходит за рамки нашего предмета, хотя их популярность в последние годы возрастает.

§ 4.6. Линейная регрессия

Пусть $(\xi,\,\eta)$ – случайный вектор, в котором, например, ξ и η – сигналы на входе и выходе некоторого устройства при наличии помех. Связь между ξ и η в общем случае является корреляционной, и возникает вопрос об отыскании среди всех функций $\hat{g}(\xi)$ такой, для которой величина $M\,(\eta - g(\xi)^2)$ минимальна. Стало быть, речь идет о наилучшем приближении в смысле метода наименьших квадратов. Функция $g(\xi)$, на которой достигается искомый минимум, называется *средне-квадратической регрессией с.в. η на с.в. ξ*. Ограничимся нахождением линейной средне-квадратической регрессии с.в. η на с.в. ξ: $g(\xi) = a\,\xi + b$. Без доказательства отметим, что так называемая нормальная средне–квадратическая регрессия η на ξ, отвечающая случаю нормальных распределений случайных величин ξ и η, обязательно является линейной функцией от ξ.

В математической статистике имеют дело с выборками $x_1,\, x_2,\, \cdots,\, x_n$ и $y_1,\, y_2,\, \cdots, y_n$ из генеральных совокупностей, отвечающих с.в. ξ и η соответственно, и задача сводится к отысканию методом наименьших квадратов функции $F(x) = a\,x + b$. В предыдущем параграфе такая задача была уже рассмотрена и для коэффициентов a и b была выписана СЛАУ (10). Решая эту систему, последовательно находим:

$$n\,b = \sum_{j=1}^{n} y_j - a\,\sum_{j=1}^{n} x_j, \tag{15}$$

то есть $b = \bar{y} - a\,\bar{x}$. Подстановка этого выражения для b во второе уравнение системы (10) и очевидные преобразования дают

$$\sum_{j=1}^{n} x_j \big(a(x_j - \bar{x}) - (y_j - \bar{y})\big) = 0.$$

Отсюда

$$a = \frac{\sum_{j=1}^{n} x_j (y_j - \bar{y})}{\sum_{j=1}^{n} x_j (x_j - \bar{x})}.$$

Более симметрично тот же результат можно записать в виде

$$a = \frac{\sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{1} (x_j - \bar{x})^2}, \tag{16}$$

так как

$$\sum_{j=1}^{n} \bar{x}\,(x_j - \bar{x}) = \sum_{j=1}^{n} \bar{x}\,(y_j - \bar{y}) = 0.$$

Введем обозначения

$$S_x^2 = \sum_{j=1}^{n} (x_j - \bar{x})^2; \quad S_y^2 = \sum_{j=1}^{n} (y_j - \bar{y})^2$$

для выборочных дисперсий и

$$r_{x,y} = \frac{\sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y})}{S_x \cdot S_y}$$

для *выборочного коэффициента корреляции*. С учетом этих обозначений формулы (15) и (16) приводят к выражениям

$$a = r_{x,y} \frac{S_y}{S_x}, \quad b = \bar{y} - r_{x,y} \frac{S_y}{S_x}\,\bar{x}.$$

Таким образом, статистическим приближением к линейному уравнению средне-квадратической регрессии η на ξ является

$$y = r_{x,y} \frac{S_y}{S_x}(x - \bar{x}) + \bar{y}. \tag{17}$$

График функции (17) в прямоугольной декартовой системе координат является прямой линией, проходящей через точку $(\bar{x}, \bar{y})$ и называется *эмпирической линией регрессии y на x*.

Аналогичный вид имеет функция, графическим изображением которой будет эмпирическая линия регрессии x на y:

$$x = r_{x,y}\frac{S_x}{S_y}(y - \bar{y}) + \bar{x}. \tag{18}$$

Уравнения обеих линий регрессии допускают симметричные формы записи. Например, уравнение (17) записывается в равносильном виде

$$\frac{y - \bar{y}}{S_y} = r_{x,y}\frac{x - \bar{x}}{S_x}.$$

* Опытные данные сведены в таблицу:

$$\xi: \quad 22, \;\; 34 \;\; 46 \;\; 54 \;\; 67 \;\; 72,$$

$$\eta: \quad 11 \;\; 23 \;\; 17 \;\; 15 \;\; 24 \;\; 23.$$

Требуется найти эмпирическое уравнение линейной регрессии ξ на η.

