

Лекция 1

Случайные события

Определение 1.1 *Элементарным исходом* (или *элементарным событием*) называют любой простейший (т.е. неделимый в рамках данного опыта) исход опыта. Множество всех элементарных исходов будем называть *пространством элементарных исходов*.

Другими словами, множество исходов опыта образует пространство элементарных исходов, если выполнены следующие требования:

- 1) в результате опыта один из исходов обязательно происходит;
- 2) появление одного из исходов опыта исключает появление остальных;
- 3) в рамках данного опыта нельзя разделить элементарный исход на более мелкие составляющие.

В дальнейшем пространство элементарных исходов будем обозначать прописной буквой Ω , а сами элементарные исходы — строчной буквой ω , снабженной, при необходимости, индексами. То, что элемент ω принадлежит Ω , записывают в виде $\omega \in \Omega$, а тот факт, что множество Ω состоит из элементов $\omega_1, \omega_2, \dots, \omega_n, \dots$, и только из них, записывают в виде $\Omega = \{\omega_1; \omega_2; \dots; \omega_n; \dots\}$ или в виде $\Omega = \{\omega_i, i = 1, 2, \dots, n, \dots\}$. В частности, Ω может содержать конечное число элементарных исходов.

Рассмотрим примеры, поясняющие понятие пространства элементарных исходов.

Пример 1.1 Пусть опыт состоит в однократном подбрасывании монеты. При математическом описании этого опыта естественно отвлечься от несущественных возможностей (например, монета встанет на ребро) и ограничиться только двумя элементарными исходами: выпадением “герба” (можно обозначить этот исход Γ , ω_Γ или ω_1) и выпадением “цифры” (Π , ω_Π или ω_2). Таким образом, $\Omega = \{\Gamma, \Pi\}$, $\Omega = \{\omega_\Gamma, \omega_\Pi\}$ или $\Omega = \{\omega_1, \omega_2\}$.

При двукратном подбрасывании монеты (или однократном подбрасывании двух монет) пространство элементарных исходов будет, очевидно, содержать 4 элемента, т.е. $\Omega = \{\omega_{\Gamma\Gamma}, \omega_{\Gamma\Pi}, \omega_{\Pi\Gamma}, \omega_{\Pi\Pi}\}$, где $\omega_{\Gamma\Gamma}$ — появление “герба” и при первом, и при втором подбрасываниях, и т.д.

Пример 1.2 При однократном бросании игральной кости возможен любой из шести элементарных исходов $\omega_1, \dots, \omega_6$, где $\omega_i, i = \overline{1, 6}$, означает появление i очков на верхней грани кости, т.е. $\Omega = \{\omega_i, i = \overline{1, 6}\}$.

При двукратном бросании игральной кости каждый из шести возможных исходов при первом бросании может сочетаться с каждым из шести исходов при втором бросании, т.е. $\Omega = \{\omega_{ij}, i, j = \overline{1, 6}\}$, где ω_{ij} — исход опыта, при котором сначала выпало i , а затем j очков.

Нетрудно подсчитать, что пространство элементарных исходов Ω содержит 36 элементарных исходов.

Пример 1.3 Пусть опыт заключается в определении числа вызовов, поступивших на телефонную станцию в течение заданного промежутка времени. Разумеется, реально это число не превышает некоторого значения (определяемого, в частности, пропускной способностью линий связи), но, поскольку это значение может быть достаточно большим, в качестве пространства элементарных исходов можно принять множество целых неотрицательных чисел, т.е. $\Omega = \{0, 1, \dots, n, \dots\}$.

Пример 1.4 Предположим, что стрелок производит единственный выстрел по плоской мишени. В этом случае Ω естественно отождествить с множеством точек на плоскости или множеством пар $(x; y)$ действительных чисел, где x — абсцисса, а y — ордината точки попадания пули в мишень в некоторой системе координат. Таким образом, $\Omega = \{(x; y) : -\infty < x < +\infty, -\infty < y < +\infty\}$.

События, действия над ними

Введем понятие случайного *события*. Поскольку в дальнейшем будем рассматривать только случайные события, то, начиная с этого момента, будем называть их, как правило, просто событиями.

Определение 1.2 Любой набор *элементарных исходов*, или, иными словами, произвольное подмножество *пространства элементарных исходов*, называют **событием**.

Элементарные исходы, которые являются элементами рассматриваемого подмножества (события), называют *элементарными исходами, благоприятствующими* данному *событию*, или *образующими* это *событие*.

События будем обозначать прописными латинскими буквами, снабжая их при необходимости индексами, например: A, B_1, C_3 и т.д.

Сразу же оговоримся, что определение 1.2 события будет уточнено в следующем параграфе в том случае, когда Ω не является счетным множеством. Здесь же мы вводим определение 1.2 по двум причинам.

Во-первых, основная цель настоящего параграфа — наглядно показать, как физическое понятие случайного события формализуется в математических понятиях теории множеств, и описать операции над событиями.

Во-вторых, определение 1.2 вполне удовлетворительно можно применять для решения практических задач, в то время как строгое определение события служит лишь для построения теории вероятностей как раздела современной математики, оперирующей логически безупречными, но сложными для неподготовленного читателя понятиями.

Часто используется следующая терминология: говорят, что событие A произошло (или наступило), если в результате опыта появился какой-либо из элементарных исходов $\omega \in A$.

Пример 1.5 В примере 1.2 было показано, что при однократном бросании игральной кости $\Omega = \{\omega_i, i = \overline{1,6}\}$, где ω_i — элементарный исход, заключающийся в выпадении i очков. Рассмотрим следующие события: A — выпадение четного числа очков; B — выпадение нечетного числа очков; C — выпадение числа очков, кратного трем. Очевидно, что $A = \{\omega_2; \omega_4; \omega_6\}$, $B = \{\omega_1; \omega_3; \omega_5\}$ и $C = \{\omega_3; \omega_6\}$.

Определение 1.3 Событие, состоящее из всех элементарных исходов, т.е. событие, которое обязательно происходит в данном опыте, называют **достоверным событием**.

Достоверное событие, как и пространство элементарных исходов, обозначают буквой Ω .

Определение 1.4 Событие, не содержащее ни одного элементарного исхода, т.е. событие, которое никогда не происходит в данном опыте, называют **невозможным событием**.

Невозможное событие будем обозначать символом \emptyset .

Пример 1.6 При бросании игральной кости достоверное событие можно описать, например, как выпадение хотя бы одного очка, а невозможное — как выпадение 7 очков. #

Часто бывает полезно наглядно представить события в виде **диаграммы Эйлера — Венна**. Изобразим все пространство элементарных исходов прямоугольником. При этом каждый элементарный исход ω соответствует точке внутри прямоугольника, а каждое событие A — некоторому подмножеству точек этого прямоугольника. Трактовкой диаграммы Эйлера — Венна может служить опыт с бросанием случайным образом частицы в прямоугольник. Тогда элементарный исход ω — это попадание частицы в точку ω прямоугольника, а событие A — в часть прямоугольника, задаваемую подмножеством A .

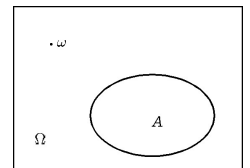


Рис. 1.1.

Рассмотрим теперь **операции (действия) над событиями**, которые, по существу, совпадают с операциями над подмножествами. Эти операции будем иллюстрировать на диаграммах Эйлера — Венна. На рис. 1.2–1.7 заштрихованы области, которые соответствуют событиям, являющимся результатами таких операций.

Определение 1.5 **Пересечением (произведением)** двух событий A и B называют событие C , происходящее тогда и только тогда, когда одновременно происходят оба события A и B , т.е. событие, состоящее из тех и только тех элементарных исходов, которые принадлежат и событию A , и событию B .

Пересечение событий A и B записывают следующим образом:

$$C = A \cap B, \quad \text{или} \quad C = AB.$$

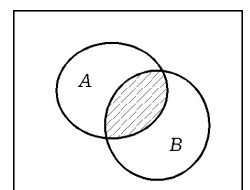


Рис. 1.2.

Определение 1.6 События A и B называют **несовместными**, или **непересекающимися**, если их пересечение является невозможным событием, т. е. если $A \cap B = \emptyset$. В противном случае события называют **совместными**, или **пересекающимися**.

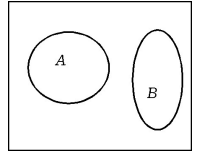


Рис. 1.3.

Определение 1.7 Объединением (суммой) двух событий A и B называют событие C , происходящее тогда и только тогда, когда происходит хотя бы одно из событий A или B , т. е. событие C , состоящее из тех элементарных исходов, которые принадлежат хотя бы одному из подмножеств A или B .

Объединение событий A и B записывают в виде

$$C = A \cup B, \quad \text{или} \quad C = A + B.$$

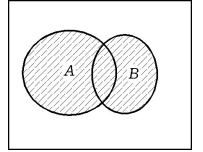


Рис. 1.4.

Аналогично определяют понятия произведения и суммы событий для любого конечного числа событий и даже для бесконечных последовательностей событий. Так, событие

$$A_1 A_2 \dots A_n \dots = \bigcap_{n=1}^{\infty} A_n$$

состоит из элементарных исходов, принадлежащих всем событиям A_n , $n \in \mathbb{N}$, а событие

$$A_1 \cup A_2 \cup \dots \cup A_n \cup \dots = \bigcup_{n=1}^{\infty} A_n$$

состоит из элементарных исходов, принадлежащих хотя бы одному из событий A_n , $n \in \mathbb{N}$. В частности, события A_1, A_2, \dots, A_n называют **парно несовместными** (**непересекающимися**), если $A_i A_j = \emptyset$ для любых $i, j = \overline{1, n}$, $i \neq j$, и **несовместными** (**непересекающимися**) в совокупности, если $A_1 A_2 \dots A_n = \emptyset$.

Определение 1.8 Разностью двух событий A и B называют событие C , происходящее тогда и только тогда, когда происходит событие A , но не происходит событие B , т.е. событие C , состоящее из тех элементарных исходов, которые принадлежат A , но не принадлежат B . Разность событий A и B записывают в виде:

$$C = A \setminus B.$$

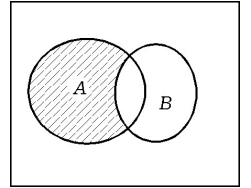


Рис. 1.5.

Определение 1.9 Дополнением события A (обычно обозначают \bar{A}) называют событие, происходящее тогда и только тогда, когда не происходит событие A . (Другими словами, $\bar{A} = \Omega \setminus A$). Событие \bar{A} называют также **событием**, **противоположным** событию A .

Если некоторое событие записано в виде нескольких действий над различными событиями, то сначала переходят к дополнениям, а затем умножают и, наконец, складывают и вычитают (слева направо) события. Так, формула

$$C = A_1 \bar{A}_2 B_1 \cup A_3 \bar{B}_2 \setminus B_3$$

эквивалентна формуле

$$C = \{ [A_1 (\bar{A}_2) B_1] \cup [A_3 (\bar{B}_2)] \} \setminus B_3.$$

Следует отметить, что все действия над событиями можно получить с помощью только двух действий — объединения и дополнения (или пересечения и дополнения). Основанием для этого утверждения служат законы де Моргана, а также соотношение $A \setminus B = \bar{A} \bar{B}$.

Кроме перечисленных выше действий над событиями нам в дальнейшем понадобится понятие включения.

Определение 1.10 Событие A включено в событие B , что записывают $A \subset B$, если появление события A обязательно влечет за собой наступление события B (, или каждый элементарный исход ω , принадлежащий A , обязательно принадлежит и событию B).

Ясно, что включение $A \subset B$ эквивалентно равенству $\bar{A} \bar{B} = A$. Используют и обратное понятие: событие B включает событие A ($B \supset A$), если $A \subset B$.

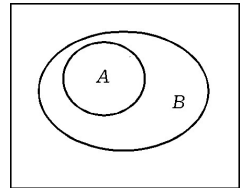


Рис. 1.7.

Пример 1.7 Рассмотрим техническое устройство (ТУ), состоящее из m элементов. В теории надежности принято говорить, что элементы соединены последовательно, если ТУ прекращает функционировать при отказе любого элемента, и соединены параллельно, если прекращение функционирования ТУ наступает только при отказе всех m элементов. Условное изображение параллельного и последовательного соединений представлено на рис. 1.8 и 1.9 соответственно. Обозначим A событие, означающее отказ ТУ, а A_i — событие, означающее отказ i -го элемента ($i = \overline{1, m}$). Тогда события A и A_i связаны соотношениями

$$A = A_1 \cup \dots \cup A_m \text{ и } A = A_1 \cap \dots \cap A_m$$

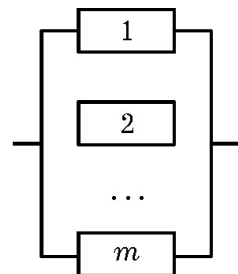


Рис. 1.8.

для последовательного соединения и параллельного соединения соответственно. Очевидно, что при параллельном соединении элементов событие A включено в каждое событие A_i , $i = \overline{1, m}$, а при последовательном соединении, наоборот, любое событие A_i , $i = \overline{1, m}$, включено в событие A . #

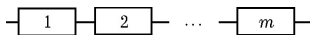


Рис. 1.9.

Приведем основные свойства операций над событиями, справедливость которых нетрудно проверить, пользуясь диаграммами Эйлера — Венна (проделайте это самостоятельно).

1. Коммутативность суммы и произведения: $A \cup B = B \cup A$, $AB = BA$.
2. Ассоциативность суммы и произведения: $A \cup B \cup C = A \cup (B \cup C)$, $(AB)C = A(BC)$.
3. Дистрибутивность относительно сложения: $(A \cup B)C = AC \cup BC$.
4. Дистрибутивность относительно умножения (новое свойство, не выполняющееся для чисел): $AB \cup C = (A \cup C)(B \cup C)$.
5. Включение A в B , т.е. $A \subset B$, влечет за собой включение \overline{B} в \overline{A} , т.е. $\overline{A} \supset \overline{B}$.
6. Совпадение двойного дополнения с исходным событием: $\overline{\overline{A}} = A$.
7. Совпадение суммы и произведения одинаковых событий с самим событием $A \cup A = AA = A$.
8. **Законы де Моргана:** $\overline{A \cup B} = \overline{A} \overline{B}$, $\overline{AB} = \overline{A} \cup \overline{B}$.

Замечание 1.1 Законы де Моргана верны для любого конечного числа событий:

$$\overline{A_1 \cup A_2 \cup \dots \cup A_n} = \overline{A_1} \overline{A_2} \dots \overline{A_n}$$

$$\overline{\overline{A_1} \overline{A_2} \dots \overline{A_n}} = A_1 \cup A_2 \cup \dots \cup A_n.$$

Сигма-алгебра событий

В предыдущем параграфе мы называли событием любое подмножество пространства элементарных исходов Ω . Такое определение допустимо, если Ω является конечным или счетным множеством. Оказывается, однако, что в случае несчетного множества элементарных исходов уже нельзя построить логически непротиворечивую теорию, называя событием произвольное подмножество множества Ω . Поэтому событиями в этом случае называют не любые подмножества элементарных исходов, а только подмножества из Ω , принадлежащие некоторому классу \mathfrak{B} . Этот класс в теории множеств принято называть сигма-алгеброй событий (пишут σ -алгебра).

С точки зрения здравого смысла *событие* — это то, что мы наблюдаем после проведения опыта. В частности, если можно после опыта установить, произошли или нет события A и B , то можно также сказать, произошли или нет события \overline{A} и \overline{B} , *объединение, пересечение и разность событий* A и B . Таким образом, σ -алгебра событий обязана быть классом подмножеств, замкнутым относительно приведенных операций над подмножествами, т.е. указанные операции над элементами (подмножествами) данного класса приводят к элементам (подмножествам) того же класса.

Дадим теперь строгое определение σ -алгебры событий.

Определение 1.11 *Сигма-алгеброй* (σ -алгеброй) событий \mathfrak{B} назовем непустую систему подмножеств пространства элементарных исходов Ω , удовлетворяющую следующим двум условиям.

1. Если подмножество A принадлежит \mathfrak{B} , то дополнение \overline{A} принадлежит \mathfrak{B} .
2. Если подмножества $A_1, A_2, \dots, A_n, \dots$ принадлежат \mathfrak{B} , то их объединение $A_1 \cup A_2 \cup \dots \cup A_n \cup \dots$ и их пересечение $A_1 A_2 \dots A_n \dots$ принадлежит \mathfrak{B} .

Поскольку $\Omega = A \cup \overline{A}$ и $\emptyset = \overline{\Omega}$, то *достоверное событие* Ω и *невозможное событие* \emptyset принадлежат \mathfrak{B} .

В случае конечного или счетного пространства элементарных исходов Ω в качестве σ -алгебры событий обычно рассматривают множество всех подмножеств Ω .

Замечание 1.2 Если в условии 2 счетное множество событий заменить на конечное, то получим определение *алгебры событий*. Любая σ -алгебра событий обязательно является алгеброй событий. Обратное утверждение, вообще говоря, не верно.

Пример 1.8 Пусть опыт состоит в подбрасывании один раз тетраэдра, каждая грань которого помечена одним из чисел 1, 2, 3 и 4.

Очевидно, что пространство элементарных исходов Ω в этом опыте имеет вид $\Omega = \{\omega_1; \omega_2; \omega_3; \omega_4\}$, где ω_i — падение тетраэдра на грань с числом i , $i = \overline{1, 4}$.

Поскольку в рассматриваемом опыте может происходить одно из следующих событий:

$$\begin{aligned} & \emptyset, \\ & \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \\ & \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\}, \{\omega_2, \omega_3\}, \{\omega_2, \omega_4\}, \{\omega_3, \omega_4\}, \\ & \{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}, \\ & \Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}, \end{aligned}$$

то алгебра событий будет содержать все подмножества Ω , включая Ω (достоверное событие) и \emptyset (невозможное событие).

Лекция 2

Вероятность

Говоря о *событиях*, мы с различной степенью уверенности относимся к возможности их наступления. Так, с большей уверенностью можно утверждать, что при однократном подбрасывании монеты выпадет “герб”, чем при однократном бросании игральной кости — 6 очков. Говорят, что первое событие более вероятно, чем второе.

Что же такое *вероятность* события? Напрашивается каждому событию A поставить в соответствие число $P(A)$, которое будет являться мерой возможности его появления. Если принять $P(\Omega) = 1$, а $P(\emptyset) = 0$ (хотя можно было взять другую единицу измерения), то тогда для любого события A естественно ожидать, что $0 \leq P(A) \leq 1$.

Определение вероятности как меры возможности появления события в современной математике вводится на основании аксиом. Но, прежде чем перейти к аксиоматическому определению, остановимся на нескольких других определениях, которые исторически возникли раньше. Они, с одной стороны, позволяют лучше понять смысл аксиоматического определения, а с другой — во многих случаях являются рабочим инструментом для решения практических задач. Приведем их, следуя хронологическому порядку появления.

Классическое определение вероятности

В классическом определении вероятности исходят из того, что *пространство элементарных исходов* Ω содержит конечное число элементарных исходов, причем все они равновозможны. Понятие равновозможности поясним следующим образом.

Элементарные исходы в некотором опыте называют *равновозможными*, если в силу условий проведения опыта можно считать, что ни один из них не является объективно более возможным, чем другие. Опыт, удовлетворяющий условию равновозможности элементарных исходов, часто называют также “*классической схемой*”.

Пусть N — общее число равновозможных элементарных исходов в Ω , а N_A — число *элементарных исходов, образующих событие* A (или, как говорят, *благоприятствующих* событию A).

Определение 2.1 *Вероятностью события* A называют отношение числа N_A благоприятствующих событию A элементарных исходов к общему числу N равновозможных элементарных исходов, т.е.

$$P(A) = \frac{N_A}{N}.$$

Данное определение вероятности события принято называть **классическим определением вероятности**.

Заметим, что наряду с названием “классическая схема” используют также названия “случайный выбор”, “равновероятный выбор” и т.д.

Пример 2.1 Из урны, содержащей $k = 10$ белых и $l = 20$ черных шаров (шары отличаются лишь цветом), наугад вынимают один шар. Требуется найти вероятность $P(A)$ события A , заключающегося в том, что из урны извлечен белый шар.

Для решения поставленной задачи заметим, что число элементарных исходов в данном опыте совпадает с общим числом шаров в урне $N = k + l = 30$, причем все исходы равновозможны, а число благоприятствующих событию A исходов $N_A = k = 10$. Поэтому в соответствии с определением классической вероятности $P(A) = \frac{k}{k+l} = \frac{1}{3}$. #

Используя классическое определение вероятности события, докажем следующие свойства.

Свойство 2.1 *Для любого события A вероятность удовлетворяет неравенству $P(A) \geq 0$.*

ДОКАЗАТЕЛЬСТВО. Свойство очевидно, так как отношение N_A/N не может быть отрицательным.

Свойство 2.2 Для достоверного события Ω (которое содержит все N элементарных исходов) $P(\Omega) = 1$.

Свойство 2.3 Если события A и B несовместны ($AB = \emptyset$), то $P(A + B) = P(A) + P(B)$.

ДОКАЗАТЕЛЬСТВО. Действительно, если событию A благоприятствуют N_1 исходов, а событию B — N_2 исходов, то в силу несовместности A и B событию $A + B$ благоприятствуют $N_1 + N_2$ исходов. Следовательно,

$$P(A + B) = \frac{N_1 + N_2}{N} = \frac{N_1}{N} + \frac{N_2}{N} = P(A) + P(B).$$

Оказывается, что эти три свойства являются основными. Из них как следствия можно получить другие полезные свойства (подробнее они будут рассмотрены ниже), например:

$$P(\bar{A}) = 1 - P(A); \quad P(\emptyset) = 0;$$

$$P(A) < P(B), \quad \text{если} \quad A \subset B.$$

Недостаток классического определения заключается в том, что оно применимо только к пространствам элементарных исходов, состоящим из конечного числа равновероятных исходов. Этим определением нельзя воспользоваться даже в тех случаях, когда пространство элементарных исходов конечно, но среди исходов есть более предпочтительные или менее предпочтительные.

Геометрическое определение вероятности

Геометрическое определение вероятности обобщает классическое на случай *бесконечного множества элементарных исходов* Ω тогда, когда Ω представляет собой подмножество пространства \mathbb{R} (числовой прямой), \mathbb{R}^2 (плоскости), \mathbb{R}^n (n -мерного евклидова пространства).

В пространстве \mathbb{R} в качестве подмножеств будем рассматривать лишь *промежутки* или их объединения, т.е. подмножества, которые имеют длину. В пространстве \mathbb{R}^2 — те подмножества, которые имеют площадь, и т.д.

Под мерой $\mu(A)$ подмножества A будем понимать его длину, площадь или объем (обобщенный объем) в зависимости от того, какому пространству принадлежит Ω : в \mathbb{R} , в \mathbb{R}^2 или в \mathbb{R}^3 (\mathbb{R}^n). Будем также считать, что пространство элементарных исходов Ω имеет конечную меру, а возможность попадания “случайно брошенной” точки в любое подмножество Ω пропорциональна мере этого подмножества и не зависит от его расположения и формы. В этом случае говорят, что рассматривается “геометрическая схема” или “точку наудачу бросают в область Ω ”.

Определение 2.2 Вероятностью события A называют число $P(A)$, равное отношению меры множества A к мере множества Ω :

$$P(A) = \frac{\mu(A)}{\mu(\Omega)},$$

где $\mu(A)$ — мера множества A .

Данное определение вероятности события принято называть **геометрическим определением вероятности**.

Заметим, что в литературе вероятность события A , определенную выше, на основе **геометрической схемы**, часто называют **геометрической вероятностью**.

Геометрическая вероятность, очевидно, сохраняет отмеченные ранее свойства вероятности $P(A)$ в условиях классической схемы.

Пример 2.2 Ромео и Джульетта договорились встретиться в определенном месте между двенадцатью часами и часом дня. Необходимо найти вероятность встречи, если приход каждого из них в течение указанного часа происходит наудачу, причем известно, что Ромео ждет Джульетту ровно 20 минут, а Джульетта Ромео — 5 минут.

Для решения задачи воспользуемся геометрической схемой вероятности. Обозначим момент прихода Ромео через x , а Джульетты через y . Тогда любой элементарный исход ω в данной задаче можно отождествить с некоторой точкой $(x; y)$ на плоскости xOy . Выберем за начало отсчета 12 часов, а за единицу измерения 1 минуту и построим на плоскости xOy пространство элементарных исходов Ω . Очевидно, что это будет квадрат со стороной 60 (см. рис. 2.1). Событие A (Ромео и Джульетта встретятся) произойдет тогда, когда разность $y - x$ не превысит $t_1 = 20$, а разность $x - y$ не превысит $t_2 = 5$, т.е. условие встречи определяет систему неравенств

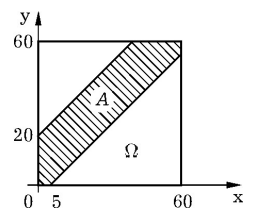


Рис. 2.1.

$$\begin{cases} y - x \leq 20; \\ x - y \leq 5. \end{cases}$$

Область A элементарных исходов, благоприятствующих этому событию, на рис. 2.1 заштрихована. Ее площадь S_A равна площади квадрата без двух угловых треугольников, т.е.

$$S_A = 60^2 - \frac{(60 - t_1)^2}{2} - \frac{(60 - t_2)^2}{2} = 1287,5.$$

Тогда, согласно определению 2.2, находим $P(A) = \frac{S_A}{S_\Omega} = \frac{1287,5}{3600} \approx 0,36$.

Статистическое определение вероятности

В основе статистического определения вероятности лежит общий принцип, в соответствии с которым методы теории вероятностей применимы только к таким испытаниям, которые могут быть, по крайней мере теоретически, повторены бесконечное число раз, и при этом имеет место свойство *устойчивости частот* появления связанных с этими испытаниями событий (см. Введение).

Пусть произведено n повторений опыта, причем в n_A из них появилось событие A . Обозначим $r_A = n_A/n$ наблюдаемую частоту события A . Практика показывает, что в тех экспериментах, для которых применимы методы теории вероятностей, частота события A с увеличением числа опытов n стабилизируется, т.е. стремится к некоторому пределу (допуская некоторую вольность речи).

Определение 2.3 *Вероятностью события* A называют (эмпирический) предел $P(A)$, к которому стремится частота r_A события A при неограниченном увеличении числа n опытов.

Данное определение вероятности события принято называть **статистическим определением вероятности**.

Можно показать, что при статистическом определении вероятности события сохраняются свойства вероятности события, справедливые в условиях классической схемы, т.е.

- 1) $P(A) \geq 0$;
- 2) $P(\Omega) = 1$;
- 3) $P(A + B) = P(A) + P(B)$, если $AB = \emptyset$.

С практической точки зрения статистическое определение вероятности является наиболее разумным. Однако с позиции теории вероятностей как раздела современной математики недостаток статистического определения очевиден: нельзя провести бесконечное число повторений опыта, а при конечном числе повторений наблюдаемая частота, естественно, будет разной при различном числе повторений.

Заметим, что связь между классическим и статистическим определениями была выявлена еще в период становления теории вероятностей как теории азартных игр. Было установлено, что при корректном использовании классического определения вероятность событий практически совпадает с их частотами при большом числе повторений эксперимента.

И хотя игроков интересовала частота определенных событий, решение задач, полученное на основе классического определения вероятности, их вполне устраивало. Иными словами, даже игроки азартных игр знали о совпадении статистического определения с другими (классическим и его обобщением — геометрическим).

Собственно говоря, задача определения связи вероятности с частотой не потеряла актуальности и в наши дни, когда в теории вероятностей повсеместно используется аксиоматическое определение вероятностей Колмогорова (см. 2.3). Это привело к появлению и широкому внедрению в практику обширного раздела теории вероятностей — математической статистики.

Аксиоматическое определение вероятности

Для того чтобы понять смысл **аксиоматического определения вероятности**, рассмотрим *классическую схему*.

В этом случае вероятность любого *элементарного исхода* ω_i , $i = \overline{1, N}$, $P(\omega_i) = 1/N$.

Вероятность любого *события* A при этом равна $P(A) = N_A/N$, где N_A — число *исходов, благоприятствующих событию* A .

Вероятность $P(A)$ можно записать также в следующем виде

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

где суммирование ведется по всем значениям индекса i , при которых элементарные исходы ω_i образуют событие A .

Однако задать вероятность события по такому принципу уже в случае *геометрической схемы* нельзя, так как при этом вероятность любого элементарного события равна нулю.

Поэтому следует дать определение вероятности события для любого пространства элементарных исходов Ω , не связанное с вероятностями элементарных исходов, а учитывающее те свойства вероятности событий, которые имеют место для всех предыдущих определений вероятности события (*классического, геометрического, статистического*).

Напомним, что этими свойствами являются следующие: 1) $P(A) \geq 0$; 2) $P(\Omega) = 1$; 3) $P(A_1 + \dots + A_m) = P(A_1) + \dots + P(A_m)$, если события A_1, \dots, A_m попарно несовместны.

Именно эти три свойства лежат в основе аксиоматического определения вероятности. При этом свойство 3 постулируется для суммы счетного множества попарно несовместных событий.

Определение 2.4 Пусть каждому событию A (т.е. подмножеству A пространства элементарных исходов Ω , принадлежащему σ -алгебре \mathfrak{B}) поставлено в соответствие число $P(A)$. Числовую функцию P (заданную на σ -алгебре \mathfrak{B}) называют **вероятностью (или вероятностной мерой)**, если она удовлетворяет следующим аксиомам:

Аксиома 1 (аксиома неотрицательности): $P(A) \geq 0$;

Аксиома 2 (аксиома нормированности): $P(\Omega) = 1$;

Аксиома 3 (расширенная аксиома сложения): для любых попарно несовместных событий A_1, \dots, A_n, \dots справедливо равенство

$$P(A_1 + \dots + A_n + \dots) = P(A_1) + \dots + P(A_n) + \dots$$

Значение $P(A)$ называют **вероятностью события A** .

Иногда вместо аксиомы 3 удобно использовать две другие аксиомы.

Аксиома 3' (аксиома сложения): для любых попарно непересекающихся событий A_1, \dots, A_n справедливо равенство

$$P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n).$$

Аксиома 4 (аксиома непрерывности): если последовательность событий A_1, \dots, A_n, \dots такова, что $A_n \subset A_{n+1}$, $n \in \mathbb{N}$, и $A_1 \cup \dots \cup A_n \cup \dots = A$, то $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.

Можно доказать, что аксиомы 3' и 4 в совокупности равносильны аксиоме 3.

Замечание 2.1 Если пространство элементарных исходов Ω является конечным или счетным множеством, то каждому элементарному исходу $\omega_i \in \Omega$, $i = 1, 2, \dots$, можно поставить в соответствие число $P(\omega_i) = p_i \geq 0$ так, что

$$\sum_{\omega_i \in \Omega} P(\omega_i) = \sum_{i=1}^{\infty} p_i = 1.$$

Тогда для любого события $A \subset \Omega$ в силу аксиомы 3 вероятность $P(A)$ равна сумме вероятностей $P(\omega_i)$ всех тех элементарных исходов, которые входят в событие A , т.е.

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

Таким образом, мы определили вероятность любого события, используя вероятности элементарных исходов. Заметим, что вероятности элементарных исходов можно задавать совершенно произвольно, лишь бы они были неотрицательными и в сумме составляли единицу. Именно в этом и состоит идея аксиоматического определения вероятности. #

В следующей теореме докажем утверждения, описывающие ряд полезных свойств вероятности.

Теорема 2.1 Вероятность удовлетворяет следующим свойствам.

1. Вероятность противоположного события $P(\bar{A}) = 1 - P(A)$.
2. Вероятность невозможного события $P(\emptyset) = 0$.
3. Если $A \subset B$, то $P(A) \leq P(B)$ ("большому" событию соответствует большая вероятность).
4. Вероятность заключена между 0 и 1: $0 \leq P(A) \leq 1$.
5. Вероятность объединения двух событий $P(A \cup B) = P(A) + P(B) - P(AB)$.
6. Вероятность объединения любого конечного числа событий

$$P(a_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n) - P(A_1 A_2) - P(A_1 A_3) - \dots - P(A_{n-1} A_n) + \\ + P(A_1 A_2 A_3) + \dots + (-1)^{n+1} P(A_1 A_2 \dots A_n).$$

ДОКАЗАТЕЛЬСТВО. Поскольку $\Omega = A + \bar{A}$, то, согласно расширенной аксиоме сложения, $P(\Omega) = P(A) + P(\bar{A})$, откуда с учетом аксиомы нормированности получаем утверждение 1.

Утверждение 2 вытекает из равенства $A = A + \emptyset$ и расширенной аксиомы сложения.

Пусть $A \subset B$. Тогда $B = A + (B \setminus A)$. В соответствии с расширенной аксиомой сложения $P(B) = P(A) + P(B \setminus A)$. Отсюда и из аксиомы неотрицательности приходим к утверждению 3.

В частности, так как всегда $A \subset \Omega$, то с учетом аксиомы неотрицательности получаем утверждение 4.

Поскольку $A \cup B = A + (B \setminus A)$, $B = (B \setminus A) + AB$, то, используя расширенную аксиому сложения, находим $P(A \cup B) = P(A) + P(B \setminus A)$ и $P(B) = P(B \setminus A) + P(AB)$. Подставляя в первое из последних двух равенств вероятность $P(B \setminus A)$, выраженную из второго равенства, приходим к утверждению 5.

Утверждение 6 можно доказать с помощью метода математической индукции по n . Так, для трех событий A , B и C

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B \cup C) - P(A(B \cup C)) = P(A) + P(B) + P(C) - P(BC) - P(AB \cup AC) = \\ &= P(A) + P(B) + P(C) - P(BC) - P(AB) - P(AC) + P(ABC). \end{aligned}$$

Для четырех и более событий это утверждение проверьте самостоятельно.

Замечание 2.2 Утверждения 5 и 6 называют *теоремами сложения вероятностей* для двух и для n событий соответственно.

Приведем пример, показывающий, что без учета того, что *события совместные*, можно прийти к неправильному результату.

Пример 2.3 Опыт состоит в двукратном подбрасывании симметричной монеты. Найдем вероятность события A , означающего появление “герба” хотя бы один раз. Обозначим A_i появление “герба” при i -м подбрасывании, $i = 1, 2$. Ясно, что $A = A_1 \cup A_2$, и в соответствии с классической схемой вероятности $P(A_1) = P(A_2) = \frac{1}{2}$. Если не учитывать, что A_1 и A_2 — совместные события, то можно получить “результат”

$$P(A) = P(A_1) + P(A_2) = \frac{1}{2} + \frac{1}{2} = 1,$$

противоречащий здравому смыслу, поскольку ясно, что *событие* A не является *достоверным*. Применяя теорему сложения для двух совместных событий и учитывая равенство $P(A_1 A_2) = \frac{1}{4}$, находим

$$P(A) = P(A_1) + P(A_2) - P(A_1 A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}. \quad \#$$

Определение 2.5 Тройку $(\Omega, \mathfrak{B}, P)$, состоящую из пространства элементарных исходов Ω , с σ -алгеброй событий \mathfrak{B} и определенной на \mathfrak{B} вероятности P , называют *вероятностным пространством*.

Таким образом, понятие вероятностного пространства объединяет хорошо известные физические понятия: исход опыта, событие, вероятность события.

Лекция 3

Условная вероятность

Рассмотрим события A и B , связанные с одним и тем же опытом. Пусть из каких-то источников нам стало известно, что событие B наступило, но не известно, какой конкретно из элементарных исходов, составляющих событие B , произошел. Что можно сказать в этом случае о вероятности события A ?

Вероятность события A , вычисленную в предположении, что событие B произошло, принято называть *условной вероятностью* и обозначать $P(A|B)$.

Понятие условной вероятности играет важнейшую роль в современной теории вероятностей. Условная вероятность позволяет учитывать дополнительную информацию при определении вероятности события. В ряде случаев при помощи условной вероятности можно существенно упростить вычисление вероятности. Понятию условной вероятности и посвящена настоящая лекция.

Определение условной вероятности

Предположим сначала, что мы находимся в рамках классической схемы. Пусть событиям A и B благоприятствуют N_A и N_B элементарных исходов соответственно. Посмотрим, что дает нам имеющаяся информация о событии B . Поскольку событие B произошло, то достоверно известно, что в результате опыта появился один из N_B элементарных исходов, составляющих событие B . Значит, теперь уже при определении степени возможности события A необходимо выбирать только из N_B возможных исходов, причем событию A благоприятствуют N_{AB} исходов, при которых происходят и событие A , и событие B , или, другими словами, происходит событие AB . При этом по-прежнему будем считать все N_B входящих в событие B исходов равновероятными. Поэтому *условную вероятность* $P(A|B)$ события A при условии события B в рамках классической схемы вероятности естественно определить как отношение числа N_{AB} исходов, благоприятствующих совместному осуществлению событий A и B , к числу N_B исходов, благоприятствующих событию B , т. е. $P(A|B) = \frac{N_{AB}}{N_B}$.

Если теперь поделить числитель и знаменатель полученного выражения на общее число N элементарных исходов, то придем к формуле $P(A|B) = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)}$.

На основании изложенного выше можно дать следующее определение.

Определение 3.1 *Условной вероятностью* события A при условии (наступлении) события B называют отношение вероятности пересечения событий A и B к вероятности события B :

$$P(A|B) = \frac{P(AB)}{P(B)}. \quad (3.1)$$

При этом предполагают, что $P(B) \neq 0$.

В связи с появлением термина “условная вероятность” будем вероятность события называть также *безусловной вероятностью* события.

Рассмотрим теперь условную вероятность $P(A|B)$ как функцию события A .

Теорема 3.1 *Условная вероятность $P(A|B)$ обладает всеми свойствами безусловной вероятности $P(A)$.*

Доказательство. Для доказательства достаточно показать, что условная вероятность $P(A|B)$ удовлетворяет аксиомам 1, 2 и 3 (см. 2.3).

Из определения 3.1 следует, что условная вероятность, удовлетворяет аксиоме неотрицательности, так как числитель дроби является неотрицательным числом, а знаменатель — положительным числом.

Поскольку $\Omega B = B$, то $\mathbf{P}(\Omega|B) = \frac{\mathbf{P}(\Omega B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B)}{\mathbf{P}(B)} = 1$, т. е. условная вероятность удовлетворяет аксиоме нормированности.

Наконец, пусть A_1, \dots, A_n, \dots — попарно непересекающиеся события. Тогда

$$(A_1 + \dots + A_n + \dots)B = A_1B + \dots + A_nB + \dots$$

и

$$\begin{aligned} \mathbf{P}(A_1 + \dots + A_n + \dots | B) &= \frac{\mathbf{P}((A_1 + \dots + A_n + \dots)B)}{\mathbf{P}(B)} = \\ &= \frac{\mathbf{P}(A_1B) + \dots + \mathbf{P}(A_nB) + \dots}{\mathbf{P}(B)} = \mathbf{P}(A_1|B) + \dots + \mathbf{P}(A_n|B) + \dots, \end{aligned}$$

где в последнем равенстве использовано свойство умножения сходящегося ряда на постоянную. Следовательно, условная вероятность удовлетворяет *расширенной аксиоме сложения* 3.

Смысл теоремы 3.1 заключается в том, что условная вероятность представляет собой безусловную вероятность, заданную на новом пространстве Ω_1 элементарных исходов, совпадающем с событием B .

Пример 3.1 Рассмотрим опыт с однократным бросанием игральной кости, но не обычной, а с раскрашенными гранями: грани с цифрами 1, 3 и 6 окрашены красным, а грани с цифрами 2, 4 и 5 — белым цветом. Введем события: A_1 — выпадение нечетного числа очков; A_2 — выпадение четного числа очков; B — появление грани красного цвета. Интуитивно ясно, что если произошло событие B , то условная вероятность события A_1 больше, чем условная вероятность события A_2 , поскольку на красных гранях нечетных чисел в два раза больше, чем четных. Заметим, что безусловные вероятности событий A_1 и A_2 при этом одинаковы и равны, очевидно, $1/2$.

Найдем условные вероятности событий A_1 и A_2 при условии события B . Очевидно, что

$$\mathbf{P}(A_1B) = \frac{N_{A_1B}}{N} = \frac{2}{6} = \frac{1}{3}, \quad \mathbf{P}(A_2B) = \frac{N_{A_2B}}{N} = \frac{1}{6}, \quad \mathbf{P}(B) = \frac{3}{6} = \frac{1}{2}.$$

Следовательно, в силу определения 3.1 условной вероятности имеем

$$\mathbf{P}(A_1|B) = \frac{1/3}{1/2} = \frac{2}{3}, \quad \mathbf{P}(A_2|B) = \frac{1/6}{1/2} = \frac{1}{3},$$

что подтверждает наше предположение.

