Оглавление

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 ω_1, $\omega_2, \ldots, \omega_n, \ldots$, и только из них, записывают в виде $\Omega = \{\omega_1; \omega_2; \ldots; \omega_n; \ldots\}$ или в виде $\Omega = \{\omega_i, \quad i = 1, 2, \ldots, n, \ldots\}$. 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 Г, $\omega_{\text{г}}$ или ω_1) и выпадением “цифры” (Ц, $\omega_{\text{ц}}$ или ω_2). Таким образом, $\Omega = \{\text{Г, Ц}\}$, $\Omega = \{\omega_{\text{г}}, \omega_{\text{ц}}\}$ 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. $\Omega = \{\omega_{\text{гг}}, \omega_{\text{гц}}, \omega_{\text{цг}}, \omega_{\text{цц}}\}$, где $\omega_{\text{гг}}$ — 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 $\omega_1, \ldots, \omega_6$, где ω_i, $i = \overline{1{,}6}$, означает появление i очков на верхней грани кости, т.е. $\Omega = \{\omega_i, \quad i = \overline{1{,}6}\}$.

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. $\Omega = \{\omega_{ij}, \quad i, j = \overline{1{,}6}\}$, 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. $\Omega = \{0, 1, \ldots, n, \ldots\}$.

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 Ω, 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 $\Omega = \{\omega_i, \quad i = \overline{1,6}\}$, 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 $A = \{\omega_2;\omega_4;\omega_6\}, \quad B = \{\omega_1;\omega_3;\omega_5\}$ 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 вданном 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.

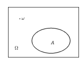
Рис 1.1.

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.

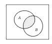
Рис 1.2.

Пересечение событий A и B записывают следующим образом:

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

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***.

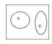
Рис 1.3.

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.

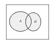
Рис 1.4.

Объединение событий A и B записывают в виде

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

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

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

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

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

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

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:

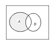
Рис 1.5.

$$C = A \setminus B.$$

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

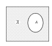
Рис 1.6.

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

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

эквивалентна формуле

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

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 $A \setminus B = A\overline{B}$.

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.

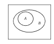
Рис 1.7.

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

Основные свойства операций над событиями

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

1. Коммутативность суммы: $A+B = B+A$.
2. Коммутативность произведения: $AB = BA$.
3. Ассоциативность суммы: $A+B+C = A+(B+C)$.
4. Ассоциативность произведения: $(AB)C = A(BC)$.
5. $A+A = A$.
6. $AA = A$.
7. $A+\Omega = \Omega$.
8. $A\Omega = A$.
9. Включение A в B, т.е. $A \subset B$, влечет за собой включение $\overline{B}$ в $\overline{A}$, т.е. $\overline{A} \supset \overline{B}$.
10. Совпадение двойного дополнения с исходным событием: $\overline{\overline{A}} = A$.
11. ***Законы де Моргана***: $\overline{A+B} = \overline{A}\,\overline{B}, \qquad \overline{AB} = \overline{A}+\overline{B}$.

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

$$\overline{A_1+A_2+\ldots+A_n} = \overline{A}_1\overline{A}_2\ldots\overline{A}_n$$
$$\overline{A_1A_2\ldots A_n} = \overline{A}_1+\overline{A}_2+\ldots+\overline{A}_n.$$

Сигма-алгебра событий

В предыдущем параграфе мы назвали 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 $\mathscr{A}$. 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 $\mathscr{A}$ назовем систему подмножеств пространства элементарных исходов Ω, удовлетворяющую следующим условиям:

1. $\Omega \in \mathscr{A}$.
2. Если подмножество A принадлежит $\mathscr{A}$, то дополнение $\overline{A}$ принадлежит $\mathscr{A}$.
3. Если подмножества $A_1, A_2, \ldots, A_n, \ldots$ принадлежат $\mathscr{A}$, то их сумма $A_1+A_2+\ldots+A_n+\ldots$ и их произведение $A_1A_2\ldots A_n\ldots$ принадлежат $\mathscr{A}$.

Заметим, что поскольку $\Omega \in A$ и $\emptyset = \overline{\Omega}$, то *невозможное событие* $\emptyset$ принадлежит $\mathscr{A}$.

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 1.7. 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:

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

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 $\mathsf{P}(A)$, которое будет являться мерой возможности его появления. Если принять $\mathsf{P}(\Omega) = 1$, а $\mathsf{P}(\emptyset) = 0$ (хотя можно было взять другую единицу измерения), то тогда для любого события A естественно ожидать, что $0 \leqslant \mathsf{P}(A) \leqslant 1$.

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***".

Назовем мощностью $|A|$ события A число элементарных исходов, принадлежащих A (или, как говорят, *благоприятствующих* событию A).

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

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|}.$$

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

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 $\mathsf{P}(A)$ 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 $k + l = 30$, причем все исходы равновозможны, а число благоприятствующих событию A исходов $k = 10$. Поэтому в соответствии с определением классической вероятности $\mathsf{P}(A) = \dfrac{k}{k+l} = \dfrac{1}{3}$. #

Используя классическое определение вероятности события, докажем следующие свойства.

Свойство 2.1. Для любого события A вероятность удовлетворяет неравенству $\mathsf{P}(A) \geqslant 0$.
Доказательство. Свойство очевидно, так как отношение $|A|/|\Omega|$ не может быть отрицательным.

Свойство 2.2. Для достоверного события Ω (которое содержит все элементарные исходы) $\mathsf{P}(\Omega) = 1$.

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

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,

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

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

$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A); \qquad \mathsf{P}(\emptyset) = 0;$$

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

Недостаток классического определения заключается в том, что оно применимо только к пространствам элементарных исходов, состоящим из конечного числа равновозможных 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 $\mathsf{P}(A)$, равное отношению меры множества A к мере множества Ω:

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

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 $\mathsf{P}(A)$ 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

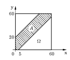

Рис 2.1.

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

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 $\mathsf{P}(A) = \dfrac{S_A}{S_\Omega} = \dfrac{1287{,}5}{3600} \approx 0{,}36.$

Статистическое определение вероятности (определение фон Мизеса)

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.

Определение 2.3. Пусть произведено n повторений опыта, причем в n_A из них появилось событие A. Число n_A/n назовем ***частотой 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 2.4. ***Вероятностью события*** A называют (эмпирический) предел $\mathsf{P}(A)$, к которому стремится частота n_A/n 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) $\mathsf{P}(A) \geqslant 0$;

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

3) $\mathsf{P}(A+B) = \mathsf{P}(A) + \mathsf{P}(B)$, 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 наблюденная 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).

Аксиоматическое определение вероятности

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

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

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

Вероятность $\mathsf{P}(A)$ можно записать также в следующем виде

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

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) $\mathsf{P}(A) \geqslant 0$;

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

3) $\mathsf{P}(A_1 + \ldots + A_m) = \mathsf{P}(A_1) + \cdots + \mathsf{P}(A_m)$, если *события* $A_1, \ldots, A_m$ *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 2.5. Пусть $\mathscr{A}$ — σ-алгебра событий на пространстве элементарных событий Ω. Назовем ***вероятностью*** числовую функцию P, определенную на $\mathscr{A}$ и удовлетворяющую следующим условиям:

Аксиома 1 (*аксиома неотрицательности*): $\mathsf{P}(A) \geqslant 0$ для всех $A \in \mathscr{A}$;

Аксиома 2 (*аксиома нормированности*): $\mathsf{P}(\Omega) = 1$;

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

$$\mathsf{P}(A_1 + \ldots + A_n + \ldots) = \mathsf{P}(A_1) + \ldots + \mathsf{P}(A_n) + \ldots$$

Определение 2.6. Число $\mathsf{P}(A)$ называют ***вероятностью события*** A.

Определение 2.7. Тройку $(\Omega, \mathscr{A}, \mathsf{P})$, состоящую из пространства элементарных исходов Ω, σ-алгебры событий $\mathscr{A}$ и определенной на $\mathscr{A}$ вероятности P, называют ***вероятностным пространством***.

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

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

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

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

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

2. Вероятность невозможного события $\mathsf{P}(\emptyset) = 0$.

3. Если $A \subset B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$ *("бóльшему" событию соответствует бóльшая вероятность).*

4. Вероятность заключена между 0 и 1: $0 \leqslant \mathsf{P}(A) \leqslant 1$.

5. Вероятность суммы двух событий $\mathsf{P}(A+B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$.

6. Вероятность суммы любого конечного числа событий

$$\mathsf{P}(A_1 + \ldots + A_n) = \mathsf{P}(A_1) + \ldots + \mathsf{P}(A_n) - \mathsf{P}(A_1A_2) - \mathsf{P}(A_1A_3) - \ldots - \mathsf{P}(A_{n-1}A_n) + \\ + \mathsf{P}(A_1A_2A_3) + \ldots + (-1)^{n+1}\mathsf{P}(A_1A_2\ldots A_n).$$

Доказательство. Поскольку $\Omega = A + \overline{A}$, то, согласно расширенной аксиоме сложения, $\mathsf{P}(\Omega) = \mathsf{P}(A) + \mathsf{P}(\overline{A})$, 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 $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$. 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 $A + B = A + (B \setminus A), \quad B = (B \setminus A) + AB$, то, используя расширенную аксиому сложения, находим $\mathsf{P}(A+B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$ и $\mathsf{P}(B) = \mathsf{P}(B \setminus A) + \mathsf{P}(AB)$. Подставляя в первое из последних двух равенств вероятность $\mathsf{P}(B \setminus A)$, 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

$$\mathsf{P}(A+B+C) = \mathsf{P}(A) + \mathsf{P}(B+C) - \mathsf{P}(A(B+C)) = \\ = \mathsf{P}(A) + \mathsf{P}(B) + \mathsf{P}(C) - \mathsf{P}(BC) - \mathsf{P}(AB+AC) = \\ = \mathsf{P}(A) + \mathsf{P}(B) + \mathsf{P}(C) - \mathsf{P}(BC) - \mathsf{P}(AB) - \mathsf{P}(AC) + \mathsf{P}(ABC).$$

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 $A = A_1 + A_2$, и в соответствии с классической схемой вероятности $\mathsf{P}(A_1) = \mathsf{P}(A_2) = \dfrac{1}{2}$. Если не учитывать, что A_1 и A_2 — совместные события, то можно получить следующее неверное (!) заключение

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

противоречащее здравому смыслу, поскольку ясно, что *событие* A не является *достоверным*. Применяя теорему сложения для двух совместных событий и учитывая равенство $\mathsf{P}(A_1A_2) = \dfrac{1}{4}$, находим

$$\mathsf{P}(A) = \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1A_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}. \quad \#$$

Лекция 3

Условная вероятность. Независимость событий. Схема Бернулли

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

Понятие условной вероятности играет важнейшую роль в современной теории вероятностей.

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 возможности появления 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*** $\mathsf{P}(A|B)$ 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. $\mathsf{P}(A|B) = \dfrac{N_{AB}}{N_B}$.

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

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:

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

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

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

Рассмотрим теперь условную вероятность $\mathsf{P}(A|B)$ как функцию события A.

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

Доказательство. Для доказательства достаточно показать, что условная вероятность $\mathsf{P}(A|B)$ удовлетворяет аксиомам 1, 2 и 3 ((см. определение 2.5)).

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. Трехтомник стихотворений расставляют на полке в случайном порядке. Определим события $A = \{$первый том попадет на первое место$\}$ и $B = \{$второй том попадет на второе место$\}$. Найдем безусловные вероятности $\mathsf{P}(A)$ и $\mathsf{P}(B)$ и условную вероятность $\mathsf{P}(A|B)$.

Понятно, что $\mathsf{P}(A) = 1/3$ и $\mathsf{P}(B) = 1/3$. Для вычисления условной вероятности применим два способа.

Первый способ. В соответствии с определением условной вероятности имеем

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \frac{1/6}{1/3} = \frac{1}{2}.$$

Второй способ. Перейдем к новому пространству Ω_1 элементарных исходов. Так как событие B произошло, то это означает, что в новом пространстве элементарных исходов всего два равновозможных исхода, т.е. $|\Omega_1| = 2$, а событию A благоприятствует при этом лишь один исход. Следовательно, $\mathsf{P}(A|B) = \dfrac{1}{2}$.

Пример 3.2. 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

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

Следовательно, в силу определения 3.1 условной вероятности имеем

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

что подтверждает наше предположение.

Геометрическая интерпретация условной вероятности

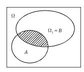

Рис 3.1.

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):

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

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

Формула умножения вероятностей

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

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

$$\mathsf{P}(A) = \mathsf{P}(A_1)\mathsf{P}(A_2|A_1)\mathsf{P}(A_3|A_1A_2)\ldots\mathsf{P}(A_n|A_1A_2\ldots A_{n-1}),$$

называемое ***формулой умножения вероятностей****.*

Доказательство. Поскольку $\mathsf{P}(A) = \mathsf{P}(A_1A_2\ldots A_n) > 0$, а $A_1A_2\ldots A_k \supseteq A_1A_2\ldots A_n$ $(k = \overline{1, n-1})$, то и $\mathsf{P}(A) = \mathsf{P}(A_1A_2\ldots A_k) > 0$. Учитывая это неравенство, согласно определению 3.1 *условной вероятности*, имеем

$$\mathsf{P}(A_n|A_1A_2\ldots A_{n-1}) = \frac{\mathsf{P}(A_1A_2\ldots A_n)}{\mathsf{P}(A_1A_2\ldots A_{n-1})}.$$

Умножая обе части этого равенства на $\mathsf{P}(A_1A_2\ldots A_{n-1})$, получаем

$$\mathsf{P}(A_1A_2\ldots A_n) = \mathsf{P}(A_1A_2\ldots A_{n-1})\mathsf{P}(A_n|A_1A_2\ldots A_{n-1}).$$

Аналогично находим $\mathsf{P}(A_1A_2\ldots A_{n-1}) = \mathsf{P}(A_1A_2\ldots A_{n-2})\mathsf{P}(A_{n-1}|A_1A_2\ldots A_{n-2})$. Тогда

$$\mathsf{P}(A_1A_2\ldots A_n) = \mathsf{P}(A_1A_2\ldots A_{n-2})\mathsf{P}(A_{n-1}|A_1A_2\ldots A_{n-2}) \times \mathsf{P}(A_n|A_1A_2\ldots A_{n-1}).$$

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. Найдем 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 $\mathsf{P}(A) = \mathsf{P}(A_1A_2A_3) = \mathsf{P}(A_1)\mathsf{P}(A_2|A_1)\mathsf{P}(A_3|A_1A_2)$. Согласно *классическому определению* 2.1 *вероятности*, имеем $\mathsf{P}(A_1) = \dfrac{1}{7}$.

Если событие A_1 произошло, то на шести оставшихся карточках буква “О” встречается два раза, поэтому условная вероятность $\mathsf{P}(A_2|A_1) = \dfrac{2}{6} = \dfrac{1}{3}$. Аналогично определяем $\mathsf{P}(A_3|A_1A_2) = \dfrac{1}{5}$. Окончательно получаем $\mathsf{P}(A) = \mathsf{P}(A_1A_2A_3) = \dfrac{1}{7}\cdot\dfrac{1}{3}\cdot\dfrac{1}{5} = \dfrac{1}{105} \approx 0{,}0095$.

Независимые и зависимые события

Из рассмотренных выше примеров видно, что *условная вероятность* $\mathsf{P}(A|B)$ *события* A при условии, что событие B произошло, может как совпадать с *безусловной вероятностью* $\mathsf{P}(A)$, 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 $\mathsf{P}(A|B)$, $\mathsf{P}(B|A)$ с безусловными.

Определение 3.2. Пусть $\mathsf{P}(B) > 0$. ***События*** A и B называют ***независимыми***, если условная вероятность A при условии B совпадает с безусловной вероятностью A, т.е.

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

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

Теорема 3.3. *События A и B, имеющие ненулевую вероятность, являются независимыми тогда и только тогда, когда*
$$\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B). \tag{3.3}$$

Доказательство. Пусть выполнено равенство (3.2). Воспользовавшись *формулой умножения вероятностей* для двух событий, получим

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

Обратно, пусть выполнено равенство (3.3). Тогда, согласно определению 3.1 условной вероятности,
$$\mathsf{P}(A|B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \mathsf{P}(A),$$
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 (3.3).

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

Замечание 3.1. Отметим, что если выполняется равенство (3.2) и $\mathsf{P}(A) > 0$, то в силу теоремы 3.3 равенство $\mathsf{P}(B|A) = \mathsf{P}(B)$ выполняется автоматически.

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

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

т.е. выполняется равенство (3.2), и поэтому события A и B независимы. #

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

Теорема 3.4. *Если события A и B независимые, то независимыми также являются пары событий $\overline{A}$ и B, A и $\overline{B}$, $\overline{A}$ и $\overline{B}$, если вероятности соответствующих событий ненулевые.*

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

Определение 3.4. ***События*** $A_1, A_2, \ldots, A_n$ называют ***независимыми в совокупности***, если для всех комбинаций индексов $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$ при любом $k = 2, 3, \ldots, n$ выполняется равенство
$$\mathsf{P}(A_{i_1} \ldots A_{i_k}) = \mathsf{P}(A_{i_1}) \ldots \mathsf{P}(A_{i_k}). \tag{3.4}$$

Другими словами 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.

Замечание 3.2. Если равенство (3.4) выполняется только для $k = 2$, только говорят о ***попарной независимости событий*** из этой совокупности.

Замечание 3.3. В силу определения независимости событий в совокупности *формула умножения вероятностей* для независимых в совокупности событий имеет вид $\mathsf{P}(A_1A_2\ldots A_n) = \mathsf{P}(A_1)\mathsf{P}(A_2)\ldots\mathsf{P}(A_n)$. #

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

$$\mathsf{P}(A_i) = \frac{2}{4} = \frac{1}{2}, \quad i = \overline{1,3}, \quad \mathsf{P}(A_2|A_1) = \frac{\mathsf{P}(A_1A_2)}{\mathsf{P}(A_1)} = \frac{1/4}{2/4} = \frac{1}{2}.$$

Аналогично $\mathsf{P}(A_i|A_j) = \dfrac{1}{2}$ 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, $\mathsf{P}(A_1|A_2A_3) = \dfrac{\mathsf{P}(A_1A_2A_3)}{\mathsf{P}(A_2A_3)} = \dfrac{1/4}{1/4} = 1 \neq \mathsf{P}(A_1)$, т. е. события A_1, A_2 и A_3 зависимы в совокупности. #

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

Замечание 3.4 (о связи между совместными и зависимыми событиями)**.** Между понятиями "несовместные" и "независимые" события имеется следующая связь:

1) если A и B — несовместные события (и $\mathsf{P}(A) \neq 0$, и $\mathsf{P}(B) \neq 0$), 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. #

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 3.5. ***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* $\overline{A}$, 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 $\mathsf{P}(A) = p$.

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

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 стрелкá по мишени можно лишь приближенно рассматривать как схему испытаний Бернулли, так как независимость результатов стрельбы может нарушаться либо из-за “пристрелки” спортсмена, либо вследствии 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 $\mathsf{P}(A_k)$ через $P_n(k)$.

Теорема 3.5. *Вероятность $P_n(k)$ того, что в n испытаниях по схеме Бернулли произойдет ровно k успехов, определяется **формулой Бернулли***

$$P_n(k) = C_n^k p^k q^{n-k}, \quad k = \overline{0,n}. \tag{3.5}$$

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 $\omega =$УНН...У можно поставить в соответствие вероятность $\mathsf{P}(\omega) = \mathsf{P}(\text{УНН...У})$. В силу независимости испытаний события У,Н,Н,...,У являются независимыми в совокупности, и потому по теореме умножения вероятностей имеем $\mathsf{P}(\omega) = p^i q^{n-i}, \quad i = \overline{0,n}$, 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 $\mathsf{P}(A_k) = P_n(k)$ формулу (3.5).

Формулу (3.5) называют также ***биномиальной***, так как ее правая часть представляет собой $(k+1)$-й член формулы бинома Ньютона.

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

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:

$$\mathsf{P}\{k_1 \leqslant k \leqslant k_2\} = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}. \tag{3.6}$$

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 (3.6) получаем формулу для вычисления вероятности хотя бы одного успеха в n испытаниях:

$$\mathsf{P}\{k \geqslant 1\} = 1 - q^n. \tag{3.7}$$

Пример 3.6. Монету (симметричную) подбрасывают $n = 10$ раз. Определим вероятность выпадения “герба”: а) ровно пять раз; б) не более пять раз; в) хотя бы один раз.

В соответствии с формулой (3.5) Бернулли имеем:

$$\text{а)}\quad P_{10}(5) = C_{10}^{5}\left(\frac{1}{2}\right)^{10} = \frac{252}{1024} = 0{,}246;$$

$$\text{б)}\quad \mathsf{P}\{k \leqslant 5\} = \frac{C_{10}^{0} + C_{10}^{1} + C_{10}^{2} + C_{10}^{3} + C_{10}^{4} + C_{10}^{5}}{1024} = \frac{638}{1024} \approx 0{,}623;$$

$$\text{в)}\quad \mathsf{P}\{k \geqslant 1\} = 1 - \left(\frac{1}{2}\right)^{10} \approx 0{,}999.$$

Лекция 4

Формула полной вероятности. Формула Байеса

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. $H_i \cdot H_j = \emptyset$ при $i \neq j$;

2) сумма этих событий есть *достоверное событие*, т.е. $H_1 + \ldots + H_n = \Omega$.

Определение 4.1. События $H_1, H_2, \dots, H_n$ удовлетворяющие условиям 1 и 2, называют ***полной группой событий*** или ***гипотезами***.

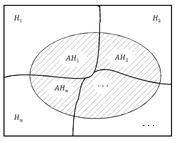

Рис 4.1

Пусть также имеется некоторое событие A и известны *вероятности* гипотез $\mathsf{P}(H_1), \dots, \mathsf{P}(H_n)$, которые предполагаются ненулевыми, и *условные вероятности* $\mathsf{P}(A|H_1), \dots, \mathsf{P}(A|H_n)$ события A при выполнении этих гипотез. Задача состоит в вычислении *безусловной вероятности* события A. Решить такую задачу можно с помощью формулы полной вероятности, которая доказана в следующей теореме.

Теорема 4.1 (*Формула полной вероятности*). *Пусть для некоторого события A и гипотез $H_1, \dots, H_n$ известны $\mathsf{P}(H_1), \dots, \mathsf{P}(H_n)$, которые положительны, и $\mathsf{P}(A|H_1), \dots, \mathsf{P}(A|H_n)$. Тогда*

$$\mathsf{P}(A) = \mathsf{P}(H_1)\mathsf{P}(A|H_1) + \ldots + \mathsf{P}(H_n)\mathsf{P}(A|H_n). \tag{4.1}$$

Доказательство. Представим событие A в виде

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

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

$$\mathsf{P}(A) = \mathsf{P}(AH_1) + \ldots + \mathsf{P}(AH_n).$$

В соответствии с *формулой умножения вероятностей* получаем

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

Поэтому

$$\mathsf{P}(A) = \mathsf{P}(H_1)\mathsf{P}(A|H_1) + \ldots + \mathsf{P}(H_n)\mathsf{P}(A|H_n).$$

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. Найдем вероятность события A — достижения путником намеченной цели.

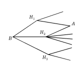

Рис 4.2.

Введем гипотезы H_i, где H_i означает, что путник выбрал в пункте B дорогу с номером i, $i = 1, 2, 3$. Ясно, что события H_i несовместные и одно из них обязательно происходит, причем в силу равновозможности выбора дорог $\mathsf{P}(H_i) = \dfrac{1}{3}$, $i = 1, 2, 3$.

Остается вычислить условные вероятности $\mathsf{P}(A|H_i)$, которые легко найти, если рассматривать новое *пространство элементарных исходов*, соответствующее выбранной гипотезе H_i.

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. $\mathsf{P}(A|H_1) = \dfrac{1}{2}$. Аналогично находим, что $\mathsf{P}(A|H_2) = \dfrac{1}{4}$ и $\mathsf{P}(A|H_3) = 0$.

Согласно формуле 4.1 полной вероятности, получаем

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

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

Пример 4.2. Студент Иванов должен выучить к экзамену N билетов, но выучил лишь $m < N$ билетов. Найти вероятность того, что Иванов вытащит счастливый билет, если: а) он идет на экзамен первым; б) он идет на экзамен вторым.

Введем событие $A = \{$Иванов вытащит счастливый билет$\}$.

а) $\mathsf{P}(A) = \frac{m}{N}$.

б) Если Иванов идет на экзамен вторым, то первый студент может вытащить счастливый билет Иванова (обозначим это событие H_1) или несчастливый билет Иванова (обозначим это событие H_2). Понятно, что события H_1 и H_2 образуют полную группу событий, и $\mathsf{P}(H_1) = \frac{m}{N}$, $\mathsf{P}(H_2) = \frac{N-m}{N}$, а $\mathsf{P}(A|H_1) = \frac{m-1}{N-1}$, $\mathsf{P}(A|H_2) = \frac{m}{N-1}$. Тогда согласно формуле полной вероятности (4.1)

$$\mathsf{P}(A) = \mathsf{P}(H_1)\mathsf{P}(A|H_1)+\mathsf{P}(H_2)\mathsf{P}(A|H_2) = \frac{m}{N}\cdot\frac{m-1}{N-1}+\frac{N-m}{N}\cdot\frac{m}{N-1} = \frac{m}{N}.$$

Формула Байеса

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

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

$$\mathsf{P}(H_i|A) = \frac{\mathsf{P}(H_i)\mathsf{P}(A|H_i)}{\mathsf{P}(H_1)\mathsf{P}(A|H_1)+\ldots+\mathsf{P}(H_n)\mathsf{P}(A|H_n)}. \tag{4.2}$$

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

Подставляя вместо вероятности $\mathsf{P}(A)$ 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*** $\mathsf{P}(H_1),\ldots,\mathsf{P}(H_n)$ обычно называют ***априорными*** (т.е. полученными “до опыта”), а условные ***вероятности*** $\mathsf{P}(H_1|A)$, $\ldots$, $\mathsf{P}(H_n|A)$ — ***апостериорными*** (т.е. полученными “после опыта”).

Пример 4.3. При посадке разбился самолет. Комиссия установила, что причина аварии — дефект навигационного прибора. Этот прибор поставляется двумя заводами. Завод № 1 поставляет 65% приборов, а завод № 2 — 35%. Надежность (вероятность безотказной работы) приборов завода № 1 равна 0.9, завода № 2 — 0.8. При сборке самолета приборы выбирались из имеющихся случайным образом. Какой из заводов (завод № 1 или завод № 2) является наиболее вероятным виновником аварии?

Введем событие $A = \{$изготовлен дефектный навигационный прибор$\}$. События $H_i = \{$навигационный прибор изготовлен заводом № $i\}$, $i = 1,2$, составляют полную группу событий. Априорные вероятности $\mathsf{P}(H_1) = 0.65$, $\mathsf{P}(H_2) = 0.35$.

Для решения задачи требуется найти апостериорные вероятности $\mathsf{P}(H_1|A)$ и $\mathsf{P}(H_2|A)$. Согласно формуле Байеса

$$\mathsf{P}(H_i|A) = \frac{\mathsf{P}(H_i)\mathsf{P}(A|H_i)}{\mathsf{P}(A)}, \qquad i = 1,2.$$

По условию $\mathsf{P}(A|H_1) = 0.1$, $\mathsf{P}(A|H_2) = 0.2$. Тогда

$$\mathsf{P}(H_1|A) = \frac{0.1 \cdot 0.65}{0.1 \cdot 0.65 + 0.2 \cdot 0.35} = \frac{13}{27},$$

$$\mathsf{P}(H_2|A) = \frac{0.2 \cdot 0.35}{0.1 \cdot 0.65 + 0.2 \cdot 0.35} = \frac{14}{27}.$$

Таким образом, $\mathsf{P}(H_2|A) > \mathsf{P}(H_1|A)$. Следовательно, наиболее вероятным виновником аварии является завод № 2.

Пример 4.4. 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: $\mathsf{P}(H_1) = 0{,}4$ и $\mathsf{P}(H_2) = 0{,}6$, а условные вероятности события A при наличии гипотез H_1 и H_2 равны 0,9 и 0,2 соответственно. Используя формулу Байеса, находим $\mathsf{P}(H_i|A) = \dfrac{0{,}4 \cdot 0{,}9}{0{,}4 \cdot 0{,}9 + 0{,}6 \cdot 0{,}2} = 0{,}75$. Итак, врач с большей уверенностью признает наличие заболевания 1.

Лекция 5

Случайные величины (СВ). Одномерные случайные величины

5.1 Определение случайной величины

Случайной величиной называют 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. Обозначать случайные величины (СВ) принято греческими буквами ξ, η, ζ т.д.

Рассмотрим примеры.

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

$$\{1;2;\ldots;6\}.$$

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

$$\begin{array}{rcl} \Omega & = & \{\omega_1 \quad \omega_2 \quad \ldots \quad \omega_6\} \\ & & \ \ \downarrow \quad\ \ \downarrow \quad\ \ \ldots \quad \downarrow \\ \xi & = & \{1 \quad\ \ 2 \quad\ \ \ldots \quad 6\}. \end{array}$$

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

Пример 5.2. Монету подбрасывают до первого появления “орла”. В этом опыте можно ввести, например, случайную величину ξ, равную количеству подбрасываний до первого появления “орла”. Множество возможных значений этой случайной величины есть $\{1;2;3;\ldots\}$. В данном опыте пространство элементарных исходов Ω можно отождествить с множеством

$$\begin{array}{rclcl} \Omega & = & \{O,\ PO,\ PPO,\ \ldots\} & = & \{\omega_1,\ \omega_2,\ \omega_3,\ \ldots\} \\ & & & & \ \ \downarrow \quad\ \ \downarrow \quad\ \ \downarrow \quad\ \ \ldots \\ \xi & = & & & \{1,\quad 2,\quad 3,\quad \ldots\} \end{array}$$

Таким образом, каждому элементарному исходу ω_m ставится в соответствие число m, равное числу подбрасываний монеты до первого появления орла.

Пример 5.3. Стрелок производит выстрел по концентрической мишени и попадает в нее. В качестве случайной величины ξ можно выбрать, например, евклидово расстояние от центра мишени до точки попадания. #

Определение 5.1. ***Случайной величиной*** ξ называют любую функцию $\xi : \Omega \to \mathbb{R}$, у которой для всякого $x \in \mathbb{R}$ множество $\{\omega \in \Omega : \xi(\omega) \leqslant x\}$ принадлежит σ-алгебре событий $\mathscr{A}$.

5.2 Функция распределения случайной величины

Определение 5.1 гарантирует, что при любом x неравенство $\xi \leqslant x$ есть событие и, следовательно, имеет смысл говорить о его вероятности.

Определение 5.2. ***Функцией распределения*** (***вероятностей***) случайной величины ξ называют функцию $F_\xi(x)$, значение которой в точке x равно вероятности события $\{\xi \leqslant x\}$, т.е.

$$F_\xi(x) = \mathsf{P}\{\xi \leqslant x\}.$$

Далее, если будет понятно, о какой случайной величине идет речь, мы для простоты записи будем обозначать $F_\xi(x) = F(x)$.

Теорема 5.1. *Функция распределения удовлетворяет следующим свойствам:*

1. $0 \leqslant F(x) \leqslant 1$.

2. $F(x_1) \leqslant F(x_2)$ *при* $x_1 < x_2$, *т.е.* $F(x)$ *— неубывающая функция.*

3. $F(-\infty) = \lim\limits_{x\to-\infty} F(x) = 0; \quad F(+\infty) = \lim\limits_{x\to+\infty} F(x) = 1.$

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

5. $F(x_0) = \lim\limits_{\substack{\varepsilon\to 0\\ \varepsilon>0}} F(x_0+\varepsilon)$, *т.е.* $F(x)$ *— непрерывная справа 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 $\{\xi \leqslant x_1\}$ включено в событие $\{\xi \leqslant x_2\}$ и, согласно свойству 3 из теоремы 2.1,

$$\mathsf{P}\{\xi \leqslant x_1\} \leqslant \mathsf{P}\{\xi \leqslant x_2\}.$$

Таким образом утверждение 2 доказано.

Событие $\{\xi \leqslant x_2\}$ при $x_1 < x_2$ представляет собой объединение двух *непересекающихся событий*: $\{\xi \leqslant x_1\}$ — случайная величина ξ приняла значение, не больше x_1, и $\{x_1 < \xi \leqslant x_2\}$ — случайная величина ξ приняла значение, лежащее в промежутке $(x_1, x_2]$. Поэтому в соответствии с *аксиомой сложения* получаем утверждение 4.

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

Ответьте на следующие вопросы. Является ли функцией распределения случайной величины функция:

а) $F(x) = \sin x$, б) $F(x) = |\sin x|$, в) $F(x) = x^2$, г) $F(x) = \begin{cases} 0, & x<0, \\ x^2, & 0 \leqslant x < 1, \\ 1, & x \geqslant 1. \end{cases}$

В теории вероятностей и смежных дисциплинах нередко употребляется термин "закон распределения вероятностей" случайной величины, определение которого приводится ниже.