Вычисления дают: $\bar{x} = 49.2$, $\bar{y} = 17.2$, $S_x = 19.1$, $S_y = 5.15$, $r_{x,y} = 0.95$.

Следовательно, $r_{x,y}\frac{S_y}{S_x} = 0.256$ и искомым уравнением будет $y = 0.256 \cdot (x - 49.2) + 17.2$.

Расчеты показали, что между ξ и η имеется тесная корреляционная связь, поскольку $r_{x,y} = 0.95$.

Возможно обобщение результатов данного параграфа на случай n–мерных случайных векторов (множественная линейная регрессия).

§ 4.7. Задачи для самостоятельного решения

1. Найти выборочное среднее и выборочную дисперсию S^2, если выборка имеет вид 20.3, 20.42, 19.85, 20.16, 19.98.

2. Построить гистограмму относительных частот n_j/n попадания значений вариационного ряда в указанные в первой строке следующей таблицы интервалы. Объем исходной выборки $n = 150$, числа n_j даны во второй строке таблицы.

$$I_j: \quad [1,3), \; [3,5) \; [5\,7) \; [7,9) \; [9,11],$$

$$n_j: \quad 10 \quad 25 \quad 35 \quad 60 \quad 20.$$

3. Показать, что выборочное среднее из генеральной совокупности с пуассоновским теоретическим законом распределения с параметром λ является несмещенной оценкой этого параметра.

4. Нобелевский лауреат Р. Милликен, впервые опытным путем измеривший заряд электрона, получил свои выводы путем статистической обработки выборки объема 58, содержащей значения зарядов мелких электрически заряженных масляных капель [7]. Во избежание громоздкости приведем здесь лишь часть выборки Милликена: $5.768 \cdot 10^{-10}$; $4.810 \cdot 10^{-10}$; $4.778 \cdot 10^{-10}$; $4.772 \cdot 10^{-10}$; $4.790 \cdot 10^{-10}$. (Заряд измерялся в гауссовой системе единиц). Найти выборочное среднее $\bar{x}$ и выборочное среднеквадратичное отклонение. Выборочное среднее и есть приближенное значение заряда электрона.

5. Случайная величина распределена по закону Максвелла, то есть имеет плотность вероятности

$$p(x) = \frac{2\,t^2}{a\sqrt{2\pi}} e^{-t^2/2},$$

где $t = x/a$, a – неизвестный параметр. По выборке $x_1, x_2, \cdots, x_n$ найти оценку максимального правдоподобия этого параметра.

6. Случайная величина распределена по закону Парето с параметром $c > 0$, то есть имеет плотность

$$p(x) = c \cdot x^{-1-c}, \text{ если } x \geqslant 1,$$

а при $x < 1$ имеем $p(x) = 0$. По данной выборке $x_1, x_2, \cdots, x_n$ из Парето-генеральной совокупности найти оценку максимального правдоподобия параметра c.

7. Найти доверительный интервал для среднего $\bar{x}$ по выборке объема 30, выписанной в последнем примере § 4.3, отвечающий доверительной вероятности $\beta = 0.96$. Дисперсия считается равной 1.

8. Решить предыдущую задачу в случае, когда дисперсия σ^2 не известна, а $\beta = 0.80$.

9. Найти доверительный интервал для σ, если выборка взята из задачи 8 и назначена доверительная вероятность $\beta = 0.80$.

10. При рентгеновском обследовании вероятность обнаружить заболевание туберкулезом у больного этой болезнью равна 0.9. Вероятность принять здорового человека за больного туберкулезом равна 0.08. Пусть

доля больных туберкулезом по отношению ко всему населению равна 0.006. Найти условную вероятность того, что человек здоров, если он был признан больным при обследовании.