Геометрическая интерпретация условной вероятности

При практическом вычислении условной вероятности события A при условии, что событие B произошло, часто удобно трактовать условную вероятность как безусловную, но заданную не на исходном пространстве Ω элементарных исходов, а на новом пространстве $\Omega_1 = B$ элементарных исходов. Действительно, используя *геометрическое определение вероятности*, получаем для безусловной и условной вероятностей события A (на рис. 3.1 заштрихованная область соответствует событию AB):

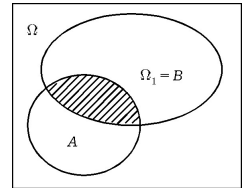


Рис. 3.1.

$$\mathbf{P}(A) = \frac{S_A}{S_\Omega} = \frac{S_{A\Omega}}{S_\Omega}, \quad \mathbf{P}(A|B) = \frac{S_{AB}/S_\Omega}{S_B/S_\Omega} = \frac{S_{AB}}{S_B} = \frac{S_{A\Omega_1}}{S_{\Omega_1}}.$$

Здесь S_A , S_Ω и т.д. обозначают соответственно площади A , Ω и т.д. Таким образом, выражение для $\mathbf{P}(A|B)$ будет совпадать с выражением для $\mathbf{P}(A)$, вычисленным в соответствии со *схемой геометрической вероятности*, если исходное пространство Ω элементарных исходов заменить новым пространством $\Omega_1 = B$.

Пример 3.2 Из урны, в которой $a = 7$ белых и $b = 3$ черных шаров, наугад без возвращения извлекают два шара. Пусть событие A_1 состоит в том, что первый извлеченный из урны шар является белым, а A_2 — белым является второй шар. Требуется найти $\mathbf{P}(A_2|A_1)$.

Приведем решение этой задачи двумя способами.

Первый способ. В соответствии с определением условной вероятности имеем (опуская пояснения):

$$\mathbf{P}(A_2|A_1) = \frac{\mathbf{P}(A_1A_2)}{\mathbf{P}(A_1)} = \frac{C_7^2/C_{10}^2}{C_7^1/C_{10}^1} = \frac{2}{3}.$$

Второй способ. Перейдем к новому пространству Ω_1 элементарных исходов. Так как событие A_1 произошло, то это означает, что в новом пространстве элементарных исходов всего равновозможных исходов $N_{\Omega_1} = a + b - 1 = 9$, а событию A_2 благоприятствует при этом $N_{A_2} = a - 1 = 6$ исходов. Следовательно, $P(A_2|A_1) = \frac{6}{9} = \frac{2}{3}$.

Формула умножения вероятностей

При решении различных задач вероятностного характера часто интересующее нас событие A можно достаточно просто выразить через некоторые события A_1, A_2, \dots, A_n с помощью операций объединения или пересечения. Если $A = A_1 A_2 \dots A_n$, то для нахождения вероятности $P(A)$ события A обычно удобно использовать следующую теорему.

Теорема 3.2 (теорема умножения вероятностей) Пусть событие $A = A_1 A_2 \dots A_n$ (т. е. A — пересечение событий A_1, A_2, \dots, A_n) и $P(A) > 0$. Тогда справедливо равенство

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1}),$$

называемое **формулой умножения вероятностей**.

ДОКАЗАТЕЛЬСТВО. Поскольку $P(A) = P(A_1 A_2 \dots A_n) > 0$, а $A_1 A_2 \dots A_k \supseteq A_1 A_2 \dots A_n$ ($k = \overline{1, n-1}$), то и $P(A) = P(A_1 A_2 \dots A_k) > 0$. Учитывая это неравенство, согласно определению 3.1 условной вероятности, имеем

$$P(A_n|A_1 A_2 \dots A_{n-1}) = \frac{P(A_1 A_2 \dots A_n)}{P(A_1 A_2 \dots A_{n-1})}.$$

Умножая обе части этого равенства на $P(A_1 A_2 \dots A_{n-1})$, получаем

$$P(A_1 A_2 \dots A_n) = P(A_1 A_2 \dots A_{n-1})P(A_n|A_1 A_2 \dots A_{n-1}).$$

Аналогично находим $P(A_1 A_2 \dots A_{n-1}) = P(A_1 A_2 \dots A_{n-2})P(A_{n-1}|A_1 A_2 \dots A_{n-2})$. Тогда

$$P(A_1 A_2 \dots A_n) = P(A_1 A_2 \dots A_{n-2})P(A_{n-1}|A_1 A_2 \dots A_{n-2}) \times P(A_n|A_1 A_2 \dots A_{n-1}).$$

Продолжая эту процедуру, получаем формулу умножения вероятностей.

Пример 3.3 На семи карточках написаны буквы, образующие слово “СОЛОВЕЙ”. Карточки перемешивают и из них наугад последовательно извлекают и выкладывают слева направо три карточки. Найдём вероятность того, что получится слово “ВОЛ” (событие A).

Введем события: A_1 — на первой выбранной карточке написана буква “В”; A_2 — на второй карточке — буква “О”; A_3 — на третьей карточке — буква “Л”. Тогда событие A есть пересечение событий A_1, A_2 и A_3 . Следовательно, в соответствии с формулой умножения вероятностей $P(A) = P(A_1 A_2 A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2)$. Согласно классическому определению 2.1 вероятности, имеем $P(A_1) = \frac{1}{7}$.

Если событие A_1 произошло, то на шести оставшихся карточках буква “О” встречается два раза, поэтому условная вероятность $P(A_2|A_1) = \frac{2}{6} = \frac{1}{3}$. Аналогично определяем $P(A_3|A_1 A_2) = \frac{1}{5}$. Окончательно получаем $P(A) = P(A_1 A_2 A_3) = \frac{1}{7} \cdot \frac{1}{3} \cdot \frac{1}{5} = \frac{1}{105} \approx 0,0095$.

Независимые и зависимые события

Из рассмотренных выше примеров видно, что условная вероятность $P(A|B)$ события A при условии, что событие B произошло, может как совпадать с безусловной вероятностью $P(A)$, так и не совпадать, т.е. наступление события B может влиять или не влиять на вероятность события A . Поэтому естественно степень связи (или степень зависимости) событий A и B оценивать путем сопоставления их условных вероятностей $P(A|B)$, $P(B|A)$ с безусловными.

Определение 3.2 События A и B , имеющие ненулевую вероятность, называют **независимыми**, если условная вероятность A при условии B совпадает с безусловной вероятностью A или если условная вероятность B при условии A совпадает с безусловной вероятностью B , т.е.

$$P(A|B) = P(A) \tag{3.2}$$

или

$$P(B|A) = P(B), \tag{3.3}$$

в противном случае события A и B называют **зависимыми**.

Теорема 3.3 События A и B , имеющие ненулевую вероятность, являются независимыми тогда и только тогда, когда

$$P(AB) = P(A)P(B). \quad (3.4)$$

ДОКАЗАТЕЛЬСТВО. Пусть выполнено равенство (3.3). Воспользовавшись формулой умножения вероятностей для двух событий, получим

$$P(AB) = P(A)P(B|A) = P(A)P(B).$$

К аналогичному выводу приходим и в случае выполнения равенства (3.2), т.е. из условия независимости событий следует (3.4).

Обратно, пусть выполнено равенство (3.4). Тогда, согласно определению 3.1 условной вероятности,

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A) \quad \text{и} \quad P(B|A) = \frac{P(AB)}{P(A)} = P(B),$$

т.е. в силу определения 3.2 события A и B независимы.

Таким образом, в качестве эквивалентного определения независимости двух событий, имеющих ненулевую вероятность, может служить следующее определение.

Определение 3.3 События A и B называют независимыми, если выполняется равенство (3.4).

Отметим, что последним определением можно пользоваться даже в том случае, когда вероятности событий A или B равны нулю.

Замечание 3.1 Из теоремы 3.3 следует, что если в определении 3.2 независимости выполняется одно из равенств (3.2) или (3.3), то выполняется автоматически и другое, т.е. в определении 3.2 достаточно потребовать выполнения любого одного из них.

Пример 3.4 Из колоды карт, содержащей $n = 36$ карт, наугад извлекают одну карту. Обозначим через A событие, соответствующее тому, что извлеченная карта будет пиковой масти, а B — событие, соответствующее появлению “дамы”. Определим, являются ли зависимыми события A и B .

После вычислений получаем

$$P(A) = \frac{9}{36} = \frac{1}{4}, \quad P(B) = \frac{4}{36} = \frac{1}{9}, \quad P(AB) = \frac{1}{36}, \quad P(B|A) = \frac{P(AB)}{P(A)} = \frac{1/36}{1/4} = \frac{1}{9} = P(B),$$

т.е. выполняется равенство (3.2), и поэтому события A и B независимы. Отметим, что в соответствии с замечанием 3.1, имеет место и равенство (3.3) $P(A|B) = \frac{1/36}{1/9} = \frac{1}{4} = P(A)$. Следовательно, события A и B независимы. #

Изменим теперь условия опыта, дополнительно добавив в колоду, допустим, $N = 100$ “пустых” карт (без рисунка). Изменится ли ответ? Имеем $P(B) = \frac{4}{136} = \frac{1}{34}$, т.е. безусловная вероятность события B уменьшилась. Однако условная вероятность $P(B|A) = \frac{P(AB)}{P(A)} = \frac{1/136}{1/4} = \frac{1}{9}$ не изменилась, т.е. события A и B стали зависимыми.

Теорема 3.4 Если события A и B независимы, то независимыми также являются пары событий \bar{A} и B , A и \bar{B} , \bar{A} и \bar{B} , если вероятности соответствующих событий ненулевые.

ДОКАЗАТЕЛЬСТВО. В силу теоремы 3.1 и независимости событий A и B имеем: $P(\bar{A}|B) = 1 - P(A|B) = 1 - P(A) = P(\bar{A})$, что означает независимость событий \bar{A} и B . Независимость остальных пар событий можно доказать аналогично.

Определение 3.4 События A_1, A_2, \dots, A_n называют **независимыми в совокупности**, если вероятность пересечения любых двух различных событий равна произведению вероятностей этих событий; вероятность пересечения любых трех событий равна произведению их вероятностей; ...; вероятность пересечения всех событий равна произведению их вероятностей.

Для событий A_1, A_2, \dots, A_n , независимых в совокупности, имеет место утверждение, аналогичное утверждению теоремы 3.4.

Теорема 3.5 Если события A_1, A_2, \dots, A_n независимы в совокупности, то и события $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ независимы в совокупности. #

Если только любые два события из данной совокупности являются независимыми, то говорят о **парной независимости событий** из этой совокупности.

Так же как и в случае двух событий, можно показать, что на вероятность каждого из независимых в совокупности событий не оказывает влияние появление или неоявление остальных событий.

Замечание 3.2 В силу определения независимости событий в совокупности *формула умножения вероятностей* для независимых в совокупности событий имеет вид $P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n)$. #

Из независимости событий с ненулевыми вероятностями в совокупности, согласно теореме 3.3, следует их попарная независимость. Однако из попарной независимости, вообще говоря, независимость в совокупности не следует, что демонстрирует следующий пример.

Пример 3.5 Опыт состоит в однократном подбрасывании тетраэдра, грани которого “пронумерованы” следующим образом: на трех гранях стоят цифры 1, 2 и 3 соответственно (одна цифра на каждой из них), а на четвертой присутствуют все цифры 1, 2 и 3.

Введем события A_i — падение тетраэдра на грань, на которой присутствует цифра i , $i = \overline{1, 3}$. Покажем, что события A_1 , A_2 и A_3 попарно независимы, но зависимы в совокупности.

Согласно классическому определению вероятности, получаем

$$P(A_i) = \frac{2}{4} = \frac{1}{2}, \quad i = \overline{1, 3}, \quad P(A_2|A_1) = \frac{P(A_1 A_2)}{P(A_1)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Аналогично $P(A_i|A_j) = \frac{1}{2}$ при любых $i, j = \overline{1, 3}$, $i \neq j$, т.е. события A_1 , A_2 и A_3 являются попарно независимыми. Однако, например, $P(A_1|A_2 A_3) = \frac{P(A_1 A_2 A_3)}{P(A_2 A_3)} = \frac{1/4}{1/4} = 1 \neq P(A_1)$, т.е. события A_1 , A_2 и A_3 зависимы в совокупности. #

Заметим, что, когда говорят о независимости событий A_1, \dots, A_n , подразумевают именно независимость событий в совокупности, в отличие от попарной независимости событий A_1, \dots, A_n .

Запишем формулу для вероятности *объединения независимых событий*. Пусть $A = A_1 \cup \dots \cup A_n$. Тогда в соответствии с *законом де Моргана* $\overline{A} = \overline{A_1} \dots \overline{A_n}$. Если события A_1, \dots, A_n независимые, то, согласно теореме 3.5, события $\overline{A_1}, \dots, \overline{A_n}$ также независимые и, значит, $P(\overline{A}) = P(\overline{A_1}) \dots P(\overline{A_n})$. Отсюда окончательно получаем *формулу для вероятности объединения независимых событий*:

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= 1 - P(\overline{A_1 \cup \dots \cup A_n}) = 1 - P(\overline{A_1} \cap \dots \cap \overline{A_n}) = \\ &= 1 - P(\overline{A_1}) \dots P(\overline{A_n}) = 1 - [1 - P(A_1)] \dots [1 - P(A_n)]. \end{aligned}$$

Замечание 3.3 (о связи между совместными и зависимыми событиями). Между понятиями “несовместные” и “независимые” события имеется следующая связь:

- 1) если A и B — несовместные события (и $P(A) \neq 0$, и $P(B) \neq 0$), то они обязательно зависимые;
- 2) если A и B — совместные события, то они могут быть и зависимыми и независимыми;
- 3) если A и B — зависимые события, то они могут быть и совместными и несовместными. #

Лекция 4

Формула полной вероятности. Формула Байеса. Схема Бернулли

Формула полной вероятности

Предположим, что в результате опыта может произойти одно из n событий H_1, H_2, \dots, H_n , которые удовлетворяют следующим двум условиям:

- 1) они являются *попарно несовместными*, т.е. $H_i H_j = \emptyset$ при $i \neq j$;
- 2) хотя бы одно из них обязательно должно произойти в результате опыта, другими словами, их объединение есть *достоверное событие*, т.е. $H_1 \cup \dots \cup H_n = \Omega$.

Определение 4.1 События H_1, H_2, \dots, H_n удовлетворяющие условиям 1 и 2, называют *гипотезами*.

Заметим, что если события удовлетворяют второму из двух указанных требований, то их совокупность называют *полной группой событий*. Таким образом, гипотезы — это попарно несовместные события, образующие полную группу событий.

Пусть также имеется некоторое событие A и известны *вероятности гипотез* $P(H_1), \dots, P(H_n)$, которые предполагаются ненулевыми, и *условные вероятности* $P(A|H_1), \dots, P(A|H_n)$ события A при выполнении этих гипотез. Задача состоит в вычислении *безусловной вероятности* события A . Для решения этой задачи используют следующую теорему.

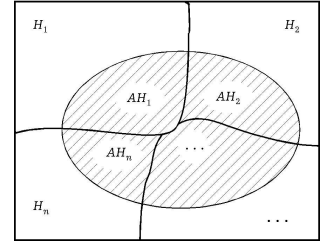


Рис. 4.1.

Теорема 4.1 Пусть для некоторого события A и гипотез H_1, \dots, H_n известны $P(H_1), \dots, P(H_n)$, которые положительны, и $P(A|H_1), \dots, P(A|H_n)$. Тогда безусловную вероятность $P(A)$ определяют по формуле

$$P(A) = P(H_1)P(A|H_1) + \dots + P(H_n)P(A|H_n), \quad (4.1)$$

которую называют *формулой полной вероятности*.

ДОКАЗАТЕЛЬСТВО. Представим событие A в виде

$$A = A\Omega = A(H_1 + \dots + H_n) = AH_1 + \dots + AH_n$$

(на рис. 4.1 область, соответствующая событию A , заштрихована). С учетом того, что события AH_i , $i = \overline{1, n}$, несовместны, имеем

$$P(A) = P(AH_1) + \dots + P(AH_n).$$

В соответствии с *формулой умножения вероятностей* получаем

$$P(AH_1) = P(H_1)P(A|H_1), \dots, P(AH_n) = P(H_n)P(A|H_n).$$

Поэтому

$$P(A) = P(H_1)P(A|H_1) + \dots + P(H_n)P(A|H_n).$$

Формула полной вероятности при всей своей простоте играет весьма существенную роль в теории вероятностей.

Пример 4.1 Путник должен попасть из пункта B в пункт A в соответствии со схемой дорог изображенной на рис. 4.2. Выбор любой дороги в любом пункте равновозможен. Найдём вероятность события A — достижения путником намеченной цели.

Для того чтобы попасть в пункт A , путник должен пройти один из промежуточных пунктов H_1 , H_2 или H_3 . Введем гипотезы H_i , где H_i означает, что путник выбрал в пункте B путь, ведущий в пункт H_i , $i = 1, 2, 3$. Ясно, что события H_i несовместны и одно из них обязательно происходит, причем в силу равновозможности выбора дорог из B в H_i $P(H_i) = \frac{1}{3}$. Остается вычислить условные вероятности $P(A|H_i)$, которые легко найти, если рассматривать новое *пространство элементарных исходов*, соответствующее выбранной гипотезе H_i .

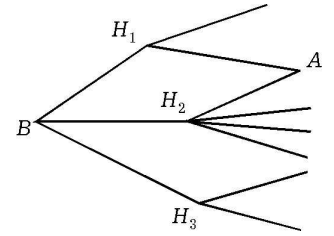


Рис. 4.2.

Например, появление H_1 означает, что есть два равновозможных исхода (из пункта H_1 выходят две дороги), из которых лишь один благоприятствует событию A , т.е. $P(A|H_1) = \frac{1}{2}$. Аналогично находим, что $P(A|H_2) = \frac{1}{4}$ и $P(A|H_3) = 0$.

Согласно формуле 4.1 полной вероятности, получаем

$$P(A) = \frac{1}{3} \cdot \left(\frac{1}{2} + \frac{1}{4} + 0 \right) = 0,25. \quad \#$$

Заметим, что данная задача может иметь техническую интерпретацию: сеть дорог — это сеть каналов передачи информации, а $P(A)$ — вероятность передачи сообщения по такой сети.

Формула Байеса

Пусть по-прежнему некоторое событие A может произойти с одним из событий H_1, \dots, H_n , образующих полную группу попарно несовместных событий, называемых, как уже отмечалось, гипотезами. Предположим, что известны вероятности гипотез $P(H_1), \dots, P(H_n)$ ($P(H_i) > 0$, $i = \overline{1, n}$) и что в результате опыта событие A произошло, т.е. получена дополнительная информация. Спрашивается, как “изменяться” вероятности гипотез, т.е. чему будут равны условные вероятности $P(H_1|A), \dots, P(H_n|A)$, если известны также условные вероятности $P(A|H_1), \dots, P(A|H_n)$ события A ? Для ответа на этот вопрос используют следующую теорему.

Теорема 4.2 Пусть для некоторого события A , $P(A) > 0$, и гипотез H_1, \dots, H_n известны $P(H_1), \dots, P(H_n)$ ($P(H_i) > 0$, $i = \overline{1, n}$) и $P(A|H_1), \dots, P(A|H_n)$. Тогда условная вероятность $P(H_i|A)$, $i = \overline{1, n}$, гипотезы H_i при условии события A определяется **формулой Байеса**

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(H_1)P(A|H_1) + \dots + P(H_n)P(A|H_n)}. \quad (4.2)$$

ДОКАЗАТЕЛЬСТВО. Согласно определению 3.1 условной вероятности, $P(H_i|A) = \frac{P(AH_i)}{P(A)}$. Выражая теперь по формуле умножения вероятностей $P(AH_i)$ через $P(A|H_i)$ и $P(H_i)$, получаем $P(AH_i) = P(H_i)P(A|H_i)$. Поэтому $P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}$.

Подставляя вместо вероятности $P(A)$ ее значение, вычисленное в соответствии с формулой (4.1) полной вероятности, приходим к утверждению теоремы.

Формула Байеса находит широкое применение в математической статистике, теории принятия решений и их приложениях. Заметим, что *вероятности* $P(H_1), \dots, P(H_n)$ обычно называют **априорными** (т.е. полученными “до опыта”), а условные *вероятности* $P(H_1|A), \dots, P(H_n|A)$ — **апостериорными** (т.е. полученными “после опыта”).

Пример 4.2 Врач после осмотра больного считает, что возможно одно из двух заболеваний, которые мы зашифруем номерами 1 и 2, причем степень своей уверенности в отношении правильности диагноза он оценивает как 40 % и 60 % соответственно. Для уточнения диагноза больного направляют на анализ, исход которого дает положительную реакцию при заболевании 1 в 90 % случаев и при заболевании 2 — в 20 % случаев. Анализ дал положительную реакцию. Как изменится мнение врача после этого?

Обозначим через A событие, означающее, что анализ дал положительную реакцию. Естественно ввести следующие гипотезы: H_1 — имеет место заболевание 1; H_2 — имеет место заболевание 2. Из условий задачи ясно, что априорные вероятности гипотез равны: $P(H_1) = 0,4$ и $P(H_2) = 0,6$, а условные вероятности события A при наличии гипотез H_1 и H_2 равны 0,9 и 0,2 соответственно. Используя формулу Байеса, находим $P(H_i|A) = \frac{0,4 \cdot 0,9}{0,4 \cdot 0,9 + 0,6 \cdot 0,2} = 0,75$. Итак, врач с большей уверенностью признает наличие заболевания 1.

Схема Бернулли

Повторные испытания — это последовательное проведение n раз одного и того же опыта или одновременное проведение n одинаковых опытов. Например, при контроле уровня надежности прибора могут либо проводить n испытаний с одним и тем же прибором, если после отказа полностью восстанавливают его исходные свойства, либо ставить на испытания n опытных образцов этого прибора, которые считают идентичными.

Определение 4.2 *Схемой Бернулли* (или *последовательностью независимых одинаковых испытаний*, или *биномиальной схемой* испытаний) называют последовательность испытаний, удовлетворяющую следующим условиям:

- 1) при каждом испытании различают лишь два *исхода*: появление некоторого *события* A , называемого “успехом”, либо появление его *дополнения* \bar{A} , называемого “неудачей”;
- 2) испытания являются *независимыми*, т.е. вероятность успеха в k -м испытании не зависит от исходов всех испытаний до k -го;
- 3) *вероятность* успеха во всех испытаниях постоянна и равна $P(A) = p$.

Вероятность неудачи в каждом испытании обозначим q , т.е. $P(\bar{A}) = 1 - p = q$.

Приведем примеры реальных испытаний, которые в той или иной степени “вписываются” в рамки сформулированной модели испытаний по схеме Бернулли.

1. Последовательное подбрасывание n раз симметричной монеты (здесь успехом является появление “герба” с вероятностью $p = 1/2$) или последовательное бросание n раз игральной кости (здесь успехом можно считать, например, появление шестерки с вероятностью $p = 1/6$). Эти две реальные схемы испытаний являются примером идеального соответствия схеме испытаний Бернулли.

2. Последовательность n выстрелов стрелка по мишени можно лишь приближенно рассматривать как схему испытаний Бернулли, так как независимость результатов стрельбы может нарушаться либо из-за “пристрелки” спортсмена, либо вследствие его утомляемости.

3. Испытания n изделий в течение заданного срока при контроле уровня их надежности, как правило, хорошо согласуются с моделью испытаний по схеме Бернулли, если на испытания поставлены идентичные образцы.

При рассмотрении схемы испытаний Бернулли основной задачей является нахождение вероятности события A_k , состоящего в том, что в n испытаниях успех наступит ровно k раз, $k = \overline{0, n}$. Для решения этой задачи используют следующую теорему, обозначая вероятность $P(A_k)$ через $P_n(k)$.

Теорема 4.3 *Вероятность $P_n(k)$ того, что в n испытаниях по схеме Бернулли произойдет ровно k успехов, определяется формулой Бернулли*

$$P_n(k) = C_n^k p^k q^{n-k}, \quad k = \overline{0, n}. \quad (4.3)$$

ДОКАЗАТЕЛЬСТВО. Результат каждого опыта можно записать в виде последовательности УНН...У, состоящей из n букв “У” и “Н”, причем буква “У” на i -м месте означает, что в i -м испытании произошел успех, а “Н” — неудача. *Пространство элементарных исходов* Ω состоит из 2^n исходов, каждый из которых отождествляется с определенной

последовательностью УНН...У. Каждому элементарному исходу $\omega = \text{УНН...У}$ можно поставить в соответствие вероятность $P(\omega) = P(\text{УНН...У})$. В силу независимости испытаний события У, Н, Н, ..., У являются независимыми в совокупности, и потому по теореме умножения вероятностей имеем $P(\omega) = p^i q^{n-i}$, $i = \overline{0, n}$, если в n испытаниях успех “У” имел место i раз, а неуспех “Н”, следовательно, $n - i$ раз.

Событие A_k происходит всякий раз, когда реализуется элементарный исход ω , в котором $i = k$. Вероятность любого такого элементарного исхода равна $p^k q^{n-k}$.

Число таких исходов совпадает с числом способов, которыми можно расставить k букв “У” на n местах, не учитывая порядок, в котором их расставляют. Число таких способов равно C_n^k .

Так как A_k есть объединение (сумма) всех указанных элементарных исходов, то окончательно получаем для вероятности $P(A_k) = P_n(k)$ формулу (4.3).

Формулу (4.3) называют также **биномиальной**, так как ее правая часть представляет собой $(k+1)$ -й член формулы бинома Ньютона.

$$1 = (p + q)^n = C_n^0 q^n + C_n^1 p^1 q^{n-1} + \dots + C_n^k p^k q^{n-k} + \dots + C_n^n p^n.$$

Набор вероятностей $P_n(k)$, $k = \overline{0, n}$, называют **биномиальным распределением вероятностей**.

Из формулы Бернулли вытекают два следствия.

1. Вероятность появления успеха (события A) в n испытаниях не более k_1 раз и не менее k_2 раз равна:

$$P\{k_1 \leq k \leq k_2\} = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}. \quad (4.4)$$

Это следует из того, что *события* A_k при разных k являются несовместными.

2. В частном случае при $k_1 = 1$ и $k_2 = n$ из (4.4) получаем формулу для вычисления вероятности хотя бы одного успеха в n испытаниях:

$$\mathbf{P}\{k \geq 1\} = 1 - q^n. \quad (4.5)$$

Пример 4.3 Монету (симметричную) подбрасывают $n = 10$ раз. Определим вероятность выпадения “герба”: а) ровно пять раз; б) не более пяти раз; в) хотя бы один раз.

В соответствии с формулой (4.3) Бернулли имеем: а) $P_{10}(5) = C_{10}^5 \left(\frac{1}{2}\right)^{10} = \frac{252}{1024} = 0,246$;

$$\text{б) } \mathbf{P}\{k \leq 5\} = \frac{C_{10}^0 + C_{10}^1 + C_{10}^2 + C_{10}^3 + C_{10}^4 + C_{10}^5}{1024} = \frac{638}{1024} \approx 0,623; \quad \text{в) } \mathbf{P}\{k \geq 1\} = 1 - \left(\frac{1}{2}\right)^{10} \approx 0,999.$$

Пример 4.4 Вероятность выигрыша на один лотерейный билет равна 0,01. Определим, сколько билетов нужно купить, чтобы вероятность хотя бы одного выигрыша в лотерее была не менее заданного значения $P_3 = 0,9$.

Пусть куплено n билетов. Предположим, что общее число билетов, разыгрывающихся в лотерее велико (во много раз больше купленных билетов). При этом можно считать, что каждый билет выигрывает независимо от остальных с вероятностью $p = 0,01$. Тогда вероятность получить k выигрышных билетов можно определить, используя формулу Бернулли. В частности, согласно (4.5), имеем при $q = 1 - p$:

$$\mathbf{P}\{k \geq 1\} = 1 - q^n = 1 - (1 - p)^n \geq P_3,$$

откуда получаем

$$n \geq \frac{\ln(1 - P_3)}{\ln(1 - p)} = \frac{\ln 0,1}{\ln 0,99} \approx 230.$$

Таким образом, нужно купить не менее 230 лотерейных билетов. #

Лекция 5

Одномерные случайные величины

Определение случайной величины

Случайной величиной естественно называть числовую величину, значение которой зависит от того, какой именно *элементарный исход* произошел в результате эксперимента со случайным исходом. Множество всех значений, которые случайная величина может принимать, называют **множеством возможных значений** этой **случайной величины**.

Следовательно, для задания случайной величины необходимо каждому элементарному исходу поставить в соответствие число — значение, которое примет случайная величина, если в результате испытания произойдет именно этот исход.

Рассмотрим примеры.

Пример 5.1 В опыте с однократным бросанием игральной кости случайной величиной является число X выпавших очков. Множество возможных значений случайной величины X имеет вид

$$\{x_1 = 1; x_2 = 2; \dots; x_6 = 6\}.$$

Если вспомнить, как выглядит *пространство элементарных исходов* в этом опыте, то будет очевидно следующее соответствие между элементарными исходами ω и значениями случайной величины X :

$$\begin{array}{cccccc} \omega & = & \omega_1 & \omega_2 & \dots & \omega_6 \\ & & \downarrow & \downarrow & \dots & \downarrow \\ X & = & 1 & 2 & \dots & 6. \end{array}$$

Иными словами, каждому элементарному исходу ω_i , $i = \overline{1, 6}$, ставится в соответствие число i .

Пример 5.2 Монету подбрасывают до первого появления “герба”. В этом опыте можно ввести, например, такие случайные величины: X — число бросаний до первого появления “герба” с множеством возможных значений $\{1; 2; 3; \dots\}$ и Y — число “цифр”, выпавших до первого появления “герба”, с множеством возможных значений $\{0; 1; 2; \dots\}$ (ясно, что $X = Y + 1$). В данном опыте пространство элементарных исходов Ω можно отождествить с множеством

$$\{\Gamma; Ц\Gamma; ЦЦ\Gamma; \dots; Ц\dots Ц\Gamma; \dots\},$$

причем элементарному исходу $Ц\dots Ц\Gamma$ ставится в соответствие число $m + 1$ или m , где m — число повторений буквы “Ц”.

Пример 5.3 На плоский экран падает частица. Будем считать, что нам известна вероятность попадания частицы в любое (измеримое, т.е. имеющее площадь) множество на экране. Случайными величинами в данном случае будут, например, расстояние X от центра экрана до точки падения, квадрат этого расстояния $Y = X^2$, угол Z в полярной системе координат и т.д. #

Определение 5.1 Скалярную функцию $X(\omega)$, заданную на пространстве элементарных исходов, называют **случайной величиной**, если для любого $x \in \mathbb{R}$ множество исходов, для которых $X(\omega) < x$, т.е. $\{\omega : X(\omega) < x\}$ — является событием.

Функция распределения случайной величины

Для исследования вероятностных свойств *случайной величины* необходимо знать правило, позволяющее находить *вероятность* того, что случайная величина примет значение из подмножества ее значений. Любое такое правило называют **законом распределения вероятностей**, или **распределением (вероятностей)** случайной величины. Законом распределения, присущим всем случайным величинам, является функция распределения.

Определение 5.2 *Функцией распределения (вероятностей)* случайной величины X называют функцию $F(x)$, значение которой в точке x равно вероятности события $\{X < x\}$, т.е. события, состоящего из тех и только тех *элементарных исходов* ω , для которых $X(\omega) < x$:

$$F(x) = \mathbf{P}\{X < x\}.$$

Теорема 5.1 *Функция распределения удовлетворяет следующим свойствам.*

1. $0 \leq F(x) \leq 1$.
2. $F(x_1) \leq F(x_2)$ при $x_1 < x_2$, т.е. $F(x)$ — неубывающая функция.
3. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$; $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$.
4. $\mathbf{P}\{x_1 \leq X < x_2\} = F(x_2) - F(x_1)$.
5. $F(x) = F(x-0)$, где $F(x-0) = \lim_{y \rightarrow x-0} F(y)$, т.е. $F(x)$ — непрерывная слева функция.

ДОКАЗАТЕЛЬСТВО. При доказательстве будем использовать свойства вероятностей событий, доказанные в теореме 2.1.

Поскольку значение функции распределения в любой точке x является вероятностью, то из свойства 4 вероятности вытекает утверждение 1.

Если $x_1 < x_2$, то событие $\{X < x_1\}$ включено в событие $\{X < x_2\}$ и, согласно свойству 3,

$$\mathbf{P}\{X < x_1\} \leq \mathbf{P}\{X < x_2\},$$

т.е. в соответствии с определением 5.2 выполнено утверждение 2.

Пусть x_1, \dots, x_n, \dots — любая возрастающая последовательность чисел, стремящаяся к $+\infty$. Событие $\{X < +\infty\}$, с одной стороны, является *достоверным*, а с другой стороны, представляет собой объединение событий $\{X < x_n\}$. Отсюда в силу *аксиомы непрерывности* следует второе равенство в утверждении 3. Аналогично доказывают и первое равенство.

Событие $\{X < x_2\}$ при $x_1 < x_2$ представляет собой объединение двух *непересекающихся* событий: $\{X < x_1\}$ — случайная величина X приняла значение, меньшее x_1 , и $\{x_1 \leq X < x_2\}$ — случайная величина X приняла значение, лежащее в промежутке $[x_1, x_2)$. Поэтому в соответствии с *аксиомой сложения* получаем утверждение 4.

Наконец, пусть x_1, \dots, x_n, \dots — любая возрастающая последовательность чисел, стремящаяся к x . Событие $\{X < x\}$ является объединением событий $\{X < x_n\}$. Снова воспользовавшись аксиомой непрерывности, приходим к утверждению 5.

Замечание 5.1 Можно показать, что любая *неубывающая* непрерывная слева функция $F(x)$, удовлетворяющая условиям $F(-\infty) = 0$ и $F(+\infty) = 1$, является функцией распределения некоторой случайной величины X . #

Дискретные случайные величины

Определение 5.3 *Случайную величину X называют дискретной*, если множество ее возможных значений конечно или счетно.

Распределение дискретной случайной величины удобно описывать с помощью ряда распределения.

Определение 5.4 *Рядом распределения (вероятностей) дискретной случайной величины X называют таблицу (табл. 5.1), состоящую из двух строк: в верхней строке перечислены все возможные значения случайной величины, а в нижней — вероятности $p_i = \mathbf{P}\{X = x_i\}$ того, что случайная величина примет эти значения.*

Для проверки правильности составления табл. 5.1 рекомендуется просуммировать вероятности p_i . В силу *аксиомы нормированности* эта сумма должна быть равна единице: $\sum_{i=1}^n p_i = 1$.

X	x_1	x_2	\dots	x_i	\dots	x_n
\mathbf{P}	p_1	p_2	\dots	p_i	\dots	p_n

Таблица 5.1.

Покажем теперь, как по ряду распределения дискретной случайной величины построить ее *функцию распределения* $F(x)$. Пусть X — дискретная случайная величина, заданная своим рядом распределения, причем значения x_1, x_2, \dots, x_n расположены в порядке возрастания. Тогда для всех $x \leq x_1$ событие $\{X < x\}$ является *невозможным* и поэтому в соответствии с определением 5.2 $F(x) = 0$. (Если $x_1 < x \leq x_2$, то событие $\{X < x\}$ состоит из тех и только тех *элементарных исходов* ω , для которых $X(\omega) = x_1$, и, следовательно, $F(x) = p_1$.)

Аналогично при $x_2 < x \leq x_3$ событие $\{X < x\}$ состоит из элементарных исходов ω , для которых либо $X(\omega) = x_1$, либо $X(\omega) = x_2$, т.е.

$$\{X < x\} = \{X = x_1\} + \{X = x_2\},$$

а следовательно, $F(x) = p_1 + p_2$ и т.д. Наконец, при $x > x_n$ событие $\{X < x\}$ достоверно и $F(x) = 1$.

Таким образом, функция распределения дискретной случайной величины является кусочно-постоянной функцией, принимающей на промежутке $(-\infty, x_1]$ значение 0, на промежутках $(x_i, x_{i+1}]$, $1 \leq i < n$, — значение $p_1 + \dots + p_i$ и на промежутке $(x_n, +\infty)$ — значение 1.

Для задания закона распределения дискретной случайной величины, наряду с рядом распределения и функцией распределения используют другие способы. Так, его можно задать аналитически в виде некоторой формулы. Например, распределение игральной кости (см. пример 5.1) описывают формулой $P\{X = i\} = \frac{1}{6}, i = \overline{1, 6}$.

Непрерывные случайные величины

Определение 5.5 *Непрерывной* называют *случайную величину* X , *функцию распределения* которой $F(x)$ можно представить в виде

$$F(x) = \int_{-\infty}^x p(y) dy. \quad (5.1)$$

Функцию $p(x)$ называют **плотностью распределения (вероятностей)** случайной величины X .

Предполагают, что несобственный интеграл в представлении (5.1) сходится.

Все реально встречающиеся плотности распределения случайных величин являются непрерывными (за исключением, быть может, конечного числа точек) функциями. Следовательно, функция распределения для непрерывной случайной величины является непрерывной на всей числовой оси и в точках непрерывности плотности распределения $p(x)$ имеет место равенство

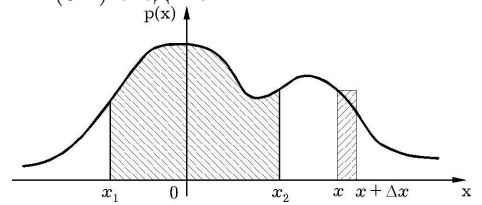


Рис. 5.1.

$$p(x) = F'(x), \quad (5.2)$$

что следует из свойств интеграла с переменным верхним пределом. Только такие случайные величины мы и будем рассматривать в дальнейшем. На рис. 5.1 изображен типичный вид плотности распределения.

Теорема 5.2 *Плотность распределения обладает следующими свойствами. 1) $p(x) \geq 0$.*

$$2) P\{x_1 \leq X < x_2\} = \int_{x_1}^{x_2} p(x) dx.$$

$$3) \int_{-\infty}^{+\infty} p(x) dx = 1.$$

$$4) P\{x \leq X < x + \Delta x\} \approx p(x)\Delta x \text{ в точках непрерывности плотности распределения.}$$

$$5) P\{X = x\} = 0.$$

Доказательство. Утверждение 1 следует из того, что плотность распределения является производной от функции распределения, в силу свойства 1 функции распределения она является *неубывающей функцией*, а производная неубывающей функции неотрицательна.

Согласно свойству 2 функции распределения, $P\{x_1 \leq X < x_2\} = F(x_2) - F(x_1)$. Отсюда в соответствии с определением непрерывной случайной величины и свойством аддитивности сходящегося несобственного интеграла имеем

$$F(x_2) - F(x_1) = \int_{-\infty}^{x_2} p(x) dx - \int_{-\infty}^{x_1} p(x) dx = \int_{x_1}^{x_2} p(x) dx,$$

что и доказывает утверждение 2.

В частности, если $x_1 = -\infty$, $x_2 = +\infty$, то событие $\{-\infty < X < \infty\}$ является *достоверным*, и поэтому справедливо утверждение 3.

Согласно свойству 4 (см. теорему 5.1),

$$P\{x \leq X < x + \Delta x\} = F(x + \Delta x) - F(x) = \Delta F(x).$$

Если Δx “мал” (см. рис. 5.1), то имеем

$$\Delta F(x) \approx dF(x) = F'(x)\Delta x = p(x)\Delta x,$$

что и доказывает утверждение 4.

Наконец, поскольку в силу определения 5.5 функция распределения случайной величины есть несобственный интеграл от плотности, то она является непрерывной, что приводит нас к утверждению 5.

Замечание 5.2 В силу свойства 2 плотности распределения вероятность попадания непрерывной случайной величины в промежуток $[x_1, x_2)$ численно равна площади криволинейной трапеции, заштрихованной на рис. 5.1.

Согласно свойству 3 площадь, заключенная под всей кривой, изображающей плотность распределения, равна единице.