Определение 5.3. Для произвольной случайной величины ξ отображение, которое ставит в соответствие множествам $B \in \mathbb{R}$ вероятность события $\{\xi \in B\}$ называют ***законом распределения вероятностей***, или ***распределением*** (***вероятностей***) случайной величины ξ.

5.3 Дискретные случайные величины

Определение 5.4. *Случайную величину* ξ называют ***дискретной***, если множество ее возможных значений конечно или счетно.

Распределение дискретной случайной величины удобно описывать с помощью ряда распределения.

Определение 5.5. ***Рядом распределения*** (***вероятностей***) ***дискретной случайной величины*** ξ называют таблицу (табл. 5.1), состоящую из двух строк: в верхней строке перечислены все возможные значения случайной величины, а в нижней — *вероятности* $p_i = \mathsf{P}\{\xi = x_i\}$ 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: $\sum\limits_{n=1}^{\infty} p_n = 1$.

ξ	x_1	x_2	$\ldots$	x_n	$\ldots$
P	p_1	p_2	$\ldots$	p_n	$\ldots$

Таблица 5.1.

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

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

$$\{\xi \leqslant x\} = \{\xi = x_1\} + \{\xi = x_2\},$$

а следовательно, $F(x) = p_1 + p_2$ и т.д. Наконец, при $x \geqslant x_n$ *событие* $\{\xi \leqslant x\}$ *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 $[x_i, x_{i+1})$, $1 \leqslant i < n$, — значение $p_1 + \ldots + p_i$ и на промежутке $[x_n, +\infty)$ — значение 1.

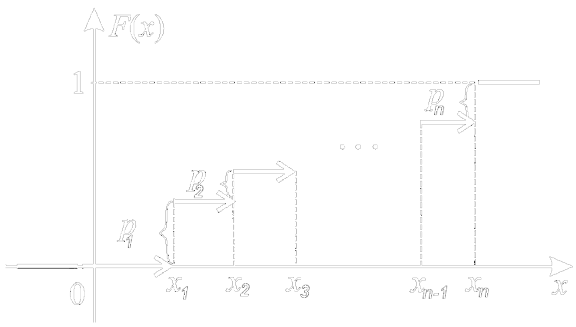

Рис 5.1.

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 $\mathsf{P}\{\xi = i\} = \dfrac{1}{6}, i = \overline{1,6}$.

5.4 Непрерывные случайные величины

Определение 5.6. ***Плотностью распределения*** (***плотностью вероятности***) случайной величины ξ называется неотрицательная кусочно-непрерывная функция $f_\xi(x)$, для которой при любом $x \in \mathbb{R}$ выполняется соотношение

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\, dt. \tag{5.1}$$

Для простоты дальнейших обозначений будем писать $f(x) = f_\xi(x)$.

Определение 5.7. Случайная величина, у которой существует плотность вероятности, называется ***абсолютно непрерывной*** (имеет ***абсолютно непрерывное распределение***).

Замечание 5.2. Отметим, что существуют непрерывные функции распределения $F(x)$, не имеющие плотностей. Такие функции распределения называют сингулярными. Пример сингулярной функции, называемой канторовой лестницей, приведен в учебнике Б.А. Севастьянов. «Курс теории вероятностей и математической статистики» (с. 89).

Замечание 5.3. Поскольку сингулярные функции распределения в данном курсе рассматриваться не будут, то в дальнейшем под непрерывными случайными величинами мы будем понимать абсолютно непрерывные случайные величины.

Пусть случайная величина ξ непрерывна. Тогда $\mathsf{P}\{\xi = x\} = 0$ для произвольного фиксированного $x \in \mathbb{R}$. Действительно,

$$\mathsf{P}\{\xi = x\} = \lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \mathsf{P}\{x-\varepsilon < \xi \leqslant x\} = \lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \Big(F(x) - F(x-\varepsilon)\Big) = 0.$$

Таким образом, для непрерывной случайной величины вероятность того, что она примет в опыте некоторое наперед заданное значение, равна 0.

Свойства $f(x)$

1) $f(x) \geqslant 0$ для всех $x \in \mathbb{R}$, т. е. выполняется *условие неотрицательности плотности*. Это свойство следует из определения 5.6.

2) $\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1$, т. е. выполняется *условие нормировки плотности*. А именно, по определению 5.6 $F(x) = \int\limits_{-\infty}^{x} f(x)\,dx$. Поэтому $\int\limits_{-\infty}^{+\infty} f(x)\,dx = F(\infty) = 1$.

3) $\int\limits_{x_1}^{x_2} f(x)\,dx = \mathsf{P}\{x_1 < \xi \leqslant x_2\}$. Действительно,

$$\int\limits_{x_1}^{x_2} f(x)\,dx = \int\limits_{-\infty}^{x_2} f(x)\,dx - \int\limits_{-\infty}^{x_1} f(x)\,dx = F(x_2) - F(x_1) = \mathsf{P}\{x_1 < \xi \leqslant x_2\}.$$

4) $F'(x) = f(x)$ в точках непрерывности плотности $f(x)$. Это свойство можно получить из определения 5.6, используя правило дифференцирования интеграла с переменным верхним пределом.

5) Справедлива следующая теорема.

Теорема 5.2. *Пусть случайная величина ξ имеет плотность $f_\xi(x)$. Пусть функция $y = \varphi(x)$ является монотонной дифференцируемой функцией. Обозначим $x = \psi(y) = \varphi^{-1}(y)$ функцию, обратную к $y = \varphi(x)$.*

Тогда плотность случайной величины $\eta = \varphi(\xi)$ есть

$$f_\eta(y) = f_\xi(\psi(y))|\psi'(y)|. \tag{5.2}$$

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

$$\mathsf{P}\{\varphi(\xi) \leqslant y\} = \mathsf{P}\{\xi \leqslant \psi(y)\}, \tag{5.3}$$

для убывающей $\varphi(x)$

$$\mathsf{P}\{\varphi(\xi) \leqslant y\} = \mathsf{P}\{\xi \geqslant \psi(y)\}. \tag{5.4}$$

Поскольку

$$F_\eta(y) = \mathsf{P}\{\eta \leqslant y\},$$

а

$$\mathsf{P}\{\xi \leqslant \psi(y)\} = F_\xi(\psi(y)) \quad \text{и} \quad \mathsf{P}\{\xi \geqslant \psi(y)\} = 1 - F_\xi(\psi(y)),$$

то окончательно получаем:

для возрастающей функции $\varphi(x)$

$$F_\eta(y) = F_\xi(\psi(y)); \tag{5.5}$$

для убывающей функции $\varphi(x)$

$$F_\eta(y) = 1 - F_\xi(\psi(y)). \tag{5.6}$$

Далее, согласно правилу дифференцирования сложной функции, имеем:

в случае возрастающей функции $\varphi(x)$

$$f_\eta(y) = F'_\eta(y) = \Big(F_\xi(x)\Big)'\Big|_{x=\psi(y)}\psi'(y) = f_\xi(\psi(y))\psi'(y);$$

в случае убывающей функции $\varphi(x)$

$$f_\eta(y) = F'_\eta(y) = -\Big(F_\xi(x)\Big)'\Big|_{x=\psi(y)}\psi'(y) = -f_\xi(\psi(y))\psi'(y).$$

Оба эти случая можно записать в виде (5.2).

Пример 5.4. Рассмотрим случайную величину $\eta = a\xi + b$, где случайная величина ξ имеет плотность $f_\xi(x)$, причем $a \neq 0$. В данном случае $\psi(y) = (y-b)/a$. Поэтому $\psi'(y) = 1/a$. Согласно теореме 5.2 получаем

$$f_\eta(y) = \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right).$$

Теорема 5.3. *Пусть случайная величина ξ имеет плотность $f_\xi(x)$. Пусть функция $y = \varphi(x)$ является кусочно-монотонной функцией. Обозначим $x = \psi_i(y)$, $i = \overline{1,k}$, прообразы точки y при отображении $y = \varphi(x)$. Если функции $\psi_i(y)$, $i = \overline{1,k}$, дифференцируемы, то плотность случайной величины $\eta = \varphi(\xi)$ есть*

$$f_\eta(y) = \sum_{i=1}^{k} f_\xi(\psi_i(y))|\psi'_i(y)|. \tag{5.7}$$

Лекция 6

Числовые характеристики случайных величин

Математическое ожидание случайной величины

Определение 6.1. ***Математическим ожиданием*** (***средним значением***) *дискретной случайной величины* ξ называют число

$$m_\xi = \mathsf{E}\xi = \sum_{i=1}^{\infty} x_i p_i,$$

где $p_i = \mathsf{P}\{\xi = x_i\}$. При этом предполагается, что $\sum\limits_{i=1}^{\infty} |x_i| p_i < \infty$.

Математическое ожидание дискретной случайной величины имеет аналог в теоретической механике. Пусть на прямой расположена система материальных точек с массами p_i, и пусть x_i — координата i-й точки. Тогда с учетом $\sum\limits_{i=1}^{\infty} p_i = 1$ центр масс системы будет иметь координату

$$x_{\text{ц}} = \frac{\sum\limits_{i=1}^{\infty} x_i p_i}{\sum\limits_{i=1}^{\infty} p_i} = \frac{\sum\limits_{i=1}^{\infty} x_i p_i}{1} = \sum_{i=1}^{\infty} x_i p_i,$$

совпадающую с математическим ожиданием $\mathsf{E}\xi$ случайной величины ξ.

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

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

Определение 6.2. Пусть плотность $f(x)$ непрерывной случайной величины ξ такова, что сходится интеграл $\int\limits_{-\infty}^{+\infty} |x| f(x)\,dx$. Тогда число $m_\xi = \mathsf{E}\xi = \int\limits_{-\infty}^{+\infty} x f(x)\,dx$ будем называть *математическим ожиданием* непрерывной случайной величины ξ.

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 $f(x)$.

Математическое ожидание функции от случайной величины.

Найдем математическое ожидание функции случайной величины. Пусть $\eta = \varphi(\xi)$ является функцией от случайной величины ξ.

Рассмотрим сначала *дискретную случайную величину* ξ, принимающую значения $x_1, \ldots, x_n, \ldots$ с вероятностями $p_n = \mathsf{P}\{\xi = x_n\}$, $n = 1, 2, \ldots$. Тогда случайная величина $\eta =$

$\varphi(\xi)$ принимает значения $\varphi(x_n)$ с вероятностями $p_n = \mathsf{P}\{\xi = x_n\}$, $n = 1,2,\ldots$, и ее математическое ожидание определяется формулой

$$\mathsf{E}\eta = \mathsf{E}\varphi(\xi) = \sum_{i=1}^{\infty} \varphi(x_i) p_i, \tag{6.1}$$

при этом требуется выполнение условия

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

Для *непрерывной случайной величины* ξ, имеющей *плотность распределения* $f(x)$, математическое ожидание случайной величины $\eta = \varphi(\xi)$ можно найти, используя аналогичную (6.1) формулу

$$\mathsf{E}\eta = \mathsf{E}\varphi(\xi) = \int_{-\infty}^{+\infty} \varphi(x) f(x)\, dx, \tag{6.3}$$

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

Дисперсия

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*** $\mathsf{D}\xi$ случайной величины ξ называют математическое ожидание квадрата отклонения случайной величины ξ от ее среднего значения, т. е. $\mathsf{D}\xi = \mathsf{E}(\xi - \mathsf{E}\xi)^2$.

Используя формулы (6.1)–(6.3), в которых положено $\varphi(x) = (x - \mathsf{E}\xi)^2$, легко написать расчетные формулы для дисперсий дискретной и непрерывной случайных величин. А именно, если дискретная случайная величина ξ принимает значения $x_1, \ldots, x_n, \ldots$ с вероятностями $p_n = \mathsf{P}\{\xi = x_n\}$, $n = 1,2,\ldots$, то

$$\mathsf{D}\xi = \sum_{i=1}^{\infty} (x_i - \mathsf{E}\xi)^2 p_i, \tag{6.4}$$

а если ξ — непрерывная случайная величина с плотностью $f(x)$, то

$$\mathsf{D}\xi = \int_{-\infty}^{+\infty} (x - \mathsf{E}\xi)^2 f(x)\, dx. \tag{6.5}$$

Определение 6.4. ***Средним квадратическим отклонением*** (СКО) случайной величины ξ называют величину $\sigma_\xi = \sqrt{\mathsf{D}\xi}$.

Пример 6.2. Пусть ξ — число выпавших очков при подбрасывании игральной кости. Так как $\mathsf{E}\xi = 3{,}5$, то

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

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

В некоторых теоретических исследованиях встречаются моменты высших порядков.

Определение 6.5. ***Начальным моментом k-го порядка*** (или ***k-м начальным моментом***) случайной величины ξ называют число $\mu_k = \mathsf{E}\xi^k$, $k = 1,2,\ldots$.

Определение 6.6. ***Центральным моментом k-го порядка*** (или ***k-м центральным моментом***) случайной величины ξ называют число $\nu_k = \mathsf{E}\left((\xi - \mathsf{E}\xi)^k\right)$, $k = 2,3,\ldots$.

Таким образом, если ξ — дискретная случайная величина, принимающая значения $x_1,\ldots,x_n,\ldots$ с вероятностями $p_1,\ldots,p_n,\ldots$ соответственно, то

$$\mu_k = \sum_{i=1}^{\infty} x_i^k p_i, \qquad \nu_k = \sum_{i=1}^{\infty} (x_i - \mathsf{E}\xi)^k p_i,$$

а если ξ — непрерывная случайная величина с плотностью $f(x)$, то

$$\mu_k = \int\limits_{-\infty}^{+\infty} x^k f(x)\,dx, \qquad \nu_k = \int\limits_{-\infty}^{+\infty} (x - \mathsf{E}\xi)^k f(x)\,dx.$$

Видно, что начальный момент первого порядка совпадает с математическим ожиданием, а центральный момент второго порядка является дисперсией.

Определение 6.7. Случайная величина $\overset{\circ}{\xi} = \xi - m_\xi$ называется *центрированной*, а случайная величина $\overset{*}{\xi} = \overset{\circ}{\xi} / \sigma_\xi$, если $\sigma_\xi > 0$, называется *нормированной*.

Свойства математического ожидания и дисперсии

1) $\mathsf{E}[c] = c$ и $\mathsf{D}[c] = 0$, если c — константа. Действительно, пусть ξ — дискретная случайная величина, принимающая с вероятностью 1 значение c, т. е. $\mathsf{P}\{\xi = c\} = 1$. Тогда $\mathsf{E}\xi = c\mathsf{P}\{\xi = c\} = c$. Аналогично, $\mathsf{D}\xi = (c - \mathsf{E}[c])^2\mathsf{P}\{\xi = c\} = 0$.

2) $\mathsf{E}[c\xi] = c\mathsf{E}\xi$, если c — константа. Действительно, пусть, например, ξ — непрерывная случайная величина, тогда

$$\mathsf{E}[c\xi] = \int_{-\infty}^{+\infty} cxf(x)\,dx = c\int_{-\infty}^{+\infty} xf(x)\,dx = cm_\xi.$$

3) $\mathsf{E}[\xi + c] = m_\xi + c$, если c — константа. Очевидно, что, например, для непрерывной случайной величины можно получить

$$\mathsf{E}[\xi + c] = \int_{-\infty}^{+\infty} (x + c)f(x)\,dx = \int_{-\infty}^{+\infty} xf(x)\,dx + \int_{-\infty}^{+\infty} cf(x)\,dx = m_\xi + c.$$

4) $\mathsf{E}\overset{*}{\xi} = 0$, $\mathsf{D}\overset{*}{\xi} = 1$. Действительно, $\mathsf{E}\overset{*}{\xi} = \mathsf{E}\left[\frac{\xi - m_\xi}{\sigma_\xi}\right]$, поэтому согласно свойствам 2) и 3) получаем $\mathsf{E}\overset{*}{\xi} = 0$ и

$$\mathsf{D}\overset{*}{\xi} = \mathsf{E}(\overset{*}{\xi} - \mathsf{E}\overset{*}{\xi})^2 = \frac{1}{\sigma_\xi^2}\mathsf{E}(\xi - m_\xi)^2 = \frac{\mathsf{D}\xi}{\sigma_\xi^2} = 1.$$

5) Пусть функция $\varphi(x) = \sum\limits_{k=1}^{n} a_k\varphi_k(x)$. Тогда, если $\mathsf{E}[\varphi_k(\xi)]$ существуют для всех $k = \overline{1,n}$, то $\mathsf{E}[\varphi(\xi)] = \sum\limits_{k=1}^{n} a_k\mathsf{E}[\varphi_k(\xi)]$. Действительно, в случае непрерывной случайной величины ξ, используя линейные свойства интеграла, получаем

$$\mathsf{E}[\varphi(\xi)] = \int_{-\infty}^{+\infty}\left(\sum_{k=1}^{n} a_k\varphi_k(x)\right)f(x)\,dx = \sum_{k=1}^{n} a_k\int_{-\infty}^{+\infty}\varphi_k(x)f(x)\,dx = \sum_{k=1}^{n} a_k\mathsf{E}[\varphi_k(\xi)]. \tag{6.6}$$

6) $\mathsf{D}\xi = \mathsf{E}[\xi^2] - m_\xi^2$. Действительно, по свойству 5) имеем

$$\mathsf{D}\xi = \mathsf{E}(\xi - m_\xi)^2 = \mathsf{E}[\xi^2] - 2\mathsf{E}\xi m_\xi + m_\xi^2 = \mathsf{E}[\xi^2] - m_\xi^2.$$

7) $\mathsf{D}[c\xi] = c^2\mathsf{D}\xi$, $\mathsf{D}[c+\xi] = \mathsf{D}\xi$, где c — константа.

8) Пусть непрерывная случайная величина ξ имеет конечное математическое ожидание и ее график плотности распределения вероятности симметричен относительно прямой $x = c$ для некоторого $c \in \mathbb{R}^1$. Тогда $\mathsf{E}\xi = c$.

Например, для случайной величины ξ с плотностью $f_\xi(x)$ рассмотрим случайную величину $\eta = \xi - c$ с плотностью $f_\eta(y) = f_\xi(y-c)$, такой что $f_\eta(y) = f_\eta(-y)$. Тогда

$$\int_{-\infty}^{0} y f_\eta(y)dy = -\int_{0}^{+\infty} y f_\eta(y)dy.$$

Таким образом $\mathsf{E}\eta = 0$. Но по свойству 3) имеем $\mathsf{E}\xi = \mathsf{E}\eta + c = c$.

9) $\mathsf{E}(\xi+\eta) = \mathsf{E}\xi + \mathsf{E}\eta$. Данное свойство будет доказано позднее.

10) $\mathsf{D}(\xi+\eta) = \mathsf{D}\xi + \mathsf{D}\eta + 2\big(\mathsf{E}(\xi\eta) - \mathsf{E}\xi\mathsf{E}\eta\big)$. Действительно,

$$\mathsf{D}(\xi+\eta) = \mathsf{E}(\xi+\eta)^2 - \big(\mathsf{E}(\xi+\eta)\big)^2 = \mathsf{E}\big(\xi^2 + 2\xi\eta + \eta^2\big) - \big((\mathsf{E}\xi)^2 + 2\mathsf{E}\xi\mathsf{E}\eta + (\mathsf{E}\eta)^2\big) =$$
$$= \big(\mathsf{E}\xi^2 - (\mathsf{E}\xi)^2\big) + \big(\mathsf{E}\eta^2 - (\mathsf{E}\eta)^2\big) + 2\big(\mathsf{E}(\xi\eta) - \mathsf{E}\xi\mathsf{E}\eta\big) = \mathsf{D}\xi + \mathsf{D}\eta + 2\big(\mathsf{E}(\xi\eta) - \mathsf{E}\xi\mathsf{E}\eta\big).$$

В дальнейшем будет показано, что для независимых случайных величин ξ и η величина $\mathsf{E}(\xi\eta) - \mathsf{E}\xi\mathsf{E}\eta$, называемая ковариацией случайных величин ξ и η, равна нулю. И, таким образом, для независимых случайных величин ξ и η имеет место равенство $\mathsf{D}(\xi+\eta) = \mathsf{D}\xi + \mathsf{D}\eta$.

Аналогично можно показать, что свойства 2), 3) и 5), доказанные выше для непрерывных случайных величин, справедливы также и для дискретных случайных величин.

Квантиль

Квантиль является одной из основных статистических характеристик, используемых в математической статистике.

Определение 6.8. Пусть ξ — случайная величина с непрерывной строго монотонной функцией распределения $F(x)$. ***Квантúлью уровня*** γ, или ***γ-квантилью***, $0 < \gamma < 1$, случайной величины ξ (распределения случайной величины ξ) называют число Q_γ, удовлетворяющее равенству

$$F(Q_\gamma) = \gamma. \tag{6.7}$$

Замечание 6.1. Учитывая (5.1), уравнение для определения γ-квантили можно записать в виде

$$\int\limits_{-\infty}^{Q_\gamma} f(x)\,dx = \gamma. \tag{6.8}$$

Определение 6.9. Пусть ξ — случайная величина с (произвольной) функцией распределения $F(x)$. ***Квантилью уровня*** γ, или ***γ-квантилью***, $0 < \gamma < 1$, случайной величины ξ (распределения случайной величины ξ) называют число Q_γ, равное минимальному значению x, при котором $F(x)$ не меньше γ, т.е.

$$Q_\gamma = \min\{x : F(x) \geqslant \gamma\}. \quad \# \tag{6.9}$$

Нетрудно заметить, что определение 6.8 является частным случаем определения 6.9.

Определение 6.10. Квантиль уровня $1/2$ называют ***медианой***.

На рис. 6.1 указаны квантили уровней α, β и γ некоторой функции распределения $F(x)$.

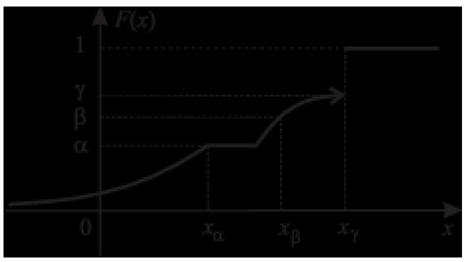

Рис 6.1

Замечание 6.2. Если плотность распределения существует и симметрична относительно оси Oy, то $x_p = -x_{1-p}$. Действительно, пусть СВ $\eta = -\xi$. При сделанных предположениях $F_\eta(x) \equiv F_\xi(x)$, поэтому $y_p = x_p$ для любого p. Далее, так как $f_\eta(x) \equiv f_\xi(x)$ и $\int_{x_{1-p}}^{\infty} f_\xi(x)\,dx = p$, то $y_p = -x_{1-p}$. Поэтому $x_p = -x_{1-p}$ (рис. 6.2).

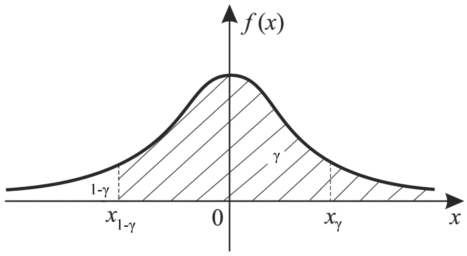

Рис 6.2

Лекция 7

Основные законы распределения случайных величин

7.1 Основные дискретные распределения

Биномиальное распределение

Определение 7.1. Случайную величину ξ назовем ***биномиальной*** с параметрами $p \in (0,1)$ и $n \in \mathbb{N}$, если она принимает значения $0, 1, 2, \dots, n$ с вероятностями, заданными формулой

$$\mathsf{P}\{\xi = k\} = C_n^k p^k q^{n-k}, \qquad q = 1 - p, \qquad k = \overline{0,n},$$

или, что тоже самое, рядом распределения, представленным в таблице 7.1. Говорят также, что эта случайная величина ***распределена по биномиальному закону***, или имеет ***биномиальное распределение***. Символически это записывается $\xi \sim Bi(n,p)$.

ξ	0	1	$\dots$	k	$\dots$	n
P	q^n	$C_n^1 p q^{n-1}$	$\dots$	$C_n^k p^k q^{n-k}$	$\dots$	p^n

Таблица 7.1.

Проверим корректность определения биномиального распределения. Действительно, $C_n^k p^k q^{n-k} > 0$ и согласно равенству

$$\sum_{k=0}^{m} C_m^k p^k q^{m-k} = (p+q)^m, \qquad m = 1,2,\dots, \tag{7.1}$$

известному как бином Ньютона,

$$\sum_{k=0}^{n} \mathsf{P}\{\xi = k\} = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1.$$

Замечание 7.1. Биномиальная случайная величина — число успехов в n испытаниях Бернулли с вероятностью успеха p и неудачи $q = 1 - p$ (см. определение 3.5).

Определение 7.2. Случайную величину, которая принимает значения 1 и 0 с вероятностями p и $1 - p$ соответственно, называют случайной величиной, имеющей распределение Бернулли с параметром $p \in (0,1)$. Другими словам, бернуллиевская случайная величина — биномиальная случайная величина с параметрами $p \in (0,1)$ и $n = 1$, т.е. $\xi \sim Bi(1,p)$.

Пример 7.1. Найдем математическое ожидание и дисперсию бернуллиевской случайной величины ξ с параметром p. Воспользовавшись определениями 7.2 и 6.1, получим

$$\mathsf{E}\xi = 0 \cdot (1-p) + 1 \cdot p = p, \tag{7.2}$$

а согласно формуле (6.4)

$$\mathsf{D}\xi = (0 - \mathsf{E}\xi)^2 \cdot (1-p) + (1 - \mathsf{E}\xi)^2 \cdot p = (0-p)^2(1-p) + (1-p)^2 p = p^2 q + q^2 p = pq. \tag{7.3}$$

Пример 7.2. Найдем математическое ожидание биномиальной случайной величины ξ с параметрами n и p. Дважды воспользовавшись равенством (7.1), получим

$$\mathsf{E}\xi = \sum_{k=0}^{n} kC_n^k p^k q^{n-k} = \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k q^{n-k} =$$

$$= \sum_{k=1}^{n} np\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k} = \Big|j=k-1\Big| = np\sum_{j=0}^{n-1} C_{n-1}^j p^j q^{n-1-j} = np(p+q)^{n-1} = np.$$

Более легкий способ вычисления математического ожидания основан на утверждении 9) на стр. 32, доказательство которой будет проведено позже и согласно которой математическое ожидание суммы конечного числа случайных величин равно сумме их математических ожиданий. Для того чтобы воспользоваться этой теоремой представим случайное число успехов ξ в n испытаниях в виде суммы $\xi = \xi_1 + \ldots + \xi_n$, где случайное число ξ_k успехов в k-ом испытании имеет распределение Бернулли, $k = \overline{0,n}$. Поэтому согласно (7.2)

$$\mathsf{E}\xi = \mathsf{E}\xi_1 + \ldots + \mathsf{E}\xi_n = \underbrace{p+p+\ldots+p}_{n\ \text{слагаемых}} = np.$$

Пример 7.3. Пусть ξ — число успехов в n испытаниях по *схеме Бернулли*. Дисперсию ξ можно подсчитать по формуле (6.4), непосредственно воспользовавшись определением 6.3 дисперсии. Однако мы поступим другим образом. Для этого представим ξ в виде суммы $\xi = \xi_1 + \ldots + \xi_n$, где ξ_k — число успехов в k-ом испытании. Дисперсия каждого слагаемого равна (см. (7.3)) pq. Согласно утверждению 10) на стр. 32 дисперсия суммы независимых случайных величин равна сумме дисперсий каждого слагаемого. Учитывая, что случайные величины ξ_k являются независимыми, получаем

$$\mathsf{D}\xi = \mathsf{D}\xi_1 + \ldots + \mathsf{D}\xi_n = npq.$$

Распределение Пуассона

Определение 7.3. Случайную величину ξ назовем ***пуассоновской*** с параметром $\lambda > 0$, если она принимает значения $0, 1, 2, \ldots, n, \ldots$ с вероятностями, заданными формулой

$$\mathsf{P}\{\xi = n\} = \frac{\lambda^n}{n!}e^{-\lambda}, \qquad n = 0, 1, \ldots,$$

или, что тоже самое, рядом распределения, представленным в таблице 7.2. Говорят также, что она ***распределена по пуассоновскому закону***, или имеет ***пуассоновское распределение***. Символически это записывается $\xi \sim \Pi(\lambda)$.

ξ	0	1	2	$\ldots$	n	$\ldots$
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!}e^{-\lambda}$	$\ldots$	$\frac{\lambda^n}{n!}e^{-\lambda}$	$\ldots$

Таблица 7.2.

Проверка корректности определения распределения Пуассона дает:

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}e^{-\lambda} = e^{-\lambda}\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda}e^{\lambda} = 1.$$

Пусть случайная величина ξ имеет *распределение Пуассона* с параметром λ. Тогда

$$\mathsf{E}\xi = \sum_{n=0}^{\infty} n\frac{\lambda^n}{n!}e^{-\lambda} = \lambda\sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda} = \lambda\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}e^{-\lambda} = \lambda e^{\lambda}e^{-\lambda} = \lambda.$$

Найдем дисперсию случайной величины $\xi \sim \Pi(\lambda)$. Вычислим второй начальный момент. С учетом $\mathsf{E}\xi = \lambda$, получим

$$\mathsf{E}\xi^2 = \sum_{n=0}^{\infty} n^2\frac{\lambda^n}{n!}e^{-\lambda} = \lambda\sum_{n=1}^{\infty} n\frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda} = \Big|k=n-1\Big| =$$

$$= \lambda\sum_{k=0}^{\infty}(k+1)\frac{\lambda^k}{k!}e^{-\lambda} = \lambda\left(\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} + \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}\right) = \lambda(\mathsf{E}\xi + 1) = \lambda^2 + \lambda.$$

Согласно свойству дисперсии 6) на стр. 31

$$\mathsf{D}\xi = \lambda^2 + \lambda - \lambda^2 = \lambda,$$

Между биномиальным распределением $Bi(n;p)$ и распределением Пуассона $\Pi(\lambda)$ имеется следующая связь.

Теорема 7.1 (Пуассона). *Пусть $n \to \infty$, $p \to 0$ и при этом $np = \lambda = const$. Тогда*

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad \text{где } P_n(k) = C_n^k p^k q^{n-k},\ k = \overline{0,n}.$$

ДОКАЗАТЕЛЬСТВО. Докажем это утверждение, пользуясь вторым замечательным пределом $(1-\lambda/n)^n \to e^{-\lambda}$ при $n \to \infty$. Так как здесь $p = \lambda/n$, $q = 1-\lambda/n$, то

$$\lim_{n\to\infty} P_n(k) = \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} =$$
$$= \frac{\lambda^k}{k!}\lim_{n\to\infty}\frac{n(n-1)\ldots(n-k+1)}{n^k}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k} = \frac{\lambda^k}{k!}e^{-\lambda}. \quad \#$$

Таким образом, при больших n и малых p (при редких явлениях) двухпараметрическое биномиальное распределение $Bi(n;p)$ можно приближенно заменить однопараметрическим распределением Пуассона $\Pi(\lambda)$, где $\lambda = np$. При этом ошибка от такой замены не превышает np^2 для всех $k = \overline{0,n}$, т. е.

$$\left| C_n^k p^k q^{n-k} - \frac{(np)^k}{k!}e^{-np} \right| \leqslant np^2.$$

Если указанная погрешность np^2 оказывается неприемлемой, то для вычисления $P_n(k)$ применяется другая приближенная формула, о которой будет рассказано позднее.

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”, т.е. маловероятное *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 7.4. Пусть некоторая система содержит 5000 независимо работающих элементов. Вероятность отказа каждого равна 0.001. Найдем вероятность отказа системы, если известно, что он происходит при отказе двух и более ее элементов. Число отказавших элементов является случайной величиной $\xi \sim Bi(5000{,}0.001)$. Поскольку значение $n = 5000$ велико, $p = 0.001$ мало $(\lambda = np = 5)$ и, кроме того, $np^2 = 0.005$ — приемлемая точность, то по теореме 7.1

$$\mathsf{P}\{X \geqslant 2\} = 1 - \mathsf{P}\{X \leqslant 1\} = 1 - P_{5000}(0) - P_{5000}(1) = 1 - \frac{5^0}{0!}e^{-5} - \frac{5^1}{1!}e^{-5} \approx 0.9596.$$

Геометрическое распределение

Определение 7.4. Дискретную случайную величину назовем ***геометрической*** с параметром $p \in (0,1)$, если она принимает значения $1, 2, \ldots$ с вероятностями, заданными формулой

$$\mathsf{P}\{\xi = n\} = q^{n-1}p, \qquad q = 1-p, \qquad n = 1,2,\ldots,$$

или, что тоже самое, рядом распределения, представленным в таблице 7.3. Говорят также, что она ***распределена по геометрическому закону***, или имеет ***геометрическое распределение***. Символически это записывается $\xi \sim G(p)$.

ξ	1	2	3	$\ldots$	n	$\ldots$
P	p	qp	q^2p	$\ldots$	$q^{n-1}p$	$\ldots$

Таблица 7.3.