11. Пользуясь критерием Пирсона, при уровне значимости 0.05 установить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с данными следующей выборки объема $n = 200$:

$$x_j: \quad 5 \quad 7 \quad 9 \quad 11 \quad 13 \quad 15 \quad 17 \quad 19 \quad 21,$$

$$n_j: \quad 15 \quad 26 \quad 25 \quad 30 \quad 26 \quad 21 \quad 24 \quad 20 \quad 13.$$

Здесь n_j – число появлений значения x_j в исходной выборке при данном j.

12. Ниже приведены таблицы среднемесячных температур в Москве и Лондоне в течение одного года:

Месяц:	I	II	III	IV	V	VI	VII	$VIII$	IX	X	XI	XII
Москва:	-16	-10	-9.4	4.0	7.5	13.8	16.4	14.8	9.2	6.4	1.0	-8.0
Лондон:	-1.1	0.4	6.9	9.7	11.6	15.9	16.3	15.9	13.8	11.7	6.7	2.8.

Рассматривая средние температуры по месяцам как выборочные значения с.в. μ и λ, найти выборочные средние, выборочные дисперсии и коэффициент корреляции.

13. В городе N с 1996 по 2002 г. включительно было выдано водительских удостоверений 65, 70, 52, 42, 48, 74, 57 последовательно по указанным годам. Зарегистрированное число выпускников средних школ в те же годы в этом городе составляло 632, 619, 598, 562, 579, 609, 580 соответственно. Исследовать линейную регрессию между первой и второй случайными величинами.

СПИСОК ЛИТЕРАТУРЫ

1. Венцель Е.С. *Теория вероятностей.* М.: Высшая школа, 2002.
2. Гихман И.И., Скороход А.В., Ядренко М.И. *Теория вероятностей и математическая статистика.* Киев: Вища школа, 1979.
3. Гмурман В.Е. *Руководство к решению задач по теории вероятностей и математической статистике.* М.: Высшая школа, 2000.
4. Жевержеев В.Ф., Кальницкий Л.А., Сапогов Н.А. *Специальный курс высшей математики для втузов.* М.: Высшая школа, 1970.
5. Коваленко И.Н., Филиппова А.А. *Теория вероятностей и математическая статистика.* М.: Высшая школа, 1982.
6. Корн Г., Корн Т. *Справочник по математике для научных работников и инженеров.* М.: Наука, 1970.
7. Мешалкин Л.Д. *Сборник задач по теории вероятностей.* М.: Изд–во МГУ, 1963.
8. Севастьянов Б.А., Чистяков В.П., Зубков А.М. *Сборник задач по теории вероятностей.* М.: Наука, 1980.
9. Смирнов Н.В., Дунин-Барковский И.В. *Курс теории вероятностей и математической статистики для технических приложений.* М.: Наука, 1969.
10. *Сборник задач по теории вероятностей, математической статистике и теории случайных функций* / Под ред. А.А. Свешникова. М.: Наука, 1970.
11. Тутубалин В.Н. *Теория вероятностей.* М.: Изд–во МГУ, 1972.
12. Феллер В. *Введение в теорию вероятностей и ее приложения.* М.: Мир, 1984.
13. Леви П. *Стохастические процессы и броуновское движение.* М.: Наука, 1972.
14. Шеретов В.Г., Щербакова С.Ю. *Российской математике – триста лет.* Тверь: НТП "Фактор", 2003.