В соответствии со свойством 4 вероятность попадания случайной величины X в некоторый “малый” промежуток $(x, x + \Delta x)$ практически пропорциональна Δx с коэффициентом пропорциональности, равным значению плотности распределения в точке x . Поэтому выражение $p(x)\Delta x$ или $p(x)dx$ называют иногда **элементом вероятности**. Можно также сказать, что непрерывная случайная величина реализует *геометрическую схему* с коэффициентом пропорциональности $p(x)$, но только в “малой” окрестности точки x .

Наконец, согласно свойству 5, вероятность попадания в любую (заданную до опыта) точку для непрерывной случайной величины равна нулю. #

Функции от случайной величины

Пусть на вероятностном пространстве $(\Omega, \mathfrak{B}, \mathbf{P})$ задана случайная величина $X = X(\omega)$. Рассмотрим действительную функцию $y = \varphi(x)$ действительного аргумента x (область определения которой включает в себя множество возможных значений случайной величины X).

Определение 5.6 Случайную величину Y , которая каждому элементарному исходу ω ставит в соответствие число $Y(\omega) = \varphi(X(\omega))$, называют **функцией $\varphi(X)$ (скалярной) от скалярной случайной величины X** .

Функция $Y = \varphi(X)$ от дискретной случайной величины также является дискретной случайной величиной, поскольку она не может принимать больше значений, чем случайная величина X . Очевидно, что если случайная величина X имеет ряд распределения, представленный в табл. 5.2, то ряд распределения случайной величины $Y = \varphi(X)$ определяется табл. 5.3.

X	x_1	x_2	\dots	x_n
\mathbf{P}	p_1	p_2	\dots	p_n

Таблица 5.2.

При этом, если в верхней строке табл. 5.3 появляются одинаковые значения $\varphi(x_i)$, соответствующие столбцы нужно объединить в один, приписав им суммарную вероятность. Функция $Y = \varphi(X)$ от непрерывной случайной величины X может быть как непрерывной, так и дискретной (если, например, множество значений функции $\varphi(X)$ конечное или счетное).

Y	$\varphi(x_1)$	$\varphi(x_2)$	\dots	$\varphi(x_n)$
\mathbf{P}	p_1	p_2	\dots	p_n

Таблица 5.3.

Теорема 5.3 Пусть случайная величина X имеет плотность $p_X(x)$. Пусть функция $y = \varphi(x)$ является монотонной, непрерывно дифференцируемой функцией. Обозначим $x = \psi(y)$ функцию, обратную к $y = \varphi(x)$. Тогда плотность случайной величины $Y = \varphi(X)$ есть

$$p_Y(y) = p_X(\psi(y))|\psi'(y)|. \quad (5.3)$$

ДОКАЗАТЕЛЬСТВО. Если функция $\varphi(x)$ является монотонной, то событие $\{\varphi(X(\omega)) < y\}$ эквивалентно событию $\{X(\omega) < \psi(y)\}$ (в случае возрастающей функции $\varphi(x)$) или событию $\{X(\omega) > \psi(y)\}$ (в случае убывающей $\varphi(x)$). Значит, для возрастающей функции $\varphi(x)$

$$\mathbf{P}\{\varphi(X) < y\} = \mathbf{P}\{X < \psi(y)\}, \quad (5.4)$$

для убывающей $\varphi(x)$

$$\mathbf{P}\{\varphi(X) < y\} = \mathbf{P}\{X > \psi(y)\}. \quad (5.5)$$

Поскольку

$$F_Y(y) = \mathbf{P}\{Y < y\},$$

а

$$\mathbf{P}\{X < \psi(y)\} = F_X(\psi(y)) \quad \text{и} \quad \mathbf{P}\{X > \psi(y)\} = 1 - F_X(\psi(y)),$$

то окончательно получаем:

для возрастающей функции $\varphi(x)$

$$F_Y(y) = F_X(\psi(y)); \quad (5.6)$$

для убывающей функции $\varphi(x)$

$$F_Y(y) = 1 - F_X(\psi(y)). \quad (5.7)$$

Далее, согласно правилу дифференцирования сложной функции, имеем: в случае возрастающей функции $Y(x)$

$$p_Y(y) = F'_Y(y) = \left(F_X(x) \right)' \Big|_{x=\psi(y)} \psi'(y) = p_X(\psi(y)) \psi'(y);$$

в случае убывающей функции $Y(x)$

$$p_Y(y) = F'_Y(y) = - \left(F_X(x) \right)' \Big|_{x=\psi(y)} \psi'(y) = -p_X(\psi(y)) \psi'(y).$$

Оба эти случая можно записать в виде (5.3).

Теорема 5.4 Пусть случайная величина X имеет плотность $p_X(x)$. Пусть функция $y = \varphi(x)$ является кусочно монотонной функцией. Обозначим $x_i = \psi_i(y)$, $i = \overline{1, k}$, прообразы точки y при отображении $y = \varphi(x)$. Если функции $\psi_i(y)$, $i = \overline{1, k}$, дифференцируемы, то плотность случайной величины $Y = \varphi(X)$ есть

$$p_Y(y) = \sum_{i=1}^k p_X(\psi_i(y)) |\psi'_i(y)|. \quad (5.8)$$

Пример 5.4 Пусть случайная величина X имеет непрерывную функцию распределения $F(x)$, которая является возрастающей функцией. Рассмотрим случайную величину $Y = F(X)$. Она принимает значение только на промежутке $[0, 1]$. Обратная функция для функции $y = F(x)$ есть

$$\psi(y) = F^{-1}(y).$$

Функция $\psi(y)$ при $y \in [0, 1]$, очевидно, удовлетворяет тождеству

$$F(\psi(y)) = y,$$

и, следовательно, в соответствии с формулой (5.6) имеем для $y \in [0, 1]$:

$$F_Y(y) = F(F^{-1}(y)) = y.$$

При $y < 0$ событие $\{Y < y\}$ является невозможным. Поэтому при $y < 0$

$$F_Y(y) = \mathbf{P}\{Y < y\} = 0.$$

При $y > 1$ событие $\{Y < y\}$ является достоверным и поэтому при $y > 1$

$$F_Y(y) = \mathbf{P}\{Y < y\} = 1.$$

Итак,

$$F_Y(y) = \begin{cases} 0, & y < 0; \\ y, & 0 \leq y \leq 1; \\ 1, & y > 1. \end{cases}$$

Таким образом, случайная величина Y имеет *равномерное распределение* на отрезке $[0, 1]$ (см. определение равномерного распределения на с. 32).

Переходя к обратной функции, видим, что случайная величина

$$X = F^{-1}(Y) \quad (5.9)$$

имеет функцию распределения $F(x)$, если случайная величина Y имеет равномерное распределение на отрезке $[0, 1]$.

Полученный результат широко применяют при моделировании случайных величин с заданной функцией распределения $F(x)$. Действительно, если нужно смоделировать такую случайную величину, то достаточно иметь датчик случайных чисел Y , распределенных равномерно на отрезке $[0, 1]$, и каждое такое число преобразовать по формуле (5.9).

Например, пусть нужно смоделировать реализацию случайной величины X с *экспоненциальной функцией* распределения (см. определение экспоненциального распределения на с. 32)

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0,$$

при известном параметре λ . Тогда, учитывая, что

$$F^{-1}(y) = -\frac{1}{\lambda} \ln(1-y),$$

реализацию X можно получить по формуле

$$X = -\frac{1}{\lambda} \ln(1-Y),$$

где Y — случайное число с равномерным в интервале $(0, 1)$ законом распределения. #

Пример 5.5 Пусть случайная величина X имеет *стандартное нормальное распределение* (см. определение нормального распределения на с. 33), т.е.

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty.$$

Найдем распределение случайной величины $Y = X^2$. Воспользуемся формулой (5.8). В данном случае $\varphi(x) = x^2$. Заметим, что функция $\varphi(x) = x^2$ принимает только неотрицательные значения и, следовательно, при $y < 0$ уравнение $\varphi(x) = y$ не имеет решений. Поэтому случайная величина $Y = Y(X) = X^2$ не может принимать отрицательные значения и, следовательно функция распределения $F_Y(y) = 0$ при $y < 0$. Отсюда дифференцированием получаем $p_Y(y) = F'_Y(y) = 0$, $y < 0$.

Далее, при $y > 0$ уравнение $y = x^2$ имеет два решения: $x = \psi_1(y) = -\sqrt{y}$ и $x = \psi_2(y) = \sqrt{y}$, причем первое решение принадлежит интервалу $(-\infty, 0)$ убывания функции $\varphi(x) = x^2$, а второе — интервалу $(0, +\infty)$ возрастания этой функции. Подставляя $\psi_1(y)$, $\psi_2(y)$ и *плотность стандартного нормального распределения* в формулу (5.8), получаем

$$p_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-\sqrt{y})^2} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad y > 0.$$

Лекция 6

Числовые характеристики случайных величин

Из результатов предыдущих лекций следует, что *вероятности* любых *событий*, связанных с каждой *случайной величиной* (в том числе *многомерной*), полностью определяются ее *законом распределения*, причем закон распределения *дискретной случайной величины* удобно задавать в виде *ряда распределения*, а непрерывной — в виде *плотности распределения*.

Однако при решении многих задач нет необходимости указывать закон распределения случайной величины, а достаточно характеризовать ее лишь некоторыми (неслучайными) числами. Такие числа (в теории вероятностей их называют числовыми характеристиками случайной величины) будут рассмотрены в настоящей лекции. Отметим, что основную роль на практике играют *математическое ожидание*, задающее “центральное” значение случайной величины, и *дисперсия*, характеризующая “разброс” значений случайной величины вокруг ее математического ожидания.

Отметим, что эти характеристики, как и все остальные, рассматриваемые в настоящей лекции, по сути дела, являются характеристиками *законов распределений* случайных величин.

Поэтому в дальнейшем вместо слов “характеристика случайной величины, имеющей некоторое *распределение* (закон распределения)” будем говорить “характеристика распределения”.

Математическое ожидание случайной величины

Определение 6.1 *Математическим ожиданием (средним значением) MX дискретной случайной величины X называют сумму произведений значений x_i случайной величины и вероятностей $p_i = P\{X = x_i\}$, с которыми случайная величина принимает эти значения:*

$$MX = \sum_i x_i p_i.$$

При этом, если *множество возможных значений* случайной величины X счетно, предполагается, что

$$\sum_{i=1}^{\infty} |x_i| p_i < +\infty,$$

т. е. ряд, определяющий математическое ожидание, сходится абсолютно; в противном случае говорят, что математическое ожидание случайной величины X не существует.

Математическое ожидание дискретной случайной величины имеет аналог в теоретической механике. Пусть на прямой расположена система материальных точек с массами p_i ($\sum_i p_i = 1$) и пусть x_i — координата i -й точки. Тогда центр масс системы будет иметь координату

$$\bar{X} = \frac{\sum_i x_i p_i}{\sum_i p_i} = \frac{\sum_i x_i p_i}{1} = \sum_i x_i p_i,$$

совпадающую с математическим ожиданием MX случайной величины X .

Пример 6.1 Пусть X — число выпавших очков при подбрасывании игральной кости. Так как $p_i = P\{X = i\} = \frac{1}{6}$, $i = \overline{1, 6}$, то

$$MX = \sum_{i=1}^6 i \frac{1}{6} = 7/2 = 3,5.$$

Определение 6.2 Математическим ожиданием (средним значением) $\mathbf{M}X$ непрерывной случайной величины называют интеграл

$$\mathbf{M}X = \int_{-\infty}^{+\infty} xp(x) dx.$$

При этом предполагается, что

$$\int_{-\infty}^{+\infty} |x|p(x) dx < +\infty,$$

т. е. несобственный интеграл, определяющий математическое ожидание, сходится абсолютно.

Так же как и в дискретном случае, математическое ожидание непрерывной случайной величины можно интерпретировать как центр масс стержня, плотность массы которого в точке x равна $p(x)$.

Пример 6.2 Случайная величина X имеет *распределение Коши*, т. е. распределение с плотностью $p(x) = \frac{1}{\pi(1+x^2)}$. Тогда $\int_{-\infty}^{+\infty} \frac{|x|dx}{\pi(1+x^2)} = +\infty$, поскольку подынтегральная функция эквивалентна $1/(\pi x)$ при $x \rightarrow +\infty$. Поэтому математическое ожидание случайной величины X не существует.

Математическое ожидание функции от случайной величины.

Найдем математическое ожидание функции случайной величины (*случайного вектора*). Пусть $Y = \varphi(X)$ является функцией от случайной величины X .

Рассмотрим сначала *дискретную случайную величину* X , принимающую значения x_1, \dots, x_n, \dots с вероятностями $p_n = \mathbf{P}\{X = x_n\}$, $n = 1, 2, \dots$. Тогда случайная величина $Y = \varphi(X)$ принимает значения $\varphi(x_n)$ с вероятностями $p_n = \mathbf{P}\{X = x_n\}$, $n = 1, 2, \dots$, и ее математическое ожидание определяется формулой

$$\mathbf{M}Y = \mathbf{M}\varphi(X) = \sum_{i=1}^{\infty} \varphi(x_i)p_i, \quad (6.1)$$

при этом требуется выполнение условия

$$\sum_{i=1}^{\infty} |\varphi(x_i)|p_i < +\infty. \quad (6.2)$$

Для *непрерывной случайной величины* X , имеющей *плотность распределения* $p(x)$, математическое ожидание случайной величины $Y = \varphi(X)$ можно найти, используя аналогичную (6.1) формулу

$$\mathbf{M}Y = \mathbf{M}\varphi(X) = \int_{-\infty}^{+\infty} \varphi(x)p(x) dx, \quad (6.3)$$

причем и здесь требуется выполнение условия $\int_{-\infty}^{+\infty} |\varphi(x)|p(x) dx < +\infty$.

Дисперсия. Моменты высших порядков

Две *случайные величины* могут иметь одинаковые *средние значения*, но их возможные значения будут по-разному рассеиваться вокруг этого среднего. Например, средний балл на экзамене в двух группах равен “4”, но в первой группе почти все студенты получили “4”, а во второй группе “четверочников” нет вообще, но есть как “пятерочники”, так и “троечники”.

Поэтому, наряду со средним значением, хотелось бы иметь и число, характеризующее “разброс” случайной величины относительно своего среднего значения. Такой характеристикой обычно служит *дисперсия*.

Определение 6.3 *Дисперсией* $\mathbf{D}X$ случайной величины X называют математическое ожидание квадрата отклонения случайной величины X от ее среднего значения, т. е. $\mathbf{D}X = \mathbf{M}(X - \mathbf{M}X)^2$.

Используя формулы (6.1)–(6.3), в которых положено $Y(x) = (x - \mathbf{M}X)^2$, легко написать расчетные формулы для дисперсий дискретной и непрерывной случайных величин соответственно:

$$\mathbf{D}X = \sum_i (x_i - \mathbf{M}X)^2 p_i \quad (6.4)$$

и

$$\mathbf{D}X = \int_{-\infty}^{+\infty} (x - \mathbf{M}X)^2 p(x) dx. \quad (6.5)$$

Пример 6.3 Пусть X — число выпавших очков при подбрасывании игральной кости. Так как $\mathbf{M}X = 3,5$, то

$$\mathbf{D}X = \sum_{i=1}^6 (i - 3,5)^2 \frac{1}{6} = \frac{35}{12}.$$

Замечание 6.1 Из определения непосредственно следует, что дисперсия любой случайной величины является неотрицательным числом.

Дисперсия $\mathbf{D}X$ представляет собой второй момент *центрированной* (имеющей нулевое математическое ожидание) *случайной величины* $\overset{\circ}{X} = X - \mathbf{M}X$. Поэтому иногда дисперсию называют **вторым центральным моментом** случайной величины.

Дисперсия имеет аналог в теоретической механике — центральный (относительно центра масс) момент инерции массы, распределенной на оси с линейной плотностью $p(x)$.

В некоторых теоретических исследованиях встречаются моменты высших порядков.

Определение 6.4 *Моментом k -го порядка m_k (k -м моментом)* случайной величины X называют математическое ожидание k -й степени случайной величины X :

$$m_k = \mathbf{M}X^k = \sum_i x_i^k p_i,$$

если X — дискретная случайная величина, и

$$m_k = \mathbf{M}X^k = \int_{-\infty}^{+\infty} x^k p(x) dx,$$

если X — непрерывная случайная величина.

Иногда k -й момент называют также **начальным моментом k -го порядка**.

Определение 6.5 *Центральным моментом k -го порядка $\overset{\circ}{m}_k$ (k -м центральным моментом)* случайной величины X называют математическое ожидание k -й степени случайной величины $\overset{\circ}{X} = X - \mathbf{M}X$:

$$\overset{\circ}{m}_k = \mathbf{M}(X - \mathbf{M}X)^k = \sum_i (x_i - \mathbf{M}X)^k p_i \quad \text{и} \quad \overset{\circ}{m}_k = \mathbf{M}(X - \mathbf{M}X)^k = \int_{-\infty}^{+\infty} (x - \mathbf{M}X)^k p(x) dx$$

для дискретной и непрерывной случайных величин X соответственно.

Момент первого порядка совпадает с математическим ожиданием, центральный момент первого порядка равен нулю, центральный момент второго порядка является дисперсией.

Квантиль

Определение 6.6 *Квантилью уровня α , или α -квантилью, ($0 < \alpha < 1$)* случайной величины X (распределения случайной величины X) называют число Q_α , удовлетворяющее неравенствам

$$\mathbf{P}\{X < Q_\alpha\} \leq \alpha \quad \text{и} \quad \mathbf{P}\{X > Q_\alpha\} \leq 1 - \alpha.$$

$1/2$ -квантиль называют также **медианой** M случайной величины X .

Для непрерывной случайной величины X α -квантиль Q_α является решением уравнения

$$F(Q_\alpha) = \alpha,$$

где $F(x)$ — функция распределения случайной величины X . Таким образом, для непрерывной случайной величины X квантиль Q_α — это такое число, меньше которого X принимает значение с вероятностью α .

Если известна плотность распределения $p(x)$ случайной величины X , то, учитывая связь между функцией распределения и плотностью распределения, уравнение для определения α -квантили можно записать в виде

$$\int_{-\infty}^{Q_\alpha} p(x) dx = \alpha.$$

Пример 6.4 Найдем α -квантиль и медиану экспоненциального распределения. В этом случае Q_α представляет собой решение уравнения

$$1 - e^{-\lambda Q_\alpha} = \alpha.$$

Поэтому

$$Q_\alpha = -\frac{\ln(1 - \alpha)}{\lambda}.$$

Медиана экспоненциального распределения

$$M = \frac{\ln 2}{\lambda}.$$

Пример 6.5 Пусть случайная величина X представляет собой число успехов в одном испытании по схеме Бернулли с вероятностью успеха p . Тогда, как видно на рис. 6.1,

$$\begin{aligned} \text{при } 0 < \alpha < q & \quad Q_\alpha = 0, \\ \text{при } q < \alpha < 1 & \quad Q_\alpha = 1, \end{aligned}$$

а q -квантилью является любое число от 0 до 1. Этот пример показывает, что, во-первых, квантили могут совпадать для разных α , а, во-вторых, для некоторых α квантили могут определяться неоднозначно.

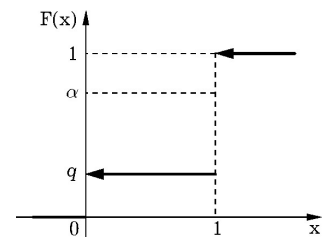


Рис. 6.1.

Лекция 7

Основные законы распределения случайных величин

Биномиальное распределение

Дискретная случайная величина X распределена по **биномиальному закону**, если она принимает значения $0, 1, 2, \dots, n$ в соответствии с распределением, заданным формулой

X	0	1	...	i	...	n
P	q^n	$C_n^1 p q^{n-1}$...	$C_n^i p^i q^{n-i}$...	p^n

Таблица 7.1.

$$P\{X = i\} = P_n(i) = C_n^i p^i q^{n-i}, i = \overline{0, n},$$

или, что тоже самое, рядом распределения, представленным в таблице 7.1, где $0 < p, q < 1$ и $p + q = 1$.

Проверим корректность определения биномиального распределения. Действительно, $P_n(i) > 0$ и

$$\sum_{i=0}^n P_n(i) = \sum_{i=0}^n C_n^i p^i q^{n-i} = (p + q)^n = 1.$$

Пример 7.1 Найдем математическое ожидание случайной величины X , распределенной по **биномиальному закону** (число успехов в n испытаниях по *схеме Бернулли* с вероятностью успеха p):

$$\begin{aligned} MX &= \sum_{i=0}^n i P_n(i) = \sum_{i=0}^n i C_n^i p^i q^{n-i} = \sum_{i=0}^n i \frac{n!}{i!(n-i)!} p^i q^{n-i} = \\ &= \sum_{i=1}^n np \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} q^{n-i} = np \sum_{j=0}^{n-1} C_{n-1}^j p^j q^{n-1-j} = np \sum_{j=0}^{n-1} P_{n-1}(j) = np. \end{aligned}$$

Биномиальное распределение является не чем иным, как распределением числа успехов X в n испытаниях по *схеме Бернулли* с вероятностью успеха p и неудачи $q = 1 - p$ (см. 4).

Пример 7.2 Пусть X — число успехов в n испытаниях по *схеме Бернулли*. Дисперсию X можно подсчитать так же, как в примере 7.1 было подсчитано математическое ожидание, а именно непосредственно воспользоваться определением 6.3 дисперсии. Однако мы поступим другим образом. Для этого представим X в виде суммы $X = X_1 + \dots + X_n$, где X_i — число успехов в i -ом испытании. Дисперсия каждого слагаемого равна:

$$DX_i = (0 - MX_i)^2 q + (1 - MX_i)^2 p = (-p)^2 q + (1 - p)^2 p = p^2 q + q^2 p = pq(p + q) = pq.$$

Позднее (см. теор. 9.2) будет доказано, что дисперсия суммы независимых случайных величин равна сумме дисперсий каждого слагаемого. Учитывая, что случайные величины X_i являются независимыми, получаем

$$DX = DX_1 + \dots + DX_n = npq.$$

Распределение Пуассона

Дискретная случайная величина X распределена по **закону Пуассона**, если она принимает целые неотрицательные значения с вероятностями

$$P\{X = i\} = P(i; \lambda) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i = 0, 1, \dots,$$

или, по-другому, с вероятностями, представленными рядом распределения в таблице 7.2, где $\lambda > 0$ — параметр распределения Пуассона.

Проверка корректности определения распределения Пуассона дает:

X	0	1	2	...	n	...
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$...	$\frac{\lambda^n}{n!} e^{-\lambda}$...

Таблица 7.2.

$$\sum_{i=0}^{\infty} P(i; \lambda) = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1.$$

Распределение Пуассона также называют **законом редких событий**, поскольку оно всегда проявляется там, где производится большое число испытаний, в каждом из которых с малой вероятностью происходит “редкое” событие. В соответствии с законом Пуассона распределены, например, число вызовов, поступивших в течение суток на телефонную станцию; число метеоритов, упавших в определенном районе; число распавшихся частиц при радиоактивном распаде вещества.

Пример 7.3 Пусть случайная величина X имеет *распределение Пуассона*. Тогда

$$MX = \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda e^{\lambda} e^{-\lambda} = \lambda.$$

Пример 7.4 Найдём дисперсию случайной величины X , распределенной по *закону Пуассона*. Для этого воспользуемся свойством 3 дисперсии. Математическое ожидание $MX = \lambda$ было найдено в примере 7.3. Определим второй момент:

$$\begin{aligned} MX^2 &= \sum_{i=0}^{\infty} i^2 \frac{\lambda^i}{i!} e^{-\lambda} = \lambda \sum_{i=1}^{\infty} i \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \\ &= \lambda \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} e^{-\lambda} = \lambda \left(\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} + \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} \right) = \lambda(MX + 1) = \lambda^2 + \lambda. \end{aligned}$$

Таким образом (см. далее теор. 9.2),

$$DX = \lambda^2 + \lambda - \lambda^2 = \lambda,$$

и, значит, дисперсия X , так же как и математическое ожидание, совпадает с параметром λ .

Геометрическое распределение

Снова рассмотрим схему Бернулли. Пусть X — число испытаний, которое необходимо провести, прежде чем появится первый успех. Тогда X — дискретная случайная величина, принимающая значения $0, 1, 2, \dots, n, \dots$. Определим вероятность события $\{X = n\}$. Очевидно, что $X = 0$, если в первом же испытании произойдет успех. Поэтому $P\{X = 0\} = p$. Далее, $X = 1$ в том случае, когда в первом испытании произошла неудача, а во втором — успех. Но вероятность такого события (см. теорему 4.3), равна qp , т.е. $P\{X = 1\} = qp$. Аналогично $X = 2$, если в первых двух испытаниях произошли неудачи, а в третьем — успех, и, значит, $P\{X = 2\} = qqp$. Продолжая эту процедуру, получаем $P\{X = i\} = pq^i$, $i = 0, 1, \dots$. Таким образом, случайная величина X имеет ряд распределения, представленный в таблице.

X	0	1	2	...	n	...
P	p	qp	q^2p	...	$q^n p$...

Таблица 7.3.

Правильность составления таблицы вытекает из равенства

$$\sum_{i=0}^{\infty} P\{X = i\} = \sum_{i=0}^{\infty} pq^i = p \sum_{i=0}^{\infty} q^i = \frac{p}{1-q} = 1.$$

Пример 7.5 Найдём математическое ожидание случайной величины X , имеющей *геометрическое распределение*:

$$MX = \sum_{i=0}^{\infty} i pq^i = pq \sum_{i=0}^{\infty} i q^{i-1} = pq \left(\sum_{i=0}^{\infty} q^i \right)'_q = pq \left(\frac{1}{1-q} \right)' = \frac{pq}{(1-q)^2} = \frac{pq}{p^2} = \frac{q}{p}.$$

Можно показать, что

$$DX = \frac{q}{p^2}.$$

Равномерное распределение

Случайная величина имеет **равномерное распределение** на отрезке $[a, b]$, если ее *плотность распределения* $p(x)$ и *функция распределения* $F(x)$ равны

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b; \\ 0, & x < a \text{ или } x > b. \end{cases}, \quad F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x \leq b; \\ 1, & x > b. \end{cases}$$

Графики плотности распределения $p(x)$ и функции распределения $F(x)$ приведены на рис. 7.1. Вероятность попадания равномерно распределенной случайной величины в интервал (x_1, x_2) , лежащий внутри отрезка $[a, b]$, равна $F(x_2) - F(x_1) = (x_2 - x_1)/(b - a)$, т.е. пропорциональна длине этого интервала. Таким образом, равномерное распределение реализует *схему геометрической вероятности* при бросании точки на отрезок $[a, b]$.

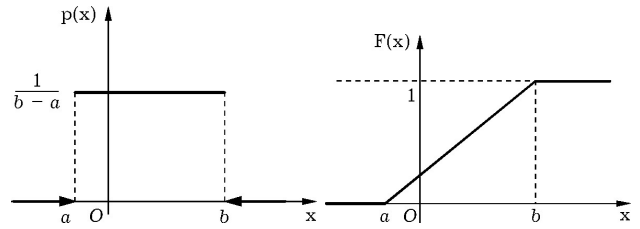


Рис. 7.1.

Пример 7.6 Найдем математическое ожидание *равномерно распределенной* на отрезке $[a, b]$ случайной величины X . Поскольку в этом случае $p(x) = 0$ при $x < a$ и $x > b$, то

$$MX = \int_{-\infty}^{+\infty} x p(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_{x=a}^{x=b} = \frac{b+a}{2}.$$

Как и следовало ожидать, MX совпадает с серединой отрезка $[a, b]$.

Пример 7.7 Дисперсия *равномерно распределенной* на отрезке $[a, b]$ случайной величины X определяется формулой

$$DX = \int_a^b \left(x - \frac{b+a}{2}\right)^2 \frac{1}{b-a} dx = \left(x - \frac{b+a}{2}\right)^3 \frac{1}{3(b-a)} \Big|_{x=a}^{x=b} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}.$$

Экспоненциальное распределение

Случайная величина распределена по **экспоненциальному (показательному) закону**, если она имеет плотность распределения $p(x)$ и функцию распределения $F(x)$

$$p(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \geq 0, \end{cases}, \quad F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases},$$

где $\lambda > 0$ — параметр экспоненциального распределения. Графики плотности распределения и функции распределения экспоненциально распределенной случайной величины приведены на и 7.2.

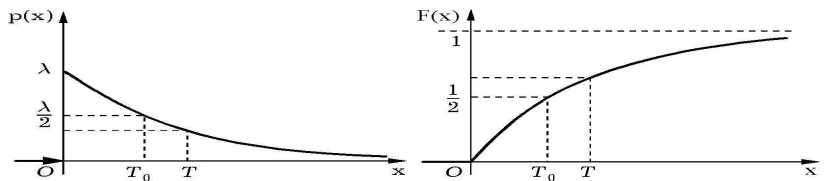


Рис. 7.2.

Пример 7.8 Найдем математическое ожидание и дисперсию *экспоненциально распределенной* случайной величины X . Применяя формулу интегрирования по частям, получим

$$MX = \int_{-\infty}^{+\infty} x p(x) dx = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda},$$

$$DX = \int_{-\infty}^{+\infty} (x - MX)^2 p(x) dx = \int_0^{+\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}.$$

Нормальное распределение

Случайная величина **распределена по нормальному** (или **гауссову**) **закону**, или имеет **нормальное** (**гауссово**) **распределение**, если ее плотность

$$\varphi_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (-\infty < m < +\infty, \sigma > 0).$$

Нормальное распределение зависит от двух параметров: m и σ . Всюду в далее запись $X \sim \mathcal{N}(\mu, \sigma^2)$ будет означать, что X — нормальная случайная величина с параметрами μ и σ .

Графики плотности $\varphi_{m,\sigma}(x)$ и функции

$$\Phi_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

нормального распределения для различных значений m и σ приведены на рис. 7.3.

Как следует из этих рисунков, параметр m определяет положение “центра симметрии” плотности нормального распределения, т.е. график плотности нормального распределения симметричен относительно прямой $x = m$, а σ — разброс значений случайной величины относительно центра симметрии. Если $m = 0$ и $\sigma = 1$, то такой **нормальный закон** называют **стандартным** и его функцию распределения обозначают $\Phi(x)$, а плотность распределения — $\varphi(x)$.

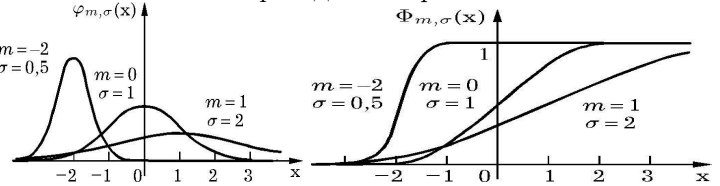


Рис. 7.3.

Как известно из курса математического анализа, интеграл $\int e^{-x^2/2} dx$ не может быть выражен через элементарные функции. Поэтому во всех справочниках и в большинстве учебников по теории вероятностей приведены таблицы значений функции стандартного нормального распределения. Покажем, как, используя эту таблицу, найти вероятность попадания случайной величины, распределенной по нормальному закону с произвольными параметрами m и σ , в интервал (a, b) .

В соответствии со свойством 2 плотности распределения (см. теорему 5.2) вероятность попадания случайной величины X , распределенной по нормальному закону с параметрами m и σ , в интервал (a, b) задается формулой

$$\mathbf{P}\{a < X < b\} = \int_a^b \varphi_{m,\sigma}(y) dy = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-m)^2/(2\sigma^2)} dy.$$

Проводя замену $x = (y - m)/\sigma$, этот интеграл можно записать в виде

$$\mathbf{P}\{a < X < b\} = \int_{(a-m)/\sigma}^{(b-m)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{(a-m)/\sigma}^{(b-m)/\sigma} \varphi(x) dx.$$

Таким образом, окончательно получаем

$$\mathbf{P}\{a < X < b\} = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right). \quad (7.1)$$

Пример 7.9 Найдем математическое ожидание случайной величины X , распределенной по нормальному закону с параметрами m и σ :

$$\mathbf{M}X = \int_{-\infty}^{+\infty} x \varphi_{m,\sigma}(x) dx = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/(2\sigma^2)} dx.$$

Делая замену $y = (x - m)/\sigma$, получаем

$$\mathbf{M}X = \int_{-\infty}^{+\infty} \frac{\sigma y + m}{\sqrt{2\pi}} e^{-y^2/2} dy = \int_{-\infty}^{+\infty} \frac{\sigma y}{\sqrt{2\pi}} e^{-y^2/2} dy + m \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y e^{-y^2/2} dy + m \int_{-\infty}^{+\infty} \varphi(y) dy.$$

Первый интеграл равен нулю в силу нечетности подынтегральной функции, а второй равен единице как интеграл от стандартной нормальной плотности. Таким образом, $\mathbf{M}X = m$, т. е. параметр m имеет смысл математического ожидания случайной величины X .

Пример 7.10 Дисперсия случайной величины X , распределенной по *нормальному закону* с параметрами m и σ , имеет вид

$$\mathbf{D}X = \int_{-\infty}^{+\infty} (x - m)^2 \varphi_{m,\sigma}(x) dx = \int_{-\infty}^{+\infty} \frac{(x - m)^2}{\sigma\sqrt{2\pi}} e^{-\frac{(x - m)^2}{2\sigma^2}} dx.$$

Делая замену $y = (x - m)/\sigma$, получаем

$$\mathbf{D}X = \sigma^2 \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Полагая $u = y/\sqrt{2\pi}$, $dv = ye^{-y^2/2} dy$ и интегрируя по частям, находим

$$\mathbf{D}X = \sigma^2 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \sigma^2 \int_{-\infty}^{+\infty} \varphi(y) dy = \sigma^2.$$

Таким образом, дисперсия нормально распределенной случайной величины совпадает с квадратом второго параметра.

Лекция 8

Случайные векторы

Функция распределения случайного вектора

Определение 8.1 Совокупность *случайных величин* $X_1 = X_1(\omega), \dots, X_n = X_n(\omega)$, заданных на одном и том же *вероятностном пространстве* $(\Omega, \mathfrak{B}, \mathbf{P})$, называют **многомерной** (***n*-мерной**) **случайной величиной**, или ***n*-мерным случайным вектором**. При этом сами случайные величины X_1, X_2, \dots, X_n называют **координатами случайного вектора**.

Пример 8.1 Отклонение точки разрыва снаряда от точки прицеливания при стрельбе по плоской цели можно задать двумерной случайной величиной $(X; Y)$, где X — отклонение по дальности, а Y — отклонение в боковом направлении.

При стрельбе по воздушной цели необходимо рассматривать трехмерный случайный вектор $(X; Y; Z)$, где X, Y, Z — координаты отклонения точки разрыва зенитного снаряда от точки прицеливания в некоторой пространственной системе координат.

Пример 8.2 При испытании прибора на надежность совокупность внешних воздействий в некоторый момент времени можно описать случайным вектором $(X; Y; Z; \dots)$. Здесь, например, X — температура окружающей среды, Y — атмосферное давление, Z — амплитуда вибрации платформы, на которой установлен прибор и т.д. Размерность этого вектора зависит от количества учитываемых факторов и может быть достаточно большой. #

Свойства многомерных случайных векторов мы будем в основном изучать на примере двумерного случайного вектора, делая, если это потребует, пояснения для случайного вектора произвольной размерности.

Напомним, что рассмотрение одномерной случайной величины начиналось с обсуждения способа задания ее *закона распределения*. В частности, закон распределения одномерной случайной величины можно задать с помощью *функции распределения*. То же можно сказать и по отношению к *n*-мерному случайному вектору. Отметим, что в дальнейшем для *пересечения событий* $\{X_1 < x_1\}, \dots, \{X_n < x_n\}$ вместо записи

$$\{X_1 < x_1\} \cap \dots \cap \{X_n < x_n\}$$

будем использовать запись

$$\{X_1 < x_1, \dots, X_n < x_n\}.$$

Определение 8.2 Функцией распределения (вероятностей)

$$F(x_1, \dots, x_n) = F_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

(*n*-мерного) **случайного вектора** $(X_1; \dots; X_n)$ называют функцию, значение которой в точке $(x_1; \dots; x_n) \in \mathbb{R}^n$ равно вероятности совместного осуществления (пересечения) событий $\{X_1 < x_1\}, \dots, \{X_n < x_n\}$, т.е.

$$F(x_1, \dots, x_n) = F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbf{P}\{X_1 < x_1, \dots, X_n < x_n\}.$$

Функцию распределения $F(x_1, \dots, x_n)$ называют также **совместной *n*-мерной функцией распределения** случайных величин X_1, \dots, X_n . Значение двумерной функции распределения в точке $(a_1; a_2)$, согласно определению 8.2, представляет собой не что иное, как вероятность попадания точки с координатами $(X_1; X_2)$ в квадрант с вершиной в точке $(a_1; a_2)$.

Теорема 8.1 Двумерная функция распределения удовлетворяет следующим свойствам.

1. $0 \leq F(x_1, x_2) \leq 1$.
2. $F(x_1, x_2)$ — неубывающая функция по каждому из аргументов x_1 и x_2 .
3. $F(-\infty, x_2) = F(x_1, -\infty) = 0$.
4. $F(+\infty, +\infty) = 1$.
5. $\mathbf{P}\{a_1 \leq X_1 < b_1, a_2 \leq X_2 < b_2\} = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)$.
6. $F(x_1, x_2)$ — непрерывная слева в любой точке $(x_1, x_2) \in \mathbb{R}^2$ по каждому из аргументов x_1 и x_2 функция.
7. $F_{X_1, X_2}(x, +\infty) = F_{X_1}(x)$, $F_{X_1, X_2}(+\infty, x) = F_{X_2}(x)$.

ДОКАЗАТЕЛЬСТВО. Утверждения 1 и 2 доказываются точно так же, как и в одномерном случае (см. теорему 5.1).

События $\{X_1 < -\infty\}$ и $\{X_2 < -\infty\}$ являются невозможными, а пересечение невозможного события с любым событием, как известно, также невозможное событие, вероятность которого равна нулю. Отсюда с учетом определения 8.2 вытекает утверждение 3.

Аналогично из того, что события $\{X_1 < +\infty\}$ и $\{X_2 < +\infty\}$ так же, как и их пересечение, являются достоверными, вероятность которых равна единице, вытекает утверждение 4.

Чтобы найти вероятность попадания двумерной случайной величины (X_1, X_2) в прямоугольник $\{a_1 \leq x_1 < b_1, a_2 \leq x_2 < b_2\}$ (на рис. 8.1 заштрихован), сначала определим вероятность попадания в полуолосу $\{x_1 < a_1, a_2 \leq x_2 < b_2\}$ (отмечена двойной штриховкой). Но эта вероятность представляет собой вероятность попадания в квадрант $\{x_1 < a_1, x_2 < b_2\}$ за вычетом вероятности попадания в квадрант $\{x_1 < a_1, x_2 < a_2\}$, т.е. $\mathbf{P}\{X_1 < a_1, a_2 \leq X_2 < b_2\} = F(a_1, b_2) - F(a_1, a_2)$. Аналогично, $\mathbf{P}\{X_1 < b_1, a_2 \leq X_2 < b_2\} = F(b_1, b_2) - F(b_1, a_2)$. Теперь осталось заметить, что вероятность попадания в прямоугольник $\{a_1 \leq x_1 < b_1, a_2 \leq x_2 < b_2\}$ совпадает с вероятностью попадания в полуолосу $\{x_1 < b_1, a_2 \leq x_2 < b_2\}$ из которой вычитается вероятность попадания в полуолосу $\{x_1 < a_1, a_2 \leq x_2 < b_2\}$, откуда следует утверждение 5.

Подобно одномерному случаю доказывается и утверждение 6.

Наконец, событие $\{X_2 < +\infty\}$ является достоверным, поэтому $\{X_1 < x_1\} \cap \{X_2 < +\infty\} = \{X_1 < x_1\}$. Аналогично $\{X_1 < +\infty\} \cap \{X_2 < x_2\} = \{X_2 < x_2\}$. Отсюда приходим к утверждению 7, которое устанавливает естественную связь между двумерной функцией распределения F_{X_1, X_2} случайного вектора (X_1, X_2) и функциями F_{X_1} и F_{X_2} , которые называют **одномерными** (говорят также **частными**, или **маргинальными**) **функциями распределения** случайных величин X_1 и X_2 .

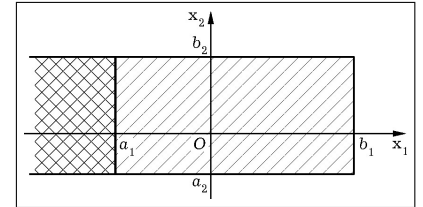


Рис. 8.1.