Корректность составления таблицы вытекает из равенства

$$\sum_{n=1}^{\infty} \mathsf{P}\{\xi = n\} = \sum_{n=1}^{\infty} q^{n-1}p = p\sum_{n=1}^{\infty} q^{n-1} = p\frac{1}{1-q} = 1.$$

Пример 7.5. Найдем математическое ожидание геометрической случайной величины ξ:

$$\mathsf{E}\xi = \sum_{n=1}^{\infty} nq^{n-1}p = p\sum_{n=1}^{\infty} nq^{n-1} = p\Big(\sum_{n=0}^{\infty} q^n\Big)'_q = p\left(\frac{1}{1-q}\right)' = p\frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Можно показать, что

$$\mathsf{D}\xi = \frac{q}{p^2}. \quad \#$$

Геометрическое распределение имеет следующее приложение. Пусть проводятся независимые испытания с двоичными исходами («успех» и «неудача») с вероятностью успеха, равной p. Тогда геометрическая случайная величина описывает количество испытаний, проведенных до первого появления успеха.

Пример 7.6. На тренировке баскетболист выполняет штрафные броски до тех пор, пока не попадет в корзину. Вероятность попадания в корзину при штрафном броске равна $p = 1/3$. Пусть ξ — количество бросков, сделанных баскетболистом. Требуется построить ряд распределения случайной величины ξ, найти ее наиболее вероятное значение, $\mathsf{E}\xi$, $\mathsf{D}\xi$.

До попадания в корзину баскетболист может сделать $1,2,3,\ldots$ бросков. Значит, множество возможных значений случайной величины ξ имеет вид $\{1,2,3,\ldots\}$. Случайное событие $\{\xi = k\}$, $k = 1,2,3,\ldots$, означает то, что первые $k-1$ бросков были неудачными, а k-й бросок — успешным. Тогда по формуле умножения вероятностей для независимых событий получаем

$$\mathsf{P}(\xi = k) = (1-p)^{k-1}p = q^{k-1}p, \quad k = 1,2,3,\ldots,$$

где $q = 1-p$.

Таким образом, ряд распределения случайной величины ξ задается табл. 7.3, т.е. $\xi \sim G(p)$ (ξ имеет *геометрическое распределение* с параметром $p = 1/3$).

Тогда согласно примеру 7.5 $\mathsf{E}\xi = 1/p = 3$, $\mathsf{D}\xi = q/p^2 = 6$.

Очевидно, что наиболее вероятное значение такой случайной величины есть $k^* = 1$, так как $\mathsf{P}(\xi = 1) = p$, а при любом $k > 1$, $\mathsf{P}\{\xi = k\} = q^{k-1}p < p$.

7.2 Основные непрерывные распределения

Равномерное распределение

Определение 7.5. Случайную величину ξ назовем ***равномерно распределенной*** на отрезке $[a,b]$ (или на интервале (a,b)), если ее плотность распределения $f(x)$ имеет вид

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b; \\ 0, & x < a \text{ или } x > b; \end{cases}$$

Говорят также, что она ***распределена по равномерному закону*** или имеет ***равномерное распределение*** и обозначают $\xi \sim R(a,b)$. Числа a и b называют параметрами распределения.

Функция распределения случайной величины $\xi \sim R(a,b)$ имеет вид

$$F(x) = \begin{cases} 0, & x < a; \\ \dfrac{x-a}{b-a}, & a \leqslant x \leqslant b; \\ 1, & x > b. \end{cases}$$

Графики плотности распределения $f(x)$ 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 $[a,b]$.

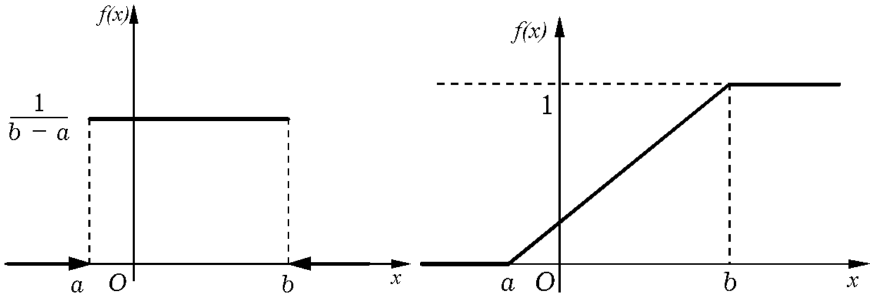

Рис 7.1.

Пример 7.7. Найдем математическое ожидание *равномерно распределенной* на отрезке $[a,b]$ случайной величины ξ. Поскольку в этом случае $f(x) = 0$ при $x < a$ и $x > b$, то

$$\mathsf{E}\xi = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\cdot\frac{x^2}{2}\bigg|_{x=a}^{x=b} = \frac{b+a}{2}.$$

Как и следовало ожидать, $\mathsf{E}\xi$ совпадает с серединой отрезка $[a,b]$.

Пример 7.8. Дисперсия *равномерно распределенной* на отрезке $[a,b]$ случайной величины ξ определяется формулой

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

Экспоненциальное распределение

Определение 7.6. Случайную величину ξ назовем ***экспоненциальной*** (или ***показательной***) с параметром $\lambda > 0$, если ее плотность распределения $f(x)$ имеет вид

$$f(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \geqslant 0; \end{cases}$$

Говорят также, что она ***распределена по экспоненциальному*** (или ***показательному***) ***закону***, или имеет ***экспоненциальное*** (***показательное***) ***распределение*** и обозначают $\xi \sim E(\lambda)$.

Функция распределения случайной величины $\xi \sim E(\lambda)$ имеет вид

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

Графики плотности распределения и функции распределения экспоненциальной случайной величины приведены на рис. 7.2.

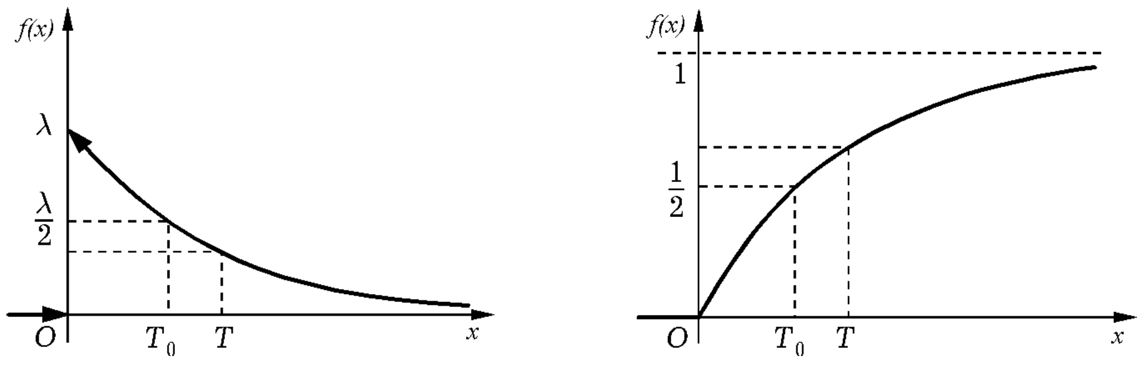

Рис 7.2

Пример 7.9. Найдем математическое ожидание и дисперсию *экспоненциально распределенной* случайной величины ξ. Применяя формулу интегрирования по частям, получим

$$\mathsf{E}\xi = \int\limits_{-\infty}^{+\infty} x f(x)\,dx = \int\limits_{0}^{+\infty} \lambda x e^{-\lambda x}\,dx = \frac{1}{\lambda},$$

$$\mathsf{D}\xi = \int\limits_{-\infty}^{+\infty} (x-\mathsf{E}\xi)^2 f(x)\,dx = \int\limits_{0}^{+\infty} \left(x-\frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x}\,dx = \frac{1}{\lambda^2}.$$

Отметим важное характеристическое свойство экспоненциальной случайной величины, которое мы докажем на семинаре.

Предложение 7.1. *Пусть случайна величина* $\xi \sim E(\lambda)$.
Тогда

$$\mathsf{P}\{\xi > t+\tau | \xi > t\} = \mathsf{P}\{\xi > \tau\}$$

для $\forall t, \tau \geqslant 0$.

Как можно интерпретировать этот результат, полагая, например, что ξ — время безотказной работы электролампы?

Нормальное распределение

Определение 7.7. Случайная величина ξ имеет ***нормальное (гауссовское) распределение*** с параметрами m и $\sigma^2 > 0$, (обозначается как $\xi \sim \mathcal{N}(m;\sigma^2)$), если ее плотность имеет вид

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}.$$

При этом случайная величина называется ***нормальной*** (***гауссовской***). График плотности нормального распределения, называемый ***кривой Гаусса***, имеет единственный максимум в точке $x = m$.

Графики плотности и функции распределения

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\,dt$$

нормальной случайной величины для различных значений m и σ приведены на рис. 7.3.

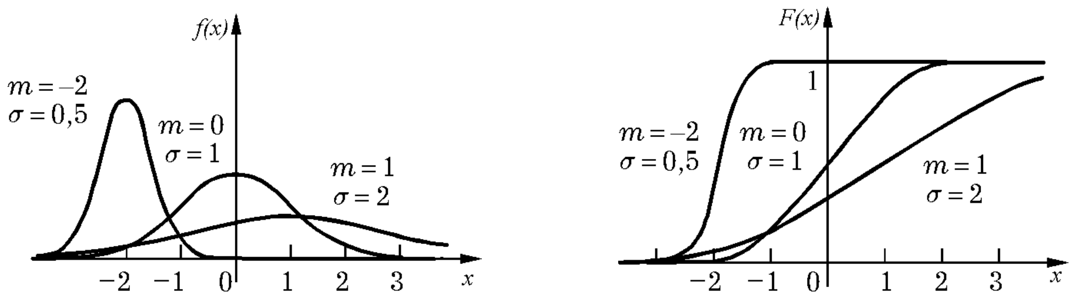

Рис 7.3

Если $m = 0$ и $\sigma = 1$, то такой ***нормальный закон*** называют ***стандартным*** и его функцию распределения обозначают $\Phi(x)$, а плотность распределения — $\varphi(x)$.

Таким образом,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt. \qquad (7.4)$$

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. Кроме того, значения $\Phi(x)$ можно вычислить с помощью различных математических пакетов.

Нередко в указанных выше таблицах и пакетах программ в действительности речь идет не о функции $\Phi(x)$, а о функции

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}}\,dt,$$

называемой функцией Лапласа.

Теорема 7.2. *Функции $\Phi(x)$ и $\Phi_0(x)$ имеют следующие свойства:*

1) $\Phi(x) = \Phi_0(x) + \frac{1}{2}$,

2) $\Phi_0(-x) = -\Phi_0(x)$, т.е. $\Phi_0(x)$ — нечетная функция,

3) $\Phi(-x) = 1 - \Phi(x)$,

4) квантиль u_p уровня p стандартного нормального распределения удовлетворяет соотношению $u_{1-p} = -u_p$.

Доказательство.

1) Несмотря на то, что неопределенный интеграл $\int e^{-x^2/2}\,dx$ не может быть выражен через элементарные функции, из курса математического анализа известно, что

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = 1,$$

откуда в силу четности подинтегральной функции

$$\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \frac{1}{2}.$$

Поэтому

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt + \int_{0}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt = \frac{1}{2} + \Phi_0(x).$$

2) Делая замену переменной $t = -s$, $dt = -ds$, получим

$$\Phi_0(-x) = \int_{0}^{-x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt = -\int_{0}^{x} \frac{1}{\sqrt{2\pi}} e^{-s^2/2}\,ds = -\Phi_0(x).$$

3) Поэтому

$$\Phi(-x) = \Phi_0(-x) + \frac{1}{2} = -\Phi_0(x) + \frac{1}{2} = -\left(\Phi(x) - \frac{1}{2}\right) + \frac{1}{2} = -\Phi(x) + 1.$$

4) В соответствии с определением 6.8 и формулой (6.7) квантиль u_p уровня p стандартного нормального распределения (другими словами квантиль нормальной случайной величины с нулевым математическим ожиданием и единичной дисперсией) определяется равенством $\Phi(u_p) = p$ для любого $p \in (0,1)$. Следовательно $\Phi(u_{1-p}) = 1 - p$. Далее, по свойству 3) $\Phi(-u_p) = 1 - \Phi(u_p) = 1 - p$, следовательно $-u_p$ есть квантиль уровня $1 - p$, т.е. $-u_p = u_{1-p}$. #

Следующая теорема проясняет смысл параметров m и σ^2 и выражает через $\Phi(x)$ вероятность попадания нормальной случайной величины в интервал (a,b).

Теорема 7.3. *Пусть ξ — нормальная случайная величина с параметрами m и σ^2. Тогда*

$$\mathsf{P}\{a<\xi<b\} = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) = \Phi_0\left(\frac{b-m}{\sigma}\right) - \Phi_0\left(\frac{a-m}{\sigma}\right), \tag{7.5}$$

$$\mathsf{E}\xi = m, \qquad \mathsf{D}\xi = \sigma^2.$$

Доказательство. В соответствии со свойством 2 плотности распределения (см. свойство 3 на стр. 27) вероятность попадания нормальной случайной величины ξ с параметрами m и σ в интервал (a,b) задается формулой

$$\mathsf{P}\{a<\xi<b\} = \int\limits_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-m)^2/(2\sigma^2)}\,dy.$$

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

$$\mathsf{P}\{a<\xi<b\} = \int\limits_{(a-m)/\sigma}^{(b-m)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = \int\limits_{(a-m)/\sigma}^{(b-m)/\sigma} \varphi(x)\,dx.$$

Поэтому

$$\mathsf{P}\{a<\xi<b\} = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

Поскольку $\Phi(x) = \Phi_0(x) + 1/2$, то $\Phi(x_2) - \Phi(x_1) = \Phi_0(x_2) - \Phi_0(x_1)$ для любых x_1 и x_2. Таким образом, окончательно получаем 7.5.

Далее

$$\mathsf{E}\xi = \int\limits_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/(2\sigma^2)}dx.$$

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

$$\mathsf{E}\xi = \int\limits_{-\infty}^{+\infty} \frac{\sigma y+m}{\sqrt{2\pi}} e^{-y^2/2}dy = \int\limits_{-\infty}^{+\infty} \frac{\sigma y}{\sqrt{2\pi}} e^{-y^2/2}dy + m\int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}dy =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} y e^{-y^2/2}dy + m\int\limits_{-\infty}^{+\infty} \varphi(y)\,dy.$$

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

По определению дисперсии случайной величины ξ

$$\mathsf{D}\xi = \int\limits_{-\infty}^{+\infty} \frac{(x-m)^2}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx.$$

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

$$\mathsf{D}\xi = \sigma^2 \int\limits_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} e^{-y^2/2}dy.$$

Воспользовавшись формулой интегрирования по частям $\int u\,dv = uv - \int v\,du$ с $u = y/\sqrt{2\pi}$, $dv = ye^{-y^2/2}dy$, находим

$$\mathsf{D}\xi = \sigma^2 \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy = \sigma^2.$$

Таким образом, дисперсия нормально распределенной случайной величины совпадает со вторым параметром распределения. #

Нормально распределенная случайная величина с большой вероятностью принимает значения, близкие к своему математическому ожиданию, что описывается «***правилом*** k ***сигм***»:

$$\mathsf{P}\{|\xi - m| \leqslant k\sigma\} = \Phi_0(k) - \Phi_0(-k) = 2\Phi_0(k) = \begin{cases} 0{,}6827, & k = 1, \\ 0{,}9545, & k = 2, \\ 0{,}9973, & k = 3. \end{cases}$$

Лекция 8

Случайные векторы

Функция распределения случайного вектора

Определение 8.1. Совокупность *случайных величин* $\xi_1 = \xi_1(\omega), \quad \ldots, \quad \xi_n = \xi_n(\omega)$, заданных на одном и том же *вероятностном пространстве* $(\Omega, \mathscr{A}, \mathsf{P})$, называют ***многомерной*** (***n-мерной***) ***случайной величиной***, или ***n-мерным случайным вектором***. При этом сами случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называют ***координатами случайного вектора***.

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

При стрельбе по воздушной цели необходимо рассматривать трехмерный случайный вектор (ξ, η, ζ), где ξ, η, ζ — 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 $(\xi, \eta, \zeta, \ldots)$. Здесь, например, ξ — температура окружающей среды, η — атмосферное давление, ζ — 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 потребуется, пояснения для случайного вектора произвольной размерности.

Напомним, что рассмотрение одномерной случайной величины начиналось с обсуждения способа задания ее *закона распределения*. Основным способом задания закона распределения одномерной случайной величины является ее *функции распределения.* То же можно сказать и по отношению к n-мерному случайному вектору. Отметим, что в дальнейшем для *пересечения событий* $\{\xi_1 \leqslant x_1\}, \ldots, \{\xi_n \leqslant x_n\}$ вместо записи

$$\{\xi_1 \leqslant x_1\} \cap \ldots \cap \{\xi_n \leqslant x_n\}$$

будем использовать запись

$$\{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\}.$$

Определение 8.2. ***Функцией распределения (вероятностей)*** (n-мерного) ***случайного вектора*** $\xi = (\xi_1, \ldots, \xi_n)$ называют функцию $F(x_1, \ldots, x_n)$, значение которой в точке $(x_1, \ldots, x_n) \in \mathbb{R}^n$ равно вероятности совместного осуществления (пересечения) событий $\{\xi_1 \leqslant x_1\}, \ldots, \{\xi_n \leqslant x_n\}$, т.е.

$$F_\xi(x_1, \ldots, x_n) = \mathsf{P}\{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\}.$$

Функцию распределения $F(x_1, \ldots, x_n)$ называют также ***совместной n-мерной функцией распределения*** случайных величин $\xi_1, \ldots, \xi_n$. 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 (ξ_1, ξ_2) в квадрант с вершиной в точке $(a_1; a_2)$.

Теорема 8.1. *Двумерная функция распределения удовлетворяет следующим свойствам.*

1. $0 \leqslant F(x_1,x_2) \leqslant 1$.

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. $\mathsf{P}\{a_1 < \xi_1 \leqslant b_1,\, a_2 < \xi_2 \leqslant b_2\} = F(b_1,b_2) - F(b_1,a_2) - F(a_1,b_2) + F(a_1,a_2)$.

6. $F(x_1,x_2)$ *— непрерывная справа в любой точке* $(x_1,x_2) \in \mathbb{R}^2$ *по каждому из аргументов* x_1 *и* x_2 *функция.*

7. $F_\xi(x,+\infty) = F_{\xi_1}(x), \quad F_\xi(+\infty,x) = F_{\xi_2}(x)$.

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

События $\{\xi_1 < -\infty\}$ и $\{\xi_2 < -\infty\}$ 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* $\{\xi_1 < +\infty\}$ и $\{\xi_2 < +\infty\}$ так же, как и их пересечение, являются *достоверными*, вероятность которых равна единице, вытекает утверждение 4.

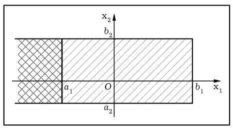

Рис 8.1.

Чтобы найти вероятность попадания двумерной случайной величины (ξ_1,ξ_2) в прямоугольник $\{a_1 < x_1 \leqslant b_1,\, a_2 < x_2 \leqslant b_2\}$ (на рис. 8.1 заштрихован), сначала определим вероятность попадания в полуполосу $\{x_1 \leqslant a_1,\ a_2 < x_2 \leqslant b_2\}$ (отмечена двойной штриховкой). Но эта вероятность представляет собой вероятность попадания в квадрант $\{x_1 \leqslant a_1,\ x_2 \leqslant b_2\}$ за вычетом вероятности попадания в квадрант $\{x_1 \leqslant a_1,\ x_2 \leqslant a_2\}$, т.е. $\mathsf{P}\{\xi_1 \leqslant a_1,\ a_2 < \xi_2 \leqslant b_2\} = F(a_1,b_2) - F(a_1,a_2)$. Аналогично, $\mathsf{P}\{\xi_1 \leqslant b_1,\, a_2 < \xi_2 \leqslant b_2\} = F(b_1,b_2) - F(b_1,a_2)$.Теперь осталось заметить, что вероятность попадания в прямоугольник $\{a_1 < x_1 \leqslant b_1,\, a_2 < x_2 \leqslant b_2\}$ совпадает с вероятностью попадания в полуполосу $\{x_1 \leqslant b_1,\, a_2 < x_2 \leqslant b_2\}$ из которой вычитается вероятность попадания в полуполосу $\{x_1 \leqslant a_1,\, a_2 < x_2 \leqslant b_2\}$, откуда следует утверждение 5.

Подобно одномерному случаю доказывается и утверждение 6.

Наконец, событие $\{\xi_2 < +\infty\}$ является достоверным, поэтому $\{\xi_1 \leqslant x_1\} \cap \{\xi_2 < +\infty\} = \{\xi_1 \leqslant x_1\}$. Аналогично $\{\xi_1 < +\infty\} \cap \{\xi_2 \leqslant x_2\} = \{\xi_2 \leqslant x_2\}$. Отсюда приходим к утверждению 7, которое устанавливает естественную связь между двумерной функцией распределения $F_\xi(x_1,x_2)$ случайного вектора $\xi = (\xi_1,\xi_2)$ и функциями $F_{\xi_1}(x_1)$ и $F_{\xi_2}(x_2)$, которые называют ***одномерными*** (говорят также ***частными***, или ***маргинальными***) ***функциями распределения*** случайных величин ξ_1 и ξ_2.

Дискретные двумерные случайные векторы

Определение 8.3. ***Двумерный случайный вектор*** (ξ,η) называют ***дискретным***, если каждая из случайных величин ξ и η 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* (ξ,η) и соответствующих *вероятностей*, с которыми эти пары значений принимают случайные величины ξ и η (для простоты ограничимся конечным *множеством возможных значений*, когда случайная величина ξ может принимать только значения $x_1, \ldots, x_m$, η — значения $y_1, \ldots, y_n$, а координаты двумерного случайного вектора (ξ,η) — пары значений (x_i,y_j), $i = \overline{1,m}$, $j = \overline{1,n}$.

ξ	η				
	y_1	y_2	$\ldots$	y_n	
x_1	p_{11}	p_{12}	$\ldots$	p_{1n}	$p_{1\bullet}$
x_2	p_{21}	p_{22}	$\ldots$	p_{2n}	$p_{2\bullet}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_m	p_{m1}	p_{m2}	$\ldots$	p_{mn}	$p_{n\bullet}$
	$p_{\bullet 1}$	$p_{\bullet 2}$	$\ldots$	$p_{\bullet n}$	

Таблица 8.1.

Такое перечисление удобно представить в виде таблицы 8.1. В этой таблице в верхней строке перечислены все возможные значения $y_1, \ldots, y_j, \ldots, y_n$ случайной величины η, а в левом столбце — значения $x_1, \ldots, x_i, \ldots, x_m$ случайной величины ξ. На пересечении столбца "y_j" со строкой "x_i" находится вероятность $p_{ij} = \mathsf{P}\{\xi = x_i, \eta = y_j\}$ совместного осуществления событий $\{\xi = x_i\}$ и $\{\eta = y_j\}$.

В таблице 8.1 по определению $p_{i\bullet} = \mathsf{P}\{\xi = x_i\}$ и $p_{\bullet j} = \mathsf{P}\{\eta = y_j\}$ для всех $i = 1,\ldots,m$, $j = 1,\ldots,n$. Поскольку события $\{\xi = x_i, \eta = y_j\}$, $i = 1,\ldots,m$, $j = 1,\ldots,n$ несовместны, то

$$p_{i\bullet} = \sum_{j=1}^{n} p_{ij}, \quad i = 1,\ldots,m, \qquad p_{\bullet j} = \sum_{i=1}^{m} p_{ij}, \quad j = 1,\ldots,n. \tag{8.1}$$

Обозначения $p_{i\bullet}$ и $p_{\bullet j}$ являются общеупотребительными, точка обозначает индекс, по которому суммируются вероятности p_{ij}.

В первом и последнем столбцах таблицы 8.1 указано частное (маргинальное) распределение случайной величины ξ , а в верхней и нижней строках — частное (маргинальное) распределение случайной величины η. Используя таблицу 8.1, нетрудно определить *совместную функцию распределения* $F(x,y)$. Ясно, что для этого необходимо просуммировать p_{ij} по всем тем значениям i и j, для которых $x_i \leqslant x$, $y_j \leqslant y$, т.е.

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

Пример 8.3. В соответствии со *схемой Бернулли* (см. определение 3.5) с вероятностью успеха p и вероятностью неудачи $q = 1 - p$ проводятся два испытания. Выпишем распределение двумерного случайного вектора (ξ, η), где ξ и η, — число успехов в в первом и втором испытаниях соответственно. Каждая из случайных величин ξ и η 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

ξ	η		
	0	1	
0	q^2	qp	q
1	pq	p^2	p
	q	p	

Таблица 8.2.

$$\mathsf{P}\{\xi = 0, \eta = 0\} = q^2,$$

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

$$\mathsf{P}\{\xi = 1, \eta = 0\} = pq.$$

Аналогично заполняем второй столбец:

$$\mathsf{P}\{\xi = 0, \eta = 1\} = qp, \qquad \mathsf{P}\{\xi = 1, \eta = 1\} = p^2.$$

Наконец, на пересечении последнего столбца и строки “0” должно стоять

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

а на пересечении последнего столбца и строки “1” —

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

Построим теперь совместную функцию распределения случайных величин ξ и η:

$$F(x,y) = \mathsf{P}\{\xi \leqslant x, \eta \leqslant y\}.$$

Поскольку при $x < 0$ или $y < 0$ нет ни одного элементарного исхода ω, для которого или $\eta(\omega) \leqslant y$, то для таких x и y событие $\{\xi \leqslant x, \eta \leqslant y\}$ невозможное, и, значит $F(x,y) = 0$ при $x < 0$ или $y < 0$.

Далее, если $0 \leqslant x < 1$ и $0 \leqslant y < 1$, то событие $\{\xi \leqslant x, \eta \leqslant y\}$ эквивалентно событию $\{\xi = 0, \eta = 0\}$, которое, как видно из табл. 8.2, происходит с вероятностью q^2, и $F(x,y) = q^2$.

Если же $0 \leqslant x < 1$, а $y \geqslant 1$, то событие $\{\xi \leqslant x, \eta \leqslant y\}$ совпадает с объединением *непересекающихся событий* $\{\xi = 0, \eta = 0\}$ и $\{\xi = 0, \eta = 1\}$. Тогда $F(x,y) = q^2 + qp = q$. Аналогично $F(x,y) = q^2 + qp = q$ при $x \geqslant 1$ и $0 \leqslant y < 1$.

Наконец, если $x \geqslant 1$ и $y \geqslant 1$, то событие $\{\xi \leqslant x, \eta \leqslant y\}$ достоверно, и, следовательно, $F(x,y) = 1$.

Непрерывные случайные векторы

Определение 8.4. ***Непрерывной двумерной случайной величиной*** (ξ,η) называют такую двумерную случайную величину (ξ,η), *совместную функцию распределения* которой $F(x,y) = \mathsf{P}\{\xi \leqslant x, \eta \leqslant y\}$ можно представить в виде сходящегося несобственного интеграла:

$$F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(s,t)\,ds\,dt = \int\limits_{-\infty}^{x} ds \int\limits_{-\infty}^{y} f(s,t)\,dt = \int\limits_{-\infty}^{y} dt \int\limits_{-\infty}^{x} f(s,t)\,ds.$$

Функцию $f(x,y)$ называют ***совместной (двумерной) плотностью распределения*** случайных величин ξ и η, или плотностью распределения случайного вектора (ξ,η).

Так же как и в одномерном случае, будем предполагать, что $f(x,y)$ 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 интеграла с переменным верхним пределом совместная плотность распределения представляет собой (в точках ее непрерывности) вторую смешанную производную совместной функции распределения:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 F(x,y)}{\partial y \partial x}. \tag{8.2}$$

Заметим, что аналогичный смысл имеет ***совместная (n-мерная) плотность распределения*** случайных величин $\xi_1,\ldots,\xi_n$, или ***плотность распределения случайного вектора*** $\xi = (\xi_1,\ldots,\xi_n)$:

$$f_\xi(x_1,\ldots,x_n) = \frac{\partial^n F_\xi(x_1,\ldots,x_n)}{\partial x_1 \ldots \partial x_n}.$$

Теорема 8.2. *Двумерная плотность распределения обладает следующими свойствами.*

1. $f(x,y) \geqslant 0$.
2. $\mathsf{P}\{a_1 < \xi \leqslant b_1, a_2 < \eta \leqslant b_2\} = \int\limits_{a_1}^{b_1} dx \int\limits_{a_2}^{b_2} f(x,y)\,dy.$
3. $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)\,dxdy = 1.$
4. $\mathsf{P}\{x < \xi \leqslant x+\Delta x, y < \eta \leqslant y+\Delta y\} \approx f(x,y)\Delta x \Delta y.$
5. $\mathsf{P}\{\xi = x, \eta = y\} = 0.$
6. $\mathsf{P}\{(\xi,\eta) \in D\} = \iint\limits_{D} f(x,y)\,dxdy.$
7. $f_\xi(x) = \int\limits_{-\infty}^{+\infty} f(x,y)\,dy.$
8. $f_\eta(y) = \int\limits_{-\infty}^{+\infty} f(x,y)\,dx.$

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:

$$F_\xi(x) = F(x,+\infty) \;=\; \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{+\infty} f(s,t)\,ds\,dt,$$

$$F_\eta(y) = F(+\infty,y) \;=\; \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{y} f(s,t)\,ds\,dt,$$

откуда, дифференцируя интегралы по переменному верхнему пределу и учитывая свойство 4 на с. 27, получаем утверждения 7 и 8 для *одномерных плотностей распределения* $f_\xi(x)$ и $f_\eta(y)$ случайных величин ξ и η.

Пример 8.4. Рассмотрим случайный вектор (ξ,η), плотность распределения которого имеет вид

$$f(x,y) = \begin{cases} A, & x^2+y^2 \leqslant R^2; \\ 0, & x^2+y^2 > R^2. \end{cases}$$

Найдем постоянную A и плотности распределения $f_\xi(x)$, $f_\eta(y)$ случайных величин ξ и η соответственно.

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

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)\,dxdy = \iint\limits_{x^2+y^2\leqslant R^2} A\,dxdy = \pi A R^2 = 1,$$

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

$$f(x,y) = \begin{cases} 0, & x^2+y^2 > R^2; \\ \dfrac{1}{\pi R^2}, & x^2+y^2 \leqslant R^2. \end{cases}$$

Согласно свойству 7 одномерная плотность распределения $f_\xi(x)$ случайной величины ξ вычисляется следующим образом:

$$f_\xi(x) = \int\limits_{-\infty}^{+\infty} f(x,y)\,dy = \int\limits_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}\,dy = \begin{cases} 0, & |x|>R; \\ \dfrac{2\sqrt{R^2-x^2}}{\pi R^2}, & |x|\leqslant R. \end{cases}$$

Аналогичное выражение можно получить и для $f_\eta(y)$.

Определение 8.5. Непрерывный случайный вектор $\xi = (\xi_1,\ldots,\xi_n)$ назовем ***равномерно распределенным в области*** $D \in \mathbb{R}^n$, если его плотность распределения $f(x_1,\ldots,x_n)$ имеет вид:

$$f(x_1,\ldots,x_n) = \begin{cases} C, & (x_1,\ldots,x_n) \in D; \\ 0, & (x_1,\ldots,x_n) \notin D. \end{cases}$$

Лекция 9

Независимые случайные величины

Определение 9.1. ***Случайные величины*** ξ и η называют ***независимыми***, если *совместная функция распределения* $F(x,y)$ является произведением *одномерных функций распределения* $F_\xi(x)$ и $F_\eta(y)$:

$$F(x,y) = F_\xi(x)F_\eta(y) \qquad \forall (x,y) \in \mathbb{R}^2.$$

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

Из определения 9.1 вытекает, что для независимых случайных величин ξ и η события $\{\xi \leqslant x\}$ и $\{\eta \leqslant y\}$ являются независимыми. Можно показать, что независимыми являются и все события $\{x < \xi \leqslant y\}$ и $\{s < \eta \leqslant t\}$.

Теорема 9.1. *Для того чтобы непрерывные случайные величины ξ и η были независимыми, необходимо и достаточно, чтобы для всех x и y*

$$f(x,y) = f_\xi(x)f_\eta(y).$$

Доказательство. Пусть случайные величины ξ и η независимые. Тогда, согласно определению 9.1, $F(x,y) = F_\xi(x)F_\eta(y)$. С учетом формул (8.2) и свойства 4 на с. 27 имеем

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \left(\frac{dF_\xi(x)}{dx}\right)\left(\frac{dF_\eta(y)}{dy}\right) = f_\xi(x)f_\eta(y).$$

Тем самым необходимость утверждения доказана.