ПРИЛОЖЕНИЕ

Таблица I. ЗНАЧЕНИЯ ФУНКЦИИ ЛАПЛАСА

$$\Phi(x) = \int_0^x e^{-u^2/2}\,d\,u$$

x :	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
$\Phi(x)$:	0.0000	0.0199	0.0398	0.0596	0.0793	0.0987	0.1179	0.1368	0.1554
x :	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
$\Phi(x)$:	0.1736	0.1915	0.2088	0.2257	0.2422	0.2580	0.2734	0.2881	0.3023
x :	0.90	0.95	1.00	1.05	1.10	1.15	1.20	1.25	1.30
$\Phi(x)$:	0.3159	0.3289	0.3413	0.3531	0.3643	0.3749	0.3849	0.3944	0.4032
x :	1.35	1.40	1.45	1.50	1.55	1.60	1.65	1.70	1.75
$\Phi(x)$:	0.4115	0.4192	0.4265	0.4332	0.4394	0.4452	0.4505	0.4554	0.4599
x :	1.80	1.85	1.90	1.95	2.00	2.20	2.40	2.70	3.00
$\Phi(x)$:	0.4641	0.4678	0.4713	0.4744	0.4772	0.4861	0.4918	0.4965	0.4986

Замечание. В таблице даны значения функции Лапласа для $x \in [0,\,3]$. Если же $x < 0$, то следует пользоваться формулой $\Phi(x) = 1 - \Phi(-x)$.

Таблица II. РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА ($t = t(\beta,\,n)$)

а) $\beta = 0.95$

n :	4	6	9	11	14	19	24	29	35	49	99
t :	2.78	2.45	2.26	2.20	2.15	2.09	2.06	2.045	2.030	2.007	1.982

б) $\beta = 0.98$

n :	4	6	9	11	14	19	24	29	35	49	99
t :	3.75	3.14	2.82	2.71	2.62	2.54	2.49	2.46	2.44	2.40	2.364

в) $\beta = 0.99$

n :	4	6	9	11	14	19	24	29	35	49	99
t :	4.60	3.71	3.25	3.11	2.98	2.86	2.78	2.76	2.73	2.67	2.62

Замечание. Значения t для степеней свободы n, не указанных в таблице, можно находить путем интерполяции.

Таблица III. χ^2–РАСПРЕДЕЛЕНИЕ ПИРСОНА

Избранные значения квантилей x для χ^2_n, отвечающих вероятности γ $\Big(P(\chi^2_n > x) = \gamma\Big)$

а) $\gamma = 0.99$

$n:$	5	9	11	14	17	19	24	29
$x:$	0.554	2.088	3.053	4.660	6.408	7.633	10.856	14.256

б) $\gamma = 0.98$

$n:$	5	9	11	14	17	19	24	29
$x:$	0.752	2.532	3.609	5.368	7.255	8.567	11.992	15.574

в) $\gamma = 0.95$

$n:$	5	9	11	14	17	19	24	29
$x:$	1.145	3.325	4.575	6.571	8.672	10.117	13.848	17.708

г) $\gamma = 0.90$

$n:$	5	9	11	14	17	19	24	29
$x:$	1.610	4.168	5.578	7.790	10.085	11.651	15.659	19.768

д) $\gamma = 0.10$

$n:$	5	9	11	14	17	19	24	29
$x:$	9.236	14.684	17.275	21.064	24.769	27.204	33.196	39.087

е) $\gamma = 0.05$

$n:$	5	9	11	14	17	19	24	29
$x:$	11.070	16.919	19.675	23.685	27.587	30.144	36.415	42.557

ж) $\gamma = 0.02$

$n:$	5	9	11	14	17	19	24	29
$x:$	13.388	19.679	22.618	26.873	30.995	33.687	40.270	46.693

з) $\gamma = 0.01$

$n:$	5	9	11	14	17	19	24	29
$x:$	15.086	21.666	24.725	29.141	33.409	36.191	42.980	49.588

ОГЛАВЛЕНИЕ

ГРИГОРЬЕВА Вера Владимировна

ШЕРЕТОВ Владимир Георгиевич

ШЕРЕТОВ Юрий Владимирович

Элементы теории вероятностей и математической статистики

Учебное пособие

Технический редактор А.В. Жильцов
Подписано в печать 09.02.2009. Формат $60 \times 84\ ^1/_{16}$.
Усл. печ. л. 5,5. Тираж 50 экз. Заказ № 45.
Тверской государственный университет
Редакционно-издательское управление
Адрес: Россия, 170100, г. Тверь, ул. Желябова, 33.
Тел. РИУ: (4822) 35-60-63.