Дискретные двумерные случайные векторы

Определение 8.3 Двумерную случайную величину $(X; Y)$ называют **дискретной**, если каждая из случайных величин X и Y является **дискретной**.

Так же как и в одномерном случае, распределение двумерной дискретной случайной величины естественно описать с помощью перечисления всевозможных пар $(x_i; y_j)$ значений координат *случайного вектора* $(X; Y)$ и соответствующих *вероятностей*, с которыми эти пары значений принимают случайные величины X и Y (для простоты ограничимся конечным *множеством возможных значений*, когда случайная величина X может принимать только значения x_1, \dots, x_n , Y — значения y_1, \dots, y_m , а координаты двумерного случайного вектора $(X; Y)$ — пары значений (x_i, y_j) , $i = \overline{1, n}$, $j = \overline{1, m}$. Такое перечисление удобно представить в виде таблицы (табл. 8.1). В этой таблице в верхней строке перечислены все возможные значения $y_1, \dots, y_j, \dots, y_m$ случайной величины Y , а в левом столбце — значения $x_1, \dots, x_i, \dots, x_n$ случайной величины X . На пересечении столбца “ y_j ” со строкой “ x_i ” находится вероятность $p_{ij} = \mathbf{P}\{X = x_i, Y = y_j\}$ совместного осуществления событий $\{X = x_i\}$ и $\{Y = y_j\}$.

В этой таблице обычно добавляют еще одну строку “ \mathbf{P}_Y ” и столбец “ \mathbf{P}_X ”.

На пересечении столбца “ \mathbf{P}_X ” со строкой “ x_i ” записывают число $p_{Xi} = p_{i1} + \dots + p_{im}$. Но p_{Xi} представляет собой не что иное, как вероятность того, что случайная величина X примет значение x_i , т.е. $p_{Xi} = \mathbf{P}\{X = x_i\}$. Таким образом, первый и последний столбцы таблицы дают нам *ряд распределения* случайной величины X .

Аналогично, в последней строке “ \mathbf{P}_Y ” помещены значения $p_{Yj} = p_{1j} + \dots + p_{nj}$, а первая и последняя строки дают ряд распределения случайной величины Y . Используя таблицу 8.1, нетрудно

X	Y				
	y_1	y_2	\dots	y_m	\mathbf{P}_X
x_1	p_{11}	p_{12}	\dots	p_{1m}	p_{X1}
x_2	p_{21}	p_{22}	\dots	p_{2m}	p_{X2}
\dots	\dots	\dots	\dots	\dots	\dots
x_n	p_{n1}	p_{n2}	\dots	p_{nm}	p_{Xn}
\mathbf{P}_Y	p_{Y1}	p_{Y2}	\dots	p_{Ym}	

Таблица 8.1.

определить *совместную функцию распределения* $F_{X,Y}(x,y)$. Ясно, что для этого необходимо просуммировать p_{ij} по всем тем значениям i и j , для которых $x_i < x$, $y_j < y$, т.е.

$$F(x,y) = \sum_{\substack{i: x_i < x \\ j: y_j < y}} p_{ij}.$$

Пример 8.3 В соответствии со *схемой Бернулли* (см. определение 4.2) с вероятностью успеха p и вероятностью неудачи $q = 1 - p$ проводятся два испытания. Выпишем распределение двумерного случайного вектора $(X_1; X_2)$, где X_i , $i = 1, 2$, — число успехов в i -м испытании. Каждая из случайных величин X_1 и X_2 может принимать два значения: 0 или 1. Числа успехов в обоих испытаниях равны нулю тогда, когда произойдут две неудачи, а это в силу независимости испытаний происходит с вероятностью q^2 . Поэтому

$$\mathbf{P}\{X_1 = 0, X_2 = 0\} = q^2,$$

и на пересечении столбца “0” со строкой “0” нужно записать q^2 (табл. 8.2). Далее, $X_1 = 1$ и $X_2 = 0$, если в первом испытании произошел успех, а во втором — неудача, и, значит,

$$\mathbf{P}\{X_1 = 1, X_2 = 0\} = pq.$$

Аналогично заполняем второй столбец:

$$\mathbf{P}\{X_1 = 0, X_2 = 1\} = qp, \quad \mathbf{P}\{X_1 = 1, X_2 = 1\} = p^2.$$

Наконец, на пересечении столбца “ \mathbf{P}_{X_1} ” и строки “0” должно стоять

$$\mathbf{P}\{X_1 = 0\} = q^2 + pq = q(q + p) = q,$$

а на пересечении столбца “ \mathbf{P}_{X_1} ” и строки “1” —

$$\mathbf{P}\{X_1 = 1\} = pq + p^2 = p(p + q) = p.$$

Построим теперь совместную функцию распределения случайных величин X_1 и X_2 :

$$F(x_1, x_2) = \mathbf{P}\{X_1 < x_1, X_2 < x_2\}.$$

Поскольку при $x_1 \leq 0$ или $x_2 \leq 0$ нет ни одного элементарного исхода ω , для которого или $X_2(\omega) < x_2$, то для таких x_1 и x_2 событие $\{X_1 < x_1, X_2 < x_2\}$ невозможное, и, значит $F(x_1, x_2) = 0$ при $x_1 \leq 0$ или $x_2 \leq 0$.

Далее, если $0 < x_1 \leq 1$ и $0 < x_2 \leq 1$, то событие $\{X_1 < x_1, X_2 < x_2\}$ эквивалентно событию $\{X_1 = 0, X_2 = 0\}$, которое, как видно из табл. 8.2, происходит с вероятностью q^2 , и $F(x_1, x_2) = q^2$.

Если же $0 < x_1 \leq 1$, а $x_2 > 1$, то событие $\{X_1 < x_1, X_2 < x_2\}$ совпадает с объединением *непересекающихся событий* $\{X_1 = 0, X_2 = 0\}$ и $\{X_1 = 0, X_2 = 1\}$. Тогда $F(x_1, x_2) = q^2 + qp = q$. Аналогично $F(x_1, x_2) = q^2 + qp = q$ при $x_1 > 1$ и $0 < x_2 \leq 1$.

Наконец, если $x_1 > 1$ и $x_2 > 1$, то событие $\{X_1 < x_1, X_2 < x_2\}$ достоверно, и, следовательно, $F(x_1, x_2) = 1$.

Непрерывные случайные векторы

Определение 8.4 *Непрерывной двумерной случайной величиной* $(X; Y)$ называют такую двумерную случайную величину $(X; Y)$, *совместную функцию распределения* которой $F(x_1, x_2) = \mathbf{P}\{X < x_1, Y < x_2\}$ можно представить в виде сходящегося несобственного интеграла:

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p(y_1, y_2) dy_1 dy_2 = \int_{-\infty}^{x_1} dy_1 \int_{-\infty}^{x_2} p(y_1, y_2) dy_2 = \int_{-\infty}^{x_2} dy_2 \int_{-\infty}^{x_1} p(y_1, y_2) dy_1.$$

Функцию $p(x_1, x_2) = p_{X,Y}(x_1, x_2)$ называют *совместной (двумерной) плотностью распределения* случайных величин X и Y , или плотностью распределения случайного вектора $(X; Y)$. Так же как и в одномерном случае, будем предполагать, что $p(x_1, x_2)$ непрерывная (или непрерывная за исключением отдельных точек или линий) функция по обоим аргументам. Тогда в соответствии с определением непрерывной случайной величины и теоремой о дифференцировании

X_1	X_2		
	0	1	\mathbf{P}_{X_1}
0	q^2	qp	q
1	pq	p^2	p
\mathbf{P}_{X_2}	q	p	

Таблица 8.2.

интеграла с переменным верхним пределом совместная плотность распределения представляет собой (в точках ее непрерывности) вторую смешанную производную совместной функции распределения:

$$p(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 F(x_1, x_2)}{\partial x_2 \partial x_1}. \quad (8.1)$$

Заметим, что аналогичный смысл имеет **совместная (n -мерная) плотность распределения** случайных величин X_1, \dots, X_n , или **плотность распределения случайного вектора** $(X_1; \dots; X_n)$:

$$p(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}.$$

Теорема 8.2 Двумерная плотность распределения обладает следующими свойствами.

1. $p(x_1, x_2) \geq 0$.

$$2. \mathbf{P}\{a_1 < X < b_1, a_2 < Y < b_2\} = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} p(x_1, x_2) dx_2.$$

$$3. \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 dx_2 = 1.$$

$$4. \mathbf{P}\{x_1 < X < x_1 + \Delta x_1, x_2 < Y < x_2 + \Delta x_2\} \approx p(x_1, x_2) \Delta x_1 \Delta x_2.$$

$$5. \mathbf{P}\{X = x_1, Y = x_2\} = 0.$$

$$6. \mathbf{P}\{(X; Y) \in D\} = \iint_D p(x_1, x_2) dx_1 dx_2.$$

$$7. p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x, y) dy.$$

$$8. p_Y(y) = \int_{-\infty}^{+\infty} p_{X,Y}(x, y) dx.$$

ДОКАЗАТЕЛЬСТВО. Свойства 1 — 5 аналогичны свойствам *одномерной плотности распределения*. Свойство 6 является обобщением свойства 2.

Докажем утверждения 7 и 8.

Из свойства 7 двумерной функции распределения и определения 8.4 двумерной плотности распределения вытекает:

$$\begin{aligned} F_X(x) = F_{X,Y}(x, +\infty) &= \int_{-\infty}^x \int_{-\infty}^{+\infty} p_{X,Y}(y_1, y_2) dy_1 dy_2, \\ F_Y(y) = F_{X,Y}(+\infty, y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^y p_{X,Y}(y_1, y_2) dy_1 dy_2, \end{aligned}$$

откуда, дифференцируя интегралы по переменному верхнему пределу и учитывая формулу (5.2), получаем утверждения 7 и 8 для *одномерных плотностей распределения* $p_X(x)$ и $p_Y(y)$ случайных величин X и Y .

Пример 8.4 Рассмотрим двумерную случайную величину, плотность распределения которой имеет вид

$$p(x_1, x_2) = \begin{cases} A, & x_1^2 + x_2^2 \leq R^2; \\ 0, & x_1^2 + x_2^2 > R^2. \end{cases}$$

Для определения коэффициента A воспользуемся свойством 3 двумерной плотности распределения. Поскольку $p(x_1, x_2) = 0$ при $x_1^2 + x_2^2 > R^2$, то

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x_1, x_2) dx_1 dx_2 = \iint_{x_1^2 + x_2^2 \leq R^2} A dx_1 dx_2 = \pi A R^2 = 1,$$

и, значит, $A = \frac{1}{\pi R^2}$. Таким образом,

$$p(x_1, x_2) = \begin{cases} 0, & x_1^2 + x_2^2 > R^2; \\ \frac{1}{\pi R^2}, & x_1^2 + x_2^2 \leq R^2. \end{cases}$$

Нетрудно найти одномерную плотность распределения случайной величины X_1 :

$$p_{X_1}(x) = \int_{-\infty}^{+\infty} p(x, y) dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2} dy = \begin{cases} 0, & |x| > R; \\ \frac{2\sqrt{R^2-x^2}}{\pi R^2}, & |x| \leq R. \end{cases}$$

Аналогичное выражение можно получить и для $p_{X_2}(y)$.

Независимые случайные величины

Определение 8.5 Случайные величины X и Y называют **независимыми**, если совместная функция распределения $F_{X,Y}(x, y)$ является произведением одномерных функций распределения $F_X(x)$ и $F_Y(y)$:

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

В противном случае **случайные величины** называют **зависимыми**.

Из определения 8.5 вытекает, что для независимых случайных величин X и Y события $\{X < x\}$ и $\{Y < y\}$ являются независимыми. Можно показать, что независимыми являются и все события $\{x_1 \leq X < x_2\}$ и $\{y_1 \leq Y < y_2\}$.

Теорема 8.3 Для того чтобы непрерывные случайные величины X и Y были независимыми, необходимо и достаточно, чтобы для всех x и y $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.

Доказательство. Пусть случайные величины X и Y независимые. Тогда, согласно определению 5.5, $F_{X,Y}(x, y) = F_X(x)F_Y(y)$. С учетом формул (8.1) и (5.2) имеем

$$p_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} = \left(\frac{dF_X(x)}{dx} \right) \left(\frac{dF_Y(y)}{dy} \right) = p_X(x)p_Y(y).$$

Тем самым необходимость утверждения доказана.

Для доказательства достаточности следует воспользоваться определением 8.4 двумерной плотности распределения и определением 5.1.

$$F_{X,Y}(x, y) = \iint_{\substack{-\infty < v < x \\ -\infty < w < y}} p_{X,Y}(v, w) dv dw = \left(\int_{-\infty}^x p_X(v) dv \right) \left(\int_{-\infty}^y p_Y(w) dw \right) = F_X(x)F_Y(y).$$

Справедлива следующая теорема.

Теорема 8.4 Если случайные величины X и Y независимыми, а функции $\varphi(t)$ и $\psi(t)$ кусочно-непрерывны, то случайные величины $\varphi(X)$ и $\psi(Y)$ также независимы.

Пример 8.5 а. Рассмотрим двумерный вектор $(X_1; X_2)$, совместная плотность распределения которого имеет вид

$$p(x_1, x_2) = \begin{cases} 1, & x_1 \in [0, 1] \text{ и } x_2 \in [0, 1]; \\ 0, & x_1 \notin [0, 1] \text{ или } x_2 \notin [0, 1]. \end{cases}$$

Легко показать, что одномерные плотности распределения $p_{X_1}(x)$ и $p_{X_2}(x)$ случайных величин X_1 и X_2 задаются формулой

$$p_{X_1}(x) = p_{X_2}(x) = \begin{cases} 1, & x \in [0, 1]; \\ 0, & x \notin [0, 1]. \end{cases}$$

Очевидно, что в данном случае совместная плотность распределения $p(x_1, x_2)$ для всех x_1, x_2 является произведением одномерных плотностей $p_{X_1}(x_1)$ и $p_{X_2}(x_1)$. Значит, случайные величины X_1 и X_2 являются независимыми.

б. Также нетрудно показать, что случайные величины X_1 и X_2 из примера 8.4 являются зависимыми.

Теорема 8.5 Дискретные случайные величины X и Y являются независимыми тогда и только тогда, когда для всех возможных значений x_i и y_j

$$p_{i,j} = \mathbf{P}\{X = x_i, Y = y_j\} = \mathbf{P}\{X = x_i\}\mathbf{P}\{Y = y_j\} = p_{X_i}p_{Y_j}. \quad \#$$

Пример 8.6 В схеме Бернулли с двумя испытаниями (см. пример 8.3)

$$\begin{aligned} \mathbf{P}\{X_1 = 0, X_2 = 0\} &= q^2 = \mathbf{P}\{X_1 = 0\}\mathbf{P}\{X_2 = 0\}, \mathbf{P}\{X_1 = 0, X_2 = 1\} = qp = \mathbf{P}\{X_1 = 0\}\mathbf{P}\{X_2 = 1\}, \\ \mathbf{P}\{X_1 = 1, X_2 = 0\} &= pq = \mathbf{P}\{X_1 = 1\}\mathbf{P}\{X_2 = 0\}, \mathbf{P}\{X_1 = 1, X_2 = 1\} = p^2 = \mathbf{P}\{X_1 = 1\}\mathbf{P}\{X_2 = 1\}. \end{aligned}$$

Таким образом, числа успехов X_1 и X_2 в первом и втором испытаниях представляют собой независимые случайные величины. Впрочем, в силу определения схемы Бернулли иного нельзя было ожидать. Убедитесь самостоятельно в том, что независимыми в совокупности являются случайные величины X_1, \dots, X_n — числа успехов в первом, втором, \dots , n -м испытаниях по схеме Бернулли. $\#$

Определение 8.6 Случайные величины X_1, \dots, X_n , заданные на одном и том же вероятностном пространстве, называют **независимыми в совокупности**, если

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n).$$

Замечание 8.1 Теоремы 8.3 и 8.5 распространяются на любое число случайных величин.

Разумеется, как и для событий, из попарной независимости не следует независимость случайных величин в совокупности.

Пример 8.7 Свяжем с бросанием тетраэдра из примера 3.5 три случайные величины: X_1 , X_2 и X_3 , каждая из которых может принимать значения 0 или 1, причем $X_i = 1$, если тетраэдр упал на грань, на которой присутствует цифра i , и $X_i = 0$ в противном случае. Аналогично X_2 характеризует наличие цифры 2, а X_3 — цифры 3. Покажем, что случайные величины X_1 , X_2 и X_3 будут попарно независимыми, но не являются независимыми в совокупности.

Действительно,

$$\mathbf{P}\{X_i = 1\} = \mathbf{P}\{X_i = 0\} = \frac{1}{2}, \quad i = 1, 2, 3,$$

и

$$\mathbf{P}\{X_i = 1, X_j = 1\} = \frac{1}{4} = \mathbf{P}\{X_i = 1\}\mathbf{P}\{X_j = 1\}, \quad i \neq j,$$

т.е. X_i попарно независимы. Однако, например,

$$\mathbf{P}\{X_1 = 1, X_2 = 1, X_3 = 1\} = \frac{1}{4} \neq \mathbf{P}\{X_1 = 1\}\mathbf{P}\{X_2 = 1\}\mathbf{P}\{X_3 = 1\} = \frac{1}{8},$$

т.е. X_i , $i = 1, 2, 3$, не являются независимыми в совокупности.

Лекция 9

Функции от случайных величин

Скалярные функции от случайного векторного аргумента

Скалярную функцию от случайного векторного аргумента определяют так же, как и функцию от одномерной случайной величины. Для простоты изложения ограничимся рассмотрением функции от двух случайных аргументов, хотя приведенные ниже выводы можно полностью перенести на случай любого числа аргументов.

Рассмотрим на вероятностном пространстве $(\Omega, \mathfrak{B}, \mathbf{P})$ двумерный случайный вектор $\vec{X} = (X_1; X_2)$ и числовую функцию $y = \varphi(x_1, x_2)$ числовых аргументов x_1 и x_2 .

Определение 9.1 Случайную величину

$$Y = \varphi(X_1, X_2) = \varphi(X_1(\omega), X_2(\omega))$$

называют *функцией (скалярной) от двумерной случайной величины (двумерного случайного вектора)* $(X_1; X_2)$.

Ясно, что функция $Y = \varphi(X_1, X_2)$ от двумерной дискретной случайной величины $(X_1; X_2)$ является дискретной случайной величиной, принимающей значения $\varphi(x_{1i}, x_{2j})$ с вероятностью

$$p_{ij} = \mathbf{P}\{X_1 = x_{1i}, X_2 = x_{2j}\},$$

где x_{1i} и x_{2j} — значения случайных величин X_1 и X_2 соответственно.

Чтобы построить ряд распределения дискретной случайной величины $Y = \varphi(X_1, X_2)$, необходимо, во-первых, не учитывать все те значения $\varphi(x_{1i}, x_{2j})$, вероятность принять которые случайной величине Y равна нулю, а во-вторых, объединить в один столбец все одинаковые значения $\varphi(x_{1i}, x_{2j})$ случайной величины Y , приписав этому столбцу суммарную вероятность.

Пример 9.1 Пусть Y — случайная величина, равная суммарному числу успехов в двух испытаниях по схеме Бернулли, а X_i — число успехов в i -м испытании, $i = 1, 2$. Тогда

$$Y = X_1 + X_2 \text{ и } \varphi(x_1, x_2) = x_1 + x_2.$$

Поскольку X_i могут принимать только значения 0 или 1, то случайная величина Y может принимать четыре значения:

$$\varphi(0, 0) = 0 + 0 = 0, \quad \varphi(1, 0) = 1 + 0 = 1, \quad \varphi(0, 1) = 0 + 1 = 1, \quad \varphi(1, 1) = 1 + 1 = 2.$$

с вероятностями q^2 , pq , qp и p^2 соответственно, где p — вероятность успеха в одном испытании, $q = 1 - p$ (табл. 9.1, см. также пример 8.3 и табл. 8.2).

Y	$\varphi(0, 0) = 0$	$\varphi(1, 0) = 1$	$\varphi(0, 1) = 1$	$\varphi(1, 1) = 2$
\mathbf{P}_Y	q^2	pq	qp	p^2

Таблица 9.1.

Заметим, что двум средним столбцам соответствует одно и то же значение 1 случайной величины Y , и их необходимо объединить. Окончательно получаем ряд распределения случайной величины Y , представленный в табл. 9.2. Как и следовало ожидать, суммарное число успехов Y в двух испытаниях имеет биномиальное распределение.

Y	0	1	2
\mathbf{P}_Y	q^2	$2pq$	p^2

Таблица 9.2.

#

В том случае, когда $(X_1; X_2)$ — двумерная непрерывная случайная величина с плотностью распределения $p_{X_1, X_2}(x_1, x_2)$, функцию распределения случайной величины $Y = \varphi(X_1, X_2)$ можно найти по формуле

$$F_Y(y) = \iint_{\varphi(x_1, x_2) \leq y} p_{X_1, X_2}(x_1, x_2) dx_1 dx_2, \quad (9.1)$$

где область интегрирования состоит из всех значений x_1 и x_2 , для которых $\varphi(x_1, x_2) < y$.

Поясним геометрически вывод формулы (9.1). Пусть поверхность, определенная функцией $y = \varphi(x_1, x_2)$, имеет вид “чаши” (см. рис. 9.1) и y — произвольное значение случайной величины $Y = \varphi(X_1, X_2)$. Проведем плоскость π , проходящую через точку $(0; 0; y)$ и ортогональную оси Oy . Обозначим через L линию пересечения плоскости π и поверхности $y = \varphi(x_1, x_2)$; L' — ее проекцию на плоскость x_1Ox_2 ; $D(y)$ — ту часть плоскости x_1Ox_2 , попадание в которую случайного вектора $(X_1; X_2)$ ведет к реализации события $\{Y < y\}$. Поскольку $Y = \varphi(X_1, X_2)$, то

$$D(y) = \{(x_1; x_2) : \varphi(x_1, x_2) < y\} = \{\varphi(x_1, x_2) < y\}.$$

События $\{Y < y\}$ и $\{(X_1; X_2) \in D(y)\}$ совпадают, и в соответствии со свойством 6 двумерной плотности распределения

$$\mathbf{P}\{Y < y\} = \mathbf{P}\{(X_1; X_2) \in D(y)\} = \iint_{\varphi(x_1, x_2) < y} p_{X_1, X_2}(x_1, x_2) dx_1 dx_2.$$

Учитывая равенство $\mathbf{P}\{Y < y\} = F_Y(y)$, приходим к формуле (9.1).

Пример 9.2 Пусть $(X_1; X_2)$ — двумерный случайный вектор, имеющий стандартное двумерное нормальное распределение (см. определение многомерного нормального распределения на с. 55). Найдем распределение случайной величины $Y = \sqrt{X_1^2 + X_2^2}$. В этом случае $\varphi(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$. Очевидно, что

$$F_Y(y) = \begin{cases} 0, & y \leq 0; \\ \iint_{\sqrt{x_1^2 + x_2^2} < y} \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)} dx_1 dx_2, & y > 0. \end{cases}$$

Переходя к полярным координатам ρ и φ , имеем

$$F_Y(y) = \int_0^y d\rho \int_0^{2\pi} \frac{1}{2\pi} e^{-\rho^2/2} \rho d\varphi = \int_0^y \rho e^{-\rho^2/2} d\rho = 1 - e^{-y^2/2}, \quad y > 0.$$

Это распределение известно как *распределение Релея*.

Математическое ожидание функции от случайного вектора

Математическое ожидание $\mathbf{M}Y$ функции $Y = \varphi(X_1, X_2)$ от дискретной двумерной случайной величины $(X_1; X_2)$ можно найти, воспользовавшись формулой

$$\mathbf{M}Y = \mathbf{M}\varphi(X_1, X_2) = \sum_{i,j} \varphi(x_i, y_j) p_{ij},$$

где $p_{ij} = \mathbf{P}\{X_1 = x_i, X_2 = y_j\}$, а функции $Y = \varphi(X_1, X_2)$ от двумерной непрерывной случайной величины (X_1, X_2) — формулой

$$\mathbf{M}Y = \mathbf{M}\varphi(X_1, X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) p_{X_1, X_2}(x, y) dx dy, \quad (9.2)$$

где $p_{X_1, X_2}(x, y)$ — совместная плотность распределения случайных величин X_1 и X_2 .

Свойства математического ожидания

Докажем теперь теорему о свойствах математического ожидания.

Теорема 9.1 Математическое ожидание удовлетворяет следующим свойствам.

1. Если случайная величина X принимает всего одно значение C с вероятностью единица (т.е., по сути дела, не является случайной величиной), то $\mathbf{M}C = C$.
2. $\mathbf{M}(aX + b) = a\mathbf{M}X + b$, где a, b — постоянные.
3. $\mathbf{M}(X_1 + X_2) = \mathbf{M}X_1 + \mathbf{M}X_2$.
4. $\mathbf{M}(X_1 X_2) = \mathbf{M}X_1 \cdot \mathbf{M}X_2$ для независимых случайных величин X_1 и X_2 .

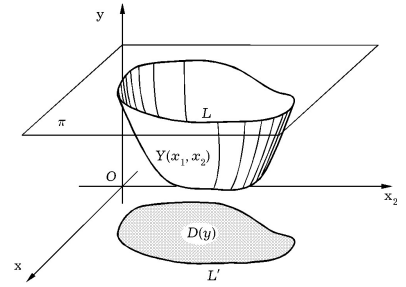


Рис. 9.1.

ДОКАЗАТЕЛЬСТВО. Если случайная величина X принимает всего одно значение C с вероятностью единица, то $\mathbf{M}C = C \cdot 1 = C$, откуда следует утверждение 1.

Доказательство свойств 2 и 4 проведем для непрерывных случайных величин (для дискретных случайных величин предлагаем читателю провести самостоятельно), а свойство 3 докажем для дискретных случайных величин (для непрерывных — доказать самостоятельно).

Найдем математическое ожидание случайной величины $Y = aX + b$ ($Y = \varphi(x) = ax + b$):

$$\mathbf{M}Y = \mathbf{M}(aX + b) = \int_{-\infty}^{+\infty} (ax + b)p_X(x)dx = a \int_{-\infty}^{+\infty} xp_X(x)dx + b \int_{-\infty}^{+\infty} p_X(x)dx = a\mathbf{M}X + b \cdot 1,$$

т.е. приходим к утверждению 2.

Пусть теперь $Y = X_1 + X_2$ ($Y = \varphi(x_1, x_2) = x_1 + x_2$). Тогда

$$\begin{aligned} \mathbf{M}Y = \mathbf{M}(X_1 + X_2) &= \sum_{i,j} (x_i + y_j)p_{ij} = \sum_{i,j} x_i p_{ij} + \sum_{i,j} y_j p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \\ &= \sum_i x_i p_{X_1 i} + \sum_j y_j p_{X_2 j} = \mathbf{M}X_1 + \mathbf{M}X_2, \end{aligned}$$

и, значит, утверждение 3 доказано.

Наконец, если X_1 и X_2 независимые случайные величины, то для математического ожидания их произведения $Y = X_1 X_2$ (воспользовавшись формулой 9.2 и теоремой 8.3) имеем:

$$\begin{aligned} \mathbf{M}Y = \mathbf{M}(X_1 X_2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 p_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 p_{X_1}(x_1) p_{X_2}(x_2) dx_1 dx_2 = \\ &= \left(\int_{-\infty}^{+\infty} x_1 p_{X_1}(x_1) dx_1 \right) \left(\int_{-\infty}^{+\infty} x_2 p_{X_2}(x_2) dx_2 \right) = \mathbf{M}X_1 \mathbf{M}X_2. \end{aligned}$$

Замечание 9.1 Свойство 4 также допускает обобщение на произведение конечного числа *независимых* (в совокупности) случайных величин:

$$\mathbf{M}(X_1 \cdot X_2 \dots X_n) = \mathbf{M}X_1 \cdot \mathbf{M}X_2 \dots \mathbf{M}X_n.$$

Пример 9.3 Представим число успехов X в n испытаниях по *схеме Бернулли* в виде $X = X_1 + \dots + X_n$, где X_i — число успехов в i -м испытании. Нетрудно видеть, что $\mathbf{M}X_i = 0 \cdot q + 1 \cdot p = p$. Значит, в силу свойства 3 $\mathbf{M}X = \mathbf{M}X_1 + \dots + \mathbf{M}X_n = np$, что совпадает с результатами примера 7.1, но получено с минимальными вычислениями. #

Определение 9.2 Вектор $\vec{m} = (\mathbf{M}X_1; \dots; \mathbf{M}X_n)$ называют **вектором математических ожиданий (средних значений)** случайного вектора \vec{X} .

Свойства дисперсии

Теорема 9.2 Дисперсия удовлетворяет следующим свойствам.

1. Если случайная величина X принимает всего одно значение C с вероятностью единица, то $\mathbf{D}C = 0$.

2. $\mathbf{D}(aX + b) = a^2 \mathbf{D}X$.

3. $\mathbf{D}X = \mathbf{M}X^2 - (\mathbf{M}X)^2$.

4. $\mathbf{D}(X + Y) = \mathbf{D}X + \mathbf{D}Y$ для независимых случайных величин X и Y .

ДОКАЗАТЕЛЬСТВО. Если случайная величина X с вероятностью единица принимает всего одно значение C , то в силу свойства 1 математического ожидания ($\mathbf{M}X = C$) получаем $\mathbf{D}X = \mathbf{M}(X - C)^2 = (C - C)^2 \cdot 1$, откуда вытекает утверждение 1.

Определим дисперсию случайной величины $Y = aX + b$. Используя свойство 2 математического ожидания, имеем

$$\begin{aligned} \mathbf{D}Y = \mathbf{M}(Y - \mathbf{M}Y)^2 &= \mathbf{M}(aX + b - \mathbf{M}(aX + b))^2 = \mathbf{M}(aX + b - a\mathbf{M}X - b)^2 = \mathbf{M}(a(X - \mathbf{M}X))^2 = \\ &= \mathbf{M}(a^2(X - \mathbf{M}X)^2) = a^2 \mathbf{M}(X - \mathbf{M}X)^2. \end{aligned}$$

Поэтому справедливо утверждение 2.

Далее, согласно свойствам 2 и 3 математического ожидания, получаем

$$\mathbf{D}X = \mathbf{M}(X - \mathbf{M}X)^2 = \mathbf{M}(X^2 - 2X\mathbf{M}X + (\mathbf{M}X)^2) = \mathbf{M}X^2 - 2(\mathbf{M}X)^2 + (\mathbf{M}X)^2 = \mathbf{M}X^2 - (\mathbf{M}X)^2,$$

т.е. приходим к утверждению 3.

Наконец, пусть X и Y — независимые случайные величины. Тогда, используя независимость случайных величин $X = X - \mathbf{M}X$ и $Y = Y - \mathbf{M}Y$, а также свойства 2–4 математического ожидания, получаем

$$\begin{aligned}\mathbf{D}(X + Y) &= \mathbf{M}(X + Y - \mathbf{M}(X + Y))^2 = \mathbf{M}((X - \mathbf{M}X) + (Y - \mathbf{M}Y))^2 = \mathbf{M}(X - \mathbf{M}X)^2 + \\ &+ 2\mathbf{M}((X - \mathbf{M}X)(Y - \mathbf{M}Y)) + \mathbf{M}(Y - \mathbf{M}Y)^2 = \mathbf{D}X + 2(\overset{\circ}{\mathbf{M}}X \cdot \overset{\circ}{\mathbf{M}}Y) + \mathbf{D}Y = \mathbf{D}X + \mathbf{D}Y,\end{aligned}$$

поскольку $\overset{\circ}{\mathbf{M}}X = 0$ и $\overset{\circ}{\mathbf{M}}Y = 0$. Значит, имеет место утверждение 4.

Замечание 9.2 Можно показать, что справедливо и свойство, обратное свойству 1, а следовательно, имеет место утверждение: дисперсия случайной величины X равна нулю тогда и только тогда, когда X с вероятностью 1 принимает всего одно значение.

Замечание 9.3 Очевидно, что свойство 4 справедливо для суммы не только двух, но и любого числа n попарно независимых случайных величин

$$\mathbf{D}(X_1 + \dots + X_n) = \mathbf{D}X_1 + \dots + \mathbf{D}X_n.$$

Нетрудно видеть, что дисперсия $\mathbf{D}X$ имеет размерность квадрата размерности случайной величины X . Для практических же целей удобно иметь величину, характеризующую разброс значений случайной величины вокруг ее математического ожидания, размерность которой совпадает с размерностью X . В качестве такой величины естественно использовать $\sigma = \sqrt{\mathbf{D}X}$, которую называют **средним квадратичным отклонением** случайной величины X .

Формула свертки

Важную роль в теории вероятностей и ее применениях играет тот случай, когда X_1 и X_2 являются независимыми случайными величинами, т.е. их двумерная плотность распределения

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$$

(мы ограничиваемся здесь только случаем *непрерывных случайных величин*), а случайная величина Y является их суммой:

$$Y = X_1 + X_2.$$

Тогда $Y = \varphi(X_1, X_2)$, где

$$\varphi(x_1, x_2) = x_1 + x_2,$$

и, согласно формуле (9.1), находим:

$$\begin{aligned}F_Y(y) &= \iint_{x_1+x_2 < y} p_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = \iint_{x_1+x_2 < y} p_{X_1}(x_1)p_{X_2}(x_2) dx_1 dx_2 = \\ &= \int_{-\infty}^{+\infty} p_{X_1}(x_1) dx_1 \int_{-\infty}^{y-x_1} p_{X_2}(x_2) dx_2 = \int_{-\infty}^{+\infty} F_{X_2}(y-x_1)p_{X_1}(x_1) dx_1.\end{aligned}$$

Дифференцируя последнюю формулу по y под знаком интеграла, получаем (с учетом переобозначения $x_1 = x$) выражение для *плотности* $p_Y(y)$ *распределения* суммы X_1 и X_2 :

$$p_Y(y) = \int_{-\infty}^{+\infty} p_{X_2}(y-x)p_{X_1}(x) dx. \quad (9.3)$$

В этом случае говорят, что плотность распределения $p_Y(y)$ случайной величины Y является **сверткой** (**композицией**) **плотностей распределения** $p_{X_1}(x)$ и $p_{X_2}(x)$ слагаемых X_1 и X_2 или что закон распределения суммы двух независимых случайных величин является **сверткой** (**композицией**) **законов распределения** слагаемых. Соотношение (9.3) условно записывают в виде $p_Y = p_{X_2} * p_{X_1}$. Формулу (9.3) называют **формулой свертки** для плотностей распределения случайных величин X_1 и X_2 .

Пример 9.4 Пусть X_1 и X_2 — независимые случайные величины, распределенные по нормальному закону со средними значениями m_1 и m_2 и средними квадратичными отклонениями σ_1 и σ_2 . Найдем плотность распределения суммы $Y = X_1 + X_2$. Воспользовавшись формулой свертки, имеем

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{+\infty} \varphi_{m_1, \sigma_1}(y-x) \varphi_{m_2, \sigma_2}(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{(y-x-m_1)^2}{2\sigma_1^2}\right) \frac{1}{\sigma_2 \sqrt{2\pi}} \exp\left(-\frac{(x-m_2)^2}{2\sigma_2^2}\right) dx = \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \times \int_{-\infty}^{+\infty} \exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2} \left(x - \frac{\sigma_1^2 m_2 - \sigma_2^2 m_1 + \sigma_2^2 y}{\sigma_1^2 + \sigma_2^2}\right)^2\right) dx. \end{aligned}$$

Делая теперь замену $z = \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2} \left(x - \frac{\sigma_1^2 m_2 - \sigma_2^2 m_1 + \sigma_2^2 y}{\sigma_1^2 + \sigma_2^2}\right)$, получаем

$$\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(y-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \exp\left(-\frac{(y-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right).$$

Таким образом, случайная величина Y также распределена по нормальному закону с параметрами $m_1 + m_2$ и $\sqrt{\sigma_1^2 + \sigma_2^2}$, т.е. композиция плотностей нормальных законов распределения является плотностью нормального закона распределения.

Лекция 10

Ковариация и коэффициент корреляции случайных величин

Пусть $(X_1; X_2)$ — двумерный случайный вектор.

Определение 10.1 Ковариацией (корреляционным моментом) $\text{cov}(X_1; X_2)$ случайных величин X_1 и X_2 называют математическое ожидание произведения случайных величин $\overset{\circ}{X}_1 = X_1 - \mathbf{M}X_1$ и $\overset{\circ}{X}_2 = X_2 - \mathbf{M}X_2$:

$$\text{cov}(X_1; X_2) = \mathbf{M}(\overset{\circ}{X}_1 \overset{\circ}{X}_2) = \mathbf{M}((X_1 - \mathbf{M}X_1)(X_2 - \mathbf{M}X_2)).$$

Запишем формулы, определяющие ковариацию.

Для дискретных случайных величин X_1 и X_2

$$\text{cov}(X_1; X_2) = \sum_{i,j} (x_i - \mathbf{M}X_1)(y_j - \mathbf{M}X_2)p_{ij},$$

для непрерывных случайных величин X_1 и X_2

$$\text{cov}(X_1; X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_1 - \mathbf{M}X_1)(x_2 - \mathbf{M}X_2)p_{X_1, X_2}(x_1, x_2) dx_1 dx_2.$$

Заметим, что введение понятия ковариации позволяет записать выражение для дисперсии суммы случайных величин и к уже имеющимся свойствам дисперсии добавить еще одно:

$$\mathbf{D}(X + Y) = \mathbf{D}X + \mathbf{D}Y + 2\text{cov}(X, Y)$$

(свойство 5 дисперсии), справедливое для произвольных, а не только независимых случайных величин X и Y . Действительно,

$$\begin{aligned} \mathbf{D}(X + Y) &= \mathbf{M}((X + Y) - \mathbf{M}(X + Y))^2 = \\ &= \mathbf{M}(X - \mathbf{M}X)^2 + 2\mathbf{M}((X - \mathbf{M}X)(Y - \mathbf{M}Y)) + \mathbf{M}(Y - \mathbf{M}Y)^2 = \\ &= \mathbf{D}X + \mathbf{D}Y + 2\text{cov}(X, Y). \end{aligned}$$

Свойство 5 дисперсии допускает обобщение на произвольное число слагаемых:

$$\mathbf{D}(X_1 + \dots + X_n) = \mathbf{D}X_1 + \dots + \mathbf{D}X_n + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

Следующая теорема устанавливает основные свойства ковариации.

Теорема 10.1 Ковариация имеет следующие свойства

1. $\text{cov}(X, X) = \mathbf{D}X$.
2. $\text{cov}(X_1, X_2) = 0$ для независимых случайных величин X_1 и X_2
3. Если $Y_i = a_i X_i + b_i$, $i = 1, 2$, то $\text{cov}(Y_1, Y_2) = a_1 a_2 \text{cov}(X_1, X_2)$.
4. $-\sqrt{\mathbf{D}X_1 \mathbf{D}X_2} \leq \text{cov}(X_1, X_2) \leq \sqrt{\mathbf{D}X_1 \mathbf{D}X_2}$, причем

$$|\text{cov}(X_1, X_2)| = \sqrt{\mathbf{D}X_1 \mathbf{D}X_2} \quad (10.1)$$

тогда и только тогда, когда случайные величины X_1 и X_2 связаны линейной зависимостью, т.е. существуют такие числа a и b , при которых

$$X_2 = aX_1 - b. \quad (10.2)$$

$$5. \text{cov}(X_1, X_2) = \mathbf{M}(X_1 X_2) - \mathbf{M}X_1 \mathbf{M}X_2.$$

ДОКАЗАТЕЛЬСТВО. Утверждение 1 вытекает из очевидного соотношения $\text{cov}(X, X) = \mathbf{M}(X - \mathbf{M}X)^2$.

Если случайные величины X_1 и X_2 являются независимыми (и имеют математические ожидания), то

$$\text{cov}(X_1, X_2) = \mathbf{M}((X_1 - \mathbf{M}X_1)(X_2 - \mathbf{M}X_2)) = \mathbf{M}(X_1 - \mathbf{M}X_1)\mathbf{M}(X_2 - \mathbf{M}X_2),$$

откуда приходим к утверждению 2.