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

$$F(x,y) = \iint\limits_{\substack{-\infty<v\leqslant x\\ -\infty<w\leqslant y}} f(v,w)\,dvdw = \left(\int\limits_{-\infty}^{x} f_\xi(v)\,dv\right)\left(\int\limits_{-\infty}^{y} f_\eta(w)\,dw\right) = F_\xi(x)F_\eta(y).$$

Справедлива следующая теорема.

Теорема 9.2. *Если случайные величины ξ и η независимыми, а детерминированные функции $\varphi(t)$ и $\psi(t)$ кусочно-непрерывны, то случайные величины $\varphi(\xi)$ и $\psi(\eta)$ также независимы.*

Пример 9.1. Рассмотрим двумерный вектор (ξ,η), совместная плотность распределения которого имеет вид

$$f(x,y) = \begin{cases} 1, & x \in [0,1] \text{ и } y \in [0,1];\\ 0, & x \notin [0,1] \text{ или } y \notin [0,1]. \end{cases}$$

Легко показать, что одномерные плотности распределения $f_\xi(x)$ и $f_\eta(x)$ случайных величин ξ и η задаются формулой

$$f_\xi(x) = f_\eta(x) = \begin{cases} 1, & x \in [0,1];\\ 0, & x \notin [0,1]. \end{cases}$$

Очевидно, что в данном случае совместная плотность распределения $f(x,y)$ для всех x, y является произведением одномерных плотностей $f_\xi(x)$ и $f_\eta(y)$. Значит, случайные величины ξ и η являются независимыми.

Пример 9.2. Вернемся к примеру 8.4. Рассмотрим случайный вектор (ξ,η) с плотностью распределения

$$f(x,y) = \begin{cases} 0, & x^2+y^2 > R^2; \\ \dfrac{1}{\pi R^2}, & x^2+y^2 \leqslant R^2. \end{cases}$$

Как было показано в примере 8.4 одномерные плотности распределения $f_\xi(x)$ и $f_\eta(x)$ случайных величин ξ и η имеют вид

$$f_\xi(x) = f_\eta(x) = \begin{cases} 0, & |x| > R; \\ \dfrac{2\sqrt{R^2-x^2}}{\pi R^2}, & |x| \leqslant R. \end{cases}$$

Видно, что $f(x,y) \neq f_\xi(x)f_\eta(y)$. Поэтому случайные величины ξ и η являются зависимыми.

Теорема 9.3. *Дискретные случайные величины ξ и η являются независимыми тогда и только тогда, когда для всех возможных значений x_i и y_j*

$$p_{ij} = \mathsf{P}\{\xi = x_i, \eta = y_j\} = \mathsf{P}\{\xi = x_i\}\mathsf{P}\{\eta = y_j\}.$$

Пример 9.3. В **схеме Бернулли** с двумя испытаниями (см. пример 8.3)

$$\begin{aligned} \mathsf{P}\{\xi = 0, \eta = 0\} = q^2 = \mathsf{P}\{\xi = 0\}\mathsf{P}\{\eta = 0\}, \\ \mathsf{P}\{\xi = 0, \eta = 1\} = qp = \mathsf{P}\{\xi = 0\}\mathsf{P}\{\eta = 1\}, \\ \mathsf{P}\{\xi = 1, \eta = 0\} = pq = \mathsf{P}\{\xi = 1\}\mathsf{P}\{\eta = 0\}, \\ \mathsf{P}\{\xi = 1, \eta = 1\} = p^2 = \mathsf{P}\{\xi = 1\}\mathsf{P}\{\eta = 1\}. \end{aligned}$$

Таким образом, числа успехов ξ и η в первом и втором испытаниях представляют собой независимые случайные величины. Впрочем, в силу определения схемы Бернулли иного нельзя было ожидать.

Определение 9.2. ***Случайные величины*** $\xi_1, \ldots, \xi_n$, заданные на одном и том же вероятностном пространстве, называют ***независимыми в совокупности***, если

$$F(x_1, \ldots, x_n) = F_{\xi_1}(x_1)\ldots F_{\xi_n}(x_n) \qquad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Замечание 9.1. Теоремы 9.1 и 9.2 распространяются на любое число случайных величин.

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

Пример 9.4. Свяжем с бросанием тетраэдра из примера 3.5 три случайные величины: ξ_1, ξ_2 и ξ_3, каждая из которых может принимать значения 0 или 1, причем $\xi_1 = 1$, если тетраэдр упал на грань, на которой присутствует цифра 1, и $\xi_1 = 0$ в противном случае. Аналогично ξ_2 характеризует наличие цифры 2, а ξ_3 — цифры 3. Покажем, что случайные величины ξ_1, ξ_2 и ξ_3 будут попарно независимыми, но не являются независимыми в совокупности.

Действительно,

$$\mathsf{P}\{\xi_i = 1\} = \mathsf{P}\{\xi_i = 0\} = \frac{1}{2}, \quad i = 1,2,3,$$

и

$$\mathsf{P}\{\xi_i = 1, \xi_j = 1\} = \frac{1}{4} = \mathsf{P}\{\xi_i = 1\}\mathsf{P}\{\xi_j = 1\}, \quad i \neq j,$$

т.е. ξ_1, ξ_2 и ξ_3 попарно независимы (согласно теореме 3.4 для попарной независимости достаточно проверить только последнее равенство). Однако, например,

$$\mathsf{P}\{\xi_1 = 1, \xi_2 = 1, \xi_3 = 1\} = \frac{1}{4} \neq \mathsf{P}\{\xi_1 = 1\}\mathsf{P}\{\xi_2 = 1\}\mathsf{P}\{\xi_3 = 1\} = \frac{1}{8},$$

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

Лекция 10

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* $(\Omega, \mathscr{A}, \mathsf{P})$ двумерный *случайный вектор* $\xi = (\xi_1, \xi_2)$ и числовую функцию $z = \varphi(x,y)$ числовых аргументов x и y.

Определение 10.1. Случайную величину

$$\eta = \varphi(\xi_1, \xi_2) = \varphi(\xi_1(\omega), \xi_2(\omega))$$

называют ***функцией (скалярной) от двумерной случайной величины (двумерного случайного вектора)*** (ξ_1, ξ_2).

Пусть (ξ_1, ξ_2) — дискретный случайный вектор с законом распределения

$$p_{ij} = \mathsf{P}\{\xi_1 = x_i, \xi_2 = y_j\}, \qquad i,j = 1,2,\ldots.$$

Ясно, что функция $\eta = \varphi(\xi_1, \xi_2)$ является дискретной случайной величиной. Если значения $\varphi(x_i, y_j)$ различны для разных пар (x_i, y_j), то случайная величина η принимает с *вероятностью* p_{ij} значения $\varphi(x_i, y_j)$. Иначе для построения *ряда распределения дискретной случайной величины* η, необходимо объединить в один столбец все одинаковые значения $\varphi(x_i, y_j)$ случайной величины η, приписав этому столбцу суммарную вероятность.

Пример 10.1. Пусть η — случайная величина, равная суммарному числу успехов в двух испытаниях *по схеме Бернулли*, а ξ_i — число успехов в i-м испытании, $i = 1,2$. Тогда

$$\eta = \xi_1 + \xi_2 \text{ и } \varphi(x,y) = x + y.$$

Поскольку ξ_i могут принимать только значения 0 или 1, то случайная величина η 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. 10.1, см. также пример 8.3 и табл. 8.2).

η	$\varphi(0,0)=0$	$\varphi(1,0)=1$	$\varphi(0,1)=1$	$\varphi(1,1)=2$
P_η	q^2	pq	qp	p^2

Таблица 10.1.

Заметим, что двум средним столбцам соответствует одно и то же значение 1 случайной величины η, и их необходимо объединить. Окончательно получаем ряд распределения случайной величины η, представленный в табл. 10.2. Как и следовало ожидать, суммарное число успехов η в двух испытаниях имеет *биномиальное распределение*. #

η	0	1	2
P_η	q^2	$2pq$	p^2

Таблица 10.2.

В том случае, когда $\xi = (\xi_1, \xi_2)$ — *двумерная непрерывная случайная величина* с *плотностью распределения* $f(x,y)$, *функцию распределения* случайной величины $\eta = \varphi(\xi_1, \xi_2)$ можно найти по формуле

$$F_\eta(z) = \iint\limits_{\varphi(x,y) \leqslant z} f(x,y)\, dx dy, \tag{10.1}$$

где область интегрирования состоит из всех значений x и y, для которых $\varphi(x,y) \leqslant z$.

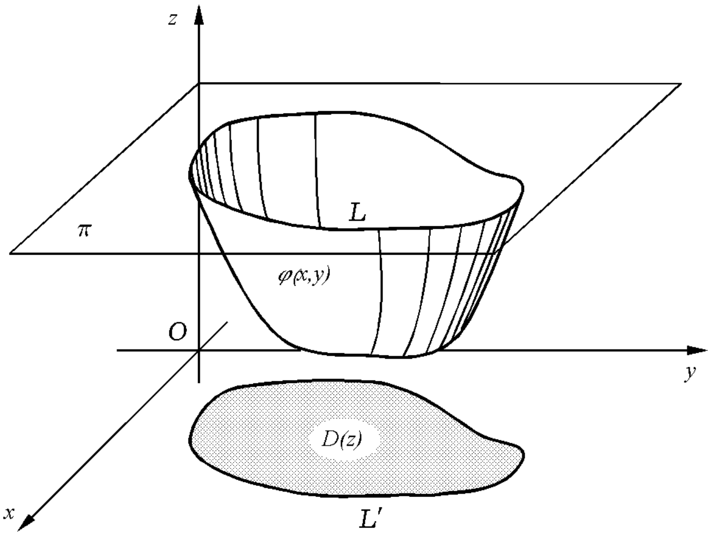

Рис 10.1.

Поясним геометрически вывод формулы (10.1). Пусть поверхность, определенная функцией $z = \varphi(x,y)$, имеет вид “чаши” (см. рис. 10.1) и z — произвольное значение случайной величины $\eta = \varphi(\xi_1, \xi_2)$. Проведем плоскость π, проходящую через точку $(0;0;z)$ и ортогональную оси Oz. Обозначим через L линию пересечения плоскости π и поверхности $z = \varphi(x,y)$; L' — ее проекцию на плоскость xOy; $D(z)$ — ту часть плоскости xOy, попадание в которую случайного вектора (ξ_1, ξ_2) ведет к реализации *события* $\{\eta \leqslant z\}$. Поскольку $\eta = \varphi(\xi_1, \xi_2)$, то

$$D(z) = \{(x,y) : \varphi(x,y) \leqslant z\} = \{\varphi(x,y) \leqslant z\}.$$

События $\{\eta \leqslant z\}$ и $\{(\xi_1, \xi_2) \in D(z)\}$ совпадают, и в соответствии со свойством 6 *двумерной плотности распределения*

$$\mathsf{P}\{\eta \leqslant z\} = \mathsf{P}\{(\xi_1, \xi_2) \in D(z)\} = \iint\limits_{\varphi(x,y) \leqslant z} f(x,y)\, dx dy.$$

Учитывая равенство $\mathsf{P}\{\eta \leqslant z\} = F_\eta(z)$, приходим к формуле (10.1).

Пример 10.2. Пусть ξ_1 и ξ_2 независимые случайные величины, имеющие стандартное нормальное распределение. Найдем распределение случайной величины $\eta = \sqrt{\xi_1^2 + \xi_2^2}$. В этом случае $\varphi(x,y) = \sqrt{x^2 + y^2}$. Очевидно, что

$$F_\eta(z) = \begin{cases} 0, & z < 0; \\ \displaystyle\iint\limits_{\sqrt{x^2+y^2} \leqslant z} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\, dx dy, & z \geqslant 0. \end{cases}$$

Переходя к полярным координатам ρ и φ, имеем

$$F_\eta(z) = \int\limits_0^z d\rho \int\limits_0^{2\pi} \frac{1}{2\pi} e^{-\rho^2/2} \rho\, d\varphi = \int\limits_0^z \rho e^{-\rho^2/2} d\rho = 1 - e^{-z^2/2}, \quad z \geqslant 0.$$

Это распределение известно как *распределение Релея*.

Математическое ожидание функции от случайного вектора

Математическое ожидание $\mathsf{E}\eta$ функции $\eta = \varphi(\xi_1, \xi_2)$ от *дискретной двумерной случайной величины* (ξ_1, ξ_2) можно найти, воспользовавшись формулой

$$\mathsf{E}\eta = \mathsf{E}\varphi(\xi_1, \xi_2) = \sum_{i,j} \varphi(x_i, y_j) p_{ij},$$

где $p_{ij} = \mathsf{P}\{\xi_1 = x_i, \xi_2 = y_j\}$, а функции $\eta = \varphi(\xi_1, \xi_2)$ от *двумерной непрерывной случайной величины* (ξ_1, ξ_2) — формулой

$$\mathsf{E}\eta = \mathsf{E}\varphi(\xi_1,\xi_2) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} \varphi(x,y) f(x,y)\, dx dy, \tag{10.2}$$

где $f(x,y)$ — совместная плотность распределения случайных величин ξ_1 и ξ_2.

Свойства математического ожидания

Некоторые свойства математического ожидания были сформулированы и доказаны в лекции 6. Дополним этот список и приведем недостающие доказательства.

Теорема 10.1. *Математическое ожидание удовлетворяет следующим свойствам.*

1. $\mathsf{E}(\xi_1 + \xi_2) = \mathsf{E}\xi_1 + \mathsf{E}\xi_2$.

2. Если случайные величины ξ_1 *и* ξ_2 *независимы, то* $\mathsf{E}(\xi_1\xi_2) = \mathsf{E}\xi_1 \cdot \mathsf{E}\xi_2$.

Доказательство.

Свойство 1 докажем для дискретных случайных величин (для непрерывных — доказать самостоятельно).

Пусть $\eta = \xi_1 + \xi_2$ ($\eta = \varphi(x,y) = x + y$). Тогда

$$\mathsf{E}\eta = \mathsf{E}(\xi_1 + \xi_2) = \sum_{i,j}(x_i + y_j)p_{ij} = \sum_{i,j} x_i p_{ij} + \sum_{i,j} y_j p_{ij} = \\ = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \sum_i x_i p_{i\bullet} + \sum_j y_j p_{\bullet j} = \mathsf{E}\xi_1 + \mathsf{E}\xi_2.$$

Таким образом, утверждение 1 доказано.

Свойство 2 докажем для для непрерывных случайных величин (для дискретных — доказать самостоятельно). Если ξ_1 и ξ_2 независимые случайные величины, то для математического ожидания их произведения $\eta = \xi_1\xi_2$ (воспользовавшись формулой 10.2 и теоремой 9.1) имеем:

$$\mathsf{E}\eta = \mathsf{E}(\xi_1\xi_2) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x_1 x_2 f(x_1,x_2)\, dx_1 dx_2 = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x_1 x_2 f_{\xi_1}(x_1) f_{\xi_2}(x_2)\, dx_1 dx_2 = \\ = \left(\int\limits_{-\infty}^{+\infty} x_1 f_{\xi_1}(x_1)\, dx_1\right)\left(\int\limits_{-\infty}^{+\infty} x_2 f_{\xi_2}(x_2)\, dx_2\right) = \mathsf{E}\xi_1 \mathsf{E}\xi_2.$$

Замечание 10.1. Свойство 2 также допускает обобщение на произведение конечного числа *независисмых* (*в совокупности*) случайных величин:

$$\mathsf{E}(\xi_1 \cdot \xi_2 \ldots \xi_n) = \mathsf{E}\xi_1 \cdot \mathsf{E}\xi_2 \ldots \mathsf{E}\xi_n.$$

Формула свертки

Важную роль в теории вероятностей и ее применениях играет тот случай, когда ξ_1 и ξ_2 являются *независимыми случайными величинами*, т.е. *их двумерная плотность распределения*

$$f(x,y) = f_{\xi_1}(x) f_{\xi_2}(y)$$

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

$$\eta = \xi_1 + \xi_2.$$

Тогда $\eta = \varphi(\xi_1, \xi_2)$, где

$$\varphi(x,y) = x + y,$$

и, согласно формуле (10.1), находим:

$$F_\eta(z) = \iint\limits_{x+y\leqslant z} f(x,y)dxdy = \iint\limits_{x+y\leqslant z} f_{\xi_1}(x)f_{\xi_2}(y)dxdy =$$

$$= \int\limits_{-\infty}^{+\infty} f_{\xi_1}(x)dx \int\limits_{-\infty}^{z-x} f_{\xi_2}(y)dy = \int\limits_{-\infty}^{+\infty} F_{\xi_2}(z-x)f_{\xi_1}(x)\,dx.$$

Дифференцируя последнюю формулу по z под знаком интеграла, получаем выражение для *плотности* $f_\eta(z)$ *распределения* суммы ξ_1 и ξ_2:

$$f_\eta(z) = \int\limits_{-\infty}^{+\infty} f_{\xi_2}(z-x)f_{\xi_1}(x)\,dx. \tag{10.3}$$

В этом случае говорят, что плотность распределения $f_\eta(z)$ случайной величины η является ***сверткой*** (***композицией***) ***плотностей распределения*** $f_{\xi_1}(x)$ и $f_{\xi_2}(y)$ слагаемых ξ_1 и ξ_2 или что закон распределения суммы двух независимых случайных величин является ***сверткой*** (***композицией***) ***законов распределения*** слагаемых. Соотношение (10.3) условно записывают в виде $f_\eta = f_{\xi_2} * f_{\xi_1}$. Формулу (10.3) называют ***формулой свертки*** для плотностей распределения случайных величин ξ_1 и ξ_2.

Пример 10.3. Пусть ξ_1 и ξ_2 — независимые случайные величины и $\xi_1 \sim N(m_1,\sigma_1^2)$, $\xi_2 \sim N(m_2,\sigma_2^2)$. Найдем плотность распределения суммы $\eta = \xi_1 + \xi_2$. Воспользовавшись формулой свертки, имеем

$$f_\eta(z) = \int\limits_{-\infty}^{+\infty} f_{\xi_2}(z-x)f_{\xi_1}(x)\,dx =$$

$$= \int\limits_{-\infty}^{+\infty} \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(z-x-m_1)^2}{2\sigma_1^2}\right)\frac{1}{\sigma_2\sqrt{2\pi}}\exp\left(-\frac{(x-m_2)^2}{2\sigma_2^2}\right)dx =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{(z-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)\times\int\limits_{-\infty}^{+\infty}\exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(x-\frac{\sigma_1^2 m_2-\sigma_2^2 m_1+\sigma_2^2 z}{\sigma_1^2+\sigma_2^2}\right)^2\right)dx.$$

Делая теперь замену

$$t = \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}\left(x-\frac{\sigma_1^2 m_2-\sigma_2^2 m_1+\sigma_2^2 z}{\sigma_1^2+\sigma_2^2}\right),$$

получаем

$$\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left(-\frac{(z-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)\int\limits_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt =$$

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

Таким образом, случайная величина $\eta \sim N(m_1+m_2,\sigma_1^2+\sigma_2^2)$. Другими словами, композиция плотностей нормальных законов распределения является плотностью нормального закона распределения.

Замечание 10.2. Выведенное в примере 10.3 свойство справедливо для суммы любого конечного числа слагаемых, распределенных по нормальному закону. А именно, если $\xi_1,\ldots,\xi_n$ — независимые нормальные случайные величины с математическими ожиданиями $m_1,\ldots,m_n$ и дисперсиями $\sigma_1^2,\ldots,\sigma_n^2$, то их сумма $eta = \xi_1+\ldots+\xi_n$ является нормальной случайной величиной с математическим ожиданием $m_1+\ldots+m_n$ и дисперсией $\sigma_1^2+\ldots+\sigma_n^2$, т.е. $\eta \sim N\left(m_1+\ldots+m_n,\sigma_1^2+\ldots+\sigma_n^2\right)$.

Лекция 11

Ковариация и коэффициент корреляции случайных величин. Числовые характеристики случайного вектора

Определение 11.1. *Ковариацией* (*корреляционным моментом*) $\mathsf{cov}(\xi,\eta)$ (или $k_{\xi\eta}$) случайных величин ξ и η называется математическое ожидание произведения центрированных случайных величин $\overset{\circ}{\xi}$ и $\overset{\circ}{\eta}$

$$\mathsf{cov}(\xi,\eta) = k_{\xi\eta} = \mathsf{E}[(\xi - m_\xi)(\eta - m_\eta)] = \mathsf{E}(\overset{\circ}{\xi}\overset{\circ}{\eta}).$$

Замечание 11.1. Математическое ожидание произведения центрированных случайных величин $\overset{\circ}{\xi}$ и $\overset{\circ}{\eta}$ называют также вторым центральным смешанным моментом этих величин.

Ковариация $k_{\xi\eta}$ непрерывных случайных величин ξ и η равна

$$k_{\xi\eta} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - m_\xi)(y - m_\eta) f(x,y)\,dx\,dy.$$

Ковариация $k_{\xi\eta}$ дискретных случайных величин ξ и η с конечными множествами возможных значений равна

$$k_{\xi\eta} = \sum_{i=0}^{n}\sum_{j=0}^{n}(x_i - m_\xi)(y_j - m_\eta)p_{ij}.$$

Величина $k_{\xi\eta}$ зависит от единиц измерения случайных величин ξ и η. Во избежание этого неудобства ковариацию часто вычисляют не для ξ и η, а для соответствующих им *нормированных* случайных величин $\overset{*}{\xi} = \frac{\xi - m_\xi}{\sigma_\xi}$, $\overset{*}{\eta} = \frac{\eta - m_\eta}{\sigma_\eta}$, если у случайных величин ξ и η существуют дисперсии $\sigma_\xi^2 > 0$ и $\sigma_\eta^2 > 0$.

Определение 11.2. Ковариация нормированных случайных величин $\overset{*}{\xi}$ и $\overset{*}{\eta}$ называется *коэффициентом корреляции*:

$$\rho_{\xi\eta} = k_{\overset{*}{\xi}\overset{*}{\eta}} = \mathsf{E}[\overset{*}{\xi}\overset{*}{\eta}] = \frac{\mathsf{E}[(\xi - m_\xi)(\eta - m_\eta)]}{\sigma_\xi\sigma_\eta} = \frac{k_{\xi\eta}}{\sigma_\xi\sigma_\eta}.$$

Определение 11.3. Случайные величины ξ и η называют *коррелированными* (*корреляционно зависимыми*), если $k_{\xi\eta} \neq 0$, и *некоррелированными*, если $k_{\xi\eta} = 0$.

Определение 11.4. Случайные величины ξ и η называют *положительно коррелированными*, если если $k_{\xi\eta} > 0$, и *отрицательно коррелированными*, если $k_{\xi\eta} < 0$.

Свойства ковариации

1. $\mathsf{cov}(\xi,\xi) = \mathsf{D}\xi$.

2. $\mathsf{D}(\xi_1 + \ldots + \xi_n) = \mathsf{D}\xi_1 + \ldots + \mathsf{D}\xi_n + 2\sum\limits_{1\leqslant i<j\leqslant n} \mathsf{cov}(\xi_i,\xi_j)$.

Для случая $n = 2$ эта формула была доказана в лекции 6. Для произвольного конечного числа слагаемых доказательство проводится по индукции.

3. $|\rho_{\xi\eta}| \leqslant 1$, т. е. $|k_{\xi\eta}| \leqslant \sigma_\xi \sigma_\eta$.

Пусть $\sigma_\xi > 0$ и $\sigma_\eta > 0$. Рассмотрим случайные величины $\overset{*}{\xi}$ и $\overset{*}{\eta}$. Для них по свойству 4) на с. 31 имеем: $\mathsf{D}[\overset{*}{\xi}] = \mathsf{D}[\overset{*}{\eta}] = 1$, и по определению $k_{\overset{*}{\xi}\overset{*}{\eta}} = \rho_{\xi\eta}$. Поэтому по свойству 2 получаем: $0 \leqslant \mathsf{D}[\overset{*}{\xi} + \overset{*}{\eta}] = \mathsf{D}[\overset{*}{\xi}] + \mathsf{D}[\overset{*}{\eta}] + 2k_{\overset{*}{\xi}\overset{*}{\eta}} = 2 + 2\rho_{\xi\eta}$. Значит, $\rho_{\xi\eta} \geqslant -1$. Аналогично получаем неравенство $0 \leqslant \mathsf{D}[\overset{*}{\xi} - \overset{*}{\eta}] = 2 - 2\rho_{\xi\eta}$. Откуда следует, что $\rho_{\xi\eta} \leqslant 1$. Объединяя эти два неравенства, получаем требуемое свойство.

4. Из независимости случайных величин ξ и η с конечными дисперсиями следует их некоррелированность.

Пусть, например, случайные величины ξ и η непрерывны, тогда согласно теореме 9.1 и свойству 4) на с. 31 имеем

$$k_{\xi\eta} = \mathsf{E}[(\xi - m_\xi)(\eta - m_\eta)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - m_\xi)(y - m_\eta) f(x,y)\,dx\,dy =$$
$$= \int_{-\infty}^{\infty} (x - m_\xi) f_\xi(x)\,dx \int_{-\infty}^{\infty} (y - m_\eta) f_\eta(y)\,dy = \mathsf{E}[\overset{\circ}{\xi}]\mathsf{E}[\overset{\circ}{\eta}] = 0.$$

5. $k_{\xi\eta} = \mathsf{E}[\xi\eta] - m_\xi m_\eta$.

Пусть, например, случайная величина $Z = (\xi,\eta)$ непрерывна. Тогда, согласно свойству 1 из теоремы 10.1, имеем

$$k_{\xi\eta} = \mathsf{E}[(\xi - m_\xi)(\eta - m_\eta)] = \mathsf{E}[\xi\eta] + m_\xi m_\eta - 2m_\xi m_\eta = \mathsf{E}[\xi\eta] - m_\xi m_\eta.$$

6. Если случайные величины ξ и η некоррелированы, то $\mathsf{D}[\xi + \eta] = \mathsf{D}[\xi] + \mathsf{D}[\eta]$.

Это свойство вытекает из свойства 2.

7. Если $\eta_i = a_i\xi_i + b_i$, $i = 1,2$, то $\mathsf{cov}(\eta_1,\eta_2) = a_1a_2\mathsf{cov}(\xi_1,\xi_2)$.

Пусть $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$. Тогда

$$\mathsf{cov}(\eta_1,\eta_2) = \mathsf{E}\big((\eta_1 - \mathsf{E}\eta_1)(\eta_2 - \mathsf{E}\eta_2)\big) =$$
$$= \mathsf{E}\big((a_1\xi_1 + b_1 - a_1\mathsf{E}\xi_1 - b_1)(a_2\xi_2 + b_2 - a_2\mathsf{E}\xi_2 - b_2)\big) = \mathsf{E}\big(a_1a_2(\xi_1 - \mathsf{E}\xi_1)(\xi_2 - \mathsf{E}\xi_2)\big).$$

Свойства коэффициента корреляции

1. $\rho(\xi,\xi) = 1$.

2. Если случайные величины ξ и η являются независимыми (и существуют $\mathsf{D}\xi > 0$ и $\mathsf{D}\eta > 0$), то $\rho(\xi,\eta) = 0$.

3. $\rho(a_1\xi_1 + b_1, a_2\xi_2 + b_2) = \pm\rho(\xi_1,\xi_2)$. При этом знак плюс нужно брать в том случае, когда a_1 и a_2 имеют одинаковые знаки, и минус — в противном случае.

4. $-1 \leqslant \rho(\xi,\eta) \leqslant 1$.

5. Если $\eta = a\xi + b$, где a, b — постоянные, $a \neq 0$, $\sigma_\xi > 0$, то

$$\rho_{\xi\eta} = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

Действительно,

$$m_\eta = \mathsf{E}[\eta] = \mathsf{E}[a\xi + b] = a\mathsf{E}[\xi] + b = am_\xi + b,$$
$$\mathsf{D}[\eta] = \mathsf{E}[(\eta - m_\eta)^2] = a^2 d_\xi, \quad \sigma_\eta = \sqrt{d_\eta} = |a|\sqrt{d_\xi} = |a|\sigma_\xi,$$
$$k_{\xi\eta} = \mathsf{E}[(\xi - m_\xi)(\eta - m_\eta)] = \mathsf{E}[a(\xi - m_\xi)^2] = a\mathsf{D}[\xi] = a\sigma_\xi^2,$$
$$\rho_{\xi\eta} = \frac{k_{\xi\eta}}{\sigma_\xi\sigma_\eta} = \frac{a\sigma_\xi^2}{\sigma_\xi^2|a|} = \frac{a}{|a|}, \quad \text{т. е.} \quad |\rho_{\xi\eta}| = 1.$$

Следовательно, если $a>0$, то $\rho_{\xi\eta}=1$ и, если $a<0$, то $\rho_{\xi\eta}=-1$.

Таким образом, по величине $\rho_{\xi\eta}$ можно судить о связи между случайными величинами ξ и η: если $\rho_{\xi\eta}=1$, то большим значениям x случайной величины ξ соответствуют большие значения y случайной величины η, а если $\rho_{\xi\eta}=-1$, то наоборот.

6. Пусть $|\rho_{\xi\eta}|=1$. Тогда случайные величины ξ и η связаны линейной зависимостью.

Рассмотрим сначала случай $\rho_{\xi\eta}=1$. Тогда

$$\mathsf{D}(\overset{*}{\xi}-\overset{*}{\eta})=\mathsf{D}\,\overset{*}{\xi}+\mathsf{D}\,\overset{*}{\eta}-2\mathsf{cov}(\overset{*}{\xi},\overset{*}{\eta})=\mathsf{D}\,\overset{*}{\xi}+\mathsf{D}\,\overset{*}{\eta}-2\rho_{\xi\eta}=1+1-2=0.$$

Отсюда следует, что величина $\overset{*}{\xi}-\overset{*}{\eta}$ является константой. Поэтому $\frac{\xi-m_\xi}{\sigma_\xi}=C+\frac{\eta-m_\eta}{\sigma_\eta}$. Таким образом, линейная зависимость между ξ и η доказана.

Пусть теперь $\rho_{\xi\eta}=-1$. Тогда

$$\mathsf{D}(\overset{*}{\xi}+\overset{*}{\eta})=\mathsf{D}\,\overset{*}{\xi}+\mathsf{D}\,\overset{*}{\eta}+2\mathsf{cov}(\overset{*}{\xi},\overset{*}{\eta})=\mathsf{D}\,\overset{*}{\xi}+\mathsf{D}\,\overset{*}{\eta}+2\rho_{\xi\eta}=1+1-2=0.$$

Отсюда следует, что величина $\overset{*}{\xi}+\overset{*}{\eta}$ является константой. Поэтому $\frac{\xi-m_\xi}{\sigma_\xi}=C-\frac{\eta-m_\eta}{\sigma_\eta}$. Таким образом, линейная зависимость между ξ и η доказана и в этом случае.

Пример 11.1. Рассмотрим две случайные величины, ξ и $\eta=\xi^2$, где $\xi\sim R(-a;a)$. Тогда ковариация между ξ и η равна

$$k_{\xi\eta}=\mathsf{E}[(\xi-m_\xi)(\eta-m_\eta)]=\mathsf{E}[\xi(\eta-m_\eta)]=\mathsf{E}[\xi(\xi^2-m_\eta)]=\frac{1}{2a}\int_{-a}^{a}x(x^2-m_\eta)\,dx=0.$$

Таким образом, случайные величины ξ и η некоррелированы. В то же время случайные величины ξ и η связаны самой «жесткой» зависимостью — функциональной.

Определение 11.5. Вектор $m=(\mathsf{E}\xi_1,\ldots,\mathsf{E}\xi_n)$ называют ***математическим ожиданием случайного вектора*** $\xi=(\xi_1,\ldots,\xi_n)$.

Определение 11.6. ***Ковариационной матрицей*** (***матрицей ковариаций***) ***случайного вектора*** $\xi=(\xi_1,\ldots,\xi_n)$ называют квадратную матрицу K размера $n\times n$ с элементами $k_{ij}=\mathsf{cov}(\xi_i,\xi_j)$, $i,j=1,\ldots,n$.

Свойства ковариационной матрицы

1. $k_{ii}=\mathsf{D}\xi_i$, $i=1,\ldots,n$.
2. K является симметричной матрицей, т.е. $k_{ij}=k_{ji}$, $i,j=1,\ldots,n$.
3. K является неотрицательно определенной матрицей, т.е. $\sum\limits_{i,j=1}^{n}a_ia_jk_{ij}\geqslant 0$ для любых действительных чисел $a_1,\ldots,a_n$.

Действительно

$$0\leqslant\mathsf{E}\left(\sum_{i=1}^{n}a_i\,\overset{\circ}{\xi}_i\right)^2=\mathsf{E}\left(\sum_{i,j=1}^{n}a_ia_j\,\overset{\circ}{\xi}_i\overset{\circ}{\xi}_j\right)=\sum_{i,j=1}^{n}a_ia_j\mathsf{E}\left(\overset{\circ}{\xi}_i\overset{\circ}{\xi}_j\right)=\sum_{i,j=1}^{n}a_ia_jk_{ij}.$$

Определение 11.7. ***Корреляционной*** (***нормированной ковариационной***) ***матрицей*** случайного вектора $\xi=(\xi_1,\ldots,\xi_n)$ называют матрицу $R=(\rho_{ij})=(\rho(\xi_i,\xi_j))$, состоящую из коэффициентов корреляций случайных величин ξ_i и ξ_j.

Многомерное нормальное распределение

Определение 11.8. Говорят, что случайный вектор $\xi=(\xi_1,\ldots,\xi_n)$ имеет ***нормальное*** (***гауссовское***) ***распределение***, $\xi\sim\mathcal{N}(m;K)$, если его плотность распределения есть

$$f_\xi(x)=\frac{1}{\sqrt{(2\pi)^n\det K}}\exp\left\{-\frac{1}{2}(x-m)^TK^{-1}(x-m)\right\},$$

где $\det K$ — определитель положительно определенной матрицы K, $m = (m_1, \ldots, m_n)$, $x = (x_1, \ldots, x_n)$.

Плотность нормального распределения можно задать также через элементы обратной матрицы $K^{-1} = C$ следующим образом:

$$f_\xi(x_1, \ldots, x_n) = \frac{\sqrt{\det C}}{\sqrt{(2\pi)^n}} \exp\left\{-\frac{1}{2}\left(\sum_{i,j=1}^{n} c_{ij}(x_i - m_i)(x_j - m_j)\right)\right\},$$

Свойства нормального распределения $\mathcal{N}(m;K)$

1) $\mathsf{E}\xi = m$, ковариационная матрица случайного вектора ξ равна $\mathsf{E}[(\xi - m)(\xi - m)^T] = K$.

2) Так как матрица K невырожденная, то каждая i-я компонента ξ_i вектора ξ распределена нормально.

3) Если $\xi \sim \mathcal{N}(m;K)$, то случайный вектор $\eta = A\xi + b$, где A — матрица размерности $n \times k$, имеющая ранг, равный k, и b — вектор размерности n, имеет распределение $\mathcal{N}(m_\eta; K_\eta)$, где $m_\eta = Am + b$, $K_\eta = AKA^T$.

4) Пусть случайная величина $\eta = a_1\xi_1 + \ldots + a_n\xi_n$, где $\xi_1, \ldots, \xi_n$ распределены нормально, а коэффициенты $a_1, \ldots, a_n \in \mathbb{R}^1$ не все равны нулю. Тогда случайная величина η распределена нормально.

5) Из попарной некоррелированности компонент случайного вектора $\xi \sim \mathcal{N}(m;K)$ вытекает их независимость.

Действительно, если компоненты случайного вектора попарно некоррелированы, то $k_{ij} = 0$ для любых $i \neq j$. Поэтому $c_{ij} = 0$ для любых $i \neq j$. Кроме того $c_{ii} = 1/\sigma_i^2$, $i = 1, \ldots, n$. Следовательно,

$$f(x_1, \ldots, x_n) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{-\frac{(x_1 - m_1)^2}{2\sigma_1^2}\right\} \times \ldots$$
$$\ldots \times \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left\{-\frac{(x_n - m_n)^2}{2\sigma_n^2}\right\} = f_1(x_1) \cdot \ldots \cdot f_n(x_n),$$

т. е. случайные величины $\xi_1, \ldots, \xi_n$ независимы.

С другой стороны согласно свойству 4 ковариации на с. 55 из независимости случайных величин следует их некоррелированность. Таким образом, для нормального распределения условия некоррелированности и независимости эквивалентны.

Лекция 12

Неравенства Чебышёва. Виды сходимости случайных последовательностей

Теорема 12.1 (неравенство Чебышева (усиленный вариант))**.** *Пусть r-й абсолютный момент случайной величины ξ конечен, т. е.* $\mathsf{E}|\xi|^r < \infty$.

Тогда для всех $\varepsilon > 0$ выполняется неравенство

$$\mathsf{P}\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\mathsf{E}|\xi|^r}{\varepsilon^r}. \tag{12.1}$$

Доказательство. Для простоты доказательства предположим, что у случайной величины ξ существует плотность распределения $f_\xi(x)$. Тогда

$$\mathsf{E}|\xi|^r = \int\limits_{-\infty}^{+\infty} |x|^r f_\xi(x)\,dx \geqslant \int\limits_{|x|\geqslant\varepsilon} |x|^r f_\xi(x)\,dx \geqslant \varepsilon^r \int\limits_{|x|\geqslant\varepsilon} f_\xi(x)\,dx = \varepsilon^r \mathsf{P}\{|\xi| \geqslant \varepsilon\},$$

откуда следует доказываемое утверждение.

Рассмотрим следующие важные частные случаи приведенного неравенства.

1. Пусть случайная величина ξ принимает только неотрицательные значения и $\mathsf{E}\xi < \infty$. Полагая в неравенстве (12.1) $r = 1$, получим, что при любом $\varepsilon > 0$ справедливо соотношение

$$\mathsf{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathsf{E}\xi}{\varepsilon}.$$

2. Пусть случайная величина ξ такова, что $\mathsf{E}\xi^2 < \infty$. Полагая в неравенстве (12.1) $r = 2$, получим, что при любом $\varepsilon > 0$ справедливо соотношение

$$\mathsf{P}\{|\xi| \geqslant \varepsilon\} \leqslant \frac{\mathsf{E}\xi^2}{\varepsilon^2}.$$

3. Применяя неравенство (12.1) для случайной величины $\xi - \mathsf{E}\xi$, получим при $r = 2$

$$\mathsf{P}\{|\xi - \mathsf{E}\xi| \geqslant \varepsilon\} \leqslant \frac{\mathsf{E}(\xi - \mathsf{E}\xi)^2}{\varepsilon^2} = \frac{\mathsf{D}\xi}{\varepsilon^2}.$$

Замечание 12.1. Последнее неравенство, позволяющее оценить сверху вероятность отклонения случайной величины от ее математического ожидания на основе информации лишь о ее дисперсии, широко используется в теории оценивания и управления стохастическими системами. В литературе чаще всего именно последнее неравенство называют неравенством Чебышева.

Пример 12.1. Суточный расход электроэнергии для личных нужд в населенном пункте составляет в среднем 4000 кВт·ч. Оценить вероятность того, что в ближайшие сутки расход электроэнергии в этом населенном пункте не превысит 10 000 кВт·ч.

Решение. Пусть случайная величина ξ — расход электроэнергии в течение суток. По условию $\mathsf{E}\xi = 4000$. Поскольку случайная величина ξ неотрицательна и $\mathsf{E}\xi$ конечно, то, применяя неравенство Чебышева для $r = 1$, получаем

$$\mathsf{P}\{\xi > \varepsilon\} \leqslant \mathsf{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathsf{E}\xi}{\varepsilon}.$$

Таким образом,

$$\mathsf{P}\{\xi > 10\,000\} \leqslant 4000/10\,000 = 0{,}4.$$

Следовательно,

$$\mathsf{P}\{\xi \leqslant 10\,000\} = 1 - \mathsf{P}\{\xi > 10\,000\} \geqslant 0{,}6.$$

Пример 12.2. Некоторый период времени на бирже сохранялся относительно стабильный курс валюты. На основании данных биржевой статистики за этот период была составлена следующая таблица возможных значений изменения курса валют.

Возможное изменение курса, %	-1	$-0{,}5$	0	0,5	1
Вероятность изменения	0,1	0,3	0,5	0,05	0,05

Таблица 12.1.

С помощью неравенства Чебышева оценить вероятность того, что произойдет изменение курса валюты меньше чем на 0,6%, и сравнить полученную оценку с точным значением вероятности.

Решение. Пусть случайная величина ξ — изменение курса валюты (в процентах). Требуется оценить вероятность $\mathsf{P}\{|\xi| < 0{,}6\}$. Воспользуемся для этого неравенством Чебышева.

$$\mathsf{P}\{|\xi| < 0{,}6\} = 1 - \mathsf{P}\{|\xi| \geqslant 0{,}6\} \geqslant 1 - \frac{\mathsf{E}\xi^2}{(0{,}6)^2}.$$

Чтобы вычислить $\mathsf{E}\xi^2$, построим ряд распределения случайной величины ξ^2 (см. табл. 12.2). Тогда $\mathsf{E}\xi^2 = 0{,}25 \cdot 0{,}35 + 1 \cdot 0{,}15 = 0{,}2375$.

ξ^2	0	0,25	1
P	0,5	0,35	0,15

Таблица 12.2.

Таким образом, имеем

$$\mathsf{P}\{|\xi| < 0{,}6\} \geqslant 1 - \frac{0{,}2375}{(0{,}6)^2} \approx 0{,}34.$$

Заметим, что можно вычислить точное значение этой вероятности, так как нам известен ряд распределения случайной величины ξ. Действительно,

$$\mathsf{P}\{|\xi| < 0{,}6\} = 1 - \mathsf{P}\Big(\{\xi = -1\} + \{\xi = 1\}\Big) = 1 - 0{,}1 - 0{,}05 = 0{,}85.$$

Отметим, что получаемая с помощью неравенства Чебышева оценка оказывается весьма грубой.

Виды сходимости случайных последовательностей

Определение 12.1. Бесконечная последовательность случайных величин ξ_n, $n = 1, 2, \ldots$, определенная на одном пространстве элементарных событий Ω, называется ***случайной последовательностью*** и обозначается $\{\xi_n\}$, $n = 1, 2, \ldots$.

Замечание 12.2. Если последовательность состоит из детерминированных величин x_n, то говорят, что последовательность сходится к величине x (это обозначается $x_n \to x$ или $\lim\limits_{n\to\infty} x_n = x$), если для любого $\varepsilon > 0$ найдется такое $N > 0$, что $|x_n - x| < \varepsilon$ для всех $n \geqslant N$. Попробуем уточнить смысл этого понятия для случайной последовательности. Так как для любого n, вообще говоря, может найтись такое $\varepsilon > 0$, что случайное событие $\{\omega : |\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\} \neq \varnothing$, то нельзя говорить о сходимости случайной последовательности ξ_n к ξ в приведенном выше детерминированном смысле. Мы рассмотрим четыре вида сходимости последовательностей случайных величин.

Определение 12.2. Случайная последовательность $\{\xi_n\}$, $n = 1,2,\ldots$, ***сходится почти наверное*** (п.н.) к случайной величине ξ при $n \to \infty$, что записывается как $\xi_n \xrightarrow{\text{п.н.}} \xi$, если

$$\mathsf{P}\left\{\omega\colon \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\right\} = 1. \quad \#$$

Очевидно, что если $\xi_n \xrightarrow{\text{п.н.}} \xi$, то вероятность события, состоящего из таких ω, что последовательность $\{x_n\}$ реализаций случайных величин $\xi_n(\omega)$ не сходится к реализации x случайной величины $\xi(\omega)$, равна нулю:

$$\mathsf{P}\left\{\omega\colon \lim_{n\to\infty} \xi_n(\omega) \neq \xi(\omega)\right\} = 0.$$

Таким образом, сходимость почти наверное случайной последовательности понимается по реализациям случайных величин ξ_n и ξ и в этом смысле похожа на сходимость детерминированной последовательности.

Определение 12.3. Случайная последовательность $\{\xi_n\}$, $n = 1,2,\ldots$, ***сходится по вероятности*** к случайной величине ξ при $n \to \infty$, что записывается как $\xi_n \xrightarrow{\text{P}} \xi$, если для всех $\varepsilon > 0$ справедливо

$$\lim_{n\to\infty} \mathsf{P}\{|\xi_n - \xi| \leqslant \varepsilon\} = 1. \quad \#$$

Очевидно, что условие сходимости $\xi_n \xrightarrow{\text{P}} \xi$ в приведенном определении эквивалентно следующему: $\lim\limits_{n\to\infty} \mathsf{P}\{|\xi_n - \xi| > \varepsilon\} = 0$.

Можно показать, что сходимость п.н. для случайной последовательности влечет за собой и сходимость по вероятности. Из сходимости по вероятности не следует сходимость п.н.

Определение 12.4. Случайная последовательность $\{\xi_n\}$, $n = 1,2,\ldots$, ***сходится*** к случайной величине ξ ***в среднем квадратическом*** при $n \to \infty$, что записывается как $\xi_n \xrightarrow{\text{с.к.}} \xi$, если $\mathsf{E}|\xi_n - \xi|^2 \to 0$ при $n \to \infty$. #

Покажем, что если $\xi_n \xrightarrow{\text{с.к.}} \xi$, то $\xi_n \xrightarrow{\text{P}} \xi$. Действительно, рассмотрим случайные величины $\eta_n = \xi_n - \xi$. В силу неравенства Чебышева для случайных величин η_n имеем

$$\mathsf{P}\{|\eta_n| > \varepsilon\} \leqslant \mathsf{P}\{|\eta_n| \geqslant \varepsilon\} \leqslant \frac{\mathsf{E}\eta_n^2}{\varepsilon^2} = \frac{\mathsf{E}|\xi_n - \xi|^2}{\varepsilon^2}.$$

Поэтому, если $\xi_n \xrightarrow{\text{с.к.}} \xi$, т. е. $\mathsf{E}|\xi_n - \xi|^2 \to 0$ при $n \to \infty$, то для любого $\varepsilon > 0$ выполняется $\mathsf{P}\{|\eta_n| > \varepsilon\} \to 0$ при $n \to \infty$, т. е. $\xi_n \xrightarrow{\text{P}} \xi$. Из сходимости по вероятности не следует сходимость в среднем квадратическом.

Определение 12.5. Пусть $F_n(x)$ — функция распределения случайной величины ξ_n, где $n = 1,2,\ldots$, и $F(x)$ — функция распределения случайной величины ξ. Случайная последовательность $\{\xi_n\}$, $n = 1,2,\ldots$, ***сходится по распределению*** к случайной величине ξ при $n \to \infty$, если последовательность функций $F_n(x)$ сходится к функции $F(x)$ в каждой точке x непрерывности функции $F(x)$, т. е. $F_n(x) \to F(x)$ при $n \to \infty$. Этот вид сходимости будем обозначать $\xi_n \xrightarrow{d} \xi$. #

Если $\xi_n \xrightarrow{\text{P}} \xi$, то можно доказать, что и $\xi_n \xrightarrow{d} \xi$. Обратное утверждение, вообще говоря, неверно.

Связь между различными видами сходимости удобно представить в виде логической схемы (рис. 12.1).

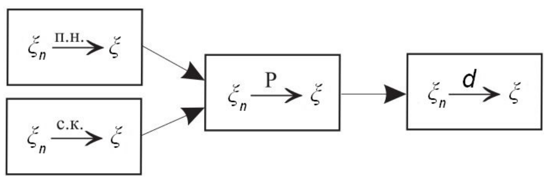

Рис 12.1.

Лекция 13

Закон больших чисел. Центральная предельная теорема. Метод Монте–Карло.

13.1 Закон больших чисел (ЗБЧ)

При определении вероятности указывался эмпирический факт, состоящий в устойчивости частоты появления события A в исследуемом опыте G при последовательном его повторении. Этот экспериментальный факт может быть обоснован математически с помощью закона больших чисел.

Определение 13.1. Будем говорить, что случайная последовательность $\{\xi_n\}$, $n = 1,2,\ldots$, является ***последовательностью независимых случайных величин*** ξ_n, если при любом n случайные величины $\xi_1,\ldots,\xi_n$ независимы.

Определение 13.2. Случайная величина $\eta_n = \frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ называется ***усредненной суммой*** случайных величин ξ_k, $k = \overline{1,n}$.

Пусть случайные величины ξ_k, $k = \overline{1,n}$, независимы. Обозначим $m_k = \mathsf{E}\xi_k$, $d_k = \mathsf{D}\xi_k$, $k = \overline{1,n}$. Тогда, используя свойства математического ожидания и дисперсии, получим

$$\mathsf{E}\eta_n = \frac{1}{n}\sum_{k=1}^{n} m_k, \quad \mathsf{D}\eta_n = \frac{1}{n^2}\sum_{k=1}^{n}\mathsf{D}\xi_k = \frac{1}{n^2}\sum_{k=1}^{n} d_k.$$

Определение 13.3. Будем говорить, что к последовательности $\{\xi_n\}$, $n = 1,2,\ldots$, независимых случайных величин применим ***закон больших чисел (ЗБЧ)*** , если $|\eta_n - \mathsf{E}\eta_n| \xrightarrow{\mathrm{P}} 0$ при $n \to \infty$.

Теорема 13.1 (теорема Маркова)**.** *Если для последовательности* $\{\xi_n\}$, $n = 1,2,\ldots$, *независимых случайных величин выполняется условие* $\lim\limits_{n\to\infty}\mathsf{D}\eta_n = 0$, *то к этой последовательности применим закон больших чисел.*
ДОКАЗАТЕЛЬСТВО. Утверждение теоремы равносильно тому, что

$$\lim_{n\to\infty}\mathsf{P}\{|\eta_n - \mathsf{E}\eta_n| > \varepsilon\} = 0.$$

Согласно неравенству Чебышева при $r = 2$ имеем

$$\mathsf{P}\{|\eta_n - \mathsf{E}\eta_n| > \varepsilon\} \leqslant \mathsf{P}\{|\eta_n - \mathsf{E}\eta_n| \geqslant \varepsilon\} \leqslant \frac{\mathsf{D}\eta_n}{\varepsilon^2}.$$

Из условия $\lim\limits_{n\to\infty}\mathsf{D}\eta_n = 0$ получаем, что $|\eta_n - \mathsf{E}\eta_n| \xrightarrow{\mathrm{P}} 0$.

Утверждение теоремы остается верным, если случайные величины $\{\xi_n\}$, $n = 1,2,\ldots$, являются лишь попарно некоррелированными.

Теорема 13.2 (теорема Чебышева)**.** *Если последовательность* $\{\xi_n\}$ *образована независимыми случайными величинами, дисперсии которых равномерно ограничены, т. е. существует такая константа* c*, что* $\mathsf{D}[\xi_n] \leqslant c$ *для всех* $n = 1,2,\ldots$*, то к этой последовательности применим закон больших чисел.*

Доказательство. Так как $\mathsf{D}\xi_k \leqslant c$ для всех $k = 1,2,\ldots$, то

$$\mathsf{D}\eta_n = \frac{1}{n^2}\sum_{k=1}^{n}\mathsf{D}\xi_k \leqslant \frac{c}{n}.$$

Но $c/n \to 0$ при $n \to \infty$, т. е. условие теоремы 13.1 выполнено и к последовательности $\{\xi_n\}$, $n = 1,2,\ldots$, применим закон больших чисел.

Теорема 13.3. *Если последовательность* $\{\xi_n\}$*,* $n = 1,2,\ldots$*, образована независимыми случайными величинами с одинаковыми распределениями и конечной дисперсией* $\mathsf{D}\xi < +\infty$*, то к этой последовательности применим закон больших чисел, причем* $\eta_n \xrightarrow{\text{P}} m_\xi$*, где* $m_\xi = m_k = \mathsf{E}\xi_k$*,* $k = 1,2,\ldots$*.*

Доказательство. В данном случае $\mathsf{D}\xi_k = d_\xi < \infty$ для всех $k = 1,2,\ldots$ Поэтому условие теоремы Чебышева выполнено. Следовательно $|\eta_n - \mathsf{E}\eta_n| \xrightarrow{\text{P}} 0$ при $n \to \infty$. Кроме того, $\mathsf{E}\xi_k = m_k = m_\xi$ для всех $k = 1,2,\ldots$ Таким образом,

$$\mathsf{E}\eta_n = \frac{1}{n}\sum_{k=1}^{n} m_k = m_\xi,$$

откуда следует, что $|\eta_n - m_\xi| \xrightarrow{\text{P}} 0$.

Определение 13.4. Будем говорить, что к последовательности $\{\xi_n\}$, $n = 1,2,\ldots$, независимых случайных величин применим ***усиленный закон больших чисел (УЗБЧ)***, если $|\eta_n - \mathsf{E}\eta_n| \xrightarrow{\text{п.н.}} 0$ при $n \to \infty$.

Из усиленного закона больших чисел следует закон больших чисел, так как из сходимости почти наверное следует сходимость по вероятности.

Теорема 13.4 (теорема Колмогорова)**.** *К последовательности* $\{\xi_n\}$*,* $n = 1,2,\ldots$*, независимых одинаково распределенных случайных величин, с конечным математическим ожиданием* m_ξ *применим усиленный закон больших чисел, т.е.* $\eta_n \xrightarrow{\text{п.н.}} m_\xi$*.*

Замечание 13.1. В данной теореме, в отличие от теоремы 13.3, не требуется существования дисперсии случайных величин ξ_n, и при этом утверждение оказывается более сильным. Но доказательство этой теоремы значительно сложнее, чем доказательство теоремы 13.3, поэтому мы не приводим его в этом курсе.

Замечание 13.2. Закон больших чисел — это, по сути, свойство случайной последовательности $\{\xi_n\}, n = 1,2,\ldots$, состоящее в том, что случайные отклонения отдельных независимых случайных величин ξ_n от их общего среднего значения m_ξ при большом n в своей массе взаимно погашаются. Поэтому, хотя сами величины ξ_n случайны, их среднее арифметическое значение при достаточно большом n практически уже неслучайно и близко к m_ξ. Таким образом, если математическое ожидание m_ξ случайных величин ξ_n заранее неизвестно, то, согласно теореме 13.4, его можно вычислить с любой «степенью точности» с помощью среднего арифметического $\eta_n = \frac{1}{n}\sum\limits_{k=1}^{n}\xi_k$. Но при этом встает вопрос о точности такого приближения. Ответ на этот вопрос будет дан ниже.

Рассмотрим опыты, проводимые по схеме Бернулли, в результате которых событие A («успех») происходит с вероятностью $p = \mathsf{P}(A)$. Рассмотрим частоту «успехов» $W_n = \xi/n$, где ξ есть число «успехов» при n испытаниях. Случайная величина ξ имеет биномиальное распределение $\mathbf{Bi}(n;p)$.

Теорема 13.5 (теорема Бернулли, усиленный вариант). *Частота «успехов» сходится почти наверное к вероятности «успеха», т. е.* $W_n \xrightarrow{\text{п.н.}} p$.

Доказательство. Так как ξ имеет биномиальное распределение, то частоту успехов $W_n = \xi/n$ можно представить в виде усредненной суммы независимых одинаково распределенных случайных величин ξ_k, $k = \overline{1,n}$, имеющих распределение Бернулли, со значениями $x_0 = 1$ и $x_1 = 0$. Причем $\mathsf{P}\{\xi_k = 1\} = p$, $\mathsf{P}\{\xi_k = 0\} = q$. Поэтому

$$W_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k, \quad \text{где } \mathsf{E}\xi_k = p,\, \mathsf{D}\xi_k = pq,\, k = 1,2,\ldots$$

Тогда по теореме 13.4, так как выполнено условие $m_\xi = p < \infty$, получаем $\xi/n \xrightarrow{\text{п.н.}} p$.

Замечание 13.3. Самому Якову Бернулли принадлежит доказательство более слабого утверждения, что $W_n \xrightarrow{\mathrm{P}} \mathsf{P}(A)$. Теорема Бернулли объясняет смысл свойства устойчивости частоты $W_n = \xi/n$, которое мы ранее принимали как экспериментальный факт. Таким образом, теорема Бернулли является «переходным мостиком» от теории вероятностей к ее приложениям.

Пример 13.1. Задана последовательность независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, причем ряд распределения случайной величины ξ_n представлен табл. 13.1. Применим ли к этой последовательности закон больших чисел?

ξ_n	$-\sqrt{n}$	0	$\sqrt{n}$
P	$1/(2n)$	$1-1/n$	$1/(2n)$

Таблица 13.1.

Решение. Проверим, выполняются ли условия теоремы Чебышева. Для этого найдем дисперсию случайной величины ξ_n. Очевидно, что математическое ожидание ξ_n равно 0, поэтому $\mathsf{D}\xi_n = \mathsf{E}\xi_n^2 = (\sqrt{n})^2 \cdot 1/n = 1$.

Таким образом, дисперсии случайных величин ξ_n, $n = 1,2,\ldots$, ограничены в совокупности константой $c = 1$. Следовательно, согласно теореме Чебышева, к последовательности $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ применим ЗБЧ.

13.2 Центральная предельная теорема

Теорема 13.6 (центральная предельная теорема). *Пусть* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *— последовательность независимых одинаково распределенных случайных величин,* $\mathsf{E}\xi_n = m$, $\mathsf{D}\xi_n = \sigma^2$, $S_n = \sum\limits_{i=1}^{n}\xi_i$. *Тогда*

$$\mathsf{P}\left\{\frac{S_n - nm}{\sqrt{n\sigma^2}} \leqslant x\right\} \underset{n\to\infty}{\longrightarrow} \Phi(x),$$

где $\Phi(x)$ *— функция распределения стандартной гауссовской величины.*

Ценность центральной предельной теоремы заключается в следующем.

Во-первых, центральная предельная теорема показывает, что сумма S_n независимых случайных величин с ростом числа слагаемых становится все больше похожа на нормальную случайную величину, в том смысле, что функция распределения $\mathsf{P}\left\{\frac{S_n - nm}{\sqrt{n\sigma^2}} \leqslant x\right\}$ нормированной суммы $\frac{S_n - nm}{\sqrt{n\sigma^2}}$ стремится с ростом n в каждой точке $x \in \mathbb{R}$ к функции распределения стандартной (т.е. с нулевым математическим ожиданием и единичной дисперсией) нормальной случайной величины. При этом слагаемые ξ_i не обязаны быть нормальными случайными величинами и, как правило, не являются таковыми. Тем самым, 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.

Отметим, что из теоремы 13.6 не вытекает сходимость последовательности случайных величин поточечно, почти наверное или по вероятности к нормальной случайной величине. Утверждается только о сходимости функций распределения членов последовательности нормированных случайных сумм к функции распределения нормальной случайной величины.

Во-вторых, центральная предельная теорема устанавливает скорость сходимости в законе больших чисел, которая пропорциональна $1/\sqrt{n}$. Из этого вытекает, например, что для того чтобы оценка $m \approx \frac{1}{n}\sum_{i=1}^{n}\xi_i$ была в k раз точнее, необходимо провести в k^2 раз больше наблюдений.

Пример 13.2. Для определения неслучайной величины m (например, скорости движения некоторого объекта) делают n измерений $\xi_1,\ldots,\xi_n$ этой величины, причем i-е измерение проводят с погрешностью ε_i, т.е. $\xi_i = m+\varepsilon_i$. Предположим, что погрешности измерений являются независимыми и одинаково распределенными случайными величинами с математическим ожиданием $\mathsf{E}\varepsilon_i = 0$ (отсутствуют систематические погрешности наблюдений) и дисперсией $\mathsf{D}\varepsilon_i = \sigma^2$.

Оценим вероятность того, что среднее арифметическое наблюдений $\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$ будет отличаться от измеряемой величины m по абсолютной величине не более чем на δ.

Из свойств математического ожидания и дисперсии вытекает, что случайная величина $\eta = \frac{(\overline{\xi}-m)\sqrt{n}}{\sigma}$ имеет нулевое математическое ожидание и единичную дисперсию, а согласно центральной предельной теореме η приближенно распределена по стандартному нормальному закону. Поэтому

$$\mathsf{P}\{|\overline{\xi}-m|<\delta\} \approx \Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)-\Phi\left(-\frac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)-1 = 2\Phi_0\left(\frac{\delta\sqrt{n}}{\sigma}\right).$$

Пример 13.3. Согласно закону больших чисел в форме Бернулли 13.5 частота $W_n = \xi/n = \widehat{p}$ появления события A в n экспериментах стремится к вероятности $p = \mathsf{P}(A)$ этого события. Следовательно, неизвестную вероятность "успеха"можно оценить с помощью частоты "успеха": $p \approx \widehat{p}$.

Оценим вероятность того, что $\widehat{p}$ будет отличаться от p по абсолютной величине не более чем на ε.

Обозначим через ξ_i появление события A в i-м эксперименте. Тогда $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$, $\mathsf{E}\xi_i = p$, $\mathsf{D}\xi_i = pq$. Поэтому из примера 13.2 вытекает, что

$$\mathsf{P}\{|\widehat{p}-p|<\varepsilon\} \approx 2\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right).$$

Следствием из центральной предельной теоремы является интегральная теорема Муавра — Лапласа.

Следствие 13.1 (интегральная теорема Муавра — Лапласа)**.** Обозначим $S_n = \xi$ 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.

$$\mathsf{P}\left\{\frac{S_n-np}{\sqrt{npq}} \leqslant x\right\} \xrightarrow[n\to\infty]{} \Phi(x).$$

Доказательство. Пусть ξ_i — число успехов в i-м испытании. Тогда $\mathsf{E}\xi_i = p$, $\mathsf{D}\xi_i = pq$. Представляя S_n в виде $S_n = \xi_1+\ldots+\xi_n$ и используя центральную предельную теорему, приходим к утверждению следствия.

Кроме того, вероятность $\mathsf{P}\{l \leqslant S_n \leqslant k\}$ может быть оценена согласно *интегральной теореме Муавра–Лапласа*:

$$\mathsf{P}\{l \leqslant S_n \leqslant k\} \approx \Phi_0\left(\frac{k-np}{\sqrt{npq}}\right)-\Phi_0\left(\frac{l-np}{\sqrt{npq}}\right),$$

где $\Phi_0(x)$ — функция Лапласа.

Поскольку $(S_n - np)/\sqrt{npq} \to U$ при $n \to \infty$, где $U \sim N(0{,}1)$, то Случайную величину $(S_n - np)/\sqrt{npq}$ при больших n можно приближенно считать стандартной нормальной величиной, а случайную величину S_n — нормальной с $\mathsf{E}S_n = np$ и $\mathsf{D}S_n = npq$.

Таким образом, мы получаем *локальную теорему Муавра–Лапласа*, в соответствии с которой

$$\mathsf{P}\{S_n = k\} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(k-np)^2}{2npq}\right\}.$$

Погрешность аппроксимации биномиальной случайной величиной S_n гауссовским распределением задается неравенством Берри—Эссена

$$\sup_x \left| \mathsf{P}\left(\frac{S_n - np}{\sqrt{npq}} \leqslant x\right) - \Phi(x)\right| \leqslant \frac{p^2+q^2}{\sqrt{npq}}.$$

Таким образом, в схеме Бернулли могут быть использованы две аппроксимации.

Пусть случайные величины $\xi \sim \mathbf{Bi}(n;p)$, тогда для вычисления соответствующих вероятностей можно пользоваться аппроксимациями, описанными в следующей таблице.

Приближение	$\boldsymbol{P_n(k)}$	$\boldsymbol{\sum_{i=l}^{k} P_n(i)}$
Пуассона	$\frac{(np)^k}{k!}e^{-np}$	$\sum_{i=l}^{k} \frac{(np)^i}{i!}e^{-np}$
Муавра–Лапласа	$\frac{\exp\left\{-\frac{(k-np)^2}{2npq}\right\}}{\sqrt{2\pi npq}}$	$\Phi_0\left(\frac{k-np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{l-np}{\sqrt{npq}}\right)$

Таблица 13.2

Погрешности этих приближений даны в табл. 13.3.

Приближение	**Погрешность**
Пуассона	np^2
Муавра–Лапласа	$\frac{p^2+q^2}{\sqrt{npq}}$

Таблица 13.3

13.3 Метод Монте–Карло (метод статистических испытаний)

Если функция $f(x)$ ограничена и интегрируема на отрезке $[0,1]$, то интеграл

$$I = \int_0^1 f(x)\,dx$$

можно рассматривать как математическое ожидание

$$I = \mathsf{E}f(\xi),$$

где ξ — случайная величина, распределенная равномерно на отрезке $[0,1]$. Следовательно, в предположении, что величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — независимы и одинаково распределены по закону $R(0,1)$, из УЗБЧ вытекает, что

$$\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) \xrightarrow{\text{п.н.}} I$$

Поэтому при достаточно большом n

$$I = \int_0^1 f(x)\,dx \approx \frac{1}{n}\sum_{k=1}^{n} f(\xi_k).$$

По ЦПТ

$$\mathsf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) - I\right| < \Delta\right) \approx 2\Phi_0\left(\Delta\frac{\sqrt{n}}{\sigma}\right),$$

где $\sigma = \sqrt{\mathsf{D}f(\xi)}$, $\xi \sim R(0,1)$.

Если известно, что $|f(x)| \leqslant C$ при $0 \leqslant x \leqslant 1$, то $\mathsf{D}f(\xi) \leqslant \mathsf{E}(f(\xi))^2 \leqslant C$. Тогда ЦПТ приводит к следующей приближенной, но вполне естественной оценке:

$$\mathsf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) - I\right| < \Delta\right) \geqslant 2\Phi_0\left(\Delta\frac{\sqrt{n}}{C}\right).$$

Эффективность описанного метода численного интегрирования возрастает при переходе от обыкновенных к кратным интегралам и становится достаточно высокой при рассмотрении интегралов большой размерности.

Отметим также, что численное интегрирование далеко не исчерпывает всех тех задач, которые решаются методом статистических испытаний.

На семинарском занятии предлагается решить следующие задачи.