Пусть $Y_1 = a_1 X_1 + b_1$, $Y_2 = a_2 X_2 + b_2$. Тогда

$$\begin{aligned} \text{cov}(Y_1, Y_2) &= \mathbf{M}((Y_1 - \mathbf{M}Y_1)(Y_2 - \mathbf{M}Y_2)) = \\ &= \mathbf{M}((a_1 X_1 + b_1 - a_1 \mathbf{M}X_1 - b_1)(a_2 X_2 + b_2 - a_2 \mathbf{M}X_2 - b_2)) = \\ &= \mathbf{M}(a_1 a_2 (X_1 - \mathbf{M}X_1)(X_2 - \mathbf{M}X_2)). \end{aligned}$$

Поэтому справедливо утверждение 3.

Рассмотрим дисперсию случайной величины $Y_x = xX_1 - X_2$, где x — произвольное число. В силу свойств дисперсии и свойства 3 ковариации

$$\mathbf{D}Y_x = \mathbf{D}(xX_1) + 2\text{cov}(xX_1, -X_2) + \mathbf{D}(-X_2) = x^2 \mathbf{D}X_1 - 2x \text{cov}(X_1, X_2) + \mathbf{D}X_2.$$

Дисперсия $\mathbf{D}Y_x$, как функции от x , представляет собой квадратный трехчлен. Но дисперсия любой случайной величины не может быть меньше нуля, а это означает, что дискриминант

$$D = (2\text{cov}(X_1, X_2))^2 - 4\mathbf{D}X_1 \mathbf{D}X_2 \quad (10.3)$$

квадратного трехчлена $\mathbf{D}Y_x$ является неположительным, т.е. $|\text{cov}(X_1, X_2)| \leq \sqrt{\mathbf{D}X_1 \mathbf{D}X_2}$.

Далее, пусть выполнено равенство (10.1). Значит, дискриминант (10.3) равен нулю, и поэтому уравнение $\mathbf{D}Y_x = 0$ имеет решение, которое обозначим a . Тогда в соответствии с замечанием 9.2 случайная величина $Y_a = aX_1 - X_2$ принимает всего одно значение (допустим, b), и, следовательно, $X_2 = aX_1 - b$, т.е. из (10.1) вытекает (10.2).

Наоборот, пусть выполнено (10.2). Тогда $Y_a = aX_1 - X_2 = b$ и в соответствии со свойством 1 дисперсии $\mathbf{D}Y_a = 0$, а значит, квадратный трехчлен $\mathbf{D}Y_x$ имеет один действительный корень. Т.к. этот же трехчлен неотрицательный, то он имеет действительный корень лишь в случае, когда его дискриминант в (10.3) равен нулю, откуда следует (10.1). Таким образом, из (10.2) вытекает (10.1). Утверждение 4 полностью доказано.

Наконец, раскрывая скобки в формуле, определяющей ковариацию, и используя свойства математического ожидания, получаем утверждение 5, которое часто бывает полезным при численном подсчете ковариации.

Замечание 10.1 Если случайные величины связаны линейной зависимостью $X_2 = aX_1 - b$, то в соответствии со свойствами 3 и 1 $\text{cov}(X_1, X_2) = a\text{cov}(X_1, X_1) = a\mathbf{D}X_1$. Поэтому знак ковариации совпадает со знаком коэффициента a и равенство (10.1) допускает следующее уточнение:

$$\begin{aligned} \text{cov}(X_1, X_2) &= \sqrt{\mathbf{D}X_1 \mathbf{D}X_2} && \text{при } a > 0; \\ \text{cov}(X_1, X_2) &= -\sqrt{\mathbf{D}X_1 \mathbf{D}X_2} && \text{при } a < 0. \end{aligned}$$

Замечание 10.2 Из свойств дисперсии и ковариации можно получить еще одно полезное при расчетах свойство дисперсии

$$\mathbf{D}(a_1 X_1 + a_2 X_2 + b) = a_1^2 \mathbf{D}X_1 + a_2^2 \mathbf{D}X_2 + 2a_1 a_2 \text{cov}(X_1, X_2).$$

Докажите это свойство самостоятельно.

Замечание 10.3 Как следует из свойства 2, ковариация независимых случайных величин равна нулю. Однако обратное, вообще говоря, неверно. Существуют зависимые и даже функционально зависимые случайные величины, ковариация которых равна нулю, что демонстрирует следующий пример.

Пример 10.1 Пусть случайная величина X имеет *равномерное* в интервале $(-1, 1)$ *распределение*, а случайная величина Y связана со случайной величиной X функциональной зависимостью $Y = X^2$. Покажем, что $\text{cov}(X, Y) = 0$, несмотря на функциональную зависимость X и Y .

Действительно, учитывая равенство $\mathbf{M}X = 0$ и свойство 5 ковариации, имеем

$$\text{cov}(X, Y) = \mathbf{M}(X \cdot Y) - \mathbf{M}X \cdot \mathbf{M}Y = \mathbf{M}X^3 = \int_{-\infty}^{+\infty} x^3 p(x) dx = \frac{1}{2} \int_{-1}^1 x^3 dx = 0.$$

Определение 10.2 Случайные величины X и Y называют **некоррелированными**, если их ковариация равна нулю, т.е. $\text{cov}(X, Y) = 0$.

Приведенный выше пример показывает, что из некоррелированности случайных величин не следует их независимость. Можно сказать, что ковариация случайных величин отражает, насколько их зависимость близка к линейной.

Рассмотрим теперь n -мерный случайный вектор $\vec{X} = (X_1; \dots; X_n)$.

Определение 10.3 Матрицей ковариаций (ковариационной матрицей) случайного вектора \vec{X} называют матрицу $\Sigma = (\sigma_{ij}) = (\text{cov}(X_i, X_j))$, состоящую из ковариаций случайных величин X_i и X_j .

Пример 10.2 Рассмотрим двумерную случайную величину $(X; Y)$, распределенную по *нормальному закону* (см. лекцию ??). Тогда

$$\text{cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{(x-m_1)(y-m_2)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp \left[-\left(\frac{(x-m_1)^2}{2\sigma_1^2(1-\rho^2)} - \frac{2\rho(x-m_1)(y-m_2)}{2\sigma_1\sigma_2(1-\rho^2)} + \frac{(y-m_2)^2}{2\sigma_2^2(1-\rho^2)} \right) \right] dx dy.$$

Делая замену $u = (x - m_1)/\sigma_1$, $v = (y - m_2)/\sigma_2$, получаем

$$\begin{aligned} \text{cov}(X, Y) &= \sigma_1\sigma_2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{uv}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2) \right\} du dv = \\ &= \sigma_1\sigma_2 \int_{-\infty}^{+\infty} \frac{u}{\sqrt{2\pi}} e^{-u^2/2} \int_{-\infty}^{+\infty} \frac{v}{2\pi(1-\rho^2)} \exp \left\{ -\frac{(v-\rho u)^2}{2(1-\rho^2)} \right\} dv du. \end{aligned}$$

Внутренний интеграл равен ρu . Поэтому

$$\text{cov}(X, Y) = \rho\sigma_1\sigma_2 \int_{-\infty}^{+\infty} \frac{u^2}{\sqrt{2\pi}} e^{-u^2/2} du = \rho\sigma_1\sigma_2.$$

Поскольку $\mathbf{D}X = \sigma_1^2$, $\mathbf{D}Y = \sigma_2^2$, то матрица Σ представляет собой матрицу ковариаций (в данном случае понятие “ковариационная матрица” мы ввели раньше, нежели выяснили смысл этого понятия). #

Существенным недостатком ковариации является то, что ее размерность совпадает с произведением размерностей случайных величин. Естественно, хотелось бы иметь безразмерную характеристику степени линейной зависимости. Но это очень просто сделать — достаточно разделить ковариацию случайных величин на произведение их средних квадратичных отклонений.

Определение 10.4 Коэффициентом корреляции случайных величин X и Y называют число $\rho = \rho(X, Y)$, определяемое равенством (предполагается, что $\mathbf{D}X > 0$ и $\mathbf{D}Y > 0$)

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\mathbf{D}X \cdot \mathbf{D}Y}}.$$

Теорема 10.2 Коэффициент корреляции имеет следующие свойства.

1. $\rho(X, X) = 1$.
2. Если случайные величины X и Y являются независимыми (и существуют $\mathbf{D}X > 0$ и $\mathbf{D}Y > 0$), то $\rho(X, Y) = 0$.
3. $\rho(a_1X_1 + b_1, a_2X_2 + b_2) = \pm\rho(X_1, X_2)$. При этом знак плюс нужно брать в том случае, когда a_1 и a_2 имеют одинаковые знаки, и минус — в противном случае.
4. $-1 \leq \rho(X, Y) \leq 1$.
5. $|\rho(X, Y)| = 1$ тогда и только тогда, когда случайные величины X и Y связаны линейной зависимостью.

ДОКАЗАТЕЛЬСТВО. Доказательство теоремы следует из свойств ковариации, и мы предлагаем провести его самостоятельно.

Коэффициент корреляции отражает “степень линейной близости” случайных величин. При $\rho > 0$ говорят о *положительной* корреляционной зависимости X и Y , при $\rho < 0$ — об *отрицательной*. Например, рост и вес человека связаны положительной корреляционной зависимостью, а температура и время сохранности продукта — отрицательной. Однако, коэффициент корреляции (ковариация) может не улавливать “степень нелинейной близости” случайных величин.

По аналогии с ковариационной матрицей для случайного вектора $\vec{X} = (X_1; \dots; X_n)$ можно ввести корреляционную матрицу.

Определение 10.5 *Корреляционной (нормированной ковариационной) матрицей* случайного вектора \vec{X} называют матрицу $P = (\rho_{ij}) = (\rho(X_i, X_j))$, состоящую из коэффициентов корреляций случайных величин X_i и X_j .

Лекция 11

Условные характеристики случайных величин

Одним из основных понятий теории вероятностей является понятие *условной вероятности*, введенное в лекции 3. Там же было показано, что условная вероятность $\mathbf{P}(A|B)$ обладает всеми свойствами *безусловной вероятности* и так же, как и безусловная вероятность, представляет собой численную меру наступления *события* A , но только при условии, что событие B произошло.

Аналогом понятия условной вероятности для двух *случайных величин* X и Y является *условный закон распределения* одной из них, допустим, X при условии, что вторая случайная величина Y приняла определенное значение. С помощью условного закона распределения вводят условные числовые характеристики. Именно эти понятия и рассматриваются в настоящей лекции.

Условные распределения

Понятие условного распределения, как обычно, введем только для случаев *дискретных* и *непрерывных случайных величин*.

В случае *двумерной дискретной случайной величины* $(X; Y)$ будем предполагать для простоты изложения, что *множества возможных значений* случайных величин X и Y являются конечными, т.е. X и Y принимают значения x_i , $i = \overline{1, n}$, и y_j , $j = \overline{1, m}$, соответственно. В этом случае, как мы знаем, *закон распределения* двумерного случайного вектора $(X; Y)$ удобно задавать набором *вероятностей*

$$p_{ij} = \mathbf{P}\{X = x_i, Y = y_j\}$$

для всех значений i и j . Напомним, что, зная вероятности p_{ij} , нетрудно найти (см. 8.1) законы распределений каждой из координат по формулам

$$p_{Xi} = \mathbf{P}\{X = x_i\} = \sum_{j=1}^m p_{ij}, \quad p_{Yj} = \mathbf{P}\{Y = y_j\} = \sum_{i=1}^n p_{ij}.$$

Определение 11.1 Для двумерной дискретной случайной величины $(X; Y)$ *условной вероятностью* π_{ij} , $i = \overline{1, n}$, $j = \overline{1, m}$, того, что случайная величина X примет значение x_i при условии $Y = y_j$, называют условную вероятность события $\{X = x_i\}$ при условии события $\{Y = y_j\}$, т.е.

$$\pi_{ij} = \mathbf{P}\{X = x_i | Y = y_j\} = \frac{\mathbf{P}\{X = x_i, Y = y_j\}}{\mathbf{P}\{Y = y_j\}} = \frac{p_{ij}}{p_{Yj}}. \quad (11.1)$$

При каждом j , $j = \overline{1, m}$, набор вероятностей π_{ij} , $i = \overline{1, n}$, определяет, с какими вероятностями случайная величина X принимает различные значения x_i , если известно, что случайная величина Y приняла значение y_j . Иными словами, набор вероятностей π_{ij} , $i = \overline{1, n}$, характеризует *условное распределение* дискретной случайной величины X при условии $Y = y_j$.

Аналогично определяют условную вероятность π_{ij}^* того, что случайная величина Y примет значение y_j при условии $X = x_i$:

$$\pi_{ij}^* = \mathbf{P}\{Y = y_j | X = x_i\} = \frac{\mathbf{P}\{X = x_i, Y = y_j\}}{\mathbf{P}\{X = x_i\}} = \frac{p_{ij}}{p_{Xi}}.$$

Пример 11.1 Условное распределение числа X_1 успехов в первом испытании по схеме Бернулли (см. пример 8.3) при условии, что число успехов во втором испытании $X_2 = j$, $j = 0, 1$, задается табл. 11.1. Из этой таблицы следует, что, независимо от числа успехов во втором испытании, 0 или 1 успех в первом испытании происходит с одними и теми же вероятностями p и q . Это очевидно, поскольку испытания по схеме Бернулли являются независимыми.

X_1	X_2		
	0	1	P_{X_1}
0	q	q	q
1	p	p	p
P_{X_2}	q	p	

Таблица 11.1.

В общем случае (т.е. когда X и Y не обязательно дискретные случайные величины) хотелось бы ввести *условную функцию распределения* случайной величины X при условии $Y = y$ по формуле

$$F_X(x|Y=y) = \frac{P\{X < x, Y = y\}}{P\{Y = y\}}. \quad (11.2)$$

Однако это не всегда возможно (например, для непрерывной случайной величины Y событие $\{Y = y\}$ имеет нулевую вероятность, т.е. $P\{Y = y\} = 0$). Поэтому воспользуемся предельным переходом, рассматривая вместо события $\{Y = y\}$ событие $\{y \leq Y < y + \Delta y\}$ и устремляя Δy к нулю.

Ограничимся случаем, когда *двумерный случайный вектор* $(X; Y)$ имеет непрерывную *совместную плотность распределения* $p(x, y)$, а следовательно (см. теорему 8.3), и плотности распределения

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy \quad \text{и} \quad p_Y(y) = \int_{-\infty}^{+\infty} p(x, y) dx,$$

случайных величин X и Y , которые также будем считать непрерывными.

Определим условную вероятность события $\{X < x\}$ при условии события $\{y \leq Y < y + \Delta y\}$:

$$P\{X < x | y \leq Y < y + \Delta y\} = \frac{P\{X < x, y \leq Y < y + \Delta y\}}{P\{y \leq Y < y + \Delta y\}} = \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)} = \frac{\int_y^{y+\Delta y} dv \int_{-\infty}^x p(u, v) du}{\int_y^{y+\Delta y} p_Y(v) dv}.$$

Можно показать, что в силу сделанных предположений функция $\int_{-\infty}^x p(u, v) du$ является непрерывной. Поэтому, согласно теореме о среднем значении,

$$\int_y^{y+\Delta y} dv \int_{-\infty}^x p(u, v) du = \Delta y \int_{-\infty}^x p(u, \xi) du, \quad \int_y^{y+\Delta y} p_Y(v) dv = p_Y(\eta) \Delta y$$

и, следовательно,

$$P\{X < x | y \leq Y < y + \Delta y\} = \frac{\int_{-\infty}^x p(u, \xi) du}{p_Y(\eta)},$$

где ξ и η — некоторые числа, заключенные между y и $y + \Delta y$.

Устремляя теперь Δy к нулю, получаем следующие выражения для *условной функции распределения* $F_X(x|Y=y)$:

$$F_X(x|Y=y) = \lim_{\Delta y \rightarrow 0} P\{X < x | y \leq Y < y + \Delta y\} = \frac{\int_{-\infty}^x p(u, y) du}{p_Y(y)} = \frac{1}{p_Y(y)} \int_{-\infty}^x p(u, y) du. \quad (11.3)$$

При сделанных предположениях о непрерывности случайного вектора $(X; Y)$ условная функция распределения $F_X(x|Y=y)$ имеет производную по x , т.е. существует условная плотность распределения случайной величины X при условии $Y = y$:

$$p_X(x|y) = \frac{p(x, y)}{p_Y(y)}. \quad (11.4)$$

Аналогично определяют условную функцию распределения $F_Y(y|X=x)$ и условную плотность распределения $p_Y(y|X=x)$ случайной величины Y при условии $X = x$:

$$F_Y(y|X=x) = \frac{1}{p_X(x)} \int_{-\infty}^y p(x, v) dv, \quad (11.5)$$

$$p_Y(y|x) = \frac{p(x,y)}{p_X(x)}. \quad (11.6)$$

Для краткости далее вместо $p_X(x|Y=y)$ и $p_Y(y|X=x)$ будем писать $p_X(x|y)$ и $p_Y(y|x)$.

Итак, для непрерывного случайного вектора $(X; Y)$ мы пришли к следующему определению условной плотности распределения.

Определение 11.2 *Условной плотностью распределения* случайной величины X , являющейся координатой двумерного случайного вектора $(X; Y)$, при условии, что другая его координата приняла некоторое фиксированное значение y , т.е. $Y=y$, называют функцию $p_X(x|y)$, определяемую соотношением (11.4). Аналогично (см. (11.6)) определяют условную плотность распределения $p_Y(y|x)$ координаты Y при условии $X=x$.

Введенные понятия — условное распределение (дискретной случайной величины), условная функция распределения и условная плотность распределения (для непрерывных случайных величин) — называют **условными законами распределения**.

Пример 11.2 Пусть случайные величины X_1 и X_2 представляют собой координаты точки падения частицы, случайным образом брошенной в круг радиуса R с центром в начале координат (см. пример 8.4). Случайный вектор $(X_1; X_2)$ имеет плотность распределения

$$p(x_1, x_2) = \begin{cases} 0, & x_1^2 + x_2^2 > R^2; \\ \frac{1}{\pi R^2}, & x_1^2 + x_2^2 \leq R^2. \end{cases}$$

Найдем условную плотность распределения абсциссы X_1 точки падения частицы при условии, что ордината X_2 приняла значение x_2 . Так как плотность распределения $p_{X_2}(x_2)$ случайной величины X_2 имеет вид

$$p_{X_2}(x_2) = \begin{cases} 0, & |x_2| > R; \\ \frac{2\sqrt{R^2 - x_2^2}}{\pi R^2}, & |x_2| \leq R, \end{cases}$$

то при $|x_2| \leq R$

$$p_{X_1}(x_1|x_2) = \frac{p(x_1, x_2)}{p_{X_2}(x_2)} = \begin{cases} 0, & |x_1| > \sqrt{R^2 - x_2^2}; \\ \frac{1}{2\sqrt{R^2 - x_2^2}}, & |x_1| \leq \sqrt{R^2 - x_2^2}. \end{cases}$$

Поэтому, случайная величина X_1 при условии $X_2 = x_2$ равномерно распределена на отрезке $[-\sqrt{R^2 - x_2^2}, \sqrt{R^2 - x_2^2}]$. Если $|x_2| > R$, то условная плотность распределения $p_{X_1}(x_1|x_2)$ не определена; но это нас не должно волновать, поскольку случайная величина X_2 не может принимать значения, по абсолютной величине большие R . #

Для проверки независимости случайных величин часто удобно пользоваться следующим критерием.

Критерий независимости случайных величин X и Y . Случайные величины X и Y являются независимыми тогда и только тогда, когда условное распределение (функция распределения, плотность распределения) случайной величины X при условии $Y=y$ совпадает с безусловным распределением (функцией распределения, плотностью распределения) случайной величины X .

В частности, дискретные величины X и Y являются независимыми тогда и только тогда, когда все условные вероятности

$$\pi_{ij} = \mathbf{P}\{X = x_i | Y = y_j\}$$

совпадают с безусловными вероятностями

$$p_{X_i} = \mathbf{P}\{X = x_i\}.$$

Пример 11.3 В двух испытаниях по схеме Бернулли (см. пример 11.1) числа успехов X_1 и X_2 в первом и втором испытаниях являются независимыми случайными величинами, поскольку в табл. 11.1 все три столбца совпадают. Этот факт нами уже был установлен другим способом в примере 8.5.

Пример 11.4 Условная плотность распределения случайной величины X_1 (абсциссы точки падения при равномерном бросании частицы в круг, см. пример 11.2) при условии $X_2 = x_2$ (ординаты точки падения) равномерна, в то время как безусловная плотность X_1 таковой не является. И в этом примере X_1 и X_2 зависимые случайные величины.

Условные числовые характеристики

Рассмотрим *двумерную случайную величину* $(X; Y)$. В соответствии с результатами предыдущего параграфа можно определить *условное распределение* случайной величины X при условии, что случайная величина Y приняла определенное значение y . Поскольку условное распределение обладает всеми свойствами обычного (безусловного) распределения, то по нему можно определить *математическое ожидание*, *дисперсию* и другие числовые характеристики, которые естественно назвать условными.

Начнем со случая *дискретной случайной величины* $(X; Y)$. Пусть случайная величина X принимает значения x_1, \dots, x_n , а случайная величина Y — значения y_1, \dots, y_m и пусть

$$\pi_{ij} = \mathbf{P}\{X = x_i | Y = y_j\} = \frac{\mathbf{P}\{X = x_i, Y = y_j\}}{\mathbf{P}\{Y = y_j\}} = \frac{p_{ij}}{p_{Yj}}, \quad i = \overline{1, n}, \quad j = \overline{1, m},$$

условные вероятности случайной величине X принять значение x_i при условии $Y = y_j$.

Определение 11.3 Для дискретной двумерной случайной величины $(X; Y)$ *значением* $\mathbf{M}(X | Y = y_j)$ *условного математического ожидания* дискретной случайной величины X при условии $Y = y_j$, называют число

$$\mathbf{M}(X | Y = y_j) = \sum_{i=1}^n x_i \pi_{ij}.$$

Далее для краткости будем писать $\mathbf{M}(X | y_j)$ вместо $\mathbf{M}(X | Y = y_j)$.

По аналогии с (безусловным) математическим ожиданием $\mathbf{M}X$ случайной величины X значение $\mathbf{M}(X | y_j)$ условного математического ожидания при условии $Y = y_j$ задает “среднее” значение случайной величины X , но при условии, что случайная величина Y приняла значение y_j .

Таким же образом интерпретируют значение $\mathbf{M}(Y | x_i) = \mathbf{M}(Y | X = x_i)$ условного математического ожидания случайной величины Y при условии $X = x_i$.

Согласно определению 11.3, значение $\mathbf{M}(X | y_j)$ условного математического ожидания зависит от значения y_j случайной величины Y , и только от него. Вспоминая понятие функции от случайной величины, приходим к следующему определению условного математического ожидания.

Определение 11.4 *Условным математическим ожиданием* $\mathbf{M}(X | Y)$ дискретной случайной величины X относительно дискретной случайной величины Y называют функцию $\mathbf{M}(X | Y) = g(Y)$ от случайной величины Y , где область определения функции $g(y)$ совпадает с множеством значений y_1, \dots, y_m случайной величины Y , а каждому значению y_j аргумента y поставлено в соответствие число $g(y_j) = \mathbf{M}(X | y_j)$.

Подчеркнем еще раз, что условное математическое ожидание $\mathbf{M}(X | Y)$ является функцией от случайной величины, т.е. также случайной величиной.

Приведем примеры.

Пример 11.5 Пусть X_1 и X_2 — числа успехов в первом и втором испытаниях по *схеме Бернулли* с вероятностью успеха p . Найдем $\mathbf{M}(X_1 | X_2)$. Воспользовавшись табл. 11.1, имеем:

$$\mathbf{M}(X_1 | 0) = 0 \cdot q + 1 \cdot p = p, \quad \mathbf{M}(X_1 | 1) = 0 \cdot q + 1 \cdot p = p.$$

Таким образом, значения $\mathbf{M}(X_1 | 0)$ и $\mathbf{M}(X_1 | 1)$ условного математического ожидания совпадают для обоих значений 0 и 1 случайной величины X_2 и равны p . Поэтому $\mathbf{M}(X_1 | X_2) \equiv p$.

Определение 11.5 Для непрерывной двумерной случайной величины $(X; Y)$ *значением* $\mathbf{M}(X | y) = \mathbf{M}(X | Y = y)$ *условного математического ожидания* непрерывной случайной величины X при условии $Y = y$ называют число

$$\mathbf{M}(X | y) = \int_{-\infty}^{+\infty} x p_X(x | y) dx,$$

где

$$p_X(x | y) = \frac{p(x, y)}{p_Y(y)}$$

является условной плотностью распределения случайной величины X при условии $Y = y$.

Определение 11.6 Для непрерывной двумерной случайной величины $(X; Y)$ *условным математическим ожиданием* $\mathbf{M}(X | Y)$ непрерывной случайной величины X относительно случайной величины Y называют функцию $g(Y) = \mathbf{M}(X | Y)$ от случайной величины Y , принимающую значение $g(y) = \mathbf{M}(X | y)$ при $Y = y$.

Проверьте самостоятельно, что свойства условного математического ожидания, выведенные для дискретного случая, справедливы и для непрерывного (исключение составляет свойство 1, поскольку непрерывная случайная величина не может принимать всего одно значение).

Резюмируя изложенное выше, можно сказать, что зависимость поведения “в среднем” случайной величины X от значения случайной величины Y характеризуется функцией $g(y) = \mathbf{M}(X|y)$.

Условное математическое ожидание, как обычное (безусловное) математическое ожидание, характеризует *центр рассеивания* случайной величины. Однако оно не дает никакой информации о степени рассеивания случайной величины относительно среднего значения.

Поскольку степень рассеивания случайной величины X можно оценить с помощью *дисперсии*, то в качестве меры рассеивания случайной величины X относительно Y можно принять *условную дисперсию*, которую естественно определить аналогично обычной дисперсии, но используя условное распределение случайной величины X при условии $Y = y$.

Определение 11.7 *Условной дисперсией* $\mathbf{D}(X|Y)$ случайной величины X относительно (случайной величины) Y называют случайную величину, задаваемую формулой

$$\mathbf{D}(X|Y) = \mathbf{M}([X - \mathbf{M}(X|Y)]^2|Y).$$

Приведенное определение применимо как для двумерной дискретной случайной величины, так и для непрерывной.

Для двумерной дискретной случайной величины $(X; Y)$ *значение* $\mathbf{D}(X|y_j)$ *условной дисперсии* X при условии $Y = y_j$ определяется формулой

$$\mathbf{D}(X|y_j) = \mathbf{M}([X - \mathbf{M}(X|y_j)]^2|Y) = \sum_{i=1}^n [x_i - \mathbf{M}(X|y_j)]^2 \pi_{ij},$$

а для двумерной непрерывной случайной величины $(X; Y)$ *значение* $\mathbf{D}(X|y)$ *условной дисперсии* X при условии $Y = y$ задается формулой

$$\mathbf{D}(X|y) = \mathbf{M}([X - \mathbf{M}(X|y)]^2|y) = \int_{-\infty}^{+\infty} [x - \mathbf{M}(X|y)]^2 p_X(x|y) dx.$$

Условная дисперсия случайной величины X так же, как и условное математическое ожидание этой случайной величины, зависит от того значения, которое приняла случайная величина Y . Поэтому условная дисперсия $\mathbf{D}(X|Y)$ является функцией от случайной величины Y , область определения которой совпадает с *множеством возможных значений* случайной величины Y .

Лекция 12

Многомерное нормальное распределение

Нормальное распределение одномерной случайной величины рассматривалось в лекции 7. Сейчас обратимся к многомерному случаю. При этом сначала введем *двумерное нормальное распределение случайного вектора* $\vec{X} = (X_1; X_2)$, а затем обобщим полученные результаты на случайный вектор \vec{X} произвольной размерности $n > 2$.

Пусть координаты X_1 и X_2 случайного вектора $\vec{X} = (X_1; X_2)$ являются случайными величинами, распределенными по нормальному закону, т.е. имеют *плотности распределения*

$$p_{X_1}(x) = \varphi_{m_1, \sigma_1}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2}\right) \text{ и } p_{X_2}(x) = \varphi_{m_2, \sigma_2}(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-m_2)^2}{2\sigma_2^2}\right)$$

Напомним, что параметры m_i и $\sigma_i > 0$, $i = \overline{1, 2}$, этих распределений называют *математическими ожиданиями* и *средними квадратическими отклонениями* случайных величин X_1 и X_2 .

Если X_1 и X_2 являются независимыми случайными величинами, то, согласно теореме 8.3,

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$$

и в этом случае плотность двумерного нормального распределения имеет вид

$$p_{X_1, X_2}(x_1, x_2) = \frac{1}{(\sqrt{2\pi})^2 \sigma_1 \sigma_2} \exp\left(-\frac{(x_1 - m_1)^2}{2\sigma_1^2} - \frac{(x_2 - m_2)^2}{2\sigma_2^2}\right).$$

В общем случае вектор $\vec{X} = (X_1; X_2)$ имеет (*невырожденное*) *двумерное нормальное распределение*, если его плотность распределения определяется формулой

$$p_{X_1, X_2}(x_1, x_2) = \frac{1}{(\sqrt{2\pi})^2 \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} e^{-\frac{1}{2} Q(x_1 - m_1, x_2 - m_2)}, \quad (12.1)$$

где функция двух переменных

$$Q(y_1, y_2) = \frac{1}{1 - \rho^2} \left(\frac{y_1^2}{\sigma_1^2} - \frac{2\rho y_1 y_2}{\sigma_1 \sigma_2} + \frac{y_2^2}{\sigma_2^2} \right), \quad (12.2)$$

$y_i = x_i - m_i$, $i = 1, 2$, есть положительно определенная квадратичная форма (т.е. $Q(y_1, y_2) > 0$ для любых $(y_1; y_2) \in \mathbb{R}$, $(y_1; y_2) \neq (0; 0)$).

Двумерное нормальное распределение зависит от пяти параметров: m_1 , m_2 , σ_1 , σ_2 , ρ . Можно показать, что $m_1 = \mathbf{M}X_1$, $m_2 = \mathbf{M}X_2$, $\sigma_1^2 = \mathbf{D}X_1$, $\sigma_2^2 = \mathbf{D}X_2$, ρ — коэффициент корреляции случайных величин X_1 и X_2 .

Последние три параметра запишем для дальнейшего обобщения на случай $n > 2$ в виде *матрицы ковариаций (ковариационной матрицы)* Σ вектора $\vec{X} = (X_1; X_2)$:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

где $\sigma_{ii} = \sigma_i^2$, $i = \overline{1, 2}$, а $\sigma_{12} = \sigma_{21} = \rho\sigma_1\sigma_2$.

Если ввести матрицу $\tilde{\Sigma}$, обратную матрице Σ , т.е.

$$\tilde{\Sigma} = \Sigma^{-1}$$

и вектор

$$\vec{y} = (y_1, y_2),$$

то квадратичную форму (12.2) можно записать в матричной форме в виде

$$Q(\vec{y}) = \vec{y} \tilde{\Sigma} \vec{y}^T, \quad (12.3)$$

где знак “ T ” означает транспонирование. Действительно, если учитывать, что

$$\tilde{\Sigma} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix},$$

то, подставляя $\tilde{\Sigma}$ в (12.3), приходим к выражению (12.2).

Далее, если заметить, что множитель

$$\sigma_1 \sigma_2 \sqrt{1 - \rho^2} = \sqrt{\det \Sigma},$$

где $\det \Sigma$ — определитель матрицы Σ , то выражение (12.1) можно записать в виде

$$p_{\vec{X}}(\vec{x}) = \frac{1}{(\sqrt{2\pi})^2 (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x} - \vec{m}) \tilde{\Sigma} (\vec{x} - \vec{m})^T}. \quad (12.4)$$

Теперь можно записать плотность (невырожденного) нормального распределения для *случайного вектора* $\vec{X} = (X_1; \dots; X_n)$ произвольной размерности $n > 2$.

Определение 12.1 Случайный вектор $\vec{X} = (X_1; \dots; X_n)$ назовем случайным вектором, имеющим нормальное распределение с вектором математических ожиданий $\vec{m} = (m_1; \dots; m_n)$ и ковариационной матрицей $\Sigma = (\sigma_{ij})$, $i, j = \overline{1, n}$, если его плотность имеет вид

$$p_{\vec{X}}(\vec{x}) = \frac{1}{(\sqrt{2\pi})^n (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x} - \vec{m}) \tilde{\Sigma} (\vec{x} - \vec{m})^T},$$

где $\tilde{\Sigma}$ — матрица, обратная к матрице Σ .

Если матрица Σ (а значит, и матрица $\tilde{\Sigma} = \Sigma^{-1}$) совпадает с единичной матрицей I , а вектор $\vec{m} = (0; \dots; 0)$, то

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)}.$$

Такую плотность по аналогии с одномерным случаем называют **плотностью стандартного многомерного (n -мерного) нормального распределения**.

Дадим геометрическую интерпретацию плотности нормального распределения.

Начнем с двумерного случая. При этом (X_1, X_2) будем трактовать как координаты брошенной случайным образом на плоскость точки.

Функция $p_{X_1, X_2}(x_1, x_2)$ задает некоторую поверхность в трехмерном пространстве. Линии уровня этой поверхности имеют уравнение $p_{X_1, X_2}(x_1, x_2) = a$, которое с учетом (12.1) можно записать в виде

$$Q(x_1 - m_1, x_2 - m_2) = b, \quad (12.5)$$

где $b = -2 \ln \{ 2\pi a (\det \Sigma)^{\frac{1}{2}} \}$, а $Q(x_1 - m_1, x_2 - m_2)$ определяется формулой (12.2).

Последнее уравнение (см. рис. 12.1) представляет собой уравнение эллипса (точнее говоря, семейства эллипсов при разных значениях b). Оси симметрии $O'x_1$ и $O'x_2$ этого эллипса проходят через точку $O'(m_1; m_2)$, а их направления совпадают с направлениями собственных векторов матрицы $\tilde{\Sigma}$. В свою очередь, собственные векторы \vec{e}_i , $i = 1, 2$, матрицы $\tilde{\Sigma}$ определяются из уравнений $\vec{e}_i \tilde{\Sigma} = \lambda_i \vec{e}_i$ где λ_i — собственные значения матрицы $\tilde{\Sigma}$, т.е. решения характеристического уравнения

$$\det(\tilde{\Sigma} - \lambda I) = 0,$$

или

$$\lambda^2 \sigma_1^2 \sigma_2^2 (1 - \rho^2) - \lambda(\sigma_1^2 + \sigma_2^2) + 1 = 0.$$

Углы α_i , $i = 1, 2$, между осями симметрии эллипса и осью Ox_1 можно найти из уравнения

$$\operatorname{tg} 2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}. \quad (12.6)$$

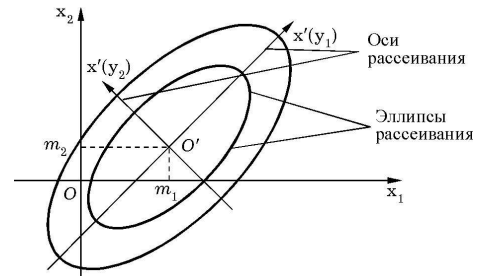


Рис. 12.1.

Это уравнение дает два значения углов: α_1 и α_2 различающиеся на $\frac{\pi}{2}$.

Оси симметрии эллипса (12.5) называют **осями рассеивания**, сам эллипс — **эллипсом рассеивания** (или **эллипсом равной вероятности**), а центр эллипса — точку $O'(m_1; m_2)$ — **центром рассеивания**.

Из формулы (12.6), в частности, следует, что при $\rho = 0$, $\sigma_1 \neq \sigma_2$ оси рассеивания параллельны координатным, при $\sigma_1 = \sigma_2 = \sigma$ эллипс рассеивания представляет собой окружность радиуса σ и в качестве осей рассеивания можно взять любые две перпендикулярные прямые, проходящие через точку O' .

Вводя новую (прямоугольную) систему координат, оси которой совпадают с осями рассеивания, т.е. каноническую систему координат для эллипса рассеивания, можно показать, что в этой системе координаты $(X'_1; X'_2)$ случайной точки имеют нормальное распределение с нулевым вектором средних значений $m' = (0; 0)$ и матрицей ковариаций $\Sigma' = \begin{pmatrix} \sigma_1'^2 & 0 \\ 0 & \sigma_2'^2 \end{pmatrix}$, где $\sigma_1' = \frac{1}{\sqrt{\lambda_1}}$, $\sigma_2' = \frac{1}{\sqrt{\lambda_2}}$.

Если изменить масштабы на осях канонической системы координат, взяв за единицы отсчета σ_1' и σ_2' соответственно, то в такой системе координат координаты случайной точки будут иметь нормальное распределение с нулевым вектором средних и единичной матрицей ковариаций, т.е. иметь двумерное стандартное нормальное распределение.

Аналогично в случае $n > 2$ уравнение $p_{\vec{X}}(\vec{x}) = a$ или эквивалентное ему уравнение

$$(\vec{x} - \vec{m})\tilde{\Sigma}(\vec{x} - \vec{m})^T = b, \quad b = -2\ln\{(\sqrt{2\pi})^n a (\det \Sigma)^{\frac{1}{2}}\}$$

в силу положительной определенности матрицы Σ представляет собой уравнение n -мерного эллипсоида, называемого **эллипсоидом рассеивания**, его оси симметрии по-прежнему называются **осями рассеивания**.

Будем трактовать n -мерный случайный вектор \vec{X} как координаты случайной точки в n -мерном пространстве. Пусть x'_1, \dots, x'_n — каноническая система координат эллипсоидов рассеивания, тогда новые координаты (X'_1, \dots, X'_n) случайной точки снова будут описываться n -мерным нормальным законом, имеющим нулевой вектор средних \vec{m}' и диагональную матрицу ковариаций Σ' , причем ее диагональные элементы $\sigma_i'^2 = 1/\lambda_i$, где λ_i , $i = \overline{1, n}$, — собственные значения матрицы $\tilde{\Sigma}$ с учетом их кратностей. Еще раз вводя новые координаты $y_i = \sigma_i' x'_i$ (т.е. изменяя масштабы на осях канонической системы координат), получаем, что в последней системе координат y_1, \dots, y_n координаты случайной точки будут распределены по стандартному нормальному закону.

Таким образом, делая обратные преобразования, можно трактовать (невырожденный) нормально распределенный вектор \vec{X} с произвольным вектором средних \vec{m} и матрицей ковариаций Σ как координаты случайной точки в некоторой (вообще говоря, не ортонормированной, но ортогональной) прямолинейной системе координат, причем эта точка имеет стандартное нормальное распределение.

Рассмотрим основные свойства многомерного нормального распределения.

1. **Закон распределения** каждой из координат случайного вектора \vec{X} , имеющего n -мерное нормальное распределение с вектором средних $\vec{m} = (m_1; \dots; m_n)$ и матрицей ковариаций $\Sigma = \sigma_{ij}$, является нормальным с параметрами m_i и σ_i .

Доказательство. Докажем это утверждение для случая $n = 2$ (общий случай требует более громоздких преобразований).

Найдем плотность распределения $p_{X_1}(x_1)$, если $p_{X_1, X_2}(x_1, x_2)$ определяется формулами (12.1) и (12.2). Воспользовавшись свойством 7 двумерной плотности распределения, имеем

$$p_{X_1}(x_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\tilde{Q}(x_1, x_2)} dx_2,$$

где

$$\tilde{Q}(x_1, x_2) = -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - m_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - m_1}{\sigma_1} \right) \left(\frac{x_2 - m_2}{\sigma_2} \right) + \left(\frac{x_2 - m_2}{\sigma_2} \right)^2 \right].$$

Делая замену

$$y = \frac{\frac{x_2 - m_2}{\sigma_2} - \rho \frac{x_1 - m_1}{\sigma_1}}{\sqrt{1 - \rho^2}},$$

после преобразований получаем

$$p_{X_1}(x_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1} e^{-\frac{y^2}{2} - \frac{(x_1 - m_1)^2}{2\sigma_1^2}} dy.$$

Поскольку

$$\int_{-\infty}^{+\infty} e^{-y^2/2} dy = \sqrt{2\pi},$$

приходим к окончательному ответу

$$p_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_1 - m_1)^2 / 2\sigma_1^2},$$

что и доказывает требуемое утверждение.

Аналогично можно показать, что

$$p_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2 - m_2)^2}{2\sigma_2^2}}.$$

2. Если ковариационная матрица Σ случайного вектора \vec{X} , распределенного по нормальному закону (невырожденному), является диагональной, то координаты вектора X_1, \dots, X_n являются независимыми случайными величинами.