Пример 13.4. Интеграл $I = \int\limits_0^1 x^2\,dx$ вычислен методом Монте-Карло на основании 10000 независимых испытаний. Найдите вероятность того, что абсолютная погрешность при вычислении интеграла не превысит 0.01.

Пример 13.5. Интеграл $I = \int\limits_0^1 x^2\,dx$ вычисляется методом Монте-Карло. Сколько испытаний надо провести, чтобы с вероятностью 0.99 гарантировать, что вычисленное значение отклонится от I по абсолютной величине не более, чем на 0.01.

Лекция 14

Математическая статистика

14.1 Основные выборочные характеристики

Основные понятия

Математическая статистика — наука о математических методах, позволяющих по статистическим данным, например по реализациям случайной величины (СВ), построить теоретико-вероятностную модель исследуемого явления. Задачи математической статистики являются, в некотором смысле, обратными к задачам теории вероятностей. Центральным понятием математической статистики является выборка.

Определение 14.1. ***Однородной выборкой (выборкой) объема*** n при $n \geqslant 1$ называется случайный вектор $Z_n = (X_1, \ldots, X_n)$, компоненты которого X_i, $i = \overline{1,n}$, называемые ***элементами выборки***, являются независимыми случайными величинами с одной и той же функцией распределения $F(x)$. Будем говорить, что выборка Z_n ***соответствует*** функции распределения $F(x)$.

Определение 14.2. ***Реализацией выборки*** называется неслучайный вектор $z_n = (x_1, \ldots, x_n)$, компонентами которого являются реализации соответствующих элементов выборки X_i, $i = \overline{1,n}$.

Из определений 14.1 и 14.2 вытекает, что реализацию выборки z_n можно также рассматривать как последовательность $x_1, \ldots, x_n$ из n реализаций одной и той же случайной величины X, полученных в серии из n независимых одинаковых опытов, проводимых в одинаковых условиях. Поэтому можно говорить, что выборка Z_n *порождена наблюдаемой* случайной величиной X, имеющей распределение $F_X(x) = F(x)$.

Определение 14.3. Если компоненты вектора Z_n независимы, но их распределения $F_1(x_1), \ldots, F_n(x_n)$ различны, то такую выборку называют ***неоднородной***.

Определение 14.4. Множество S всех реализаций выборки Z_n называется ***выборочным пространством***.

Выборочное пространство может быть всем n-мерным евклидовым пространством $\mathbb{R}^n$ или его частью, если случайная величина X непрерывна, а также может состоять из конечного или счетного числа точек из $\mathbb{R}^n$, если случайная величина X дискретна.

На практике при исследовании конкретного эксперимента распределения $F_1(x_1), \ldots, F_n(x_n)$ случайных величин $X_1, \ldots, X_n$ редко бывают известны полностью. Часто априори (до опыта) можно лишь утверждать, что распределение $F_{Z_n}(z_n) = F_1(x_1) \cdot \ldots \cdot F_n(x_n)$ случайного вектора Z_n принадлежит некоторому классу (семейству) $\mathscr{F}$.

Определение 14.5. Пара $(S, \mathscr{F})$ называется ***статистической моделью*** описания серии опытов, порождающих выборку Z_n.

Определение 14.6. Если распределения $F_{Z_n}(z_n, \theta)$ из класса $\mathscr{F}$ определены с точностью до некоторого векторного параметра $\theta \in \Theta \subset \mathbb{R}^s$, то такая статистическая модель называется ***параметрической*** и обозначается $(S_\theta, F_{Z_n}(z_n, \theta))$, $\theta \in \Theta \subset \mathbb{R}^s$.

В некоторых случаях выборочное пространство может не зависеть от неизвестного параметра θ распределения $F_{Z_n}(z_n, \theta)$.

В зависимости от вида статистической модели в математической статистике формулируются соответствующие задачи по обработке информации, содержащейся в выборке.

Определение 14.7. Случайная величина $Z = \varphi(Z_n)$, где $\varphi(z_n)$ — произвольная функция, определенная на выборочном пространстве S и не зависящая от распределения $F_{Z_n}(z_n,\theta)$, называется ***статистикой***.

Вариационный ряд

Определение 14.8. Упорядочим элементы реализации выборки $x_1,\ldots,x_n$ по возрастанию: $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$, где нижний индекс соответствует номеру элемента в упорядоченной последовательности. Обозначим через $X_{(k)}$ случайную величину, которая при каждой реализации z_n выборки Z_n принимает k-е (по нижнему номеру) значение $x_{(k)}$, $k = \overline{1,n}$. Упорядоченную последовательность случайных величин $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ называют ***вариационным рядом выборки***.

Определение 14.9. Элементы $X_{(k)}$ вариационного ряда называются ***порядковыми статистиками***, а крайние члены вариационного ряда $X_{(1)}, X_{(n)}$ — ***экстремальными порядковыми статистиками***.

Например, для $k = 1$ функция $\varphi(z_n)$ для статистики $X_{(1)} = \varphi(Z_n)$ определяется следующим образом:

$$\varphi(z_n) = \min\left\{x_k\colon\ k = \overline{1,n}\right\}.$$

Определение 14.10. Порядковая статистика $X_{([np]+1)}$ с номером $[np]+1$, где $[\cdot]$ — целая часть числа, называется *выборочной квантилью уровня* p.

Выборочная (эмпирическая) функция распределения

Пусть $X_1,\ldots,X_n$ независимые наблюдения случайной величины X с функцией распределения $F(x)$. Обозначим через $\nu_n(x)$ случайную величину, реализация которой для каждого $x \in \mathbb{R}$ и соответствующей реализации $z_n = (x_1,\ldots,x_n)$ выборки $Z_n = (X_1,\ldots,X_n)$ равна количеству наблюдений (т.е. количеству чисел $x_1,\ldots,x_n$), оказавшихся не больше x.

Отметим, что $\nu_n(x)$ при фиксированном x является биномиальной случайной величиной с вероятностью “успеха” $p = \mathsf{P}\{X \leqslant x\} = F(x)$.

Определение 14.11. Функцию

$$\hat{F}_n(x) = \frac{\nu_n(x)}{n}, \quad x \in \mathbb{R}, \tag{14.1}$$

будем называть ***выборочной (эмпирической) функцией распределения*** случайной величины X.

Для каждого фиксированного $x \in \mathbb{R}^1$ случайная величина $\widehat{F}_n(x)$ является статистикой, реализациями которой являются числа 0, $1/n$, $2/n,\ldots,n/n$, и при этом

$$\mathsf{P}\left\{\widehat{F}_n(x) = \frac{k}{n}\right\} = \mathsf{P}\{\nu_n(x) = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = \overline{0,n}, \quad p = \mathsf{P}\{X \leqslant x\} = F(x).$$

Любая реализация выборочной функции $\widehat{F}_n(x)$ является ступенчатой функцией, характерный вид которой показан на рис. 14.1. В точках $x_{(1)} < \ldots < x_{(n)}$, где, напомним, $x_{(k)}$ — реализация порядковой статистики $X_{(k)}$, реализация выборочной функции $\widehat{F}_n(x)$ имеет скачки величиной $1/n$ и является непрерывной справа.

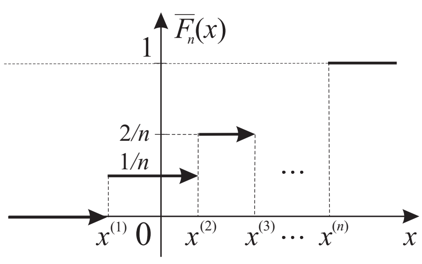

Рис 14.1.

Свойства $\widehat{F}_n(x)$

1) $\mathsf{E}\left[\widehat{F}_n(x)\right] = F(x)$, для любого $x \in \mathbb{R}^1$ и любого $n \geqslant 1$.

Действительно, при фиксированном x выборочная функция распределения $\widehat{F}_n(x)$ является частотой $\frac{\nu_n}{n}$, математическое ожидание которой равно $p = F(x)$.

2) $\sup\limits_{x\in\mathbb{R}^1} \left|\widehat{F}_n(x) - F(x)\right| \xrightarrow{\text{п.н.}} 0$ при $n \to \infty$. Доказательство данного свойства, называемого ***теоремой Гливенко–Кантелли***.

Эти свойства свидетельствуют о том, что при увеличении числа испытаний n происходит сближение выборочной функции распределения $\widehat{F}_n(x)$ с функцией распределения $F(x)$ случайной величины X. Свойство 2 обобщает усиленный вариант теоремы Бернулли.

Гистограмма

$[\alpha_1,\alpha_2)$	$\cdots$	$[\alpha_{l-1},\alpha_l]$
$\overline{p}_1$	$\cdots$	$\overline{p}_{l-1}$

Таблица 14.1.

Рассмотрим процедуру ***группировки*** выборки. Для этого действительную ось $\mathbb{R}^1 = (-\infty,\infty)$ разделим точками $\alpha_0,\ldots,\alpha_{l+1}$ на $l+1$ непересекающихся полуинтервалов (***разрядов***) $\Delta_k = [\alpha_k,\alpha_{k+1})$, $k = \overline{0,l}$, таким образом, что $-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_l < \alpha_{l+1} = +\infty$, $\alpha_1 \leqslant x_{(1)}$, $\alpha_l \geqslant x_{(n)}$. Обычно длина разрядов Δ_k, $k = \overline{1,l-1}$, выбирается одинаковой, т. е. равной $h_k = (\alpha_l - \alpha_1)/(l-1)$. Используя реализацию вариационного ряда $x_{(1)} < \ldots < x_{(n)}$, для каждого k-го разряда Δ_k, $k = \overline{1,l-1}$, вычислим частоту попадания элементов реализации выборки в этот разряд. Получаем $\overline{p}_k = n_k/n$, где n_k — число элементов реализации выборки z_n, попавших в k-й разряд. Если рассмотреть выборку Z_n и случайное число N_k элементов этой выборки, попавших в k-й разряд, то получим набор случайных величин $\widehat{p}_k = N_k/n$.

Определение 14.12. Последовательность пар $(\Delta_k,\widehat{p}_k)$, $k = \overline{1,l-1}$, называется ***статистическим рядом***, а его реализация $(\Delta_k,\overline{p}_k)$, $k = \overline{1,l-1}$, представляется в виде табл. 14.1.

Изобразим графически статистический ряд.

Определение 14.13. На оси Ox отложим разряды и на них, как на основании, построим прямоугольники с высотой, равной $\overline{p}_k/h_k$, $k = \overline{1,l-1}$. Тогда площадь каждого прямоугольника будет равна $\overline{p}_k$. Полученная фигура называется ***столбцовой диаграммой***, а кусочно-постоянная функция $\overline{f}_n(x)$, образованная верхними гранями полученных прямоугольников, — ***гистограммой*** (рис. 14.2).

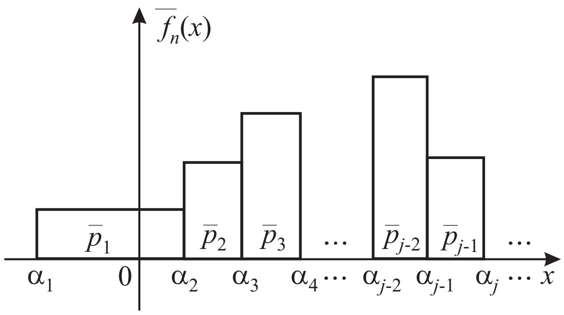

Рис 14.2.

При этом полагают $\overline{f}_n(x) = 0$ для всех $x < \alpha_1$ и $x \geqslant \alpha_l$, так как $n_0 = 0$ и $n_l = 0$.
Пусть плотность распределения $f(x)$ непрерывна и ограничена. При выборе числа интервалов группировки можно воспользоваться формулой Стерджесса $l = 1 + [3.32\lg n]$, где [] — целая часть числа. Тогда выборочная плотность распределения $\widehat{f}_n(x)$, реализациями которой служат гистограммы $\overline{f}_n(x)$, сходится по вероятности к плотности $f(x)$ наблюдаемой случайной величины, т. е. $\widehat{f}_n(x) \xrightarrow{\text{P}} f(x)$ при $n \to \infty$ для любого $x \in \mathbb{R}^1$. Таким образом, при достаточно «мелком» разбиении отрезка $[\alpha_1,\alpha_l]$ и при большом объеме выборки n высоты построенных прямоугольников можно принимать в качестве приближенных значений плотности $f(x)$ в средних точках соответствующих интервалов. Из этого следует, что гистограмму можно рассматривать как статистический аналог плотности распределения наблюдаемой случайной величины X. Используя гистограмму, неизвестную плотность можно аппроксимировать кусочно-постоянной функцией.

Определение 14.14. Сглаженную гистограмму в виде ломанной, у которой прямые линии последовательно соединяют середины верхних граней прямоугольников, образующих столбцовую диаграмму, называют *полигоном частот*.

Пример 14.1. В 1889–1890 гг. был измерен рост 1000 взрослых мужчин (рабочих московских фабрик). Результаты измерений представлены в табл. 14.2. По имеющимся наблюдениям требуется построить гисторамму.

Рост [см]	143–146	146–149	149–152	152–155	155–158
Число мужчин	1	2	8	26	65
Рост [см]	158–161	161–164	164–167	167–170	170–173
Число мужчин	120	181	201	170	120
Рост [см]	173–176	176–179	179–182	182–185	185–188
Число мужчин	64	28	10	3	1

Таблица 14.2

Решение. Пусть случайная величина X — рост взрослого мужчины.

В данной задаче группировка выборки уже проведена: действительная ось разделена на 17 полуинтервалов Δ_k, $k=\overline{0,16}$, где $\Delta_0 = (-\infty,143)$, $\Delta_{16} = [188,+\infty)$, а остальные 15 полуинтервалов имеют одинаковую длину $h=3$. Во второй строке таблицы приведены числа n_k, $k=\overline{1,15}$, равные количеству элементов выборки, попавших в k-й разряд. Вычислим частоту попадания в k-й полуинтервал, $k=\overline{1,15}$: $\overline{p}_k = n_k/1000$, и построим реализацию статистического ряда (см. табл. 14.3).

Δ_k	143–146	146–149	149–152	152–155	155–158
$\overline{p}_k$	0,001	0,002	0,008	0,026	0,065
Δ_k	158–161	161–164	164–167	167–170	170–173
$\overline{p}_k$	0,12	0,181	0,201	0,17	0,12
Δ_k	173–176	176–179	179–182	182–185	185–188
$\overline{p}_k$	0,064	0,028	0,01	0,003	0,001

Таблица 14.3

Теперь на оси OX отложим разряды Δ_k, $k=\overline{1,15}$, и на них, как на основании, построим прямоугольники высотой $\overline{p}_k/h$ (рис. 14.3).

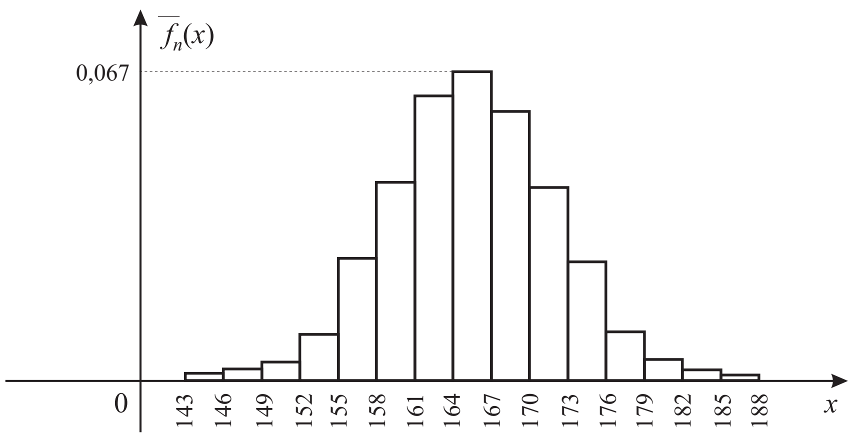

Рис 14.3.

Выборочные моменты

Пусть имеется выборка $Z_n = (X_1,\ldots,X_n)$, которая порождена случайной величиной X с функцией распределения $F_X(x)$.

Определение 14.15. Для выборки Z_n объема n ***выборочными начальными*** и ***центральными моментами порядка*** r случайной величины X называются следующие случайные величины:

$$\widehat{\mu}_r = \frac{1}{n}\sum_{k=1}^{n}(X_k)^r, \quad r=1,2,\ldots;$$

$$\widehat{\nu}_r = \frac{1}{n}\sum_{k=1}^{n}(X_k-\widehat{\mu}_1)^r, \quad r=2,3,\ldots$$

Определение 14.16. ***Выборочным средним*** и ***выборочной дисперсией*** случайной величины X называются соответственно

$$\overline{X} = \widehat{\mu}_1 = \frac{1}{n}\sum_{k=1}^{n} X_k,$$
$$s_X^2 = \widehat{\nu}_2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2.$$

В дальнейшем мы будем также использовать обозначения $\widehat{m}_X = \overline{X}$, $\widehat{d}_X = s_X^2$. Пусть имеется также выборка $V_n = (Y_1, \ldots, Y_n)$, порожденная случайной величиной Y с функцией распределения $F_Y(y)$.

Определение 14.17. ***Выборочной ковариацией*** случайных величин X и Y называют

$$\widehat{k}_{XY} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y}).$$

Определение 14.18. ***Выборочным коэффициентом корреляции*** случайные величины X и Y называют

$$\widehat{\rho}_{XY} = \frac{\widehat{k}_{XY}}{s_X s_Y}.$$

Пусть существуют исследуемые моменты μ_r, ν_r. Тогда справедливы следующие свойства.

Свойства выборочных моментов

1) $\mathsf{E}[\widehat{\mu}_r] = \mu_r$ для любого $n \geqslant 1$ и для всех $r = 1, 2, \ldots$ Действительно,

$$\mathsf{E}[\widehat{\mu}_r] = \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}[X_k^r] = \frac{1}{n}(n\mu_r) = \mu_r.$$

2) $\widehat{\mu}_r \xrightarrow{\text{п.н.}} \mu_r$ при $n \to \infty$ для всех $r = 1, 2, \ldots$ Это свойство вытекает из теоремы Колмогорова.

3) $\widehat{\nu}_r \xrightarrow{\text{п.н.}} \nu_r$ при $n \to \infty$ для всех $r = 2, 3, \ldots$ Используя разложение бинома Ньютона, получим

$$\widehat{\nu}_r = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^r = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=0}^{r} C_r^i X_k^i (-\overline{X})^{r-i} = = \sum_{i=0}^{r} C_r^i (-\overline{X})^{r-i}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^i\right) = \sum_{i=0}^{r} C_r^i (-\overline{X})^{r-i}\widehat{\mu}_i.$$

Используя свойство 2), устанавливаем, что $\widehat{\mu}_i \xrightarrow{\text{п.н.}} \mu_i$, $i = 1, \ldots, r$. Проводя обратное преобразование по биному Ньютона, получаем требуемое утверждение.

4) $\mathsf{D}\left[\overline{X}\right] = d_X/n$, где $d_X = \mathsf{D}[X]$. В самом деле, воспользовавшись независимостью случайных величин X_k, $k = 1, \ldots, n$, находим

$$\mathsf{D}\left[\overline{X}\right] = \mathsf{D}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] = \frac{1}{n^2}\sum_{k=1}^{n}\mathsf{D}[X_k] = \frac{1}{n^2}(nd_X) = \frac{d_X}{n}.$$

5) $\mathsf{E}\left[s^2\right] = \frac{n-1}{n} d_X$.

Докажем это свойство на семинаре.

6) $\left(\overline{X} - m_X\right) \Big/ \sqrt{d_X/n} \xrightarrow{d} U_1$ при $n \to \infty$, где случайная величина U_1 имеет распределение $\mathcal{N}(0;1)$. Поскольку последовательность X_i, $i = 1, 2, \ldots$, образована независимыми одинаково распределенными случайными величинами, то $\mathsf{E}\left[\overline{X}\right] = m_X$ по свойству 1), и $\mathsf{D}\left[\overline{X}\right] = d_X/n$ по свойству 4). Таким образом, к последовательности

$$S_n = \frac{1}{\sqrt{d_X n}}\sum_{k=1}^{n}(X_k - m_X) = \frac{1}{\sqrt{d_X/n}}\left(\overline{X} - m_X\right)$$

применима ЦПТ.

Второе и третье свойства указывают на то, что с увеличением объема выборки выборочные моменты будут сколь угодно близки к соответствующим теоретическим моментам.

Из первого свойства вытекает, что математические ожидания выборочных начальных моментов совпадают с соответствующими значениями начальных моментов случайной величины X, т. е. в этом смысле обладают свойством «несмещенности». А математическое ожидание выборочной дисперсии s^2 не совпадает с дисперсией d_X случайной величины X, т. е. в этом смысле случайная величина s^2 является «смещенной» выборочной характеристикой d_X. Поэтому часто вместо s^2 используют «***исправленную***» выборочную дисперсию $\widetilde{s}_X^2 = \frac{n}{n-1}s^2$, для которой $\mathsf{E}\left[\widetilde{s}_X^2\right] = d_X$.

Пример 14.2. В метеорологии принято характеризовать температуру месяца ее средним значением (среднее значение температуры месяца равно сумме температур всех дней данного месяца, деленной на число дней в этом месяце). В табл. 14.4 приведены значения средней температуры января в Саратове и Алатыре.

Год	1891	1892	1893	1894	1895	1896	1897
Саратов	−19,2	−14,8	−19,6	−11,1	−9,4	−16,9	−13,7
Алатырь	−21,8	−15,4	−20,8	−11,3	−11,6	−19,2	−13,0
Год	1899	1911	1912	1913	1914	1915	
Саратов	−4,9	−13,9	−9,4	−8,3	−7,9	−5,3	
Алатырь	−7,4	−15,1	−14,4	−11,1	−10,5	−7,2	

Таблица 14.4

Требуется по данным реализациям найти:

а) выборочное среднее и выборочную дисперсию средней температуры января в Саратове и Алатыре;

б) выборочный коэффициент корреляции средней температуры января в Саратове и средней температуры января в Алатыре.

Р е ш е н и е. Пусть случайная величина X — средняя температура января в Саратове, а случайная величина Y — средняя температура января в Алатыре. В табл. 14.4 приведена реализация $x_1,\ldots,x_{13}$ выборки $X_1,\ldots,X_{13}$, порожденной случайной величиной X, и реализация $y_1,\ldots,y_{13}$ выборки $Y_1,\ldots,Y_{13}$, порожденной случайнойвеличиной Y. Выборочное среднее $\overline{X}$ случайной величины X для данной реализации $x_1,\ldots,x_{13}$ равно

$$\overline{X} = \frac{1}{13}\sum_{i=1}^{13} x_i \approx -11{,}87,$$

а выборочное среднее $\overline{Y}$ случайной величины Y равно

$$\overline{Y} = \frac{1}{13}\sum_{i=1}^{13} y_i \approx -13{,}75.$$

Выборочная дисперсия s_X^2 случайной величины X для данной реализации $x_1,\ldots,x_{13}$ равна

$$s_X^2 = \frac{1}{13}\sum_{i=1}^{13}(x_i-(-11{,}87))^2 \approx 22{,}14,$$

а выборочная дисперсия s_Y^2 случайной величины Y равна

$$s_Y^2 = \frac{1}{13}\sum_{i=1}^{13}(y_i-(-13{,}75))^2 \approx 20{,}09.$$

Выборочный коэффициент корреляции случайной величины X и Y:

$$\widehat{\rho}_{XY} = \frac{\sum_{i=1}^{13}(x_i-(-11{,}87))\,(y_i-(-13{,}75))}{13\sqrt{22{,}14}\sqrt{20{,}09}} \approx 0{,}95.$$

О т в е т. $\overline{X} \approx -11{,}87$, $\overline{Y} \approx -13{,}75$, $s_X^2 \approx 22{,}14$, $s_Y^2 \approx 20{,}09$, $\widehat{\rho}_{XY} \approx 0{,}95$.

14.2 Основные распределения в статистике

Распределение хи-квадрат

Определение 14.19. Пусть U_k, $k = \overline{1,n}$, — набор из n независимых нормально распределенных случайных величин, $U_k \sim \mathscr{N}(0;1)$. Тогда случайная величина

$$X_n = \sum_{k=1}^{n} U_k^2$$

имеет ***распределение хи-квадрат*** (***χ^2-распределение***) с n ***степенями свободы***, что обозначается как $X_n \sim \chi^2(n)$.

Свойства распределения хи-квадрат $\chi^2(n)$

1) Случайная величина X_n имеет следующую плотность распределения:

$$f(x,n) = \begin{cases} \frac{1}{2^{(n/2)}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}, & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

где $\Gamma(m) = \int_0^{+\infty} y^{m-1} e^{-y}\, dy$ — гамма-функция. Графики функций $f(x,n)$ (рис. 14.4), называемые ***кривыми Пирсона***, асимметричны и начиная с $n > 2$ имеют один максимум в точке $x = n-2$.

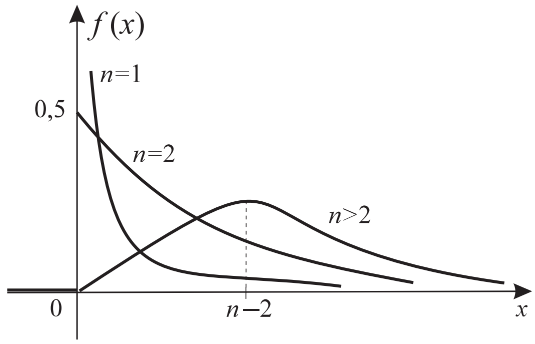

Рис 14.4

2) Случайная величина $X_n \sim \chi^2(n)$ имеет следующие моменты:

$$\mathsf{E}[X_n] = n, \qquad \mathsf{D}[X_n] = 2n.$$

3) Сумма любого числа m независимых случайных величин X_k, $k = \overline{1,m}$, имеющих распределение хи-квадрат с n_k степенями свободы, имеет распределение хи-квадрат с $n = \sum_{k=1}^{m} n_k$ степенями свободы.

4) Распределение хи-квадрат обладает свойством ***асимптотической нормальности***:

$$\frac{X_n - n}{\sqrt{2n}} \xrightarrow{d} U \qquad \text{при } n \to \infty,$$

где случайная величина U имеет распределение $\mathscr{N}(0;1)$. Это означает, что при достаточно большом объеме n выборки можно приближенно считать $X_n \sim \mathscr{N}(n;2n)$. Фактически эта аппроксимация имеет место уже при $n \geqslant 30$.

Пример 14.3. Приведем пример, в котором возникает распределение хи-квадрат. Пусть выборка Z_n соответствует нормальному распределению $\mathscr{N}(m;\sigma^2)$. Рассмотрим выборочную дисперсию

$$s_X^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})^2,$$

где $\overline{X}$ — выборочное среднее. Тогда случайная величина $Y_n = ns_X^2 / \sigma^2$ имеет распределение $\chi^2(n-1)$ и не зависит от $\overline{X}$.

Распределение Стьюдента

Определение 14.20. Пусть U и X_n — независимые случайные величины, $U \sim \mathscr{N}(0;1)$, $X_n \sim \chi^2(n)$. Тогда случайная величина $T_n = U \Big/ \sqrt{X_n/n}$ имеет ***распределение Стьюдента с*** n ***степенями свободы***, что обозначают как $T_n \sim \mathbf{t}(n)$.

Свойства распределения Стьюдента $\mathbf{t}(n)$

1) Случайная величина T_n имеет плотность распределения

$$f(t,n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

Графики плотностей $f(t,n)$ (рис. 14.5), называемые ***кривыми Стьюдента***, симметричны при всех $n = 1,2,\ldots$ относительно оси ординат.

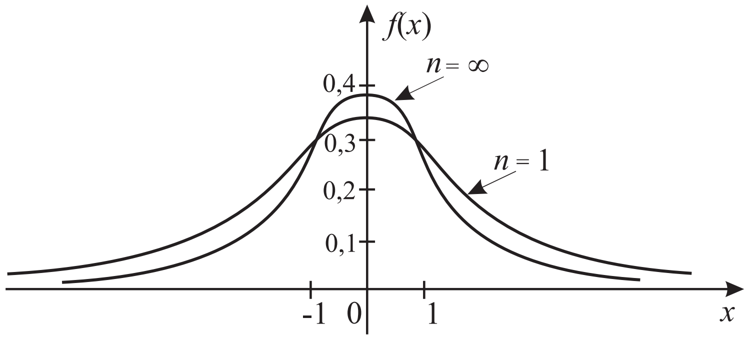

Рис 14.5

2) Случайная величина T_n имеет математическое ожидание, равное $\mathsf{E}[T_n] = 0$ для всех $n \geqslant 2$, и дисперсию $\mathsf{D}[T_n] = n/(n-2)$ при $n > 2$. При $n = 2$ дисперсия $\mathsf{D}[T_n] = +\infty$.

3) При $n = 1$ распределение Стьюдента называется ***распределением Коши***, плотность которого равна

$$f(t,1) = \frac{1}{\pi}\frac{1}{1+t^2}.$$

Математическое ожидание и дисперсия случайной величины T_1, имеющей распределение Коши, не существуют, так как бесконечен предел

$$\lim_{a\to\infty} I(a) = +\infty,$$

где $I(a) = \frac{1}{\pi}\int\limits_0^a \frac{t}{t^2+1}\,dt$.

4) Можно показать, что при $n \to \infty$ распределение $\mathbf{t}(n)$ асимптотически нормально, т. е. $T_n \xrightarrow{d} U$, где случайная величина U имеет распределение $\mathcal{N}(0;1)$. При $n \geqslant 30$ распределение Стьюдента $\mathbf{t}(n)$ практически не отличается от $\mathcal{N}(0;1)$.

Пример 14.4. Приведем пример, в котором встречается распределение Стьюдента. Пусть выборка Z_n соответствует нормальному распределению $\mathcal{N}(m;\sigma^2)$. Пусть $\overline{X}$ — выборочное среднее, а s_X^2 — выборочная дисперсия. Тогда случайная величина

$$T_n = \sqrt{n-1}\,\frac{\overline{X}-m}{\sqrt{s_X^2}}$$

имеет распределение Стьюдента $\mathbf{t}(n-1)$.

Распределение Фишера

Определение 14.21. Пусть независимые случайные величины X_n и X_m имеют распределения хи-квадрат соответственно с n и m степенями свободы. Тогда случайная величина $V_{n,m} = \frac{X_n/n}{X_m/m}$ имеет ***распределение Фишера с*** n ***и*** m ***степенями свободы***, что записывают как $V_{n,m} \sim \mathbf{F}(n;m)$.

Свойства распределения Фишера $\mathbf{F}(n;m)$

1) Случайная величина $V_{n,m}$ имеет плотность $f(v,n,m) = 0$ при $v \leqslant 0$ и

$$f(v,n,m) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\,n^{\frac{n}{2}}m^{\frac{m}{2}}\,\frac{v^{\frac{n}{2}-1}}{(m+nv)^{\frac{n+m}{2}}} \quad \text{при } v > 0.$$

Графики функции $f(v,n,m)$, называемые ***кривыми Фишера***, асимметричны и при $n > 2$ достигают максимальных значений в точках $v = \frac{(n-2)m}{(m+2)n}$, близких к единице при больших значениях m и n. Типовой вид кривой Фишера приведен на рис. 14.6.

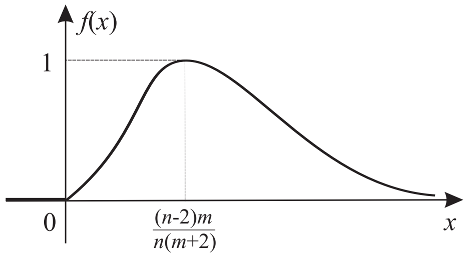

Рис 14.6

2) Случайная величина $V_{n,m}$ имеет следующие моменты:

$$bM[V_{n,m}] = \frac{m}{m-2} \text{ при } m > 2, \; \mathsf{D}[V_{n,m}] = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)} \text{ при } m > 4.$$

Пример 14.5. Пусть $Z_n = (X_1, \ldots, X_n)$ — выборка объема n, порожденная случайнойвеличиной X с нормальным распределением $\mathscr{N}(m_X; \sigma^2)$, а $W_n = (Y_1, \ldots, Y_m)$ — выборка объема m, порожденная случайнойвеличиной Y с нормальным распределением $\mathscr{N}(m_Y; \sigma^2)$, и случайные величины Z_n и W_n независимы. Тогда случайная величина $V_{n,m}$, образованная отношением «исправленных» выборочных дисперсий случайных величин X и Y, т. е.

$$V_{n,m} = \frac{\frac{1}{n-1} \sum\limits_{k=1}^{n} \left(X_k - \overline{X}\right)^2}{\frac{1}{m-1} \sum\limits_{j=1}^{m} \left(Y_j - \overline{Y}\right)^2},$$

имеет распределение $\mathbf{F}(n-1; m-1)$.

Лекция 15

Точечные оценки

15.1 Определения и свойства оценок

Определение 15.1. ***Параметром распределения*** $\theta \in \Theta \subset \mathbb{R}^1$ случайной величины X называется любая числовая характеристика этой случайной величины (математическое ожидание, дисперсия и т. п.) или любая константа, явно входящая в выражение для функции распределения.