ДОКАЗАТЕЛЬСТВО. Действительно, матрица $\tilde{\Sigma} = \Sigma^{-1}$ также является диагональной и имеет вид

$$\tilde{\Sigma} = \begin{pmatrix} \sigma_1^{-2} & 0 & \dots & 0 \\ 0 & \sigma_2^{-2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^{-2} \end{pmatrix}, \quad |\Sigma| = \sigma_1^2 \dots \sigma_n^2,$$

и, следовательно, формула (12.1) для совместной (n -мерной) плотности распределения имеет вид

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n \sigma_1 \dots \sigma_n} e^{-\left[\frac{(x_1 - m_1)^2}{2\sigma_1^2} + \dots + \frac{(x_n - m_n)^2}{2\sigma_n^2} \right]} = p_{X_1}(x_1) \dots p_{X_n}(x_n),$$

т.е. случайные величины X_1, \dots, X_n являются независимыми (см. замечание 8.1).

Заметим, что если $\sigma_{ij} = 0$ для некоторых i и j или, что то же самое, коэффициент корреляции $\rho_{ij} = 0$, то говорят, что **случайные величины** X_i и X_j являются **некоррелированными**.

Таким образом, из некоррелированности координат случайного вектора, распределенного по нормальному закону, следует (в силу теоремы 8.3) их независимость. Поскольку независимые случайные величины являются некоррелированными, то для нормально распределенных случайных векторов некоррелированность координат равносильна их независимости.

3. Если вектор $\vec{X} = (X_1; \dots; X_n)$ имеет нормальный закон распределения с вектором средних $\vec{m} = (m_1; \dots; m_n)$ и матрицей ковариаций Σ , то вектор $\vec{X}' = (X_1; \dots; X_{n-1})$ также распределен по нормальному закону с вектором средних $\vec{m}' = (m_1; \dots; m_{n-1})$ и матрицей ковариаций Σ' , полученной из матрицы Σ вычеркиванием последних строки и столбца.

ДОКАЗАТЕЛЬСТВО. Это свойство доказывается так же, как и свойство 1, но в силу громоздкости вывода оно здесь не приводится.

Из свойства 3 методом математической индукции можно показать, что любой набор координат n -мерного случайного вектора $\vec{X} = (X_1; \dots; X_n)$, распределенного по нормальному закону, снова имеет нормальное распределение. В частности, двумерный случайный вектор (X_1, X_2) распределен по нормальному закону с вектором средних (m_1, m_2) и матрицей ковариаций $\Sigma' = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$.

Пример 12.1 Пусть двумерный случайный вектор $(X; Y)$ имеет нормальное распределение с вектором средних значений $(m_1; m_2)$ и матрицей ковариаций

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad (\sigma_1, \sigma_2 > 0, \quad -1 < \rho < 1).$$

Найдем условную плотность распределения случайной величины X при условии $Y = y$.

Как известно (см. (12.1)–(12.2)), совместная двумерная плотность распределения случайных величин X и Y

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{(x-m_1)^2}{\sigma_1^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2} \right) \right\},$$

а плотность распределения случайной величины Y

$$p_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-m_2)^2/(2\sigma_2^2)}.$$

Значит,

$$p_X(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}} \exp \left\{ -\frac{1}{2\sigma_1^2(1-\rho^2)} \left[x - \left(m_1 + \frac{\rho\sigma_1(y-m_2)}{\sigma_2} \right) \right]^2 \right\}.$$

Таким образом, условное распределение X при условии $Y = y$ также является нормальным со *средним значением* (которое обозначим $g(y)$)

$$g(y) = m_1 + \rho \frac{\sigma_1}{\sigma_2}(y - m_2) \quad (12.7)$$

и *средним квадратичным отклонением* (обозначим его $\sigma_{X|y}$)

$$\sigma_{X|y} = \sigma_1 \sqrt{1 - \rho^2}. \quad (12.8)$$

Аналогично условное распределение Y при условии $X = x$ является нормальным со *средним значением* (которое обозначим $h(x)$)

$$h(x) = m_2 + \rho \frac{\sigma_2}{\sigma_1}(x - m_1) \quad (12.9)$$

и *средним квадратичным отклонением* (обозначим его $\sigma_{Y|x}$)

$$\sigma_{Y|x} = \sigma_2 \sqrt{1 - \rho^2}. \quad (12.10)$$

Пример 12.2 Известно, что рост X_1 и вес X_2 взрослого мужчины (и женщины), проживающего в одном регионе, достаточно хорошо описывается *двумерным нормальным законом распределения*. В частности, рост (в сантиметрах) и вес (в килограммах) мужчин некоторой страны Нормалии подчинены нормальному закону с *вектором средних значений* $\vec{m} = (172, 74)$ и *матрицей ковариаций* $\Sigma = \begin{pmatrix} 45 & 28 \\ 28 & 40 \end{pmatrix}$.

Пусть известно, что вес случайно встреченного нормальца равен x_2 . Тогда его рост будет иметь нормальное распределение со *средним значением* (в см)

$$g(x_2) = 172 + \frac{0,66\sqrt{45}(x_2 - 74)}{\sqrt{40}} \approx 120 + 0,70x_2$$

и *средним квадратичным отклонением*

$$\sigma_{X_1|x_2} = \sqrt{45}\sqrt{1 - 0,66^2} \approx 5,0.$$

Таким образом,

$$p_{X_1}(x_1|x_2) = \frac{1}{\sqrt{50\pi}} e^{-(x_1 - 120 - 0,7x_2)^2/50}.$$

В частности, весу 70 кг соответствует среднее значение роста 169 см, весу 75 кг — около 173 см и т.д.

Отметим, что в отличие от среднего роста $g(x_2)$, зависящего линейно от x_2 , среднее квадратичное отклонение роста $\sigma_{X_1|x_2}$ является постоянным, т.е. не зависит от x_2 . Графическое изображение зависимости роста от веса приведено на рис. 12.2. Здесь по оси абсцисс отложены значения роста нормальца, а по оси ординат — его веса. Прямая линия $x_1 = g(x_2)$ показывает зависимость среднего роста от веса. Условная плотность распределения $p_{X_1}(x_1|x_2)$ роста, как функции от веса x_2 , изображена в виде “срезов”.

Аналогичные вычисления показывают, что условная плотность распределения $p_{X_2}(x_2|x_1)$ веса нормальца X_2 в зависимости от его роста x_1 является плотностью нормального распределения с параметрами $h(x_1) \approx 0,62x_1 - 33$ и $\sigma_{X_2|x_1} \approx 4,8$, т.е. имеет вид

$$p_{X_2}(x_2|x_1) = \frac{1}{\sqrt{46\pi}} e^{-(x_2 + 33 - 0,62x_1)^2/46}.$$

Графическое изображение зависимости веса от роста приведено на рис. 12.3.

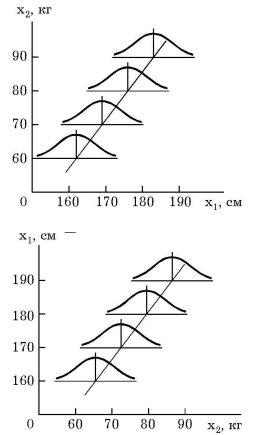


Рис. 12.3.

Лекция 13

Предельные теоремы теории вероятностей

С самого начала изучения курса теории вероятностей мы говорили о том, что практическое применение методов этой математической дисциплины основывается на законе предельного постоянства частоты события, установленном эмпирически. Согласно этому закону, если один и тот же опыт повторяется многократно, то частота появления конкретного случайного события теряет свойства случайности и приближается к некоторому пределу, который в соответствии со *статистическим определением вероятности* (см. лекцию 2) и называют *вероятностью*.

Однако для того чтобы теория согласовывалась с практикой, при *аксиоматическом определении вероятности*, которое мы использовали, этот закон предельного постоянства частоты должен быть обоснован теоретически. Иначе говоря, он должен быть сформулирован и доказан в виде одной или нескольких теорем. В теории вероятностей теоремы такого типа обычно называют различными формами *закона больших чисел*. В этой лекции мы докажем некоторые формы этого закона, которые, в частности, поясняют смысл *математического ожидания случайной величины*, и то, почему его называют также *средним значением*.

Далее доказывается простейший вариант *центральной предельной теоремы*, уточняющей закон больших чисел. Центральная предельная теорема, в свою очередь, объясняет то широкое распространение, которое получило на практике *нормальное распределение*.

Определение 13.1 Если последовательность $X_1, X_2, \dots, X_n, \dots$ случайных величин для любого $\varepsilon > 0$ удовлетворяет условию

$$\lim_{n \rightarrow \infty} P\{|X_n - X| < \varepsilon\} = 1,$$

то говорят о *сходимости* этой последовательности к случайной величине X *по вероятности*. Сходимость к X по вероятности записывается в виде

$$X_n \xrightarrow[n \rightarrow \infty]{P} X.$$

Определение 13.2 Если последовательность $X, X_1, X_2, \dots, X_n, \dots$ случайных величин удовлетворяет условию

$$P\{\lim_{n \rightarrow \infty} X_n = X\} = 1,$$

то говорят о *сходимости* этой последовательности к X *с вероятностью 1 или почти наверное* и обозначают

$$X_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} X.$$

Неравенства Чебышева. Закон больших чисел

Прежде чем приступить к рассмотрению *закона больших чисел*, докажем два *неравенства Чебышева*. Заметим, что неравенства Чебышева представляют и самостоятельный интерес, поскольку в современной теории вероятностей широко используются неравенства такого типа.

Теорема 13.1 Для каждой неотрицательной случайной величины X , имеющей математическое ожидание MX , при любом $\varepsilon > 0$ справедливо соотношение

$$P\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon},$$

называемое *первым неравенством Чебышева*.

ДОКАЗАТЕЛЬСТВО. Доказательство проведем для *непрерывной случайной величины* X с *плотностью распределения* $p(x)$. Поскольку случайная величина X является неотрицательной, то

$$\mathbf{M}X = \int_0^{+\infty} xp(x) dx.$$

Так как подынтегральное выражение неотрицательное, то при уменьшении области интегрирования интеграл может только уменьшиться. Поэтому

$$\mathbf{M}X = \int_0^{\varepsilon} xp(x) dx + \int_{\varepsilon}^{+\infty} xp(x) dx \geq \int_{\varepsilon}^{+\infty} xp(x) dx.$$

Заменяя в подынтегральном выражении множитель x на ε , имеем

$$\int_{\varepsilon}^{+\infty} xp(x) dx \geq \varepsilon \int_{\varepsilon}^{+\infty} p(x) dx.$$

Остается заметить, что последний интеграл (равный площади области, заштрихованной на 13.1) представляет собой вероятность события $X \geq \varepsilon$, и, значит, $\mathbf{M}X \geq \varepsilon \mathbf{P}\{X \geq \varepsilon\}$, откуда и вытекает первое неравенство Чебышева. Аналогично первое неравенство Чебышева доказывается и для дискретной случайной величины, при этом нужно только заменить интеграл суммой.

Ясно, что применять первое неравенство Чебышева имеет смысл только тогда, когда $\varepsilon > \mathbf{M}X$; в противном случае оно дает тривиальную оценку.

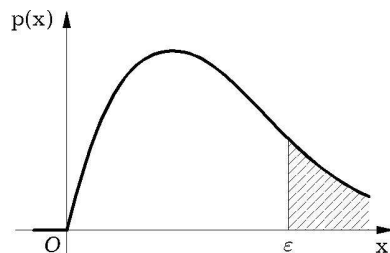


Рис. 13.1.

Пример 13.1 Пусть X — время опоздания студента на лекцию, причем известно, что $\mathbf{M}X = 1$ мин. Воспользовавшись первым неравенством Чебышева, оценим вероятность $\mathbf{P}\{X \geq 5\}$ того, что студент опоздает не менее, чем на 5 мин.

Имеем

$$\mathbf{P}\{X \geq 5\} \leq \frac{\mathbf{M}X}{5} = 0,2.$$

Таким образом, искомая вероятность не более 0,2, т.е. в среднем из каждых пяти студентов опаздывает, по крайней мере, на 5 мин не более чем один студент. #

Рассмотрим теперь случайную величину X , имеющую *дисперсию* $\mathbf{D}X = \sigma^2$. Мы уже говорили, что дисперсия является показателем разброса X вокруг математического ожидания $\mathbf{M}X$. Однако с точки зрения исследователя разброс естественнее характеризовать вероятностью $\mathbf{P}\{|X - \mathbf{M}X| \geq \varepsilon\}$ отклонения случайной величины X от $\mathbf{M}X$ на величину, большую некоторого заданного ε . Следующее неравенство позволяет оценить эту вероятность с помощью дисперсии σ^2 .

Теорема 13.2 Для каждой случайной величины X , имеющей дисперсию $\mathbf{D}X = \sigma^2$, при любом $\varepsilon > 0$ справедливо **второе неравенство Чебышева**

$$\mathbf{P}\{|X - \mathbf{M}X| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

ДОКАЗАТЕЛЬСТВО. Для доказательства воспользуемся утверждением первого неравенства Чебышева. Применяя к случайной величине $Y = (X - \mathbf{M}X)^2$ это неравенство, в котором ε заменено на ε^2 , получаем

$$\mathbf{P}\{|X - \mathbf{M}X| \geq \varepsilon\} = \mathbf{P}\{(X - \mathbf{M}X)^2 \geq \varepsilon^2\} = \mathbf{P}\{Y \geq \varepsilon^2\} \leq \frac{\mathbf{M}Y}{\varepsilon^2} = \frac{\mathbf{D}X}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2},$$

что и доказывает второе неравенство Чебышева.

Геометрический смысл второго неравенства Чебышева понятен из рис. 13.2.

Второе неравенство Чебышева имеет содержательный смысл лишь при $\varepsilon > \sigma$.

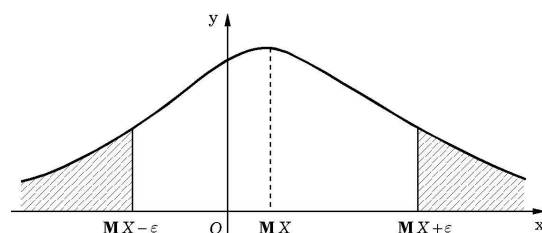


Рис. 13.2.

Пример 13.2 Пусть в условиях предыдущего примера известно дополнительно, что $\sigma = \sqrt{DX} = 1$. Оценим минимальное значение x_0 , при котором вероятность опоздания студента на время не менее x_0 не превышает заданного значения $P_3 = 0,1$.

Для решения поставленной задачи воспользуемся вторым неравенством Чебышева. Тогда

$$P_3 \leq \mathbf{P}\{X \geq x_0\} = \mathbf{P}\{X - \mathbf{M}X \geq x_0 - \mathbf{M}X\} \leq \mathbf{P}\{|X - \mathbf{M}X| \geq x_0 - \mathbf{M}X\} \leq \frac{\sigma^2}{(x_0 - \mathbf{M}X)^2}.$$

Значит, $(x_0 - \mathbf{M}X)^2 \leq \frac{\sigma^2}{P_3}$ и $x_0 \leq \mathbf{M}X + \sqrt{\frac{\sigma^2}{P_3}}$. Подставляя конкретные значения, имеем $x_0 \leq 1 + \sqrt{\frac{1}{0,1}} \approx 4,16$. Таким образом, вероятность опоздания студента на время более 4,16 мин не более 0,1.

Сравнивая полученный результат с результатом примера 13.1, видим, что дополнительная информация о дисперсии времени опоздания позволяет дать более точную оценку искомой вероятности.

Рассмотрим некоторые формы закона больших чисел.

Пусть $X_1, X_2, \dots, X_n, \dots$ — последовательность случайных величин, имеющих математические ожидания $m_i = \mathbf{M}X_i$.

Определение 13.3 Последовательность $X_1, X_2, \dots, X_n, \dots$ случайных величин удовлетворяет *закону больших чисел (слабому)*, если для любого $\varepsilon > 0$

$$\mathbf{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n m_i\right| \geq \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 0,$$

т.е.

$$\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n m_i \xrightarrow[n \rightarrow \infty]{\mathbf{P}} 0.$$

Иными словами, выполнение закона больших чисел отражает предельную устойчивость средних арифметических случайных величин: при большом числе испытаний они практически перестают быть случайными и совпадают со своими средними значениями.

Очевидно, что последовательность $X_1, X_2, \dots, X_n, \dots$ удовлетворяет закону больших чисел тогда и только тогда, когда среднее арифметическое случайных величин $X_1 - m_1, X_2 - m_2, \dots, X_n - m_n$ сходится по вероятности к нулю при $n \rightarrow \infty$.

Теорема 13.3 Если последовательность $X_1, X_2, \dots, X_n, \dots$ независимых случайных величин такова, что существуют $\mathbf{M}X_i = m_i$ и $DX_i = \sigma_i^2$, причем дисперсии σ_i^2 ограничены в совокупности (т.е. $\sigma_i^2 \leq C < +\infty$), то для последовательности $X_1, X_2, \dots, X_n, \dots$ выполнен закон больших чисел.

При этом говорят также, что к последовательности $X_1, X_2, \dots, X_n, \dots$ случайных величин применим *закон больших чисел в форме Чебышева*.

ДОКАЗАТЕЛЬСТВО. Теорема является элементарным следствием второго неравенства Чебышева. Действительно, в силу свойств математического ожидания и дисперсии

$$\mathbf{M}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n m_i, \quad \mathbf{D}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \leq \frac{Cn}{n^2} = \frac{C}{n}.$$

Применяя теперь второе неравенство Чебышева к случайным величинам $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$, получаем для любого $\varepsilon > 0$

$$\mathbf{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n m_i\right| \geq \varepsilon\right\} \leq \frac{C}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Таким образом, мы показали, что для последовательности $X_1, X_2, \dots, X_n, \dots$ выполняется закон больших чисел.

Следствие 13.1 Если случайные величины $X_i, i = 1, 2, \dots$, в условиях теоремы 13.3 являются также одинаково распределенными (в этом случае $m_i = m$ и $\sigma_i^2 = \sigma^2$), то последовательность $X_1, X_2, \dots, X_n, \dots$ случайных величин удовлетворяет закону больших чисел в следующей форме:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\mathbf{P}} m.$$

ДОКАЗАТЕЛЬСТВО. Доказательство проведите самостоятельно.

Теорема 13.4 Пусть $X_1, X_2, \dots, X_n, \dots$ последовательность независимых одинаково распределенных случайных величин с $\mathbf{M}X_i = m$ и $\mathbf{M}|X_i| < \infty$. Тогда

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{n.n.} m.$$

Это утверждение называют усиленным законом больших чисел в форме Колмогорова.

Теорема 13.5 Пусть проводится n испытаний по схеме Бернулли и Y_n — общее число успехов в n испытаниях. Тогда наблюдаемая частота успехов

$$r_n = \frac{Y_n}{n}$$

сходится по вероятности к вероятности p успеха в одном испытании

$$r_n \xrightarrow[n \rightarrow \infty]{\mathbf{P}} p.$$

Доказательство. Обозначим X_i число успехов в i -м испытании Бернулли. Тогда частоту успехов в n испытаниях можно определить в виде $r_n = \frac{1}{n} \sum_{i=1}^n X_i$, причем $\mathbf{M}X_i = p$ и $\mathbf{D}X_i = pq$. Значит, выполняются все условия следствия 13.1, из которого вытекает утверждение теоремы.

Теорему 13.5 называют также *теоремой Бернулли*, или *законом больших чисел в форме Бернулли*. Из хода доказательства теоремы 13.5 видно, что закон больших чисел в форме Бернулли является частным случаем закона больших чисел в форме Чебышева.

Пример 13.3 Пусть дана последовательность $X_1, X_2, \dots, X_n, \dots$ независимых случайных величин, причем ряд распределения случайной величины X_n представлен в табл. 13.1. Покажем, что к этой последовательности применим закон больших чисел в форме Чебышева. Для этого вычислим дисперсию $\mathbf{D}X_n$. Имеем

X_n	$-5n$	0	$5n$
\mathbf{P}	$\frac{1}{2n^2}$	$1 - \frac{1}{n^2}$	$\frac{1}{2n^2}$

Таблица 13.1.

$$\mathbf{M}X_n = (-5n) \cdot \frac{1}{2n^2} + 0 \cdot \left(1 - \frac{1}{n^2}\right) + 5n \cdot \frac{1}{2n^2} = 0,$$

$$\mathbf{D}X_n = \mathbf{M}X_n^2 - (\mathbf{M}X_n)^2 = \mathbf{M}X_n^2 = (-5n)^2 \frac{1}{2n^2} + 0^2 \left(1 - \frac{1}{n^2}\right) + (5n)^2 \frac{1}{2n^2} = 25.$$

Итак, дисперсии $\mathbf{D}X_n$ ограничены в совокупности, и к последовательности $X_1, X_2, \dots, X_n, \dots$ применим закон больших чисел в форме Чебышева.

Центральная предельная теорема

Рассмотрим последовательность $X_1, X_2, \dots, X_n, \dots$ независимых одинаково распределенных случайных величин, имеющих математическое ожидание $\mathbf{M}X_n = m$. Предположим также, что существует дисперсия $\mathbf{D}X_n = \sigma^2$. Закон больших чисел (слабый) для этой последовательности можно представить в следующей форме:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mathbf{M}X_i) = \frac{1}{n} (S_n - nm) \xrightarrow[n \rightarrow \infty]{\mathbf{P}} 0,$$

где

$$S_n = \sum_{i=1}^n X_i \quad -$$

суммарное значение первых n случайных величин последовательности, а сходимость можно понимать как в смысле *сходимости по вероятности* (слабый закон больших чисел), так и в смысле сходимости с вероятностью 1 (усиленный закон больших чисел).

Однако сразу возникает вопрос: поскольку случайные величины X_n имеют не только математическое ожидание, но и дисперсию, то нельзя ли доказать более “тонкую” предельную теорему, позволяющую точнее описать предельное поведение распределений величин $S_n - nm$? Такая теорема существует, ее называют центральной предельной теоремой.

Теорема 13.6 (центральная предельная теорема) Пусть $X_1, X_2, \dots, X_n, \dots$ — последовательность независимых одинаково распределенных случайных величин, $\mathbf{M}X_n = m$, $\mathbf{D}X_n = \sigma^2$. Тогда

$$\mathbf{P} \left\{ \frac{S_n - nm}{\sqrt{n\sigma^2}} < x \right\} \xrightarrow[n \rightarrow \infty]{} \Phi(x),$$

где $\Phi(x)$ — функция стандартного нормального распределения.

Центральная предельная теорема выявляет ту особую роль, которую играет нормальное распределение на практике. Нормальный закон всегда имеет место в тех ситуациях, когда случайная величина порождена большим количеством случайных факторов, действующих независимо друг от друга. Уже само название “нормальный закон” объясняется тем широким распространением, которое он находит в самых различных областях научных исследований.

Следствием из центральной предельной теоремы является интегральная теорема Муавра — Лапласа.

Теорема 13.7 (интегральная теорема Муавра — Лапласа) Обозначим S_n суммарное число успехов в n испытаниях по схеме Бернулли с вероятностью успеха p и вероятностью неудачи $q = 1 - p$. Тогда с ростом n последовательность функций распределения случайных величин $(S_n - np)/\sqrt{npq}$ сходится к функции стандартного нормального распределения, т.е.

$$\mathbf{P}\left\{\frac{S_n - np}{\sqrt{npq}} < x\right\} \xrightarrow{n \rightarrow \infty} \Phi(x).$$

ДОКАЗАТЕЛЬСТВО. Пусть X_i — число успехов в i -м испытании. Тогда $\mathbf{M}X_i = p$, $\mathbf{D}X_i = pq$. Представляя S_n в виде $S_n = X_1 + \dots + X_n$ и используя центральную предельную теорему, приходим к утверждению теоремы.

Пример 13.4 Для определения скорости v движения объекта делают n измерений v_1, \dots, v_n , причем i -е измерение проводят с погрешностью X_i (т.е. $v_i = v + X_i$), при этом погрешности измерений являются независимыми и одинаково распределенными случайными величинами с математическим ожиданием $\mathbf{M}X_i = 0$ (отсутствуют систематические погрешности наблюдений) и дисперсии $\mathbf{D}X_i = \sigma^2$.

Оценим вероятность того, что средняя наблюдаемая скорость

$$v_{\text{ср}} = \frac{v_1 + \dots + v_n}{n}$$

будет отличаться от истинной скорости v не более чем на ε . Имеем

$$\mathbf{P}\{|v_{\text{ср}} - v| < \varepsilon\} = \mathbf{P}\{-\varepsilon < \frac{v_1 + \dots + v_n - nv}{n} < \varepsilon\} = \mathbf{P}\left\{-\varepsilon\sqrt{\frac{n}{\sigma^2}} < \frac{v_1 + \dots + v_n - nv}{\sqrt{n\sigma^2}} < \varepsilon\sqrt{\frac{n}{\sigma^2}}\right\}.$$

Считая теперь, что число n измерений велико, воспользуемся центральной предельной теоремой, согласно которой случайная величина

$$\frac{v_1 + \dots + v_n - nv}{\sqrt{n\sigma^2}}$$

приблизительно распределена по стандартному нормальному закону. Значит,

$$\mathbf{P}\{|v_{\text{ср}} - v| < \varepsilon\} \approx \Phi\left(\varepsilon\sqrt{\frac{n}{\sigma^2}}\right) - \Phi\left(-\varepsilon\sqrt{\frac{n}{\sigma^2}}\right) = 2\Phi_0\left(\varepsilon\sqrt{\frac{n}{\sigma^2}}\right).$$

Лекция 14

Основные понятия выборочной теории

Определение 14.1 В математической статистике множество возможных значений случайной величины X называют **генеральной совокупностью** случайной величины X или просто генеральной совокупностью X .

Под **законом распределения** (**распределением**) **генеральной совокупности** X будем понимать закон распределения вероятностей случайной величины X .

Исходным материалом для изучения свойств генеральной совокупности (т.е. некоторой случайной величины) являются **экспериментальные (статистические) данные**, под которыми понимают значения случайной величины, полученные в результате повторений случайного эксперимента (наблюдений над случайной величиной).

Предполагаем, что эксперимент хотя бы теоретически может быть повторен сколько угодно раз в одних и тех же условиях. Под словами “в одних и тех же условиях” будем понимать, что распределение случайной величины X_i , $i = 1, 2, \dots$, заданной на множестве исходов i -го эксперимента, не зависит от номера испытания и совпадает с распределением генеральной совокупности X . В этом случае принято говорить о **независимых повторных экспериментах (испытаниях)** или о **независимых повторных наблюдениях** над случайной величиной.

Определение 14.2 Совокупность независимых случайных величин X_1, \dots, X_n , каждая из которых имеет то же распределение, что и случайная величина X , будем называть **случайной выборкой** из генеральной совокупности X и записывать $\vec{X}_n = (X_1; \dots; X_n)$ (иногда просто X_1, \dots, X_n). При этом число n называют **объемом случайной выборки**, а случайные величины X_i — **элементами случайной выборки**.

Часто используется и другая терминология. Так, если $F(x)$ (или $p(x)$) — функция распределения (или плотность) случайной величины X , говорят, что \vec{X}_n — случайная выборка из $F(x)$ (или из $p(x)$), или что X — случайная выборка из распределения X . Если X — нормальная случайная величина с $MX = \mu$, $DX = \sigma^2$, то будем писать, что \vec{X}_n — случайная выборка из $\mathcal{N}(\mu, \sigma^2)$.

Определение 14.3 Любое возможное значение $\vec{x}_n = (x_1; \dots; x_n)$ случайной выборки \vec{X}_n будем называть **выборкой** из генеральной совокупности X (также **реализацией случайной выборки** \vec{X}_n). Число n характеризует **объем выборки**, а числа x_i , $i = \overline{1, n}$, представляют собой **элементы выборки** \vec{x}_n .

Выборку \vec{x}_n можно интерпретировать как совокупность n чисел x_1, \dots, x_n , полученных в результате проведения n повторных независимых наблюдений над случайной величиной X .

Для краткости случайную выборку часто называют просто выборкой, если это не приводит к путанице.

Множество возможных значений $\mathcal{X}_n \in \mathbb{R}^n$ случайной выборки \vec{X}_n называют **выборочным пространством**. Любую функцию $\varphi(\vec{X}_n)$ от случайной выборки \vec{X}_n будем называть **статистикой**.

Одним из самых простых преобразований статистических данных является их упорядочивание по величине. Пусть $(x_1; \dots; x_n)$ — выборка объема n из генеральной совокупности X . Ее можно упорядочить, расположив значения в неубывающем порядке:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(n)}, \quad (14.1)$$

где $x_{(1)}$ — наименьший, $x_{(n)}$ — наибольший из элементов выборки.

Определение 14.4 Последовательность чисел

$$x_{(1)}, \quad x_{(2)}, \quad \dots, x_{(i)}, \quad \dots, \quad x_{(n)},$$

удовлетворяющих условию (14.1), называют **вариационным рядом выборки**, или, для краткости, просто **вариационным рядом**; число $x_{(i)}$, $i = \overline{1, n}$, называют i -м **членом вариационного ряда**.

Обозначим $X_{(i)}$, $i = \overline{1, n}$, случайную величину, которая при каждой *реализации случайной выборки* \vec{X}_n принимает значение, равное i -му члену вариационного ряда.

Определение 14.5 Последовательность случайных величин

$$X_{(1)}, \quad X_{(2)}, \quad \dots, \quad X_{(i)}, \quad \dots, \quad X_{(n)}$$

называют **вариационным рядом случайной выборки**. При этом $X_{(i)}$, $i = \overline{1, n}$, называют i -м **членом вариационного ряда случайной выборки**.

Среди элементов выборки x_1, \dots, x_n (а значит и среди членов вариационного ряда $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$) могут быть одинаковые. Так бывает либо когда наблюдаемая случайная величина X — дискретная, либо когда X — непрерывная, но ее значения при измерениях округляют.

Пусть среди элементов выборки x_1, \dots, x_n выделены $r < n$ их различных значений, расположенных в порядке возрастания. Обозначим их $z_{(1)}, \dots, z_{(r)}$. Предположим, что, каждое из них повторяется соответственно n_1, \dots, n_r раз, причем, разумеется $\sum_{k=1}^r n_k = n$.

Определение 14.6 **Статистическим рядом** для выборки называют таблицу 14.1, где в первой строке расположены элементы $z_{(1)}, \dots, z_{(r)}$ (напомним, что $z_{(1)} < \dots < z_{(r)}$), а во второй — числа их повторений. Число n_k , $k = \overline{1, r}$, показывающее, сколько раз встречался элемент $z_{(k)}$ в выборке, называют **частотой**, а отношение n_k/n — **относительной частотой** этого значения.

$z_{(1)}$	$z_{(2)}$	\dots	$z_{(r)}$
n_1	n_2	\dots	n_r

Таблица 14.1.

Пример 14.1 В течение суток измеряют напряжение X тока в электросети (в вольтах). В результате опыта получена выборка объема $n = 30$:

$z_{(k)}$	216	217	218	219	220	221	222	223
n_k	1	3	4	6	9	4	2	1

217, 218, 220, 219, 220, 221, 219, 220, 221, 217, 218, 219, 220, 217, 220, 219, 221, 221, 220, 219, 222, 223, 220, 216, 220, 219, 220, 218, 222. Построим статистический ряд для данной выборки. Наименьшее значение в выборке $x_{(1)} = 106$, наибольшее — $x_{(8)} = 223$. Подсчитываем частоту n_k , $k = \overline{1, 8}$, каждого из восьми различных значений в выборке и строим таблицу 14.2. #

Таблица 14.2.

Рассмотрим функцию $n(x, \vec{X}_n)$, которая для каждого значения $x \in \mathbb{R}$ и каждой реализации \vec{x}_n случайной выборки \vec{X}_n принимает значение $n(x, \vec{x}_n)$, равное числу элементов в выборке \vec{x}_n , меньших x .

Определение 14.7 Функцию

$$\hat{F}(x, \vec{X}_n) = \frac{n(x, \vec{X}_n)}{n}, \quad x \in \mathbb{R}, \quad (14.2)$$

где n — объем случайной выборки, будем называть **выборочной функцией распределения**.

Согласно определению 14.7, при любом фиксированном x функция $\hat{F}(x; \vec{X}_n)$ есть случайная величина, которая принимает одно из значений $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, \frac{n}{n} = 1$ и имеет биномиальное распределение с параметром p , равным значению функции распределения генеральной совокупности X в точке x , т.е. $p = F(x)$.

Теорема 14.1 Для любого фиксированного x последовательность случайных величин $\{\hat{F}(x; \vec{X}_n)\}$ сходится по вероятности при $n \rightarrow \infty$ к значению $F(x)$ функции распределения генеральной совокупности X в точке x .

Доказательство. При любом фиксированном x выборочная функция распределения $\hat{F}(x; \vec{X}_n)$ есть относительная частота события $\{X < x\}$. В соответствии с законом больших чисел в форме Бернулли, относительная частота при $n \rightarrow \infty$ сходится по вероятности к вероятности события $\{X < x\}$. Следовательно,

$$\hat{F}(x; \vec{X}_n) \xrightarrow[n \rightarrow \infty]{P} P\{X < x\} = F(x).$$

Если все выборочные значения x_1, \dots, x_n различны, то функцию $\hat{F}(x, \vec{x})$ можно записать в виде

$$F_n(x) = \begin{cases} 0, & x \leq x_{(1)}; \\ \frac{i}{n}, & x_{(i)} < x \leq x_{(i+1)}, \quad i = \overline{1, n-1}; \\ 1, & x > x_{(n)}, \end{cases}$$

т.е. в каждой точке $x_{(i)}$ функция $F_n(x)$ имеет скачок величиной $1/n$. График функции $F_n(x)$ изображен на рис. 14.1.

При больших объемах выборки n ($n > 50$) обычно производят группирование исходных данных следующим образом. Промежуток $J = [x_{(1)}, x_{(n)}]$, содержащий все выборочные значения, разбивают на m полуинтервалов J_1, \dots, J_m , как правило, одинаковой длины Δ и таких, что каждый из них, кроме последнего, содержит левую границу, а последний содержит обе границы, и подсчитывают число n_i элементов выборки, попавших в i -ый промежуток J_i , $i = \overline{1, m}$, $n = n_1 + \dots + n_m$, а результаты представляют в виде следующей таблицы 14.2, которую называют **интервальным статистическим рядом**.

Иногда в верхней строке таблицы 14.3 указывают не интервал, а его середину \tilde{x}_i , а в нижней строке вместо частоты n_i записывают относительную частоту n_i/n .

Число промежутков m , на которые разбивают промежуток $J = [x_{(1)}, x_{(n)}]$, содержащий все выборочные значения, выбирают в зависимости от объема выборки n . Для ориентировочной оценки величины m можно пользоваться следующей приближенной формулой.

$$m \approx \log_2 n + 1,$$

Определение 14.8 График функции

$$p_n(x) = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i; \\ 0, & x \notin J, \end{cases} \quad (14.3)$$

представляющий собой кусочно постоянную функцию называют **гистограммой** (см. рис. 14.2).

Часто гистограммой называют диаграмму, составленную из прямоугольников с основанием Δ и высотами $n_i/(n\Delta)$, $i = \overline{1, m}$. Нетрудно увидеть, что суммарная площадь всех прямоугольников, образующих такую диаграмму, равна 1, так как $\sum_{i=1}^m \frac{n_i}{n\Delta} \Delta = \frac{1}{n} \sum_{i=1}^m n_i = 1$. Площадь каждого прямоугольника n_i/n есть частота попадания элементов выборки в соответствующий интервал J_i статистического ряда.

Рассмотрим случайную величину $n_i(\vec{X}_n)/n$, которая для каждой реализации \vec{x}_n случайной выборки \vec{X}_n равна частоте n_i/n . В соответствии с законом больших чисел в форме Бернулли $n_i(\vec{X}_n)/n$ при $n \rightarrow \infty$ будет сходиться по вероятности к вероятности попадания случайной величины X в промежуток J_i , $i = \overline{1, m}$, т.е.

$$\frac{n_i(\vec{X}_n)}{n} \xrightarrow[n \rightarrow \infty]{P} P\{X \in J_i\} = \int_{J_i} p(x) dx,$$

где $p(x)$ — плотность распределения генеральной совокупности X . Если длина Δ промежутков достаточно мала и объем выборки n велик, то с вероятностью, близкой к 1, можно утверждать, что

$$\frac{n_i}{n} \approx p(\tilde{x}_i)\Delta,$$

или

$$\frac{n_i}{n\Delta} \approx p(\tilde{x}_i),$$

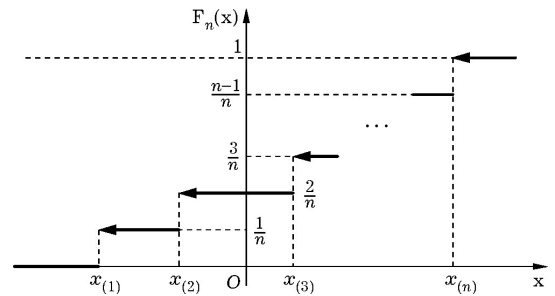


Рис. 14.1.

J_1	J_2	\dots	J_m	
n_1	n_2	\dots	n_m	$\sum_{i=1}^m n_i$

Таблица 14.3.

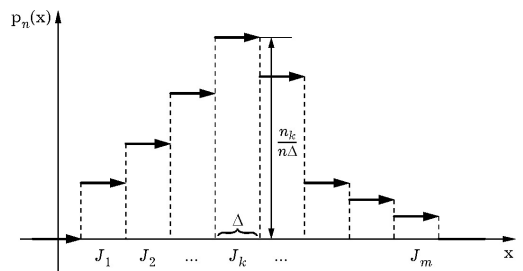


Рис. 14.2.

где \tilde{x}_i — середина промежутка J_i , $i = \overline{1, m}$. Таким образом, при большом объеме выборки n и достаточно малом Δ с вероятностью, близкой к 1, можно считать, что $p_n(x) \approx p(x)$. Иными словами, функция $p_n(x)$ является статистическим аналогом плотности распределения $p(x)$, наблюдаемой в эксперименте случайной величины X , а гистограмма выглядит приблизительно как график плотности случайной величины X .

Пусть \vec{X}_n — случайная выборка из генеральной совокупности X с функцией распределения $F(x)$ (и плотностью распределения $p(x)$ в случае непрерывной статистической модели).

Случайную величину

$$\hat{\mu}_k(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (14.4)$$

называют **выборочным начальным моментом k -го порядка**. В частности, выборочный начальный момент первого порядка $\bar{X} = \hat{\mu}_1(\vec{X}_n)$ называют **выборочным средним**.

Случайную величину

$$\hat{\nu}_k(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad (14.5)$$

называют **выборочным центральным моментом k -го порядка**. В частности, выборочный центральный момент 2-го порядка $\hat{\sigma}^2(\vec{X}_n) = \hat{\nu}_2(\vec{X}_n)$ называют **выборочной дисперсией**.

Выборочную характеристику $\hat{\sigma}(\vec{X}_n) = \sqrt{\hat{\sigma}^2(\vec{X}_n)}$ называют **выборочным средним квадратичным отклонением**.

Случайную величину

$$\hat{K}(\vec{X}_n, \vec{Y}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (14.6)$$

называют **выборочным корреляционным моментом**.

Выборочную характеристику

$$\hat{\rho}(\vec{X}_n, \vec{Y}_n) = \frac{\hat{K}(\vec{X}_n, \vec{Y}_n)}{\hat{\sigma}_x(\vec{X}_n) \hat{\sigma}_y(\vec{Y}_n)}, \quad (14.7)$$

где

$$\hat{\sigma}_x^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \hat{\sigma}_y^2(\vec{Y}_n) = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

называют **выборочным коэффициентом корреляции**.

Значения выборочных моментов (неслучайные числа) будем для краткости называть теми же терминами, что и сами моменты (случайные величины).

Основное свойство выборочных моментов состоит в том, что при увеличении объема выборки n они сходятся по вероятности к соответствующим теоретическим моментам. В частности, при $n \rightarrow \infty$ имеем $\bar{X} \xrightarrow[n \rightarrow \infty]{P} \mathbf{M}X$, а $\hat{\sigma}^2(\vec{X}_n) \xrightarrow[n \rightarrow \infty]{P} \mathbf{D}X$.

Лекция 15

Точечные оценки

Одной из задач математической статистики является *оценка* неизвестных параметров выбранной *параметрической модели*.

Предположим, что закон распределения *генеральной совокупности* принадлежит множеству $\{F(x; \vec{\theta}) : \vec{\theta} \in \Theta\}$, где вид функции распределения задан, а вектор параметров $\vec{\theta} = (\theta_1; \dots; \theta_r)$ неизвестен. Требуется найти оценку для $\vec{\theta}$ или некоторой функции от него (например, математического ожидания, дисперсии) по *случайной выборке* $(X_1; \dots; X_n)$ из генеральной совокупности X .

Например, предположим, что масса X детали имеет нормальный закон распределения, но его параметры $\theta_1 = \mathbf{M}X$ и $\theta_2 = \mathbf{D}X$ неизвестны. Нужно найти приближенное значение параметров по результатам наблюдений x_1, \dots, x_n , полученным в эксперименте (по реализации случайной выборки).

Пусть $\vec{X}_n = (X_1; \dots; X_n)$ — *случайная выборка* из *генеральной совокупности* X , функция распределения $F(x; \theta)$ которой известна, а θ — неизвестный параметр, т.е. рассматривается *параметрическая модель* $\{F(x; \vec{\theta}) : \vec{\theta} \in \Theta\}$ (для простоты изложения будем считать пока, что θ — скаляр).

Требуется построить *статистику* $\hat{\theta}(\vec{X}_n)$, которую можно было бы принять в качестве *точечной оценки* параметра θ .

Определение 15.1 *Точечной оценкой* параметра $\theta \in \Theta$ назовем любую функцию от наблюдений (т.е. любую статистику) $\hat{\theta}(\vec{X}_n)$.

Определение 15.2 Статистику $\hat{\theta}(\vec{X}_n)$ называют *состоятельной оценкой* параметра $\theta \in \Theta$, если с ростом *объема выборки* n она сходится по вероятности к оцениваемому параметру θ , т.е.

$$\hat{\theta}(\vec{X}_n) \xrightarrow[n \rightarrow \infty]{\mathbf{P}} \theta.$$

Определение 15.3 Статистику $\hat{\theta}(\vec{X}_n)$ называют *несмещенной оценкой* параметра θ , если ее математическое ожидание совпадает с θ , т.е. $\mathbf{M}\hat{\theta}(\vec{X}_n) = \theta$ для любого фиксированного n .

Замечание 15.1 Можно показать, что статистика

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

является несмещенной и состоятельной оценкой дисперсии $\mathbf{D}X$ генеральной совокупности. Ее называют *исправленной выборочной дисперсией*.

Замечание 15.2 Можно доказать, что *выборочные начальные и центральные моменты* являются состоятельными оценками соответствующих теоретических моментов, если только они существуют. Однако эти оценки, кроме \bar{X} , являются смещенными.

Пример 15.1 Пусть X_1, \dots, X_n — случайная выборка из генеральной совокупности X , имеющей нормальное распределение с неизвестным средним значением θ и известной дисперсией σ^2 .

Оценка $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) = X_1$ является несмещенной для θ , ибо $\mathbf{M}X_1 = \mathbf{M}X = \theta$, но не является состоятельной, так как, во-первых, X_1 не зависит от объема выборки и, следовательно, ее распределение не меняется с ростом n , а во-вторых,

$$\mathbf{P}\{|X_1 - \theta| < \varepsilon\} = \frac{2}{\sigma\sqrt{2\pi}} \int_0^\varepsilon e^{-\frac{t^2}{2\sigma^2}} dt \neq 1.$$

Метод моментов

Метод моментов был предложен английским статистиком К. Пирсоном и является одним из первых общих методов оценивания. Он состоит в следующем.

Пусть имеется случайная выборка $\vec{X}_n = (X_1; \dots; X_n)$ из генеральной совокупности X , распределение которой $p(x; \vec{\theta})$ известно с точностью до вектора параметров $\vec{\theta} = (\theta_1; \dots; \theta_r)$. Требуется найти *оценку* параметра $\vec{\theta}$ по случайной выборке \vec{X}_n .

Обозначим

$$\mu_k(\vec{\theta}) = \mathbf{M}(X)^k, \quad \nu_k(\vec{\theta}) = \mathbf{M}(X - \mathbf{M}X)^k -$$

начальный и центральный моменты порядка $k, k = 1, 2, \dots$. В методе моментов в качестве точечной оценки $\hat{\vec{\theta}}(\vec{X}_n) = (\hat{\theta}_1(\vec{X}_n); \dots; \hat{\theta}_r(\vec{X}_n))$ вектора параметров $\vec{\theta}$ берут решение системы r уравнений

$$\begin{cases} \hat{\mu}_{i_\alpha}(\vec{X}_n) = \mu_{i_\alpha}(\vec{\theta}), & \alpha = \overline{1, k}, \\ \hat{\nu}_{j_\beta}(\vec{X}_n) = \nu_{j_\beta}(\vec{\theta}), & \beta = \overline{1, l}, \end{cases}, \quad k + l = r, \quad (15.1)$$

относительно неизвестных $\theta_1, \dots, \theta_r$. Индексы i_α и j_β выбирают так, чтобы система уравнений решалась как можно проще. Можно показать, что при условии непрерывной зависимости решения этой системы от $\hat{\mu}_{i_\alpha}$ и $\hat{\nu}_{j_\beta}$, оценка, полученная методом моментов, является состоятельной и имеет **асимптотически нормальное распределение**, т.е. ее распределение при $n \rightarrow \infty$ стремится к нормальному. При этом уравнения (15.1) во многих случаях просты и их решение не вызывает больших вычислительных сложностей.

Понятно, что метод моментов не применим, когда моменты генеральной совокупности нужного порядка не существуют (например, для *распределения Коши*, у которого не существует даже начальный момент первого порядка — математическое ожидание).

Пример 15.2 Пусть дана случайная выборка $(X_1; \dots; X_n)$ объема n из генеральной совокупности X , имеющей равномерный закон распределения

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b); \\ 0, & x \notin (a, b), \end{cases}$$

с неизвестными параметрами a и b . Найдем методом моментов точечные оценки этих параметров.

Известно, что для равномерно распределенной случайной величины X

$$\mathbf{M}X = \frac{a+b}{2}, \quad \mathbf{D}X = \frac{(b-a)^2}{12}.$$

Выборочное среднее \bar{X} и выборочная дисперсия $\hat{\sigma}^2(\vec{X}_n)$ вычисляются по формулам

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Составляем систему двух уравнений

$$\begin{cases} \frac{a+b}{2} = \bar{x}, \\ \frac{(b-a)^2}{12} = \hat{\sigma}^2(\vec{x}_n). \end{cases}$$

Решая систему, получаем

$$\hat{b} = \bar{x} + \sqrt{3} \hat{\sigma}(\vec{x}_n), \quad \hat{a} = \bar{x} - \sqrt{3} \hat{\sigma}(\vec{x}_n).$$

Окончательно имеем

$$\hat{b}(\vec{X}_n) = \bar{X} + \sqrt{3} \hat{\sigma}(\vec{X}_n), \quad \hat{a}(\vec{X}_n) = \bar{X} - \sqrt{3} \hat{\sigma}(\vec{X}_n).$$

Метод максимального правдоподобия

Одним из наиболее универсальных методов оценивания параметров является **метод максимального правдоподобия** (предложенный Р. Фишером), суть которого состоит в следующем.

Рассмотрим *функцию правдоподобия* случайной выборки \vec{X}_n из генеральной совокупности X , плотность распределения $p(x; \vec{\theta})$ которой известна с точностью до параметра $\vec{\theta} \in \Theta$:

$$L(X_1, \dots, X_n, \vec{\theta}) = \prod_{i=1}^n p(X_i; \theta).$$

По определению, *оценкой максимального правдоподобия* параметра θ называют статистику $\hat{\theta}(\vec{X}_n)$, значения $\hat{\theta}$ которой для любой выборки \vec{x}_n удовлетворяют условию

$$L(\vec{x}_n; \hat{\theta}) = \max_{\vec{\theta} \in \Theta} L(\vec{x}_n; \vec{\theta}), \quad (15.2)$$

т.е. для выборки функция правдоподобия, как функция аргумента $\vec{\theta}$, достигает максимума.

Если функция $L(\vec{x}_n; \vec{\theta})$ дифференцируема как функция аргумента $\vec{\theta}$ при любом значении \vec{x}_n из множества \mathcal{X}_n значений случайной выборки \vec{X}_n и максимум $L(\vec{x}_n; \vec{\theta})$ достигается во внутренней точке из Θ , то значение точечной оценки максимального правдоподобия в случае скалярного параметра удовлетворяет уравнению (необходимому условию экстремума)

$$\frac{\partial L(\vec{x}_n; \theta)}{\partial \theta} = 0, \quad \text{или} \quad \frac{\partial \ln L(\vec{x}_n; \theta)}{\partial \theta} = 0, \quad (15.3)$$

так как при логарифмировании точки экстремума остаются теми же, а уравнение, как правило, упрощается.

Если распределение случайной величины X зависит от вектора параметров $\vec{\theta} = (\theta_1; \dots; \theta_r)$, то второе из уравнений (15.3) заменяется системой уравнений

$$\frac{\partial \ln L(\vec{x}_n; \theta)}{\partial \theta_k} = 0, \quad k = \overline{1, r}. \quad (15.4)$$

Уравнения (15.3) и (15.4) называют *уравнениями правдоподобия*. Для наиболее важных семейств распределений $p(x; \vec{\theta})$ уравнение правдоподобия имеет единственное решение $\hat{\theta} = \hat{\theta}_1; \dots; \hat{\theta}_r$. Во многих случаях решение системы (15.4), являющейся, как правило, нелинейной, приходится искать численными методами.

Пример 15.3 Пусть \vec{X}_n — случайная выборка из $\mathcal{N}(\theta_1, \theta_2^2)$. Методом максимального правдоподобия найдем оценку вектора параметров $\vec{\theta} = (\theta_1; \theta_2)$.

В этом случае функция правдоподобия

$$L(\vec{X}_n; \theta_1, \theta_2) = \frac{1}{(\theta_2 \sqrt{2\pi})^n} \exp\left(-\frac{1}{2\theta_2^2} \sum_{i=1}^n (X_i - \theta_1)^2\right)$$

и, как следствие,

$$\ln L(\vec{x}_n; \theta_1, \theta_2) = -n \ln \sqrt{2\pi} - n \ln \theta_2 - \frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2.$$

Поскольку число неизвестных параметров $r = 2$, система уравнений правдоподобия (15.4) будет состоять из двух уравнений:

$$\begin{cases} \frac{\partial}{\partial \theta_1} \ln L = \frac{1}{\theta_2^2} \sum_{i=1}^n (x_i - \theta_1) = 0, \\ \frac{\partial}{\partial \theta_2} \ln L = -\frac{n}{\theta_2} + \frac{1}{\theta_2^3} \sum_{i=1}^n (x_i - \theta_1)^2 = 0. \end{cases}$$

Решая систему, получаем

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{\theta}_2^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Следовательно, оценками максимального правдоподобия для математического ожидания $\mathbf{M}X = \theta_1$ и дисперсии $\mathbf{D}X = \theta_2^2$ случайной величины, распределенной по нормальному закону, являются соответственно выборочное среднее

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

и *выборочная дисперсия*

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad \#$$

Лекция 16

Интервальные оценки и доверительные интервалы

Некоторые важные распределения

Обозначим $\Gamma(p)$ — гамма-функцию, определяемую формулой

$$\Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt,$$

а $B(x, y)$ — бета-функцию, определяемую формулой

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 u^{p-1}(1-u)^{q-1} du.$$

Определение 16.1 Случайную величину с плотностью

$$p(x) = \begin{cases} \frac{1}{2^{m/2}\Gamma(m/2)} x^{m/2-1} e^{-x/2}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

называют случайной величиной, имеющей **распределение χ^2 (хи-квадрат)** или **χ^2 -распределение** с m степенями свободы.

Определение 16.2 Случайную величину с плотностью

$$p(x) = \frac{1}{\sqrt{\pi m}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \frac{1}{(1+x^2/m)^{m/2}}, \quad -\infty < x < \infty,$$

называют случайной величиной, имеющей **распределение Стьюдента** с n степенями свободы.

При $n \rightarrow \infty$ плотность распределения Стьюдента стремится к плотности стандартного нормального распределения $\mathcal{N}(0, 1)$. Для квантилей $t_p(m)$ распределения Стьюдента уровня p с m степенью свободы справедливо соотношение

$$t_p(m) = -t_{1-p}(m),$$

которое следует из четности плотности распределения Стьюдента.

Определение 16.3 Случайную величину с плотностью

$$p(x) = \begin{cases} \frac{\left(\frac{n}{m}\right)^{n/2}}{B\left(\frac{n}{2}, \frac{m}{2}\right)} \frac{x^{\frac{n}{2}-1}}{\left(1 + \frac{nx}{m}\right)^{\frac{n+m}{2}}}, & x \geq 0; \\ 0, & x < 0, \end{cases}$$

назовем случайной величиной, имеющей **распределение Фишера** (F -распределение, распределение Снедекора) с числом степеней свободы n и m .

Для квантилей $F_p(n, m)$ распределения Фишера уровня p с n и m степенями свободы имеет место соотношение

$$F_p(n, m) = \frac{1}{F_{1-p}(m, n)}.$$

Справедливы следующие теоремы.

Теорема 16.1 Пусть (X_1, \dots, X_n) — выборка из распределения $\mathcal{N}(\mu, \sigma^2)$. Тогда выборочное среднее $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ и исправленная выборочная дисперсия $S^2(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ независимы; при этом случайная величина $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$ имеет стандартное нормальное распределение, случайная величина $\frac{\sqrt{n}(\bar{X} - \mu)}{S(\vec{X}_n)}$ — распределение Стьюдента с $n-1$ степенью свободы, а случайная величина $\frac{(n-1)S^2(\vec{X}_n)}{\sigma^2}$ — χ^2 -распределение с $n-1$ степенью свободы.

Теорема 16.2 Пусть $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_m)$ — две независимые выборки из распределений $\mathcal{N}(\mu_1, \sigma_1^2)$ и $\mathcal{N}(\mu_2, \sigma_2^2)$ соответственно,

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, & S^2(\vec{X}_n) &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \\ \bar{Y} &= \frac{1}{m} \sum_{i=1}^m Y_i, & S^2(\vec{Y}_n) &= \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2. \end{aligned}$$

Тогда случайная величина $\frac{S^2(\vec{X}_n)}{S^2(\vec{Y}_n)}$ имеет распределение Фишера с числом степеней свободы $n-1$ и $m-1$.

Понятия интервальной оценки и доверительного интервала

При оценивании неизвестных параметров наряду с рассмотренными выше *точечными оценками* используются также *интервальные оценки*. В отличие от точечной оценки интервальная оценка позволяет получить вероятностную характеристику точности оценивания неизвестного параметра.

Пусть \vec{X}_n — случайная выборка объема n из генеральной совокупности X с функцией распределения $F(x; \theta)$, зависящей от параметра θ , значение которого неизвестно. Предположим, что для параметра θ построен интервал $(\underline{\theta}(\vec{X}_n), \bar{\theta}(\vec{X}_n))$, где $\underline{\theta}(\vec{X}_n)$ и $\bar{\theta}(\vec{X}_n)$ являются функциями случайной выборки \vec{X}_n , такими, что выполняется равенство

$$\mathbf{P}\{\underline{\theta}(\vec{X}_n) < \theta < \bar{\theta}(\vec{X}_n)\} = \gamma. \quad (16.1)$$

В этом случае интервал $(\underline{\theta}(\vec{X}_n), \bar{\theta}(\vec{X}_n))$ называют **интервальной оценкой** для параметра θ с **коэффициентом доверия** γ (или, сокращенно, **γ -доверительной интервальной оценкой**), а $\underline{\theta}(\vec{X}_n)$ и $\bar{\theta}(\vec{X}_n)$ соответственно **нижней** и **верхней границами** интервальной оценки.

Интервальная оценка $(\underline{\theta}(\vec{X}_n), \bar{\theta}(\vec{X}_n))$ представляет собой интервал со случайными границами, который с заданной вероятностью γ накрывает неизвестное истинное значение параметра θ . Таким образом, для различных *реализаций случайной выборки* \vec{X}_n , т.е. для различных элементов *выборочного пространства* \mathcal{X}_n , *статистики* $\underline{\theta}(\vec{X}_n)$ и $\bar{\theta}(\vec{X}_n)$ могут принимать различные значения. Более того, согласно (16.1), существует подмножество $\mathcal{K} \subset \mathcal{X}_n$, такое, что если $\vec{x}_n \in \mathcal{K}$, то $\theta \notin (\underline{\theta}(\vec{x}_n), \bar{\theta}(\vec{x}_n))$.

Вероятностной характеристикой точности оценивания параметра θ является случайная величина

$$l(\vec{X}_n) = \bar{\theta}(\vec{X}_n) - \underline{\theta}(\vec{X}_n),$$

которая для любой реализации \vec{x}_n случайной выборки \vec{X}_n есть длина интервала $(\underline{\theta}(\vec{x}_n), \bar{\theta}(\vec{x}_n))$.

Интервал $(\underline{\theta}(\vec{x}_n), \bar{\theta}(\vec{x}_n))$ называют **доверительным интервалом** для параметра θ с коэффициентом доверия γ или **γ -доверительным интервалом**.

Наряду с термином “коэффициент доверия” широко используют также термины **доверительная вероятность** и **уровень доверия**. При этом коэффициент доверия γ чаще всего выбирают равным 0,9, 0,95 или 0,99, т.е. близким к 1.

В некоторых ситуациях (например, при рассмотрении дискретных случайных величин) вместо равенства (16.1) удастся обеспечить лишь неравенство

$$\mathbf{P}\{\underline{\theta}(\vec{X}_n) < \theta < \bar{\theta}(\vec{X}_n)\} \geq \gamma,$$

т.е. построить интервальную оценку для параметра θ с коэффициентом доверия, не меньшим γ .

Примеры построения интервальных оценок для параметров нормальной случайной величины

Доверительная оценка для математического ожидания при известной дисперсии

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — случайная выборка объема n из генеральной совокупности X , распределенной по нормальному закону с параметрами μ и σ^2 , σ^2 известна.

В данном случае статистика

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

имеет стандартное нормальное распределение $\mathcal{N}(0, 1)$. Поэтому

$$\mathbf{P} \left\{ -u_{1-\alpha} < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u_{1-\alpha} \right\} = 1 - 2\alpha,$$

где $u_{1-\alpha}$ — квантиль уровня $1 - \alpha$ стандартного нормального распределения. Умножая все части двойного неравенства на $-\frac{\sigma}{\sqrt{n}}$, а затем прибавляя \bar{X} , получим

$$\mathbf{P} \left\{ \bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} \right\} = 1 - 2\alpha,$$

т.е. случайный интервал $\left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} \right)$ накрывает неизвестное математическое ожидание μ с необходимой доверительной вероятностью $1 - 2\alpha$.

Доверительная оценка для математического ожидания при неизвестной дисперсии

При неизвестной дисперсии σ^2 статистика

$$\frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n}$$

имеет *распределение Стьюдента* с $n - 1$ степенями свободы. Поэтому

$$\mathbf{P} \left\{ t_\alpha(n-1) < \frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n} < t_{1-\alpha}(n-1) \right\} = 1 - 2\alpha,$$

где $t_q(n-1)$ — квантиль уровня q распределения Стьюдента с $n - 1$ степенями свободы. Поскольку плотность распределения Стьюдента — четная функция, то $t_q(n-1) = -t_{1-q}(n-1)$. Умножая все части двойного неравенства на $-\frac{S(\vec{X}_n)}{\sqrt{n}}$, а затем прибавляя \bar{X} , заключаем, что нижняя и верхняя границы интервальной оценки с коэффициентом доверия $\gamma = 1 - 2\alpha$ для параметра μ в случае с неизвестной дисперсией можно определить по формулам

$$\underline{\mu}(\vec{X}_n) = \bar{X} - \frac{S(\vec{X}_n)}{\sqrt{n}} t_{1-\alpha}(n-1), \quad \bar{\mu}(\vec{X}_n) = \bar{X} + \frac{S(\vec{X}_n)}{\sqrt{n}} t_{1-\alpha}(n-1).$$

Доверительная оценка для разности математических ожиданий нормальных случайных величин с известными дисперсиями

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ и $\vec{Y}_m = (Y_1, \dots, Y_m)$ — две независимые выборки из распределений $\mathcal{N}(\mu_1, \sigma_1^2)$ и $\mathcal{N}(\mu_2, \sigma_2^2)$ соответственно,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i.$$

Нетрудно показать, что случайная величина

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

имеет распределение $\mathcal{N}(0, 1)$. Поэтому

$$\mathbf{P} \left\{ -u_{1-\alpha/2} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} < u_{1-\alpha/2} \right\} = 1 - \alpha,$$

где $u_{1-\alpha}$ — квантиль уровня $1 - \alpha$ стандартного нормального распределения. Умножая все части двойного неравенства на $-\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$, а затем прибавляя $\bar{X} - \bar{Y}$, получим

$$\mathbf{P} \left\{ \bar{X} - \bar{Y} - \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} u_{1-\alpha} < \mu_1 - \mu_2 < \bar{X} - \bar{Y} + \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} u_{1-\alpha} \right\} = 1 - 2\alpha,$$

т.е. случайный интервал $\left(\bar{X} - \bar{Y} - \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} u_{1-\alpha}, \bar{X} - \bar{Y} + \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} u_{1-\alpha} \right)$ покрывает неизвестную разность математических ожиданий $\mu_1 - \mu_2$ с необходимой доверительной вероятностью $1 - 2\alpha$.

Доверительная оценка для разности математических ожиданий нормальных случайных величин с неизвестными, но равными дисперсиями

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ и $\vec{Y}_n = (Y_1, \dots, Y_m)$ — две независимые выборки из распределений $\mathcal{N}(\mu_1, \sigma^2)$ и $\mathcal{N}(\mu_2, \sigma^2)$ соответственно, μ_2 и σ^2 неизвестны. Обозначим

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, & S^2(\vec{X}_n) &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \\ \bar{Y} &= \frac{1}{m} \sum_{i=1}^m Y_i, & S^2(\vec{Y}_n) &= \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2. \end{aligned}$$

Можно показать, что случайная величина

$$\sqrt{\frac{mn(m+n-2)}{m+n}} \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(n-1)S^2(\vec{X}_n) + (m-1)S^2(\vec{Y}_n)}}$$

имеет распределение Стьюдента с $m+n-2$ степенями свободы. Поэтому

$$\mathbf{P} \left\{ -t_{1-\alpha}(m+n-2) < \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(n-1)S^2(\vec{X}_n) + (m-1)S^2(\vec{Y}_n)}} < t_{1-\alpha}(m+n-2) \right\} = 1 - 2\alpha,$$

где $t_{1-\alpha}(m+n-2)$ — квантиль уровня $1 - \alpha$ распределения Стьюдента с $m+n-2$ степенями свободы. Умножая все части двойного неравенства на

$$-\sqrt{\frac{(m+n)((n-1)S^2(\vec{X}_n) + (m-1)S^2(\vec{Y}_n))}{mn(m+n-2)}}$$

а затем прибавляя $\bar{X} - \bar{Y}$, заключаем, что нижняя $\underline{\theta}(\vec{X}_n, \vec{Y}_n)$ и верхняя $\bar{\theta}(\vec{X}_n, \vec{Y}_n)$ границы интервальной оценки с коэффициентом доверия $\gamma = 1 - 2\alpha$ для разности $\mu_1 - \mu_2$ в случае с неизвестными, но равными дисперсиями можно определить по формулам

$$\begin{aligned} \underline{\theta}(\vec{X}_n, \vec{Y}_n) &= \bar{X} - \bar{Y} - t_{1-\alpha}(m+n-2) \sqrt{\frac{(m+n)((n-1)S^2(\vec{X}_n) + (m-1)S^2(\vec{Y}_n))}{mn(m+n-2)}}, \\ \bar{\theta}(\vec{X}_n, \vec{Y}_n) &= \bar{X} - \bar{Y} + t_{1-\alpha}(m+n-2) \sqrt{\frac{(m+n)((n-1)S^2(\vec{X}_n) + (m-1)S^2(\vec{Y}_n))}{mn(m+n-2)}}. \end{aligned}$$

Замечание 16.1 Можно показать, что все четыре интервальных оценки являются самыми короткими среди всех интервальных оценок с таким же уровнем доверия.

Доверительная оценка для дисперсии при неизвестном математическом ожидании

Статистика

$$\frac{(n-1)S^2(\vec{X}_n)}{\sigma^2}$$

имеет χ^2 -распределение с $n-1$ степенью свободы. Поэтому

$$\mathbf{P}\left\{\chi_{\alpha}^2(n-1) < \frac{(n-1)S^2(\vec{X}_n)}{\sigma^2} < \chi_{1-\alpha}^2(n-1)\right\} = 1 - 2\alpha,$$

где $\chi_q^2(n-1)$ — квантиль уровня q χ^2 -распределения с $n-1$ степенью свободы. Деля все части двойного неравенства на $(n-1)S^2(\vec{X}_n)$, а затем переходя к неравенству для обратных величин, получим

$$\mathbf{P}\left\{\frac{(n-1)S^2(\vec{X}_n)}{\chi_{1-\alpha}^2(n-1)} < \sigma^2 < \frac{(n-1)S^2(\vec{X}_n)}{\chi_{\alpha}^2(n-1)}\right\} = 1 - 2\alpha,$$

т.е.

$$\left(\frac{(n-1)S^2(\vec{X}_n)}{\chi_{1-\alpha}^2(n-1)}, \frac{(n-1)S^2(\vec{X}_n)}{\chi_{\alpha}^2(n-1)}\right) -$$

интервальная оценка для дисперсии σ^2 уровня доверия $1 - 2\alpha$.

Приближенные интервальные оценки для математического ожидания случайной величины

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — случайная выборка объема n из распределения случайной величины X с математическим ожиданием $\mu = \mathbf{M}X$ и дисперсией $\sigma^2 = \mathbf{D}X$.

В соответствии с центральной предельной теоремой функция распределения случайной величины

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

стремится к функции распределения $\Phi(x)$ стандартной нормальной случайной величины $\mathcal{N}(0,1)$. Поэтому

$$\mathbf{P}\left\{-u_{1-\alpha} < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u_{1-\alpha}\right\} \approx 1 - 2\alpha,$$

где $u_{1-\alpha}$ — квантиль уровня $1 - \alpha$ стандартного нормального распределения. Умножая все части двойного неравенства на $-\frac{\sigma}{\sqrt{n}}$, а затем прибавляя \bar{X} , получим

$$\mathbf{P}\left\{\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right\} \approx 1 - 2\alpha.$$

Заменяя неизвестную величину σ ее оценкой, например, случайной величиной

$$S(\vec{X}_n) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

получим приближенную интервальную оценку

$$\left(\bar{X} - \frac{S(\vec{X}_n)}{\sqrt{n}} u_{1-\alpha}, \bar{X} + \frac{S(\vec{X}_n)}{\sqrt{n}} u_{1-\alpha}\right)$$

математического ожидания μ с доверительной вероятностью $1 - 2\alpha$.

Доверительная оценка вероятности успеха в схеме Бернулли

Пусть X_i число успехов в i -ом, $i = \overline{1, n}$, испытании по схеме Бернулли с вероятностью успеха p и вероятностью неудачи $q = 1 - p$. Тогда $\mathbf{M}X_i = p$, $\mathbf{D}X_i = pq$. Оценкой дисперсии $\mathbf{D}X_i$ будет случайная величина $\hat{p}\hat{q}$, где $\hat{p} = \bar{X}$ и $\hat{q} = 1 - \hat{p}$ — доли успехов и неудач соответственно. Поэтому

$$\left(\hat{p} - \frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}} u_{1-\alpha}, \hat{p} + \frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}} u_{1-\alpha}\right) -$$

приближенная интервальная оценка вероятности успеха p с доверительной вероятностью $1 - 2\alpha$.

Доверительная оценка параметра распределения Пуассона

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — случайная выборка пуассоновской случайной величины X с параметром λ . Так как $\mathbf{M}X = \lambda$, $\mathbf{D}X = \lambda$, то дисперсию $\mathbf{D}X$ можно оценить как и математическое ожидание $\mathbf{M}X$ случайной величиной \bar{X} . Поэтому приближенной интервальной оценкой параметра λ с доверительной вероятностью $1 - 2\alpha$ будет интервал

$$\left(\bar{X} - u_{1-\alpha} \sqrt{\frac{\bar{X}}{n}}, \bar{X} + u_{1-\alpha} \sqrt{\frac{\bar{X}}{n}} \right).$$

Доверительная оценка выборочного коэффициента корреляции

Пусть (X_i, Y_i) , $i = \overline{1, n}$ — выборка из распределения нормального случайного вектора (X, Y) . Обозначим

$$\hat{\rho}(\vec{X}_n, \vec{Y}_n) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

выборочный коэффициент корреляции, являющийся состоятельной оценкой коэффициента корреляции ρ между X и Y .

Р. Фишер показал, что случайная величина

$$Z = \frac{1}{2} \ln \frac{1 + \hat{\rho}(\vec{X}_n, \vec{Y}_n)}{1 - \hat{\rho}(\vec{X}_n, \vec{Y}_n)}$$

уже для небольших значений n приблизительно распределена по нормальному закону с параметрами

$$\mathbf{M}Z \approx \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} + \frac{\rho}{2(n-1)}, \quad \mathbf{D}Z = \frac{1}{n-3}.$$

Отсюда следует, что интервальная оценка $(\underline{\rho}, \bar{\rho})$ для ρ уровня доверия $1 - \alpha$ имеет вид

$$(\underline{\rho}, \bar{\rho}) = (\text{th } \underline{z}, \text{th } \bar{z}), \quad (16.2)$$

где

$$\underline{z} = \frac{1}{2} \ln \frac{1 + \hat{\rho}}{1 - \hat{\rho}} + \frac{\hat{\rho}}{2(n-1)} - \frac{u_{1-\alpha/2}}{\sqrt{n-3}}; \quad (16.3)$$

$$\bar{z} = \frac{1}{2} \ln \frac{1 + \hat{\rho}}{1 - \hat{\rho}} + \frac{\hat{\rho}}{2(n-1)} + \frac{u_{1-\alpha/2}}{\sqrt{n-3}}, \quad (16.4)$$

а $u_{1-\alpha/2}$ — квантиль уровня $1 - \alpha/2$ стандартного нормального распределения. Равенствами (16.3), (16.4) можно пользоваться и в тех случаях, когда вектор (X, Y) не является нормальным. Но в этом случае увеличивается длина интервала $(\text{th } \underline{z}, \text{th } \bar{z})$, а значит, ухудшается точность оценивания.

Номер наблюдения	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Рост, см	165	171	182	165	183	180	183	166	173	172	174	170	164	168	184
Масса, кг	72,9	48,4	66,3	64,1	62,7	76,0	73,8	50,6	52,3	56,5	66,8	61,6	72,8	52,6	68,6

Таблица 16.1.

Пример 16.1 Вычислим значение $\hat{\rho}$ для пары случайных величин (X, Y) , где X — рост (в см), а Y — масса тела (в кг) наугад выбранного студента-первокурсника. Выборка объема $n = 15$ представлена в табл. 16.1. Имеем

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = \frac{2620}{15} = 173,3; \quad \bar{y} = \frac{1}{15} \sum_{i=1}^{15} y_i = \frac{945}{15} = 63,1.$$

$$\sum_{i=1}^{15} (x_i - \bar{x})^2 = 747,33; \quad \sum_{i=1}^{15} (y_i - \bar{y})^2 = 1171,4; \quad \sum_{i=1}^{15} (x_i - \bar{x})(y_i - \bar{y}) = 293,3.$$

Таким образом,

$$\hat{\rho} = \frac{293,3}{\sqrt{747,33 \cdot 1171,4}} = 0,313.$$

Найдем значения $\underline{\rho}$ и $\bar{\rho}$ при $1 - \alpha = 0,9$. Определив по таблице квантилей нормального распределения (см. 16.1) значение $u_{1-\alpha/2} = u_{0,95} = 1,65$ и воспользовавшись формулой (16.2), получим

$$\underline{\rho} = \text{th } \underline{z} \approx -0,162, \quad \bar{\rho} = \text{th } \bar{z} \approx 0,658.$$

Лекция 17

Проверка гипотез. Параметрические модели

Основные понятия

Пусть имеется выборка \vec{x}_n , являющаяся реализацией случайной выборки \vec{X}_n из генеральной совокупности X , функция распределения которой $F(t; \theta)$ зависит от неизвестного параметра θ .

Определение 17.1 *Статистической гипотезой* называют любое утверждение о функции распределения случайной величины X , при этом слово “статистическая” для краткости обычно опускают. *Параметрической гипотезой* называют любое утверждение о параметре θ функции распределения $F(t; \theta)$ случайной величины X . При этом если θ — скаляр, то речь идет об *однопараметрических гипотезах*, а если вектор, — то о *многопараметрических гипотезах*. *Статистическую гипотезу H* называют *простой*, если она имеет вид $H: \vec{\theta} = \vec{\theta}_0$, где $\vec{\theta}_0$ — некоторое заданное значение параметра. *Статистическую гипотезу* называют *сложной*, если она имеет вид $H: \vec{\theta} \in D$, где D — некоторое множество значений параметра θ , состоящее более чем из одного элемента.

Пример 17.1 Пусть \vec{X}_n — случайная выборка объема n из генеральной совокупности X , распределенной по нормальному закону с неизвестным математическим ожиданием μ и известной дисперсией σ^2 . Тогда гипотеза $H: \mu = \mu_0$, где μ_0 — некоторое заданное значение параметра μ , является простой.

Гипотезы $H_1: \mu \geq \mu_0$; $H_2: \mu \leq \mu_0$; $H: \mu_0 \leq \mu \leq \mu_1$ являются сложными.

Пример 17.2 Пусть в примере 17.1 оба параметра μ и σ неизвестны. В этом случае гипотеза $H: \mu = \mu_0$ становится сложной, так как ей соответствует множество значений двумерного вектора $\vec{\theta} = (\mu; \sigma)$, для которых $\mu = \mu_0$, $0 < \sigma < \infty$.

Проверка двух простых гипотез

Рассмотрим сначала случай, когда проверяются две *простые статистические гипотезы* вида

$$H_0: \theta = \theta_0, \quad H_1: \theta = \theta_1,$$

где θ_0, θ_1 — два заданных (различных) значения параметра. Первую *гипотезу H_0* обычно называют *основной*, или *нулевой*, а вторую H_1 — *альтернативной*, или *конкурирующей гипотезой*. По данным выборки \vec{x}_n исследователю нужно решить, можно ли принять выдвинутую гипотезу или ее нужно отклонить как противоречащую результатам эксперимента и принять некоторую альтернативную гипотезу (например, $\theta \neq \theta_0$).

Критерием, или **статистическим критерием**, проверки гипотез называют правило, по которому по данным выборки \vec{x}_n принимается решение о справедливости либо первой, либо второй гипотезы.

Критерий задают с помощью **критического множества** $W \in \mathbb{R}^n$, являющегося подмножеством *выборочного пространства \mathcal{X}_n* случайной выборки \vec{X}_n . Решение принимают следующим образом:

- если выборка \vec{x}_n принадлежит критическому множеству W , то отвергают основную гипотезу H_0 и принимают альтернативную гипотезу H_1 ;

- если выборка \vec{x}_n не принадлежит критическому множеству W (т.е. принадлежит дополнению \bar{W} множества W до выборочного пространства \mathcal{X}_n), то отвергают альтернативную гипотезу H_1 и принимают основную гипотезу H_0 . Множество \bar{W} называют **доверительным множеством**.

При использовании любого критерия возможны ошибки следующих видов:

- принять гипотезу H_1 , когда верна H_0 — **ошибка первого рода**;
- принять гипотезу H_0 , когда верна H_1 — **ошибка второго рода**.

Вероятности совершения ошибок первого и второго рода обозначают α и β :

$$\alpha = \mathbf{P}\{\vec{X}_n \in W \mid H_0\}, \quad \beta = \mathbf{P}\{\vec{X}_n \in \bar{W} \mid H_1\}.$$

Здесь $\mathbf{P}\{A \mid H_j\}$ — вероятность события A при условии, что справедлива гипотеза H_j , $j = 0, 1$. Указанные вероятности вычисляют с использованием функции плотности $p(t; \theta)$ распределения случайной выборки \vec{X}_n :

$$\alpha = \int \dots \int_W \prod_{k=1}^n p(t_k; \theta_0) dt_1 \dots dt_n, \quad \beta = \int \dots \int_{\bar{W}} \prod_{k=1}^n p(t_k; \theta_1) dt_1 \dots dt_n.$$

Вероятность совершения ошибки первого рода α называют также **уровнем значимости критерия**.

Величину $1 - \beta$, равную вероятности отвергнуть основную гипотезу H_0 , когда она неверна, называют **мощностью критерия**.

Критерий Неймана — Пирсона

При построении критерия для проверки статистических гипотез, как правило, исходят из необходимости максимизации его мощности $1 - \beta$ (минимизации вероятности совершения ошибки второго рода) при фиксированном уровне значимости α критерия (вероятности совершения ошибки первого рода). Для упрощения дальнейших рассуждений будем считать, что \vec{X}_n — случайная выборка объема n из генеральной совокупности непрерывной случайной величины X , плотность распределения вероятностей которой $p(t; \theta)$ зависит от неизвестного параметра θ , и рассмотрим две простые гипотезы $H_0: \theta = \theta_0$ и $H_1: \theta = \theta_1$.

Введем функцию случайной выборки \vec{X}_n :

$$\varphi(\vec{X}_n) = \frac{L(\vec{X}_n; \theta_1)}{L(\vec{X}_n; \theta_0)}, \quad L(\vec{X}_n; \theta) = \prod_{i=1}^n p(X_i; \theta).$$

Статистика $\varphi(\vec{X}_n)$ представляет собой отношение функций правдоподобия при истинности альтернативной и основной гипотез соответственно. Ее называют **отношением правдоподобия**. Для построения **оптимального** (наиболее мощного) при заданном уровне значимости α **критерия Неймана — Пирсона** в критическое множество W включают те элементы \vec{x}_n выборочного пространства \mathcal{X}_n случайной выборки \vec{X}_n , для которых выполняется неравенство

$$\varphi(\vec{x}_n) \geq C_\varphi,$$

где константу C_φ выбирают из условия

$$\mathbf{P}\{\varphi(\vec{X}_n) \geq C_\varphi \mid H_0\} = \alpha,$$

которое обеспечивает заданное значение уровня значимости α и может быть записано в виде

$$\int \dots \int_{\varphi(t_1, \dots, t_n) \geq C_\varphi} L(t_1, \dots, t_n; \theta_0) dt_1 \dots dt_n = \alpha.$$

При этом вероятность ошибки второго рода не может быть уменьшена при данном значении вероятности ошибки первого рода α .

Рассмотрим примеры построения оптимального критерия Неймана — Пирсона при проверке простых гипотез относительно параметров основных, наиболее часто используемых распределений.

Пример 17.3 Построение оптимального критерия Неймана — Пирсона для параметра μ нормального закона распределения с известной дисперсией σ^2 проведем для случая двух простых гипотез

$$H_0: \mu = \mu_0, \quad H_1: \mu = \mu_1,$$

где μ_0 и μ_1 — некоторые заданные значения, связанные неравенством $\mu_0 < \mu_1$.