В общем случае будем предполагать, что параметр распределения θ может быть векторным, т. е. $\theta \in \Theta \subset \mathbb{R}^s$.

В случае параметрической статистической модели $(S_\theta, F_{Z_n}(z_n, \theta))$ таким параметром распределения может служить неизвестный вектор $\theta \in \Theta \subset \mathbb{R}^s$, характеризующий распределение $F_{Z_n}(z_n, \theta)$.

Пусть имеется выборка $Z_n = (X_1, \ldots, X_n)$ с реализацией $z_n = (x_1, \ldots, x_n)$.

Определение 15.2. ***Точечной*** (***выборочной***) ***оценкой*** неизвестного параметра распределения $\theta \in \Theta \subset \mathbb{R}^s$ называется произвольная статистика $\widehat{\theta}(Z_n)$, построенная по выборке Z_n и принимающая значения в множестве Θ.

Ясно, что существует много разных способов построения точечной оценки, которые учитывают тип статистической модели. Для параметрической и непараметрической моделей эти способы могут быть различны. Рассмотрим некоторые свойства, которые характеризуют качество введенной оценки.

Определение 15.3. Оценка $\widehat{\theta}(Z_n)$ параметра θ называется ***несмещенной***, если ее математическое ожидание равно θ, т. е. $\mathsf{E}\left[\widehat{\theta}(Z_n)\right] = \theta$ для любого $\theta \in \Theta$.

Определение 15.4. Оценка $\widehat{\theta}(Z_n)$ параметра θ называется ***состоятельной***, если она сходится по вероятности к θ, т. е. $\widehat{\theta}(Z_n) \xrightarrow{\mathrm{P}} \theta$ при $n \to \infty$ для любого $\theta \in \Theta$.

Определение 15.5. Оценка $\widehat{\theta}(Z_n)$ параметра θ называется ***сильно состоятельной***, если она сходится почти наверное к θ, т. е. $\widehat{\theta}(Z_n) \xrightarrow{\text{п.н.}} \theta$ при $n \to \infty$ для любого $\theta \in \Theta$.

Очевидно, что если оценка сильно состоятельная, то она является также состоятельной.

Пример 15.1. Оценка $\widehat{\theta}_1(Z_n) = \overline{X}$ неизвестного математического ожидания $\theta_1 = m_X$ случайной величины X является несмещенной (по свойству 1 выборочных моментов).
А оценка $\widehat{\theta}_2(Z_n) = s_X^2$ неизвестной дисперсии $\theta_2 = d_X$ — смещенной, так как $\mathsf{E}\left[s_X^2\right] = \frac{n-1}{n} d_X$ (по свойству 5 выборочных моментов). Оценка $\widehat{\theta}_2(Z_n) = \widetilde{s}_X^2$ является несмещенной оценкой d_X по определению $\widetilde{s}_X^2$.
Если существуют моменты m_X, d_X, то сильная состоятельность оценок $\overline{X}$, $\widehat{d_X}$ гарантируется свойствами 2), 3) выборочных моментов.

Свойствами состоятельности и несмещенности могут обладать сразу несколько оценок неизвестного параметра θ.

Определение 15.6. Несмещенная оценка $\widehat{\theta}^*(Z_n)$ скалярного параметра θ называется ***эффективной***, если $\mathsf{D}\left[\widehat{\theta}^*(Z_n)\right] \leqslant \mathsf{D}\left[\widehat{\theta}(Z_n)\right]$ для всех несмещенных оценок $\widehat{\theta}(Z_n)$ параметра

θ, т. е. ее дисперсия минимальна по сравнению с дисперсиями других несмещенных оценок при одном и том же объеме n выборки Z_n.

Вообще говоря, дисперсии несмещенных оценок могут зависеть от параметра θ. В этом случае под эффективной оценкой понимается такая, для которой вышеприведенное неравенство является строгим хотя бы для одного значения параметра θ.

15.2 Методы оценивания параметров

На практике часто удается предсказать вид распределения наблюдаемой случайной величины с точностью до неизвестных параметров $\theta = (\theta_1, \ldots, \theta_s)$, т. е. для непрерывной случайной величины X оказывается известной плотность $f(x, \theta)$, а для дискретной случайной величины X — вероятности $p(x_i, \theta) = \mathsf{P}_\theta\{X = x_i\}$, $i = \overline{1,m}$, где x_i, $i = \overline{1,m}$, — возможные значения случайной величины X. Например, может быть $\theta_1 = m_X$, $\theta_2 = d_X$ при $s = 2$. Эти неизвестные параметры требуется оценить по имеющейся выборке Z_n. Расскажем о двух распространённых методах оценивания параметров - методе максимального правдоподобия (ММП) и методе моментов.

Метод максимального правдоподобия

Определение 15.7. ***Функцией правдоподобия*** для неизвестного параметра $\theta \in \Theta \subset \mathbb{R}^s$ называется: в случае непрерывной наблюдаемой случайной величины X — плотность распределения

$$L(z_n, \theta_1, \ldots, \theta_s) = f_{Z_n}(z_n, \theta_1, \ldots, \theta_s) = \prod_{k=1}^{n} f_X(x_k, \theta_1, \ldots, \theta_s),$$

где $f_X(x, \theta_1, \ldots, \theta_s)$ — плотность распределения случайной величины X, а в случае дискретной наблюдаемой случайной величины X — произведение вероятностей

$$L(z_n, \theta_1, \ldots, \theta_s) = \prod_{k=1}^{n} P_X(x_k, \theta_1, \ldots, \theta_s),$$

где $P_X(x_k, \theta_1, \ldots, \theta_s)$ — вероятность события $\{X = x_k\}$.

Рассмотрим параметрическую статистическую модель $(S_\theta, F_{Z_n}(z_n, \theta))$, $\theta \in \Theta \subset \mathbb{R}^s$, для которой известна функция правдоподобия $L(z_n, \theta_1, \ldots, \theta_s)$.

Определение 15.8. ***Оценкой максимального правдоподобия*** (МП-***оценкой***) параметра $\theta \in \Theta$ называется статистика $\widehat{\theta}(Z_n)$, максимизирующая для каждой реализации z_n функцию правдоподобия, т. е.

$$\widehat{\theta}(z_n) = \arg\max_{\theta \in \Theta} L(z_n, \theta).$$

Способ построения МП-оценки называется ***методом максимального правдоподобия***.

Поскольку функция правдоподобия $L(z_n, \theta)$ и её логарифм $\ln L(z_n, \theta)$ достигают максимума при одних и тех же значениях θ, то часто вместо $L(z_n, \theta)$ рассматривают логарифмическую функцию правдоподобия $\ln L(z_n, \theta)$.

Определение 15.9. ***Логарифмической функцией правдоподобия*** для неизвестного параметра $\theta \in \Theta \subset \mathbb{R}^s$ называется функция $\ln L(z_n, \theta_1, \ldots, \theta_s)$.

В случае дифференцируемости функции $\ln L(z_n, \theta)$ по θ МП-оценку можно найти, решая относительно $\theta_1, \ldots, \theta_s$ систему ***уравнений правдоподобия***

$$\begin{cases} \frac{\partial \ln L(z_n, \theta_1, \ldots, \theta_s)}{\partial \theta_1} = 0, \\ \ldots \\ \frac{\partial \ln L(z_n, \theta_1, \ldots, \theta_s)}{\partial \theta_s} = 0. \end{cases}$$

Пример 15.2. Если, например, случайная величина X имеет нормальное распределение $\mathcal{N}(m_X; \sigma_X^2)$ с неизвестным математическим ожиданием $\theta = m_X$, то легко установить, что оценкой максимального правдоподобия параметра $\theta = m_X$ при любых σ_X является выборочное среднее $\overline{X}$. Действительно, в этом случае

$$L(z_n,m_X) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x_k - m_X)^2}{2\sigma_X^2}\right\} = \frac{1}{(2\pi)^{n/2}\sigma_X^n} \exp\left\{-\sum_{k=1}^{n} \frac{(x_k - m_X)^2}{2\sigma_X^2}\right\}.$$

Дифференцируя функцию $\ln L(z_n,m_X)$ по m_X и приравнивая нулю получаемое выражение, находим уравнение, решение которого

$$\widehat{\theta}(Z_n) = \frac{1}{n}\sum_{k=1}^{n} X_k = \overline{X}.$$

Пример 15.3. Пусть случайная величина X имеет равномерное распределение $\mathscr{R}(a;b)$ с неизвестными параметрами $\theta_1 = a$, $\theta_2 = b$. В данном случае плотность распределения $f(x,a,b) = 1/(b-a)$, если $x \in [a,b]$, и $f(x,a,b) = 0$, если $x \notin [a,b]$. Оценим параметры a и b методом максимального правдоподобия. Функция правдоподобия в данном случае имеет вид

$$L(x_1,\ldots,x_n,a,b) = \prod_{k=1}^{n} f(x_k,a,b) = \left(\frac{1}{b-a}\right)^n,$$

если $a \leqslant x_k \leqslant b$ для всех $k = \overline{1,n}$, и $L(x_1,\ldots,x_n,a,b) = 0$ в остальных случаях. Таким образом, функция правдоподобия отлична от нуля, если неизвестные параметры a и b удовлетворяют неравенствам

$$b \geqslant x_{(n)} = \max_{k=\overline{1,n}} x_k, \qquad a \leqslant x_{(1)} = \min_{k=\overline{1,n}} x_k.$$

При этом функция $L(x_1,\ldots,x_n,a,b)$ достигает максимума по a и b, когда разность $b-a$ оказывается минимально возможной, не нарушающей полученные неравенства, т. е. в случае достижения в них равенств. Таким образом получаем МП-оценки неизвестных параметров a и b:

$$\widehat{a}(Z_n) = X_{(1)}, \qquad \widehat{b}(Z_n) = X_{(n)},$$

где $X_{(n)}$ и $X_{(1)}$ — крайние члены вариационного ряда.

Пример 15.4. Пусть случайная величина X имеет распределение Пуассона $\Pi(a)$ с неизвестным параметром $\theta = a$. Построим МП-оценку параметра a. Функция правдоподобия в этом случае равна

$$L(z_n,a) = \prod_{k=1}^{n} \frac{a^{x_k} e^{-a}}{x_k!},$$

поэтому логарифмическая функция правдоподобия равна

$$\ln L(z_n,a) = \sum_{k=1}^{n} x_k \ln a - na - \ln(x_1! \cdot \ldots \cdot x_n!).$$

Решая соответствующее уравнение правдоподобия

$$\frac{\partial}{\partial a} \ln L(z_n,a) = 0,$$

находим МП-оценку неизвестного параметра a:

$$\widehat{a}(Z_n) = \frac{1}{n}\sum_{k=1}^{n} X_k = \overline{X}.$$

Пример 15.5. Найдем МП-оценку для вероятности p «успеха» в схеме испытаний Бернулли. В этом случае имеем распределение Бернулли $\mathbf{Bi}(1;p)$ с неизвестным параметром $\theta = p$. Поэтому

$$L(z_n,a) = \prod_{k=1}^{n} p^{x_k}(1-p)^{1-x_k},$$

где $x_k = 1$, если в k-м испытании был «успех», и $x_k = 0$ — в противном случае. Решая уравнение правдоподобия

$$\frac{\partial}{\partial p}\ln L(z_n,p) = \sum_{k=1}^{n}\left(\frac{x_k}{p} - \frac{1-x_k}{1-p}\right) = 0$$

относительно параметра p, находим МП-оценку:

$$\widehat{p}(Z_n) = \frac{1}{n}\sum_{k=1}^{n} X_k = \overline{X}.$$

Метод моментов

Исторически первым для оценивания неизвестных параметров был предложен следующий метод. Пусть имеется параметрическая статистическая модель $(S_\theta, F_{Z_n}(z_n,\theta))$, $\theta \in \Theta \subset \mathbb{R}^s$. Предположим, что у наблюдаемой случайной величины X, порождающей выборку Z_n, существуют начальные моменты $\mu_i = \mathsf{E}[X^i]$, $i = \overline{1,s}$. Тогда в общем случае от неизвестных параметров будут зависеть и начальные моменты, т. е. $\mu_i = \mu_i(\theta)$.

Пусть $\widehat{\mu}_i$, $i = \overline{1,s}$, — выборочные начальные моменты. Рассмотрим систему уравнений

$$\mu_i(\theta) = \widehat{\mu}_i, \quad i = \overline{1,s},$$

и предположим, что её можно разрешить относительно параметров $\theta_1,\ldots,\theta_s$, т. е. найти функции $\widehat{\theta}_i = \varphi_i(\widehat{\mu}_1,\ldots,\widehat{\mu}_s)$, $i = \overline{1,s}$.

Определение 15.10. Решение полученной системы уравнений $\widehat{\theta}_i = \varphi_i(\widehat{\mu}_1,\ldots,\widehat{\mu}_s)$, $i = \overline{1,s}$, называется ***оценкой*** параметра θ, найденной по ***методу моментов***, или ММ-***оценкой***.

Если функции $\varphi_1(\cdot),\ldots,\varphi_s(\cdot)$ непрерывны, то ММ-оценки являются состоятельными.

Замечание 15.1. Уравнения метода моментов часто оказываются более простыми по сравнению с уравнениями правдоподобия, и их решение не связано с большими вычислительными трудностями.

Пример 15.6. Пусть $Z_n = (X_1,\ldots,X_n)$ — выборка, соответствующая нормальному распределению $\mathcal{N}(m;\sigma^2)$ с неизвестными параметрами $\theta_1 = m$ и $\theta_2 = \sigma^2$. Оценим параметры m и σ^2 с помощью метода моментов. В данном случае $\mu_1 = m$, $\mu_2 = m^2 + \sigma^2$ и система уравнений для метода моментов принимает вид

$$\begin{cases} m = \widehat{\mu}_1 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k, \\ m^2 + \sigma^2 = \widehat{\mu}_2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2. \end{cases}$$

Решая эту систему, находим ММ-оценки:

$$\widehat{\theta}_1(Z_n) = \widehat{\mu}_1, \qquad \widehat{\theta}_2(Z_n) = \widehat{\mu}_2 - \widehat{\mu}_1^2.$$

Пример 15.7. Пусть $Z_n = (X_1,\ldots,X_n)$ — выборка, соответствующая равномерному распределению $\mathcal{R}(a;b)$ с неизвестными параметрами $\theta_1 = a$ и $\theta_2 = b$. Оценим параметры a и b с помощью метода моментов. Поскольку для данного распределения

$$\mu_1 = \frac{a+b}{2}, \qquad \mu_2 = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2,$$

то система уравнений метода моментов принимает вид

$$\begin{cases} \dfrac{a+b}{2} = \widehat{\mu}_1 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k, \\ \dfrac{(b-a)^2}{12} + \left(\dfrac{a+b}{2}\right)^2 = \widehat{\mu}_2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k^2. \end{cases}$$

Решая эту систему, получаем ММ-оценки неизвестных параметров:

$$\widehat{a}(Z_n) = \widehat{\mu}_1 - \sqrt{3s_X^2}, \qquad \widehat{b}(Z_n) = \widehat{\mu}_1 + \sqrt{3s_X^2},$$

где $s_X^2 = \frac{1}{n}\sum\limits_{k=1}^{n}(X_k - \overline{X})^2$.

Оценки, полученные в примере 15.6 с помощью метода моментов, совпадают с МП-оценками, найденными в примере 15.2 из п. 15.2, а ММ-оценки в примере 15.7 не совпадают с МП-оценками, построенными в примере 15.3 из того же пункта.

Лекция 16

Интервальные оценки

Основные понятия

Пусть имеется параметрическая статистическая модель $(S_\theta, F_{Z_n}(z_n,\theta))$, $\theta \in \Theta \subset \mathbb{R}^1$, и по выборке $Z_n = (X_1,\ldots,X_n)$, соответствующей распределению $F(x,\theta)$ наблюдаемой случайной величины X, требуется оценить неизвестный параметр θ. Вместо точечных оценок, рассмотренных ранее, рассмотрим другой тип оценок неизвестного параметра $\theta \in \Theta \subset \mathbb{R}^1$.

Определение 16.1. Интервал $[\theta_1(Z_n),\theta_2(Z_n)]$ со случайными концами, «накрывающий» с вероятностью $1-\alpha$, $0<\alpha<1$, неизвестный параметр θ, т. е.

$$\mathsf{P}\{\theta_1(Z_n) \leqslant \theta \leqslant \theta_2(Z_n)\} = 1-\alpha,$$

называется ***доверительным интервалом*** (или ***интервальной оценкой***) уровня надежности $1-\alpha$ параметра θ.

Аналогично определяется доверительный интервал для произвольной функции от параметра θ.

Определение 16.2. Число $\delta = 1-\alpha$ называется ***доверительной вероятностью*** или ***уровнем доверия (надежности)***.

Определение 16.3. Доверительный интервал $[\theta_1(Z_n),\theta_2(Z_n)]$ называется ***центральным***, если выполняются следующие условия:

$$\mathsf{P}\{\theta \geqslant \theta_2(Z_n)\} = \frac{\alpha}{2}, \qquad \mathsf{P}\{\theta_1(Z_n) \geqslant \theta\} = \frac{\alpha}{2}.$$

Часто вместо двусторонних доверительных интервалов рассматривают односторонние доверительные интервалы, полагая $\theta_1(Z_n) = -\infty$ или $\theta_2(Z_n) = +\infty$.

Определение 16.4. Интервал, границы которого удовлетворяют условию

$$\mathsf{P}\{\theta \geqslant \theta_2(Z_n)\} = \alpha \qquad (\text{или} \quad \mathsf{P}\{\theta_1(Z_n) \geqslant \theta\} = \alpha),$$

называется соответственно ***правосторонним*** (или ***левосторонним***) ***доверительным интервалом***.

Рассмотрим метод построения доверительных интервалов, основанный на центральной (опорной) статистике.

Использование центральной статистики

Определение 16.5. Функция $G(Z_n,\theta)$ случайной выборки Z_n, такая что её распределение $F_G(y)$ не зависит от параметра θ и при любом значении z_n функция $G(z_n,\theta)$ является непрерывной и строго монотонной по θ, называется ***центральной (опорной) статистикой*** для параметра θ.

Зная распределение $F_G(y)$ центральной статистики $G(Z_n,\theta)$, можно найти числа g_1 и g_2, удовлетворяющие условию

$$\mathsf{P}\{g_1 \leqslant G(Z_n,\theta) \leqslant g_2\} = 1-\alpha.$$

Тогда границы доверительного интервала $[\theta_1(Z_n),\theta_2(Z_n)]$ для параметра θ могут быть найдены, если разрешить, учитывая свойства функции $G(z_n,\theta)$, следующие неравенства:

$$g_1 \leqslant G(Z_n,\theta) \leqslant g_2.$$

В частности, если $G(z_n,\theta)$ — монотонно возрастающая по θ функция, то

$$\theta_1(Z_n) = G^{-1}(Z_n,g_1), \qquad \theta_2(Z_n) = G^{-1}(Z_n,g_2),$$

где $G^{-1}(z_n,g_1)$ — функция, обратная по отношению к $G(z_n,\theta)$. Если $G(z_n,\theta)$ — монотонно убывающая по θ функция, то

$$\theta_1(Z_n) = G^{-1}(Z_n,g_2), \qquad \theta_2(Z_n) = G^{-1}(Z_n,g_1).$$

Применим данный метод для построения доверительных интервалов неизвестных параметров нормального распределения $\mathcal{N}(m_X;\sigma_X^2)$. С этой целью сформулируем утверждение, с помощью которого можно определить центральные статистики для неизвестых параметров m_X, σ_X^2.

Теорема 16.1 (теорема Фишера)**.** *Пусть* $Z_n = (X_1,\ldots,X_n)$ *— выборка, порожденная случайнойвеличиной* $X \sim \mathcal{N}(m_X;\sigma_X^2)$, *а* $\overline{X}$ *и* s_X^2 *— выборочные среднее и дисперсия.*

Тогда

1) случайная величина $\frac{(\overline{X}-m_X)\sqrt{n}}{\sigma_X}$ *имеет распределение* $\mathcal{N}(0;1)$*;*

2) случайная величина ns_X^2/σ_X^2 *имеет распределение* $\chi^2(n-1)$*;*

3) случайные величины $\overline{X}$ *и* s_X^2 *независимы;*

4) $\frac{(\overline{X}-m_X)\sqrt{n-1}}{s_X}$ *имеет распределение Стьюдента* $\mathbf{t}(n-1)$.

Пример 16.1. По выборке Z_n из нормального распределения требуется построить доверительный интервал для неизвестного математического ожидания m_X при известной дисперсии σ_X^2. Из приведенного выше утверждения следует, что случайная величина $\frac{(\overline{X}-m_X)\sqrt{n}}{\sigma_X}$ имеет нормальное распределение $\mathcal{N}(0;1)$, которое не зависит от m_X, и, кроме того, функция $G(Z_n,m_X) = (\overline{X}-m_X)\sqrt{n}/\sigma_X$ является непрерывной и убывающей по m_X. Это значит, что указанная случайная величина является центральной статистикой. Поэтому доверительный интервал для неизвестного m_X можно построить, если найти такие величины g_1 и g_2, что

$$\mathsf{P}\left(g_1 \leqslant \frac{(\overline{X}-m_X)\sqrt{n}}{\sigma_X} \leqslant g_2\right) = 1-\alpha.$$

Заметим, что данное условие неоднозначно определяет g_1, g_2. Выберем доверительный интервал минимальной длины. Учитывая симметрию относительно оси OY плотности стандартного нормального распределения, можно показать, что такой интервал будет иметь минимальную длину, если положить $g_1 = -g_2$, и при этом он оказывается центральным. Таким образом, получаем следующий доверительный интервал:

$$\left(\overline{X} - \frac{\sigma_X}{\sqrt{n}}u_\gamma; \overline{X} + \frac{\sigma_X}{\sqrt{n}}u_\gamma\right),$$

где u_γ — квантиль уровня $\gamma = 1-\alpha/2$ стандартного нормального распределения $\mathcal{N}(0;1)$. В данном случае длина доверительного интервала равна $\Delta = 2u_\gamma\sigma_X/\sqrt{n}$ и не случайна. Поэтому, задавшись значениями любых двух из трех величин Δ, α, n, можно определить значение третьей величины.

Пример 16.2. Теперь по выборке Z_n из нормального распределения построим доверительный интервал для неизвестного математического ожидания m_X при неизвестной дисперсии σ_X^2.Используя утверждение 4) теоремы 16.1, получаем, что распределение случайной величины $\frac{(\overline{X}-m_X)\sqrt{n-1}}{s_X}$ не зависит от параметра m_X, а функция $G(Z_n,m_X) = (\overline{X}-m_X)\sqrt{n-1}\big/ s_X$ является непрерывной и убывающей по m_X. Следовательно, функция $G(Z_n,m_X) = (\overline{X}-m_X)\sqrt{n-1}\big/ s_X$ является центральной статистикой. По аналогии с предыдущим примером получим центральный доверительный интервал для неизвестного m_X и в случае, когда величина σ_X неизвестна:

$$\left(\overline{X} - \frac{s_X}{\sqrt{n-1}}t_{\gamma,n-1};\overline{X} + \frac{s_X}{\sqrt{n-1}}t_{\gamma,n-1}\right),$$

где $t_{\gamma,n-1}$ — квантиль уровня $\gamma = 1-\alpha/2$ распределения Стьюдента $\mathbf{t}(n-1)$.

В отличие от предыдущего примера, длина доверительного интервала случайна и зависит от случайной величины s_X.

Заметим также, что при $n \geqslant 30$ распределение Стьюдента близко к стандартному нормальному распределению, и поэтому при больших объёмах выборки ($n \geqslant 30$) в последней формуле можно использовать квантили стандартного нормального распределения.

Пример 16.3. Центральный доверительный интервал для неизвестного параметра σ_X^2 случайной величины $X \sim \mathcal{N}(m_X;\sigma_X^2)$ при известном m_X имеет следующий вид

$$\left(\frac{\sum\limits_{i=1}^{n}(x_i-m_X)^2}{\chi^2_{1-\alpha/2;n}};\frac{\sum\limits_{i=1}^{n}(x_i-m_X)^2}{\chi^2_{\alpha/2;n}}\right),$$

где $\chi^2_{1-\alpha/2;,n}$ и $\chi^2_{\alpha/2;,n}$ — квантили уровней $1-\alpha/2$ и $\alpha/2$ распределения $\chi^2(n)$. Покажите это самостоятельно.

Пример 16.4. Центральный доверительный интервал для неизвестного параметра σ_X^2 случайной величины $X \sim \mathcal{N}(m_X;\sigma_X^2)$ при неизвестном m_X можно получить, используя утверждение 2) теоремы 16.1,

$$\left(\frac{n}{\chi^2_{1-\alpha/2;n-1}}s_X^2;\frac{n}{\chi^2_{\alpha/2;,n-1}}s_X^2\right),$$

где $\chi^2_{1-\alpha/2;,n-1}$ и $\chi^2_{\alpha/2;,n-1}$ — квантили уровней $1-\alpha/2$ и $\alpha/2$ распределения $\chi^2(n-1)$.

Пример 16.5. Пусть $Z_m = (X_1,\ldots,X_m)$ и $V_n = (Y_1,\ldots,Y_n)$ — две независимые выборки, порождённые случайными величинами с распределениями $\mathcal{N}(\mu_1,\sigma_1^2)$ и $\mathcal{N}(\mu_2,\sigma_2^2)$ соответственно. Дисперсии σ_1^2 и σ_1^2 известны, а математические ожидания μ_1 и μ_2 — неизвестны.Требуется построить доверительный интервал для параметра $\theta = \mu_1-\mu_2$. Покажите, что случайная функция

$$G(Z_m,V_n,\theta) = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}}$$

является центральной статистикой для параметра θ. Найдите распределение этой статистики и постройте центральный доверительный интервал уровня надёжности $1-\alpha$ для параметра θ.

Теорема 16.2. *Пусть $Z_m = (X_1,\ldots,X_m)$ и $V_n = (Y_1,\ldots,Y_n)$ — две независимые выборки, порождённые случайными величинами с распределениями $\mathcal{N}(\mu_1,\sigma^2)$ и $\mathcal{N}(\mu_2,\sigma^2)$ соответственно. Параметры μ_1, μ_2 и σ^2 неизвестны.Обозначим*

$$\overline{X} = \frac{1}{m}\sum_{i=1}^{m}X_i, \qquad S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i-\overline{X})^2,$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}Y_i, \qquad S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2,$$

$$S_{X,Y} = \sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}.$$

Тогда:

1) случайная величина $\frac{S_X^2}{S_Y^2}$ *имеет распределение Фишера с числом степеней свободы* $m-1$ *и* $n-1$*;*

2) случайная величина

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_{X,Y}\sqrt{\frac{1}{m}+\frac{1}{n}}}$$

имеет распределение Стьюдента с $m+n-2$ *степенями свободы.*

Пример 16.6. Пользуясь утверждением 2 теоремы16.2, постройте центральный доверительный интервал для разности средних значений $\theta = \mu_1 - \mu_2$ двух независимых гауссовских величин с одинаковыми, но неизвестными дисперсиями.

Пример 16.7. Построим доверительный интервал для неизвестного параметра b равномерного распределения $\mathscr{R}(0;b)$. Можно показать, что случайная величина $G(Z_n,b) = \left(\frac{X^{(n)}}{b}\right)^n$ имеет распределение $\mathscr{R}(0;1)$ для любого $b > 0$. Кроме того, функция $G(z_n,b)$ — убывающая по b. Следовательно, $G(Z_n,b)$ является центральной статистикой. Тогда получаем условие

$$\mathsf{P}\left\{g_1 \leqslant \left(\frac{X^{(n)}}{b}\right)^n \leqslant g_2\right\} = 1-\alpha$$

для некоторых чисел g_1, g_2. Разрешая двойное неравенство в этом вероятностном условии относительно b, получаем следующий доверительный интервал:

$$\left(\frac{X^{(n)}}{g_2^{1/n}};\frac{X^{(n)}}{g_1^{1/n}}\right).$$

Отметим, что этот интервал будет иметь наименьшую длину, если $g_2 = 1$, а g_1 является квантилью уровня α распределения $\mathscr{R}(0;1)$, т. е. $g_1 = \alpha$.

Типовые задачи

Пример 16.8. Используя данные примера 14.2, построить центральный доверительный интервал уровня доверия 0,95 для неизвестного математического ожидания средней температуры января в Саратове (случайная величина X). Будем предполагать, что случайная величина X имеет гауссовское распределение.

Решение. Согласно примеру 16.2 центральный доверительный интервал для неизвестного математического ожидания m_X уровня доверия 0,95 определяется из следующего соотношения

$$\mathsf{P}(\overline{X} - \sqrt{\frac{s_X^2}{n-1}}t_{0{,}975,n-1} \leqslant m_X \leqslant \overline{X} + \sqrt{\frac{s_X^2}{n-1}}t_{0{,}975,n-1}) = 0{,}95,$$

где объем выборки $n = 13$, $\overline{X} = -11.87$, $s_X^2 = 22.14$ (результат решения примера 14.2), а квантиль распределения Стьюдента $t_{0.975,12} = 2.18$ находится по таблице.

Сделав необходимые вычисления, получаем доверительный интервал для m_X : $[-14.43,-8.9]$.

Ответ. $[-14.43,-8.9]$.

Пример 16.9. Используя данные типовой примера 14.2, построить центральный доверительный интервал уровня доверия 0.95 для неизвестной дисперсии средней температуры января в Саратове (случайная величина X). Будем предполагать, что случайная величина X имеет гауссовское распределение.

Р е ш е н и е. Согласно примеру 16.4 центральный доверительный интервал для неизвестной дисперсии σ_X^2 уровня доверия 0.95 найдём, используя следующее соотношение

$$\mathsf{P}(\frac{n}{\chi^2_{0.975,n-1}} s_X^2 \leqslant \sigma_X^2 \leqslant \frac{n}{\chi^2_{0.025,n-1}} s_X^2) = 0.95,$$

где объем выборки $n = 13$, $s_X^2 = 22.14$ (результат решения примера 14.2), а $\chi^2_{0.975,12} = 23.3$, $\chi^2_{0.025,12} = 4.4$ — квантили распределения $\chi^2(12)$, которые находятся по таблице.

Таким образом, получаем доверительный интервал для σ_X^2 : $[11.4, 60.38]$.

О т в е т. $[11.4, 60.38]$.

Лекция 17

Проверка статистических гипотез

Основные понятия

Определение 17.1. ***Статистической гипотезой*** H (или просто ***гипотезой***) называется любое предположение относительно вида распределения,параметров распределения или свойств закона распределения наблюдаемой в эксперименте случайной величины X.

Определение 17.2. Проверяемая гипотеза называется ***основной*** (или ***нулевой***) и обозначается H_0. Гипотеза, конкурирующая с H_0, называется ***альтернативной*** и обозначается H_1.

Определение 17.3. Статистическая гипотеза H_0 называется ***простой***, если она однозначно определяет параметр или распределение случайной величины X. В противном случае гипотеза H_0 называется ***сложной***.

Пример 17.1. По выборке Z_n требуется проверить гипотезу H_0 о том, что $m_X = m_0$, где m_0 — некоторое фиксированное число, против гипотезы H_1 о том, что $m_X \neq m_0$. Или проверить гипотезу H_0 против гипотезы H_2 о том, что $m_X > m_0$.

Одна из основных задач математической статистики состоит в проверке соответствия результатов эксперимента предполагаемой гипотезе H_0. С этой целью выбирается некоторая статистика $Z = \varphi(Z_n)$, для которой предполагается известным распределение $F_Z(z)$ в том случае, когда проверяемая гипотеза H_0 верна. Обозначим это распределение $F_Z(z|H_0)$. С помощью этой статистики строится процедура (правило) проверки гипотезы.

Определение 17.4. ***Статистическим критерием*** (***критерием согласия, критерием значимости*** или ***решающим правилом***) проверки гипотезы H_0 называется правило, в соответствии с которым по реализации $z = \varphi(z_n)$ статистики Z гипотеза H_0 принимается или отвергается.

Определение 17.5. ***Критической областью*** $\overline{G}$ статистического критерия называется область реализаций z статистики Z, при которых гипотеза H_0 отвергается.

Определение 17.6. ***Доверительной областью*** G статистического критерия называется область значений z статистики Z, при которых гипотеза H_0 принимается.

Например, в качестве статистического критерия можно использовать правило:

1) если значение $z = \varphi(z_n)$ статистики $Z = \varphi(Z_n)$ лежит в критической области $\overline{G}$, то гипотеза H_0 отвергается и принимается альтернативная гипотеза H_1;

2) если реализация $z = \varphi(z_n)$ статистики $Z = \varphi(Z_n)$ лежит в доверительной области G, то гипотеза H_0 принимается.

При реализации этого правила возникают ошибки двух видов.

Определение 17.7. ***Ошибкой 1-го рода*** называется событие, состоящее в том, что гипотеза H_0 отвергается, когда она верна.