В рассматриваемом случае функция правдоподобия имеет вид

$$LX_1, \dots, X_n \mu = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \right),$$

а отношение правдоподобия —

$$\varphi(\vec{X}_n) = \frac{L(X_1, \dots, X_n; \mu_1)}{L(X_1, \dots, X_n; \mu_0)} = \exp \left(\frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n X_i \right) \exp \left(-\frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2} \right).$$

В данном случае неравенство

$$\varphi(\vec{x}_n) = \exp \left(\frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n x_i \right) \exp \left(-\frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2} \right) \geq C_\varphi$$

равносильно неравенству

$$\sum_{i=1}^n x_i \geq C, \quad (17.1)$$

где константу C выбирают из условия обеспечения заданного уровня значимости α :

$$\mathbf{P} \left\{ \sum_{i=1}^n X_i \geq C \mid \mu = \mu_0 \right\} = \alpha. \quad (17.2)$$

Действительно,

$$\ln \left(\exp \left(\frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n x_i \right) \exp \left(-\frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2} \right) \right) = \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n x_i - \frac{n(\mu_1 - \mu_0)^2}{2\sigma^2} \geq \ln C_\varphi,$$

откуда следует, что

$$\sum_{i=1}^n x_i \geq \frac{\sigma^2}{\mu_1 - \mu_0} \left(\ln C_\varphi - \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2} \right) = C.$$

Случайная величина $X_1 + \dots + X_n$ имеет нормальное распределение с математическим ожиданием $n\mu$ и дисперсией $n\sigma^2$. Поэтому условие (17.2) можно записать в виде

$$1 - \Phi \left(\frac{C - n\mu_0}{\sigma\sqrt{n}} \right) = \alpha, \quad (17.3)$$

или

$$\frac{C - n\mu_0}{\sigma\sqrt{n}} = u_{1-\alpha}.$$

Таким образом, константа C , задающая критическую область в (17.1), определяется равенством

$$C = n\mu_0 + u_{1-\alpha}\sigma\sqrt{n}. \quad (17.4)$$

При этом вероятность совершения ошибки второго рода

$$\beta = \mathbf{P} \left\{ \sum_{i=1}^n X_i < C \mid \mu = \mu_1 \right\} = \Phi \left(\frac{C - n\mu_1}{\sigma\sqrt{n}} \right) \quad (17.5)$$

является минимально возможной при данном значении α .

Пример 17.4 Если в условиях примера 17.3 неравенство $\mu_0 < \mu_1$ заменить неравенством $\mu_1 < \mu_0$, то в этом случае критическое множество W задается неравенством

$$\sum_{i=1}^n x_i \leq C,$$

где константу C выбирают из условия

$$\mathbf{P}\left\{\sum_{i=1}^n X_i \leq C \mid \mu = \mu_0\right\} = \alpha.$$

Таким образом,

$$\Phi\left(\frac{C - n\mu_0}{\sigma\sqrt{n}}\right) = \alpha$$

или, что то же самое,

$$\frac{C - n\mu_0}{\sigma\sqrt{n}} = u_\alpha = -u_{1-\alpha}.$$

Из последнего равенства находим $C = n\mu_0 - u_{1-\alpha}\sigma\sqrt{n}$.

Сложные параметрические гипотезы

Предположим, что требуется проверить две *сложные гипотезы*

$$H_0: \theta \in \Theta_0, \quad H_1: \theta \in \Theta_1, \quad (17.6)$$

где Θ_0, Θ_1 — некоторые непересекающиеся области значений параметра θ . Например, области Θ_0, Θ_1 могут быть заданы неравенствами $\theta \leq \theta_0$ и $\theta \geq \theta_1$, где θ_0 и θ_1 — некоторые фиксированные значения параметра, удовлетворяющие неравенству $\theta_0 < \theta_1$.

Критерий проверки сложных гипотез (17.6) по-прежнему задается с помощью *критического множества* W *реализаций случайной выборки* \vec{X}_n , на основе которого решение принимают следующим образом:

- если реализация \vec{x}_n случайной выборки \vec{X}_n принадлежит критическому множеству W , тогда *основную гипотезу* H_0 отвергают и принимают *альтернативную гипотезу* H_1 ;
- если реализация \vec{x}_n случайной выборки \vec{X}_n не принадлежит критическому множеству W , тогда отвергают альтернативную гипотезу H_1 и принимают основную гипотезу H_0 .

Вероятности совершения *ошибок первого и второго рода* в случае сложных гипотез имеют прежний смысл и определяются выражениями

$$\alpha(\theta) = \mathbf{P}\{(X_1; \dots; X_n) \in W \mid \theta\}, \quad \theta \in \Theta_0; \quad \beta(\theta) = \mathbf{P}\{(X_1; \dots; X_n) \in \bar{W} \mid \theta\}, \quad \theta \in \Theta_1.$$

В отличие от случая *простых гипотез*, величины $\alpha(\theta), \beta(\theta)$ являются некоторыми функциями от параметра θ .

Максимально возможное значение вероятности совершения ошибки первого рода

$$\alpha = \max_{\theta \in \Theta_0} \alpha(\theta)$$

называют **размером критерия**.

Функцию

$$M(\theta) = \mathbf{P}\{(X_1; \dots; X_n) \in W \mid \theta\},$$

определяющую значение вероятности отклонения основной гипотезы H_0 в зависимости от истинного значения параметра θ , называют **функцией мощности критерия**. Если существует критерий, который при данном фиксированном размере α максимизирует функцию мощности $M(\theta)$ по всем возможным критериям одновременно при всех θ из множества Θ_1 , то такой **критерий** называют **равномерно наиболее мощным**. Равномерно наиболее мощные критерии существуют лишь в некоторых частных случаях при проверке гипотез относительно одномерных параметров (см. примеры 17.5–17.7).

Вероятности совершения ошибок первого и второго рода связаны с функцией мощности следующими соотношениями:

$$\alpha(\theta) = M(\theta), \quad \theta \in \Theta_0; \quad (17.7)$$

$$\beta(\theta) = 1 - M(\theta), \quad \theta \in \Theta_1. \quad (17.8)$$

Тем самым равномерно наиболее мощный критерий, если он существует, минимизирует вероятность совершения ошибки второго рода $\beta(\theta)$ (при фиксированном размере α) одновременно при всех $\theta \in \Theta_1$.

Построение критериев для проверки сложных параметрических гипотез проиллюстрируем далее для случая нормальной модели.

Пример 17.5 Рассмотрим проверку простой гипотезы $H_0: \mu = \mu_0$ против сложной гипотезы $H_1: \mu > \mu_0$ относительно параметра — среднего μ нормального распределения при известной дисперсии σ^2 .

При любом $\mu_1 > \mu_0$ критическая область оптимального наиболее мощного критерия Неймана — Пирсона размера α для простых гипотез $\mu = \mu_0$ против $\mu = \mu_1$ имеет вид (17.1), где константу C выбирают из условия (17.2) или (17.3). Поэтому она не зависит от μ_1 . Это означает, что построенный уже выше для указанных простых гипотез критерий с критическим множеством, задаваемым неравенством (17.1)

$$\sum_{i=1}^n x_i \geq C = n\mu_0 + u_{1-\alpha}\sigma\sqrt{n}, \quad (17.9)$$

является равномерно наиболее мощным критерием размера α для данной задачи со сложной альтернативной гипотезой $H_1: \mu > \mu_0$.

Пример 17.6 В условиях предыдущего примера рассмотрим проверку простой гипотезы $H_0: \mu = \mu_0$ против сложной гипотезы $H_1: \mu < \mu_0$.

В этом случае, используя результаты, полученные при рассмотрении примера 17.4, приходим к выводу, что равномерно наиболее мощный критерий размера α для данной задачи задается критическим множеством, определяемым неравенством

$$\sum_{i=1}^n x_i \leq C = n\mu_0 - u_{1-\alpha}\sigma\sqrt{n}.$$

Пример 17.7 В условиях примера 17.5 рассмотрим проверку двух сложных гипотез вида

$$H_0: \mu \leq \mu_0, \quad H_1: \mu \geq \mu_1, \quad (17.10)$$

где $\mu_0 < \mu_1$.

Заметим, что для критерия с критическим множеством (17.9) вероятность совершения ошибки первого рода

$$\alpha(\mu) = \mathbf{P}\left\{\sum_{i=1}^n X_i \geq C \mid \mu\right\} = 1 - \Phi\left(u_{1-\alpha} + (\mu_0 - \mu)\frac{\sqrt{n}}{\sigma}\right)$$

есть возрастающая функция переменного μ . Тем самым максимальное значение вероятности совершения ошибки первого рода, определяемое как

$$\alpha = \max_{\mu \leq \mu_0} \alpha(\mu),$$

достигается в точке $\mu = \mu_0$, откуда следует, что данный критерий, применяемый к сложным гипотезам (17.10), имеет размер $\alpha = \alpha(\mu_0)$.

Рассуждая далее так же, как в примере 17.5, получаем, что указанный критерий с критической областью (17.9) является равномерно наиболее мощным критерием для данной задачи со сложными гипотезами.

Пример 17.8 Рассмотрим проверку гипотез относительно параметра нормального распределения μ следующего вида:

$$H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$$

(по-прежнему предполагаем, что дисперсия σ^2 известна).

В этом случае основная гипотеза H_0 является простой, а альтернативная гипотеза H_1 является сложной. При $\mu = \mu_0$ рассмотрим статистику

$$\frac{\bar{X} - \mu_0}{\sigma}\sqrt{n},$$

которая имеет стандартное нормальное распределение. Критическое множество для проверки указанных гипотез H_0, H_1 определим следующим образом:

$$\frac{|\bar{x} - \mu_0|}{\sigma}\sqrt{n} \geq u_{1-\alpha/2}.$$

Соответствующий критерий по построению имеет вероятность совершения ошибки первого рода α .

Пример 17.9 Рассмотрим проверку двух сложных гипотез

$$H_0: \mu = \mu_0, \quad H_1: \mu > \mu_0 \quad (17.11)$$

относительно параметра μ нормального закона распределения в случае, когда дисперсия σ^2 неизвестна.

В отличие от примера 17.5 гипотеза H_0 также является сложной. При $\mu = \mu_0$ статистика

$$\frac{\bar{X} - \mu_0}{S(\vec{X}_n)} \sqrt{n} \quad (17.12)$$

имеет *распределение Стьюдента* с $n - 1$ *степенями свободы*. Исходя из этого получаем, что критерий с уровнем значимости α для гипотез (17.11) задается критическим множеством

$$\frac{\bar{x} - \mu_0}{S(\vec{x}_n)} \sqrt{n} \geq t_{1-\alpha}(n-1),$$

где $t_{1-\alpha}(n-1)$ — квантиль уровня $1 - \alpha$ распределения Стьюдента с $n - 1$ степенями свободы.

Аналогично на основе статистики (17.12) строят критерий для проверки сложных гипотез

$$H_0: \mu = \mu_0, \quad H_1: \mu < \mu_0 \quad (17.13)$$

или

$$H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0. \quad (17.14)$$

Для гипотез (17.13) критерий размера α задается критическим множеством, определяемым неравенством

$$\frac{\bar{x} - \mu_0}{S(\vec{x}_n)} \sqrt{n} \leq -t_{1-\alpha}(n-1).$$

Для гипотез вида (17.14) критерий размера α задают критическим множеством, определяемым неравенством

$$\frac{|\bar{x} - \mu_0|}{S(\vec{x}_n)} \sqrt{n} \geq t_{1-\alpha/2}(n-1).$$

Пример 17.10 Рассмотрим проверку гипотез о равенстве математических ожиданий для двух различных нормальных распределений.

Пусть определены две случайные выборки $(X_1; \dots; X_n)$ и $(Y_1; \dots; Y_m)$ объемов n и m из *генеральных совокупностей* независимых случайных величин $X \sim N(\mu_1, \sigma_1^2)$ и $Y \sim N(\mu_2, \sigma_2^2)$ соответственно. Рассмотрим следующие задачи проверки сложных гипотез относительно параметров μ_1, μ_2 в случае, когда дисперсии σ_1^2, σ_2^2 известны:

$$H_0: \mu_1 = \mu_2, \quad H_1: \mu_1 > \mu_2; \quad (17.15)$$

$$H_0: \mu_1 = \mu_2, \quad H_1: \mu_1 < \mu_2; \quad (17.16)$$

$$H_0: \mu_1 = \mu_2, \quad H_1: \mu_1 \neq \mu_2. \quad (17.17)$$

Разность *выборочных средних* $\bar{X} - \bar{Y}$ имеет нормальное распределение с математическим ожиданием $\mu_1 - \mu_2$ и дисперсией $\sigma_1^2/n + \sigma_2^2/m$. Отсюда следует, что при справедливости основной гипотезы, т.е. при $\mu_1 = \mu_2$, статистика

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \quad (17.18)$$

имеет стандартное нормальное распределение. Исходя из этого, заключаем, что критерии размера α для указанных задач задаются критическими множествами

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \geq u_{1-\alpha}; \quad \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \leq -u_{1-\alpha}; \quad \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \geq u_{1-\alpha/2}. \quad \#$$

Рассмотрим также задачу проверки гипотез (17.15)–(17.16) о равенстве средних двух нормальных распределений в предположении, что их дисперсии не известны, но равны между собой: $\sigma_1 = \sigma_2 = \sigma$. Обозначим через

$$S^2(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S^2(\vec{Y}_m) = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2$$

соответствующие исправленные оценки дисперсии. Статистики $(n-1)S^2(\vec{X}_n)/\sigma^2$ и $(m-1)S^2(\vec{Y}_m)/\sigma^2$ имеют χ^2 -распределения с $n-1$ и $m-1$ степенями свободы. Тем самым статистика

$$\frac{(n-1)S^2(\vec{X}_n)}{\sigma^2} + \frac{(m-1)S^2(\vec{Y}_m)}{\sigma^2}$$

имеет также χ^2 -распределение с $n + m - 2$ степенями свободы . Учитывая, что случайная величина (17.18) при $\mu_1 = \mu_2$ имеет стандартное нормальное распределение, получаем, что статистика

$$\tilde{T}(\vec{X}_n, \vec{Y}_m) = \frac{(\bar{X} - \bar{Y})\sqrt{n+m-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{(n-1)S^2(\vec{X}_n) + (m-1)S^2(\vec{Y}_m)}}$$

имеет распределение Стьюдента с $n + m - 2$ степенями свободы . Поэтому критерии размера α для проверки гипотез (17.15)–(17.16) задаются с помощью критических множеств, определяемых следующими неравенствами:

$$\tilde{T}(\vec{x}_n, \vec{y}_m) \geq t_{1-\alpha}(n+m-2), \quad (17.19)$$

$$\tilde{T}(\vec{x}_n, \vec{y}_m) \leq -t_{1-\alpha}(n+m-2), \quad (17.20)$$

$$|\tilde{T}(\vec{x}_n, \vec{y}_m)| \geq t_{1-\alpha/2}(n+m-2). \quad (17.21)$$

Лекция 18

Проверка непараметрических гипотез

Критерий согласия χ^2 . Простая гипотеза

Пусть наблюдается дискретная случайная величина X , принимающая r различных значений u_1, \dots, u_r с положительными вероятностями p_1, \dots, p_r :

$$\mathbf{P}\{X = u_k\} = p_k, \quad k = \overline{1, r}, \quad \sum_{k=1}^r p_k = 1.$$

Допустим, что в выборке $\vec{x}_n = (x_1; \dots; x_n)$ число u_k встретилось $n_k(\vec{x}_n)$ раз, $k = \overline{1, r}$. Отметим, что $\sum_{k=1}^r n_k(\vec{x}_n) = n$, т.е. случайные величины $n_1(\vec{X}_n), \dots, n_r(\vec{X}_n)$ зависимы.

Теорема 18.1 (теорема Пирсона) *Распределение случайной величины*

$$\sum_{k=1}^r \frac{(n_k(\vec{X}_n) - np_k)^2}{np_k}$$

при $n \rightarrow \infty$ сходится к χ^2 -распределению с $r - 1$ степенями свободы. $\#$

Этой теоремой можно воспользоваться для проверки *простой гипотезы*

$$H_0: p_1 = p_{10}, \dots, p_r = p_{r0}, \quad (18.1)$$

где p_{10}, \dots, p_{r0} — известные величины, против альтернативной гипотезы

$$H_1: \text{существуют такие } k, \text{ что } p_k \neq p_{k0}, \quad k = \overline{1, r}. \quad (18.2)$$

Если истинной является гипотеза H_0 , то по закону больших чисел

$$\frac{n_k(\vec{X}_n)}{n} - p_{k0} \rightarrow 0, \quad k = \overline{1, r},$$

а если верна H_1 , то

$$\frac{n_k(\vec{X}_n)}{n} - p_{k0} \rightarrow p_k - p_{k0} \neq 0, \quad \text{для некоторых } k = \overline{1, r}.$$

Поэтому при H_1 случайная величина

$$\chi^2(\vec{X}_n) = \sum_{k=1}^r \frac{(n_k(\vec{X}_n) - np_{k0})^2}{np_{k0}} = n \sum_{k=1}^r \frac{\left(\frac{n_k(\vec{X}_n)}{n} - p_{k0}\right)^2}{p_{k0}} \quad (18.3)$$

стремится к бесконечности и, следовательно, в эксперименте, как правило, принимает большие значения, чем при H_0 , когда ее распределение стремится к распределению χ^2 с $r - 1$ степенями свободы.

Таким образом, становится естественным следующее определение **критерия согласия** χ^2 (хи-квадрат). Этот критерий при больших n на уровне значимости α отклоняет гипотезу H_0 в пользу альтернативной гипотезы H_1 , если

$$\chi^2(\vec{x}_n) > \chi_{1-\alpha}^2(r-1),$$

где $\chi_{1-\alpha}^2(r-1)$ — квантиль уровня $1 - \alpha$ χ^2 -распределения с $r - 1$ степенями свободы, а $\chi^2(\vec{x}_n)$ — реализация случайной величины (18.3).

Если же

$$\chi^2(\vec{x}_n) \leq \chi_{1-\alpha}^2(r-1),$$

то делается вывод о том, что гипотеза H_0 не противоречит статистическим данным и ее следует принять.

Критерием χ^2 при небольших объемах выборки n пользоваться нельзя. На практике при небольших r необходимо, чтобы выполнялись условия $np_k \geq 10$, $k = \overline{1, r}$, а если r велико ($r \geq 20$), достаточно, чтобы было $np_k \geq 5$, $k = \overline{1, r}$.

Критерий χ^2 можно использовать и тогда, когда случайная величина X непрерывна или дискретна, но принимает счетное множество значений с положительными вероятностями. В этом случае множество M возможных значений X разбивают на r непересекающихся подмножеств M_k , $k = \overline{1, r}$, таким образом, чтобы вероятность p_k , $k = \overline{1, r}$, попадания случайной величины X в k подмножество M_k удовлетворяла условию $np_k \geq 5$ или $np_k \geq 10$, $k = \overline{1, r}$. Если X — непрерывная случайная величина, то в качестве M_k , $k = \overline{1, r}$, обычно берут множества вида

$$(-\infty, s_1), [s_1, s_2), \dots, [s_{r-2}, s_{r-1}), [s_{r-1}, \infty),$$

где $s_1 < s_2 < \dots < s_{r-1}$, $s_k \in \mathbb{R}$, $k = \overline{1, r-1}$.

Определим дискретную случайную величину X' , принимающую значение k тогда и только тогда, когда $X \in M_k$, $k = \overline{1, r}$. В этом случае исходная задача проверки статистических гипотез сводится к проверке основной гипотезы (18.1) при альтернативной гипотезе (18.2), где в случае непрерывности случайной величины X

$$p_{k0} = \int_{M_k} p_0(t) dt \quad -$$

вероятность попадания случайной величины X в множество M_k , а $p_0(t)$ — плотность X при H_0 .

Если X — дискретная случайная величина, имеющая счетное множество возможных значений z_1, z_2, \dots , и $\mathbf{P}\{X = z_j\} = q_j > 0$, $j = 1, 2, \dots$, то вместо проверки гипотезы

$$H_0: q_j = q_{j0}, \quad j = 1, 2, \dots,$$

где q_{j0} , $j = 1, 2, \dots$, — известные числа, при альтернативной гипотезе

$$H_1: \text{существуют такие } j, \text{ что } q_j \neq q_{j0}, \quad j = 1, 2, \dots,$$

проверяют гипотезу (18.1) при альтернативной гипотезе (18.2), где вероятности p_{k0} , $k = \overline{1, r}$, вычисляют по формулам

$$p_{k0} = \sum_{z_j \in M_k} q_{j0}, \quad k = \overline{1, r}.$$

Далее для выборки \vec{x}_n находят число $n_k(\vec{x}_n)$ ее элементов, принадлежащих множеству M_k , $k = \overline{1, r}$. Затем, подставляя \vec{x}_n вместо \vec{X} в формулу (18.3), определяют реализацию $\chi^2(\vec{x}_n)$ случайной величины $\chi^2(\vec{X})$. Гипотеза H_0 отклоняется в пользу гипотезы H_1 , если $\chi^2(\vec{x}_n) > \chi_{1-\alpha}^2(r-1)$ и принимается в противном случае.

Недостатком использования критерия χ^2 для случайных величин, принимающих бесконечное множество значений, является некоторая потеря информации при переходе от X к случайной величине X' с конечным числом значений.

Пример 18.1 При 4040 бросаниях монеты французский естествоиспытатель Ж.Л.Л. Бюффон (1707–1788) получил 2048 выпадений “герба” и 1992 выпадений “решки”. Совместимо ли это с гипотезой о том, что вероятность выпадения “герба” при одном бросании равна $1/2$?

Здесь $n = 4040$, $r = 2$, $n_1(\vec{x}_n) = 2048$, $n_2(\vec{x}_n) = 1992$, $p_{10} = p_{20} = 0,5$, число степеней свободы $r - 1 = 1$, и при $\alpha = 0,05$ находим $\chi_{0,95}^2(1) = 3,841$.

Проверим гипотезу H_0 о том, что вероятности p_1 и p_2 выпадения “герба” и “решки” равны $1/2$. На основании (18.3) получаем

$$\chi^2(\vec{x}_n) = \frac{(2048 - 4040 \cdot 0,5)^2}{4040 \cdot 0,5} + \frac{(1992 - 4040 \cdot 0,5)^2}{4040 \cdot 0,5} = 0,776.$$

Так как $0,776 < 3,841$, то статистические данные не противоречат гипотезе H_0 .

Критерий χ^2 для сложной гипотезы

Пусть функция распределения дискретной случайной величины X , принимающей конечное множество значений u_1, \dots, u_r , зависит от неизвестного d -мерного вектора параметров θ . Тогда вероятность p_k того, что X примет возможное значение u_k , зависит от θ , т.е. $p_k = p_k(\theta)$, $k = \overline{1, r}$. А так как вероятности $p_1(\theta), \dots, p_r(\theta)$ полностью определяют функцию распределения случайной величины X , то в рассматриваемом случае *основная гипотеза* принимает следующий вид:

$$H_0: \mathbf{P}\{X = u_k\} = p_k(\theta), \quad k = \overline{1, r}, \quad \theta \in \Theta \subset \mathbb{R}^d.$$

Эту сложную гипотезу можно проверить при помощи модификации критерия χ^2 Пирсона.

Пусть $\hat{\theta}(\vec{x}_n)$ — значение оценки $\hat{\theta}(\vec{X}_n)$ максимального правдоподобия для θ , а $n_k(\vec{x}_n)$ — количество элементов выборки \vec{x}_n , равных u_k , $k = \overline{1, r}$. Оценку $\hat{\theta}(\vec{X}_n)$ получают в результате минимизации логарифма *функции правдоподобия*

$$L(\vec{X}_n; \theta) = \frac{n!}{n_1! \dots n_r!} \prod_{k=1}^r p_k^{n_k(\vec{X}_n)}(\theta), \quad \sum_{i=1}^r n_i(\vec{X}_n) = n,$$

как решение системы уравнений

$$\sum_{k=1}^r \frac{n_k(\vec{X}_n)}{p_k(\theta)} \frac{\partial p_k(\theta)}{\partial \theta_j} = 0, \quad j = \overline{1, d}.$$

Можно показать, что при некоторых предположениях о гладкости функций $p_k(\theta)$, $k = \overline{1, r}$, распределение случайной величины при $n \rightarrow \infty$

$$\chi^2(\vec{X}_n) = \sum_{i=1}^r \frac{(n_i(\vec{X}_n) - np_i(\hat{\theta}(\vec{X}_n)))^2}{np_i(\hat{\theta}(\vec{X}_n))}$$

сходится к χ^2 -распределению с $r - d - 1$ степенями свободы.

Если X — непрерывная случайная величина с функцией распределения $F(t)$, то, разбивая множество возможных значений X на конечное число непересекающихся подмножеств и переходя к дискретной случайной величине X' , можно проверить сложную гипотезу

$$H_0: F(t) \in \{F(t; \theta), \theta \in \Theta \subset \mathbb{R}^d\}.$$

Необходимо только помнить, что оценку максимального правдоподобия $\hat{\theta}(\vec{X}_n)$ следует строить не по наблюдениям X_1, \dots, X_n случайной величины X , а по значениям частот $n_1(\vec{x}'_n), \dots, n_r(\vec{x}'_n)$ случайной величины X' , что, как правило, гораздо труднее. Построение такой оценки для наиболее распространенных параметрических семейств распределений (нормального, экспоненциального, пуассоновского и т.д.) можно найти в специальной литературе (См. Г. Крамер).

Критерий независимости, основанный на выборочном коэффициенте корреляции

Пусть (X_i, Y_i) , $i = \overline{1, n}$ — выборка из распределения нормального случайного вектора (X, Y) . Обозначим

$$\hat{\rho}(\vec{X}_n, \vec{Y}_n) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

выборочный коэффициент корреляции.

При проверке статистической гипотезы $H_0: \rho = 0$ (т.е. гипотезы о том, что нормально распределенные случайные величины независимы) используют *статистику*

$$t = \frac{\hat{\rho}(\vec{X}_n, \vec{Y}_n) \sqrt{n-2}}{\sqrt{1 - \hat{\rho}^2(\vec{X}_n, \vec{Y}_n)}}, \quad (18.4)$$

которая имеет *распределение Стьюдента* с $n - 2$ степенями свободы. Если окажется, что

$$\frac{|\hat{\rho}| \sqrt{n-2}}{\sqrt{1 - \hat{\rho}^2}} \leq t_{1-\alpha/2}(n-2),$$

то гипотезу H_0 принимают при *уровне значимости* α , где $t_{1-\alpha/2}(n-2)$ — квантиль уровня $1 - \alpha/2$ распределения Стьюдента с $n - 2$ степенями свободы.

Пример 18.2 В примере 16.1 лекции 15 найдено значение точечной оценки $\hat{\rho} = 0,313$. Проверим гипотезу $H_0: \rho = 0$ на уровне значимости $\alpha = 0,1$.

По таблице квантилей распределения Стьюдента находим квантиль $t_{0,95}(13) = 1,77$ и сравниваем со значением

$$\hat{\rho} \frac{\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} = 0,313 \frac{\sqrt{13}}{\sqrt{0,902}} = 1,19.$$

Поскольку $1,19 < 1,77$, то гипотезу $\rho = 0$ принимаем.

Таблицы сопряженности признаков и критерий χ^2

Пусть имеется случайная выборка

$$(\vec{X}_n; \vec{Y}_n) = ((X_1, Y_1); \dots; (X_n, Y_n))$$

из генеральной совокупности двумерной дискретной случайной величины $(X; Y)$, где случайная величина X может принимать значения u_1, \dots, u_r , а случайная величина Y — значения v_1, \dots, v_s . Определим случайную величину $n_{ij}(\vec{X}_n, \vec{Y}_n)$, реализация n_{ij} которой равна количеству элементов выборки $(\vec{x}_n, \vec{y}_n) = ((x_1, y_1); \dots; (x_n, y_n))$, совпадающих с элементом $(u_i; v_j)$, $i = \overline{1, r}$, $j = \overline{1, s}$.

Введем случайные величины $n_i(\vec{X}_n, \vec{Y}_n)$ и $n_j(\vec{X}_n, \vec{Y}_n)$, значения n_i и n_j которых определим по формулам

$$n_i = \sum_{j=1}^s n_{ij}, \quad n_j = \sum_{i=1}^r n_{ij}.$$

При этом n_i — количество элементов выборки $(\vec{x}_n; \vec{y}_n)$, в которых встретилось значение u_i , а n_j — количество элементов выборки $(\vec{x}_n; \vec{y}_n)$, в которых встретилось значение v_j . Кроме того, имеют место очевидные равенства

$$\sum_{i=1}^r n_i = \sum_{j=1}^s n_j = \sum_{i=1}^r \sum_{j=1}^s n_{ij} = n.$$

В рассматриваемом случае результаты наблюдений удобно оформлять в виде таблицы, называемой **таблицей сопряженности признаков** (18.1). Пусть далее

$$p_{ij} = \mathbf{P}\{X = u_i, Y = v_j\}, \quad p_i = \mathbf{P}\{X = u_i\}, \quad p_j = \mathbf{P}\{Y = v_j\}, \quad i = \overline{1, r}, \quad j = \overline{1, s}.$$

Дискретные случайные величины X и Y независимы тогда и только тогда, когда

$$\mathbf{P}\{X = u_i, Y = v_j\} = \mathbf{P}\{X = u_i\} \mathbf{P}\{Y = v_j\}, \quad i = \overline{1, r}, \quad j = \overline{1, s}.$$

Поэтому основную гипотезу о независимости дискретных случайных величин X и Y можно представить в следующем виде:

$$H_0: p_{ij} = p_i p_j, \quad i = \overline{1, r}, \quad j = \overline{1, s}. \quad (18.5)$$

При этом, как правило, в качестве альтернативной используют гипотезу

$$H_1: p_{ij} \neq p_i p_j \text{ для некоторых } i = \overline{1, r}, \quad j = \overline{1, s}. \quad (18.6)$$

Для проверки основной гипотезы (18.5) при альтернативной гипотезе (18.6) К. Пирсон предложил использовать статистику $\hat{\chi}^2(\vec{X}_n, \vec{Y}_n)$, называемую **статистикой Фишера — Пирсона**, реализация $\hat{\chi}^2(\vec{x}_n, \vec{y}_n)$ которой определяется формулой

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = n \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_i n_j}{n}\right)^2}{n_i n_j}. \quad (18.7)$$

Из закона больших чисел следует, что при $n \rightarrow \infty$

$$\frac{n_{ij}(\vec{X}_n, \vec{Y}_n)}{n} \rightarrow p_{ij}, \quad \frac{n_i(\vec{X}_n, \vec{Y}_n)}{n} \rightarrow p_i, \quad \frac{n_j(\vec{X}_n, \vec{Y}_n)}{n} \rightarrow p_j, \quad i = \overline{1, r}, \quad j = \overline{1, s}.$$

Поэтому при истинности гипотезы H_0 и больших объемах выборки (\vec{x}_n, \vec{y}_n) должно выполняться приближенное равенство

$$n_{ij} \approx n_i n_j, \quad i = \overline{1, r}, \quad j = \overline{1, s},$$

и, следовательно, значения (18.7) статистики $\hat{\chi}^2(\vec{X}_n, \vec{Y}_n)$ должны быть “не слишком велики”. “Слишком большие” значения должны свидетельствовать о том, что H_0 неверна.

Ответ на вопрос о том, какие значения нужно считать слишком большими, а какие — нет, дает следующая теорема.

Теорема 18.2 Если истинна гипотеза H_0 , то распределение статистики $\hat{\chi}^2(\vec{X}_n, \vec{Y}_n)$ при $n \rightarrow \infty$ сходится к случайной величине, имеющей χ^2 -распределение с числом степеней свободы $k = (r - 1)(s - 1)$:

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\hat{\chi}^2(\vec{X}_n, \vec{Y}_n) < z\} = \int_0^z \frac{t^{\frac{k}{2}-1}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} e^{-\frac{t}{2}} dt, \quad z > 0. \quad \#$$

В соответствии с теоремой 18.2 **критерий независимости χ^2** отклоняет гипотезу H_0 на уровне значимости $1 - \alpha$, если

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) > \chi_{1-\alpha}^2((r-1)(s-1)),$$

где $\chi_{1-\alpha}^2((r-1)(s-1))$ — квантиль уровня значимости $1 - \alpha$ χ^2 -распределения с числом степеней свободы $(r-1)(s-1)$. При этом считается, что критерий χ^2 можно использовать, если $n_i n_j / n \geq 5$.

Правую часть равенства (18.7) можно преобразовать к форме, более удобной для практического использования:

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_i n_j} - 1 \right). \quad (18.8)$$

В частном, но очень распространенном случае таблиц сопряженности при $r = s = 2$ формула (18.7) для вычисления $\hat{\chi}^2(\vec{x}_n, \vec{y}_n)$ имеет еще более простой вид:

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 n_2 n_1 n_2}. \quad (18.9)$$

Для таблиц сопряженности при $r = s = 2$, как правило, используют статистику $\tilde{\chi}^2(\vec{X}_n, \vec{Y}_n)$ с реализациями

$$\tilde{\chi}^2(\vec{x}_n, \vec{y}_n) = \frac{(n|n_{11}n_{22} - n_{12}n_{21}| - n/2)^2}{n_1 n_2 n_1 n_2}, \quad (18.10)$$

называемую **статистикой Фишера — Пирсона с поправкой Йейтса на непрерывность**, распределение которой лучше согласуется с χ^2 -распределением.

Пример 18.3 В табл. 18.2 приведены данные о распределении цвета волос на голове и бровей у 46542 человек. Проверим на уровне значимости $\alpha = 0,05$ гипотезу о независимости этих признаков. Здесь $n = 46592$, $r = s = 2$, $n_{11} = 30472$, $n_{12} = 3238$, $n_{21} = 3364$, $n_{22} = 9468$, $n_1 = 33710$, $n_2 = 12832$, $n_1 = 33836$, $n_2 = 12706$,

Цвет бровей	Цвет волос на голове		Сумма
	светлые	темные	
Светлые	30472	3238	33710
Темные	3364	9468	12832
Сумма	33836	12706	46542

число степеней свободы $(r-1)(s-1) = 1$. Из (18.9) получаем $\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = 19,288$. По таблице квантилей χ^2 -распределения находим $\chi_{0,95}^2(1) = 3,84$. Так как $19,288 > 3,84$, то гипотезу о независимости признаков следует отклонить.

Таблица 18.2.

Лекция 19

Метод наименьших квадратов

Рассмотрим задачу о подборе функции одного переменного - подборе по неточным наблюдениям (измерениям). Предположим, что переменные y и x_1, \dots, x_p связаны линейным соотношением

$$y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p,$$

где коэффициенты $\theta = (\theta_1, \dots, \theta_p)$ неизвестны. При некоторых значениях $x_{i1}, x_{i2}, \dots, x_{ip}$, $i = \overline{1, n}$, переменных x_1, \dots, x_p (называемых обычно факторами) были произведены измерения переменной y (называемой откликом) со случайной ошибкой ε_i , так что вместо неслучайных величин

$$y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_p x_{ip}, \quad i = \overline{1, n},$$

наблюдались случайные величины

$$Y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_p x_{ip} + \varepsilon_i, \quad i = \overline{1, n}. \quad (19.1)$$

Возникает задача оценивания неизвестных коэффициентов $\theta = (\theta_1, \dots, \theta_p)$ по наблюдениям $Y = (y_1, y_2, \dots, y_n)^T$ и элементам x_{ij} матрицы X размера $n \times m$.

Основное предположение об ошибках состоит в том, что случайные величины $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ считаются независимыми и $E\varepsilon_i = 0$, т.е. систематических ошибок при измерении отклика нет. Менее важные предположения заключаются в том, что ε_i распределены одинаково и по нормальному закону $N(0, \sigma^2)$. Величина σ обычно считается неизвестной. Она численно выражает неточность (изменчивость) измерений, т.е. масштаб случайных ошибок.

Систему (19.1) можно записать в матричном виде

$$Y = X\theta + \varepsilon. \quad (19.2)$$

Один из способов оценивания коэффициентов $\theta = (\theta_1, \dots, \theta_p)$, называемый методом наименьших квадратов состоит в следующем.

Определение 19.1 Оценкой $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$ параметра $\theta = (\theta_1, \dots, \theta_p)$ по методу наименьших квадратов называется точка минимума функции

$$S(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T(Y - X\theta) = \sum_{i=1}^n (Y_i - \theta_1 x_{i1} - \theta_2 x_{i2} - \dots - \theta_p x_{ip})^2.$$

Теорема 19.1 Предположим, что ранг матрицы X равен p . Тогда оценка наименьших квадратов имеет вид

$$\hat{\theta} = (X^T X)^{-1} X^T Y. \quad (19.3)$$

Теорема 19.2 Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ — независимые одинаково распределенные случайные величины с $M\varepsilon_i = 0$ и конечной дисперсией $D\varepsilon_i = \sigma^2$. Тогда оценка наименьших квадратов

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

является несмещенной и состоятельной оценкой параметра $\theta = (\theta_1, \dots, \theta_p)$.

Обозначим

$$S(\theta) = (Y - X\theta)^T(Y - X\theta),$$

(d_1, d_2, \dots, d_p) — диагональные элементы матрицы $(X^T X)^{-1}$.

Теорема 19.3 Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ — независимые одинаково распределенные нормальные случайные величины с $M\varepsilon_i = 0$ и конечной дисперсией $D\varepsilon_i = \sigma^2$. Тогда оценка наименьших квадратов

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

является несмещенной, состоятельной оценкой параметра $\theta = (\theta_1, \dots, \theta_p)$ и нормальным случайным вектором с математическим ожиданием $\theta = (\theta_1, \dots, \theta_p)$ и ковариационной матрицей $\sigma^2(X^T X)^{-1}$. Интервальная оценка для θ_j уровня доверия $1 - \alpha$ имеет вид $(\hat{\theta}_j - \Delta, \hat{\theta}_j + \Delta)$, где

$$\Delta = t_{1-\alpha}(n-p) \sqrt{\frac{d_j}{n-p} S(\hat{\theta})},$$

а $t_{1-\alpha}(n-p)$ — квантиль распределения Стьюдента уровня $1 - \alpha$ с $n - p$ степенями свободы.

Рассмотрим теперь задачу оценивания зависимости

$$y = \theta_1 \varphi_1(t) + \theta_2 \varphi_2(t) + \dots + \theta_p \varphi_p(t),$$

считая функции $\varphi_1, \varphi_2, \dots, \varphi_p$ известными, по измерениям $Y = (Y_1, Y_2, \dots, Y_n)$ величины y в неслучайных точках t_1, t_2, \dots, t_n со случайными ошибками $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$:

$$Y_i = \theta_1 \varphi_1(t_i) + \theta_2 \varphi_2(t_i) + \dots + \theta_p \varphi_p(t_i) + \varepsilon_i, \quad i = \overline{1, n}. \quad (19.4)$$

Обозначив

$$x_{ij} = \varphi_j(t_i), \quad i = \overline{1, n}, j = \overline{1, p},$$

сведем модель (19.4) к модели (19.2).

Пример 19.1 В “Основах химии”

Д. И. Менделеев приводит следующие данные о количестве y азот-

натриевой

соли

Таблица 19.1.
NaNO₃,

которое

можно растворить в 100 г воды в зависимости от температуры t (см. таб. 19.1). Построим по этим данным приближенную эмпирическую формулу вида

$$y = \theta_1 + \theta_2 t + \theta_3 t^2,$$

описывающую зависимость между рассматриваемыми величинами.

Оценим коэффициенты $(\theta_1, \theta_2, \theta_3)$ по $n = 9$ наблюдениям (y_1, y_2, \dots, y_n) случайных величин (Y_1, Y_2, \dots, Y_n) . В этом случае

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 4 & 10 & 15 & 21 & 29 & 36 & 51 & 68 \\ 0 & 16 & 100 & 225 & 441 & 841 & 1296 & 2601 & 4624 \end{pmatrix},$$

$$X^T X = \begin{pmatrix} 9 & 234 & 10144 \\ 234 & 10144 & 531828 \\ 10144 & 531828 & 30788836 \end{pmatrix},$$

$$(X^T X)^{-1} = \begin{pmatrix} 0.4878864808 & -0.0299495315 & 0.0003565864 \\ -0.0299495315 & 0.0028828545 & -0.0000399292 \\ 0.0003565864 & -0.0000399292 & 0.0000006047 \end{pmatrix},$$

$$\hat{\theta} = (66.71, 0.9604, -0.001359),$$

$$y \approx 66.71 + 0.9604t - 0.001359t^2.$$

Оглавление

1	Случайные события	1
2	Вероятность	6
3	Условная вероятность	11
4	Формула полной вероятности. Формула Байеса. Схема Бернулли	16
5	Одномерные случайные величины	20
6	Числовые характеристики случайных величин	26
7	Основные законы распределения случайных величин	30
8	Случайные векторы	35
9	Функции от случайных величин	41
10	Ковариация и коэффициент корреляции случайных величин	46
11	Условные характеристики случайных величин	50
12	Многомерное нормальное распределение	55
13	Предельные теоремы теории вероятностей	60
14	Основные понятия выборочной теории	65
15	Точечные оценки	69
16	Интервальные оценки и доверительные интервалы	72
17	Проверка гипотез. Параметрические модели	79
18	Проверка непараметрических гипотез	86
19	Метод наименьших квадратов	91