Определение 17.8. ***Ошибкой 2-го рода*** называется событие, состоящее в том, что принимается гипотеза H_0, когда верна гипотеза H_1.

Определение 17.9. ***Уровнем значимости*** статистического критерия называется вероятность ошибки 1-го рода $\alpha = \mathsf{P}\{Z \in \overline{G}|H_0\}$. Вероятность ошибки 1-го рода α может быть вычислена, если известно распределение $F(z|H_0)$ статистики Z.

Вероятность ошибки 2-го рода равна $\beta = \mathsf{P}\{Z \in G|H_1\}$ и может быть вычислена, если известно распределение $F(z|H_1)$ статистики Z при справедливости гипотезы H_1.

Ясно, что с уменьшением вероятности α ошибки 1-го рода возрастает вероятность β ошибки 2-го рода, и наоборот, т. е. при выборе критической и доверительной областей должен достигаться определенный компромисс. Поэтому часто при фиксированной вероятности ошибки 1-го рода критическая область выбирается таким образом, чтобы вероятность ошибки 2-го рода была минимальна.

Проверка статистической гипотезы может быть подразделена на следующие этапы:

1) сформулировать проверяемую гипотезу H_0 и альтернативную к ней гипотезу H_1;

2) выбрать уровень значимости α;

3) выбрать статистику Z для проверки гипотезы H_0;

4) найти распределение $F(z|H_0)$ статистики Z, при условии что гипотеза H_0 верна;

5) построить, в зависимости от формулировки гипотезы H_1 и уровня значимости α, критическую область $\overline{G}$;

6) получить реализацию выборки наблюдений $x_1,\ldots,x_n$ и вычислить реализацию $z = \varphi(x_1,\ldots,x_n)$ статистики Z критерия;

7) принять статистическое решение на уровне доверия $1-\alpha$: если $z \in \overline{G}$, то отклонить гипотезу H_0 как не согласующуюся с результатами наблюдений, а если $z \in G$, то принять гипотезу H_0 как не противоречащую результатам наблюдений.

Проверка гипотезы о значении параметра

Пусть имеется параметрическая статистическая модель $(S_\theta, F_{Z_n}(z_n,\theta))$, $\theta \in \Theta \subset \mathbb{R}^1$, т. е. выборка Z_n соответствует распределению $F(x,\theta)$ с неизвестным параметром θ. Проверим простую гипотезу H_0, состоящую в том, что $\theta = \theta_0$, где θ_0 — некоторое фиксированное число из множества Θ.

Формулировка альтернативной гипотезы H_1 и уровень значимости α определяют размер и положение критической области $\overline{G}$ на множестве значений статистики Z. Например, если альтернативная гипотеза H_1 формулируется как $\theta > \theta_0$ (или $\theta < \theta_0$), то критическая область размещается на правом (или левом) «хвосте» распределения статистики Z, т. е.

$$\overline{G} = \{z > z_{1-\alpha}\} \quad (\text{или } \overline{G} = \{z < z_\alpha\}),$$

где $z_{1-\alpha}$ и z_α — квантили уровней $1-\alpha$ и α соответственно распределения $F_Z(z|H_0)$ статистики Z, при условии что верна гипотеза H_0. В этом случае статистический критерий называется ***односторонним***. Если альтернативная гипотеза H_1 формулируется как $\theta \neq \theta_0$, то критическая область $\overline{G}$ размещается на обоих «хвостах» распределения статистики Z, т. е. определяется совокупностью неравенств

$$\overline{G} = \left\{z < z_{\alpha/2}\right\} \cup \left\{z > z_{1-\alpha/2}\right\},$$

где $z_{\alpha/2}$ и $z_{1-\alpha/2}$ — квантили уровней $\alpha/2$ и $1-\alpha/2$ соответственно распределения $F(z|H_0)$. В этом случае критерий называется ***двусторонним***.

Пример 17.2. Пусть известно, что случайная величина X имеет нормальное распределение. Требуется, используя реализацию выборки z_n, проверить гипотезу H_0, состоящую в том, что $m_X = m_0$ (m_0 — некоторое фиксированное число), против альтернативной гипотезы H_1 о том, что $m_X \neq m_0$. Возможны два случая: дисперсия σ_X^2 известна и неизвестна. Статистики Z для обоих случаев можно выбрать, используя утверждение из п. 16.4. Представим эти случаи в виде табл. 17.1.

Предположение	Статистика Z критерия	Распределение $F(z\|H_0)$	Доверительная область G критерия
σ_X^2 известна	$\dfrac{(\overline{X}-m_0)\sqrt{n}}{\sigma_X}$	$\mathcal{N}(0;1)$	$[-u_\gamma, u_\gamma]$
σ_X^2 неизвестна	$\dfrac{(\overline{X}-m_0)\sqrt{n-1}}{\sqrt{s^2}}$	$\mathbf{t}(n-1)$	$\left[-t_{\gamma,n-1}, t_{\gamma,n-1}\right]$

Таблица 17.1

Для каждого случая в соответствии с утверждениями 1) и 4) теоремы 16.1 получаем свою доверительную область, где u_γ, $t_{\gamma,n-1}$ — квантили уровня $\gamma = 1-\alpha/2$ распределений $\mathcal{N}(0;1)$ и $\mathbf{t}(n-1)$ соответственно.

Пример 17.3. Пусть случайная величина X нормально распределена, а ее дисперсия неизвестна. Требуется на основе реализации z_n выборки Z_n, порожденной случайной величиной X, проверить гипотезу H_0 о том, что $\sigma_X^2 = \sigma_0^2$ (σ_0 — некоторое фиксированное число), против альтернативной гипотезы H_1, состоящей в том, что $\sigma_X^2 \neq \sigma_0^2$.

Возможны два случая: m_X — известно или m_X — неизвестно. Представим эти случаи в виде следующей таблицы:

Предположение	Статистика Z критерия	Распределение $F(z\|H_0)$	Доверительная область G критерия
m_X известно	$\dfrac{\sum\limits_{k=1}^{n}(X_k-m_X)^2}{\sigma_0^2}$	$\chi^2(n)$	$[\chi^2_{1-\gamma,n}, \chi^2_{\gamma,n}]$
m_X неизвестно	$\dfrac{ns^2}{\sigma_0^2}$	$\chi^2(n-1)$	$[\chi^2_{1-\gamma,n-1}, \chi^2_{\gamma,n-1}]$

Таблица 17.2

Здесь $\chi^2_{\gamma,k}$, $\chi^2_{1-\gamma,k}$ — квантили уровня $\gamma = 1-\alpha/2$ и $1-\gamma$ распределения $\chi^2(k)$ с k степенями свободы, где $k = n, n-1$.

Пример 17.4. Используя данные примера 14.2, проверить на уровне значимости $\alpha = 0.05$ гипотезу H_0, состоящую в том, что математическое ожидание m_X средней температуры января в Саратове (случайная величина X) равно -13.75, т. е. что $m_X = -13.75$, против альтернативной гипотезы H_1, состоящей в том, что $m_X \neq -13.75$. Будем предполагать, что случайная величина X имеет гауссовское распределение.

Решение. Дисперсия σ_X^2 случайной величины X неизвестна, поэтому выберем для проверки гипотезы H_0 статистику

$$Z = \frac{(\widehat{m}_X - m_0)\sqrt{n-1}}{s_X}.$$

В данной задаче $n = 13$, $m_0 = -13.75$ а $\widehat{m}_X = -11.87$ и $s_X^2 = 22.14$ (результат решения примера 14.2). Согласно примеру 17.2 распределением $F(z|H_0)$ статистики Z при справедливости H_0 является распределение Стьюдента $\mathbf{S}(n-1)$.

Таким образом, доверительная область G имеет вид

$$G = \left[-t_{0.975,12}, t_{0.975,12}\right] = [-2.18, 2.18].$$

Вычисляя значение z статистики Z для данной реализации выборки, имеем

$$z = \frac{(-11.87+13.75)\sqrt{12}}{\sqrt{22.14}} \approx 1.38.$$

Таким образом, значение z статистики Z попадает в доверительную область G, и, следовательно, на уровне доверия $1-\alpha = 0.95$ можно считать, что результаты наблюдений не противоречат гипотезе H_0, состоящей в том, что $m_X = -13.75$.

Ответ. Гипотеза H_0 на уровне значимости $\alpha = 0.05$ принимается.

Пример 17.5. Используя данные примера 14.2, проверить на уровне доверия $1-\alpha=0.95$ гипотезу H_0, состоящую в том, что дисперсия σ_X^2 средней температуры января в г. Саратове (случайная величина X) равна 20, т. е. что $\sigma_X^2=20$, против альтернативной гипотезы H_1, состоящей в том, что $\sigma_X^2 \neq 20$. Будем предполагать, что случайная величина X имеет гауссовское распределение.

Р е ш е н и е. Математическое ожидание m_X случайной величины X неизвестно, поэтому выберем для проверки гипотезы H_0 статистику

$$Z=\frac{n\widehat{d}_X}{\sigma_0^2}.$$

В данной задаче $n=13$, $\sigma_0^2=20$, а $s_X^2=22.14$ (результат решения примера 14.2). Согласно примеру 17.3 распределением $F(z|H_0)$ статистики Z при справедливости H_0 является распределение $\chi^2(12)$.

Таким образом, доверительная область G имеет вид

$$G=\left[\chi^2_{0.025,12},\chi^2_{0.975,12}\right]=[4.4,23.3].$$

Вычисляя значение z статистики Z для данной реализации выборки, имеем

$$z=\frac{13\cdot 22.14}{20}=14.39.$$

Таким образом, значение z статистики Z попадает в доверительную область G, и, следовательно, на уровне доверия $1-\alpha=0.95$ можно считать, что результаты наблюдений не противоречат гипотезе H_0, состоящей в том, что $\sigma_X^2=20$.

О т в е т. Гипотеза H_0 на уровне значимости $\alpha=0.05$ принимается.

Пример 17.6. Пусть $Z_m=(X_1,\ldots,X_m)$ и $V_n=(Y_1,\ldots,Y_n)$ — две независимые выборки, порождённые случайными величинами с распределениями $\mathcal{N}(\mu_1,\sigma^2)$ и $\mathcal{N}(\mu_2,\sigma^2)$ соответственно. Параметры μ_1, μ_2 неизвестны, а параметр σ^2 известен.Опишите процедуру проверки гипотезы H_0 о равенстве средних значений $\mu_1=\mu_2$ против альтернативной гипотезы $H_1:\mu_1\neq\mu_2$.

Пример 17.7. Пусть $Z_m=(X_1,\ldots,X_m)$ и $V_n=(Y_1,\ldots,Y_n)$ — две независимые выборки, порождённые случайными величинами с распределениями $\mathcal{N}(\mu_1,\sigma^2)$ и $\mathcal{N}(\mu_2,\sigma^2)$ соответственно. Параметры μ_1, μ_2 и σ^2 неизвестны.Используя утверждение теоремы16.2, опишите процедуру проверки гипотезы H_0 о равенстве средних значений $\mu_1=\mu_2$ против альтернативной гипотезы $H_1:\mu_1\neq\mu_2$.

Замечание 17.1. На практике обычно задают $\alpha\in[0.01,0.05]$.

Лекция 18

Проверка гипотезы о виде закона распределения. Проверка гипотезы о независимости двух случайных величин

18.1 Проверка гипотезы о виде закона распределения

Пусть имеется реализация z_n выборки Z_n, порожденной случайной величиной X с неизвестной функцией распределения $F(x)$. Требуется проверить гипотезу H_0, состоящую в том, что случайная величина X имеет определенный закон распределения $\overline{F}(x,\theta)$ (например, нормальный, равномерный и т. д.). Истинный закон распределения $F(x)$ неизвестен. Закон распределения $\overline{F}(x,\theta)$ часто называют гипотетическим (т.е. соответствующим основной гипотезе H_0). Для проверки такой гипотезы можно использовать ***статистический критерий хи-квадрат*** (***критерий Пирсона***). Правило проверки заключается в следующем.

1) Формулируется гипотеза H_0, состоящая в том, что случайная величина X имеет распределение определенного вида $\overline{F}(x,\theta_1,\ldots,\theta_s)$ с s неизвестными параметрами $\theta_1,\ldots,\theta_s$ (например, m и σ^2 для нормального распределения, a и b — для равномерного и т. д.).

2) По реализации z_n выборки Z_n методом максимального правдоподобия находятся оценки $\widehat{\theta}_1,\ldots,\widehat{\theta}_s$ неизвестных параметров $\theta_1,\ldots,\theta_s$.

3) Действительная ось $\mathbb{R}^1$ разбивается на $l+1$ непересекающихся полуинтервалов (разрядов) $\Delta_0,\ldots,\Delta_l$ следующим образом. Действительная ось $\mathbb{R}^1=(-\infty,\infty)$ разделяется точками $\alpha_0,\ldots,\alpha_{l+1}$, образуя таким образом $l+1$ непересекающихся полуинтервалов $\Delta_k=[\alpha_k,\alpha_{k+1})$, $k=\overline{0,l}$, при этом $-\infty=\alpha_0<\alpha_1<\ldots<\alpha_l<\alpha_{l+1}=+\infty$. Обычно выбирают $\alpha_1\leqslant x_{(1)}$, $\alpha_l\geqslant x_{(n)}$ так, как это сделано при построении гистограммы в п. 14.11. Подсчитывается число n_k элементов выборки, попавших в каждый k-й разряд Δ_k, $k=\overline{1,l-1}$, за исключением Δ_0 и Δ_l. Полагается $n_0=n_l=0$.

4) Вычисляются гипотетические вероятности p_k попадания случайной величины X в полуинтервалы Δ_k, $k=\overline{0,l}$.

$$p_k=\overline{F}\left(\alpha_{k+1},\widehat{\theta}_1,\ldots,\widehat{\theta}_s\right)-\overline{F}\left(\alpha_k,\widehat{\theta}_1,\ldots,\widehat{\theta}_s\right)$$

Если у распределения $\overline{F}(x,\theta_1,\ldots,\theta_s)$ имеется плотность $\overline{f}(x,\theta_1,\ldots,\theta_s)$, то вероятности p_k могут быть вычислены следующим образом:

$$p_k=\int\limits_{\alpha_k}^{\alpha_{k+1}}\overline{f}\left(x,\widehat{\theta}_1,\ldots,\widehat{\theta}_s\right)dx,$$

где $\alpha_0=-\infty$, $\alpha_{l+1}=+\infty$.

5) Вычисляется реализация статистики критерия хи-квадрат по формуле

$$z=\varphi(z_n)=np_0+\sum_{k=1}^{l-1}\frac{(n_k-np_k)^2}{np_k}+np_l.$$

6) Известно, что при соблюдении некоторых естественных условий регулярности и достаточно большом объеме n выборки Z_n распределение $F(z|H_0)$ статистики $Z = \varphi(Z_n)$ хорошо аппроксимируется распределением $\chi^2(l-s)$ с $l-s$ степенями свободы, где s — количество неизвестных параметров предполагаемого закона распределения $\overline{F}(x,\theta_1,\ldots,\theta_s)$, а $l+1$ — количество разрядов, вероятность попадания в которые ненулевая. Тогда критическая область $\overline{G}$ принимает вид $\overline{G} = (x_{1-\alpha}(l-s),+\infty)$, где $x_{1-\alpha}(l-s)$ — квантиль уровня $1-\alpha$ распределения $\chi^2(l-s)$, α — заданный уровень значимости (обычно $\alpha = 0.05$).

7) В соответствии с критерием хи-квадрат гипотеза H_0 принимается (т. е. реализация выборки z_n согласуется с гипотезой H_0) на уровне надежности $1-\alpha$, если $\varphi(z_n) \in G = [0,x_{1-\alpha}(l-s)]$. Если же $\varphi(z_n) \in \overline{G}$, то гипотеза H_0 отвергается.

Замечание 18.1. Если при разбиении на полуинтервалы Δ_k оказалось, что $np_k < 5$ для $k = \overline{1,l-1}$, то рекомендуется объединить соседние полуинтервалы.

Если при обработке наблюдений имеется только реализация статистического ряда, то вычисляя выборочные моменты считают все выборочные значения, попавшие в k-й интервал, равными середине этого интервала. Это вносит известную ошибку, особенно заметную при малом числе интервалов. Для уменьшения ошибок, вносимых группировкой, применяют ***поправки Шеппарда***. Если все интервалы Δ_k имеют длину, равную h, то с учетом поправок Шеппарда первые четыре выборочных момента $\widehat{\nu}_i'$, $i = \overline{1,4}$, соответственно равны

$$\widehat{\nu}_1' = \widehat{\nu}_1, \qquad \widehat{\nu}_2' = \widehat{\nu}_2 - \frac{1}{12}h^2,$$

$$\widehat{\nu}_3' = \widehat{\nu}_3 - \frac{1}{4}\widehat{\nu}_1 h^2, \qquad \widehat{\nu}_4' = \widehat{\nu}_4 - \frac{1}{2}\widehat{\nu}_2 h^2 + \frac{7}{240}h^4.$$

Пример 18.1. В течение Второй мировой войны на южную часть Лондона упало 535 снарядов. Территория южного Лондона была разделена на 576 участков площадью 0,25 км^2. В следующей таблице приведено количество участков n_k, на каждый из которых упало по k снарядов:

k	0	1	2	3	4	5
n_k	299	211	93	35	7	1

Таблица 18.1

Требуется с помощью критерия хи-квадрат проверить на уровне доверия $1-\alpha = 0{,}95$ гипотезу H_0, состоящую в том, что случайная величина X (число снарядов, упавших на один участок) распределена по закону Пуассона.

Р е ш е н и е. Распределение Пуассона $\mathbf{\Pi}(\theta)$ имеет один параметр, МП-оценка этого параметра $\widehat{\theta} = \widehat{m}_X \approx 0{,}93$ (см. пример 15.4).

В данной задаче действительная ось естественным образом разбивается на 8 непересекающихся полуинтервалов Δ_k: $(-\infty,0)$, $[0,1)$, $[1,2)$,..., $[5,6)$, $[6,+\infty)$.

Гипотетические вероятности p_k попадания пуассоновской случайной величины X в k-й полуинтервал вычисляются по следующим формулам:

$$p_k = \frac{e^{-\widehat{\theta}}\widehat{\theta}^{k-1}}{(k-1)!} \qquad \text{для } k = \overline{1,6},$$

$$p_7 = 1 - \sum_{k=1}^{6} p_k, \qquad p_0 = \mathsf{P}\{X < 0\} = 0.$$

Вычисленные значения вероятностей указаны в табл. 18.2.

k	1	2	3	4	5	6	7
Δ_k	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,\infty)$
p_k	0,394	0,366	0,171	0,053	0,012	0,002	0,002

Таблица 18.2

Вычисляя значения z статистики критерия хи-квадрат, получим

$$z = \sum_{k=1}^{6} \frac{(n_k - np_k)^2}{np_k} + np_7 \approx 2{,}2.$$

При справедливости гипотезы H_0 статистика Z имеет распределение $\chi^2(5)$. Тогда критическая область $\overline{G}$ имеет вид $\overline{G} = (x_{0,95}(5),+\infty) = (11{,}1,+\infty)$, а доверительная область — $G = [0{,}11{,}1]$.

Так как вычисленное по выборке значение статистики попадает в доверительную область G, то с вероятностью 0,95 можно утверждать, что опытные данные согласуются с гипотезой H_0.

О т в е т. Гипотеза H_0 принимается на уровне значимости $\alpha = 0.05$.

Пример 18.2. Используя данные примера 14.1, проверить на уровне значимости $\alpha = 0{,}95$ гипотезу H_0, состоящую в том, что рост взрослого мужчины (случайная величина X) имеет нормальное распределение.

Решение. Нормальное распределение $\mathbf{N}(\theta_1;\theta_2^2)$ имеет два неизвестных параметра: $\theta_1 = m_X$ и $\theta_2^2 = \sigma_X^2$. МП-оценкой параметра θ_1 является выборочное среднее $\widehat{m}_X$, а параметра θ_2^2 — выборочная дисперсия $\widehat{d}_X$. С учетом поправок Шеппарда

$$\widehat{m}_X = 165{,}77, \qquad \widehat{d}_X = 34{,}24.$$

Вычислим гипотетические вероятности p_k, $k = \overline{1{,}15}$, попадания гауссовской случайной величины X в полуинтервалы Δ_k (которые определены в задаче 14.1) по приближенной формуле

$$p_k = h \cdot \frac{1}{\sqrt{2\pi}\sqrt{\widehat{d}_X}} \exp\left\{-\frac{(\bar{x}_k - \widehat{m}_X)^2}{2\widehat{d}_X}\right\},$$

где $\bar{x}_k$ — середина интервала Δ_k, а h — длина интервала Δ_k, которая в данной задаче равна 3 для всех $k = \overline{1{,}15}$. Результаты вычислений для $k = \overline{1{,}15}$ приведены в табл. 18.3.

Δ_k	143–146	146–149	149–152	152–155	155–158
p_k	0,0003	0,002	0,007	0,023	0,058
Δ_k	158–161	161–164	164–167	167–170	170–173
p_k	0,115	0,175	0,204	0,184	0,127
Δ_k	173–176	176–179	179–182	182–185	185–188
p_k	0,067	0,027	0,009	0,002	0,0004

Таблица 18.3

Вероятность попадания в интервал $\Delta_0 = (-\infty, 143)$ равна $p_0 = 4 \times 10^{-5}$, а в интервал $\Delta_{16} = [188, +\infty)$ — $p_{16} = 4 \cdot 10^{-5}$. Вычисляя реализацию z статистики Z, получим

$$z = np_0 + \sum_{k=1}^{15} \frac{(n_k - np_k)^2}{np_k} + np_{16} \approx 7{,}177.$$

При справедливости гипотезы H_0 статистика Z имеет распределение $\chi^2(14)$. Тогда критическая область $\overline{G}$ имеет вид $\overline{G} = (x_{0,95}(14), +\infty) = (23{,}7, +\infty)$, а доверительная область — $G = [0, 23{,}7]$.

Так как вычисленное по выборке значение статистики попадает в доверительную область G, то с вероятностью 0,95 можно утверждать, что опытные данные согласуются с гипотезой H_0.

Ответ. Гипотеза H_0 принимается на уровне значимости $\alpha = 0.05$.

18.2 Проверка гипотезы о независимости двух случайных величин

Пусть имеется случайный вектор $V = (X, Y)$ с функцией распределения $F_V(x,y)$. Компонентами V являются случайные величины X и Y с функциями распределения $F_X(x)$ и $F_Y(y)$ соответственно. Пусть имеется выборка $Z_n = (V_1, \ldots, V_n)$, где $V_k = (X_k, Y_k)$. Здесь выборка $X_1, \ldots, X_n$ соответствует распределению $F_X(x)$ случайной величины X, а выборка $Y_1, \ldots, Y_n$ соответствует распределению $F_Y(y)$ случайной величины Y. Требуется проверить гипотезу H_0 о независимости случайных величин X и Y, то есть

$$H_0 : F_V(x,y) = F_X(x)F_Y(y) \qquad \text{для всех } x \in \mathbb{R}^1,\ y \in \mathbb{R}^1. \tag{18.1}$$

против альтернативной гипотезы

$$H_1 : \exists x, y \in R^1 \text{ такие, что } F_V(x,y) \neq F_X(x)F_Y(y). \tag{18.2}$$

Для проверки этой гипотезы используем ***критерий хи-квадрат***, процедура применения которого состоит в следующих действиях.

1) Множество значений случайной величины X разбивается на s непересекающихся интервалов (разрядов) $\Delta_{x,1}, \ldots, \Delta_{x,s}$, а множество значений случайной величины Y — на r непересекающихся интервалов $\Delta_{y,1}, \ldots, \Delta_{y,r}$.

2) Для каждого $i = \overline{1,s}$ и каждого $j = \overline{1,r}$ вычисляется число n_{ij} элементов выборки z_n, принадлежащих прямоугольнику $\Delta_{x,i} \times \Delta_{y,j}$.

3) Вычисляется суммарное число $n_{x,i}$ элементов выборки z_n, первая компонента которой попала в i-й разряд $\Delta_{x,i}$ для случайной величины X, и аналогично — число $n_{y,j}$ элементов той же выборки z_n, вторая компонента которой попала в j-й разряд $\Delta_{y,j}$ для случайной величины Y:

$$n_{i\bullet} = \sum_{j=1}^{r} n_{ij}, \; n_{\bullet j} = \sum_{i=1}^{s} n_{ij}, \; n = \sum_{i=1}^{s}\sum_{j=1}^{r} n_{ij}.$$

4) Вычисляется значение статистики критерия хи-квадрат по формуле

$$z = \varphi(z_n) = n\left(\sum_{i=1}^{s}\sum_{j=1}^{r}\frac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}} - 1\right).$$

5) При справедливости гипотезы H_0 и достаточно большом n распределение статистики $Z = \varphi(Z_n)$ хорошо аппроксимируется распределением хи-квадрат с $m = (s-1)(r-1)$ степенями свободы. Поэтому критическая область имеет вид

$$\overline{G} = (\chi^2_{1-\alpha}(m), +\infty),$$

где $\chi^2_{1-\alpha}(m)$ — квантиль уровня $1-\alpha$ распределения $\chi^2(m)$.

6) Принимается статистическое решение на уровне значимости α: отклонить гипотезу H_0, если $\varphi(z_n) \in \overline{G}$, и принять гипотезу H_0 — в противном случае.

Типовые задачи

Пример 18.3. По переписи населения Швеции 1936 г. из совокупности всех супружеских пар была получена выборка в 25 263 пары, вступивших в брак в течение 1931–1936 гг. В следующей таблице приведено распределение годовых доходов (в тыс. крон) и количество детей у супружеских пар в этой выборке. Требуется установить на уровне значимости $\alpha = 0{,}05$ являются ли зависимыми количество детей в семье (случайная величина X) и уровень годового дохода этой семьи (случайная величина Y).

Доходы / Число детей	0–1	1–2	2–3	> 3	Сумма
0	2 161	3 577	2 184	1 636	9 558
1	2 755	5 081	2 222	1 052	11 110
2	936	1 753	640	306	3 635
3	225	419	96	38	778
$\geqslant 4$	39	98	31	14	182
Сумма	6 116	10 928	5 173	3 016	25 263

Таблица 18.4

Решение. Для проверки гипотезы H_0 о независимости случайной величины X и случайной величины Y воспользуемся критерием хи-квадрат. В данной задаче множество значений случайной величины X (количество детей в семье) разбивается на 5 разрядов $\Delta_{x,i}$, $i = \overline{1,5}$, естественным образом: 0, 1, 2, 3 и не менее 4 детей (см. первый столбец табл. 18.4). Множество значений случайной величины Y (годовой доход семьи) разбито на 4 разряда $\Delta_{y,j}$, $j = \overline{1,4}$, которые указаны в первой строке табл. 18.4). Вычисляя значение z (см. п. 18.2) по формуле

$$z = n\left(\sum_{i=1}^{s}\sum_{j=1}^{r}\frac{n_{ij}^2}{n_{i\bullet}n_{j\bullet}} - 1\right),$$

где $n = 25263$, $s = 5$, $r = 4$, числа $n_{i\bullet}$, $i = \overline{1,5}$, приведены в последнем столбце таблицы, а числа $n_{\bullet j}$, $j = \overline{1,4}$, приведены в последней строке таблицы, получим $z = 568{,}5$.

При справедливости гипотезы H_0 статистика Z имеет распределение хи-квадрат с числом степеней свободы $m = (s-1)(r-1) = 4 \times 3 = 12$. Тогда критическая область $\overline{G}$ имеет вид $\overline{G} = (\chi^2_{0,95}(12), +\infty) = (21, +\infty)$, а доверительная область — $G = [0, 21]$.

Так как вычисленное по выборке значение статистики попадает в критическую область $\overline{G}$, то на уровне значимости 0,05 гипотеза H_0 о независимости случайной величины X (количество детей в семье) и случайной величины Y (уровень годового дохода семьи) отвергается.

Ответ. Гипотеза H_0 о независимости случайной величины X (количество детей в семье) и случайной величины Y (уровень годового дохода семьи) отвергается на уровне значимости 0,05.

Часто в социологических, психологических и медицинских исследованиях возникают задачи выявления зависимостей таких показателей, которые измеряются не в количественной, а в качественной шкале.Например, показатель А может принимать значения $A_1,\ldots,A_s$ (программист, математик,...продавец), а показатель В - значения $B_1,\ldots,B_r$ (добрый, весёлый,... умный). Обозначим $p_{ij} = p(A = A_i, B = B_j)$, $p_{i\cdot} = p(A = A_i)$, $p_{\cdot j} = p(B = B_j)$. Тогда гипотеза H_0 о независимости признаков A и B формулируется следующим образом:

$$p_{ij} = p_{i\cdot}p_{\cdot j},\ \forall i = 1,\ldots,s,\ j = 1,\ldots,r. \tag{18.3}$$

Для проверки такой гипотезы также может быть применён представленный выше критерий хи-квадрат. Разбиение выборочных данных на разряды,о которых говорится в пункте 1, естественным образом может быть представлено с помощью таблицы

Таблица 9.3

$A\backslash B$	B_1	$\cdots$	B_r	
A_1	n_{11}	$\cdots$	n_{1r}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_s	n_{s1}	$\cdots$	n_{sr}	$n_{s\cdot}$
	$n_{\cdot 1}$	$\cdots$	$n_{\cdot r}$	n

Такую таблицу принято называть *таблицей сопряженности признаков*.

Пример 18.4. В таблице приведены 2663 случая, классифицированных по двум признакам: А — наличие прививки против холеры, В — отсутствие заболевания. Проверить гипотезу о независимости признаков А и В на уровне значимости 0,05.

$A\backslash B$	B	$\overline{B}$
A	1625	5
$\overline{A}$	1022	11

18.3 Проверка гипотезы о некоррелированности случайных величин

Пусть справедливо предположение о том, что вектор $V = (X,Y)$ — гауссовский, причем $\mathsf{D}X > 0$ и $\mathsf{D}Y > 0$.

Тогда гипотеза H_0 вида (18.1) о независимости СВ X и Y эквивалентна гипотезе

$$H_0:\ \rho_{XY} = 0, \tag{18.4}$$

где $\rho_{XY} = \frac{k_{XY}}{\sqrt{\mathsf{D}X\mathsf{D}Y}}$ — коэффициент корреляции СВ X и Y, а k_{XY} — их ковариация.

Эквивалентность (18.1) и (18.4) следует из того, что компоненты гауссовского вектора X и Y независимы тогда и только тогда, когда X и Y некоррелированы.

Оценкой неизвестного коэффициента корреляции ρ_{XY} является *выборочный коэффициент корреляции*

$$\widehat{\rho}_{XY}(n) = \frac{\widehat{k}_{XY}}{s_X s_Y}, \tag{18.5}$$

где s_X^2, s_Y^2 — выборочные дисперсии, построенные по выборкам $X_1,\ldots,X_n$ и $Y_1,\ldots,Y_n$ соответственно, а $\widehat{k}_{XY}$ — выборочная ковариация, построенная по двумерной выборке V.

Можно показать, что при выполнении сделанного предположения и справедливости гипотезы H_0 вида (18.4), статистика

$$Z(V) = \frac{\sqrt{n-2}\,\widehat{\rho}_{XY}}{\sqrt{1-\widehat{\rho}_{XY}^2}} \tag{18.6}$$

имеет распределение Стьюдента $t(n-2)$ с $n-2$ степенями свободы.

При $n \to \infty$ и справедливости H_0 статистика вида

$$\widetilde{Z}(V) = \sqrt{n}\,\widehat{\rho}_{XY} \tag{18.7}$$

асимптотически имеет стандартное нормальное распределение.

Критические области уровня значимости α для критериев, основанных на статистиках (18.6) и (18.7), приведены в таблице 9.1, где $t_\gamma(l)$, u_γ — квантили уровня γ распределений $t(l)$ и $N(0;1)$ соответственно.

Таблица 9.1

H_A	Критические области для $Z(V)$	Критические области для $\widetilde{Z}(V)$
$\rho_{XY} < 0$	$(-\infty; t_{\alpha,n-2})$	$(-\infty; u_\alpha)$
$\rho_{XY} > 0$	$(t_{1-\alpha,n-2}; +\infty)$	$(u_{1-\alpha}; +\infty)$
$\rho_{XY} \neq 0$	$(-\infty; t_{\frac{\alpha}{2},n-2}) \cup (t_{1-\frac{\alpha}{2},n-2}; +\infty)$	$(-\infty; u_{\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$

Важно отметить, что если вектор $V = (X,Y)$ — гауссовский, то утверждение о том, что СВ X и Y зависимы справедливо тогда и только тогда, когда $\rho_{XY} \neq 0$. Таким образом, в рассматриваемом случае альтернативная гипотеза общего вида

$$H_1 : \ \exists x,y \in R^1 \text{ такие, что } F_V(x,y) \neq F_X(x)F_Y(y) \tag{18.8}$$

эквивалентна гипотезе

$$H_A : \ \rho_{XY} \neq 0.$$