

Оглавление

1	Случайные события. Вероятность	3
1.1	Случайные события	3
1.2	Вероятность	8
2	Условная вероятность. Формула полной вероятности. Формула Байеса. Схема Бернулли	15
2.1	Условная вероятность	15
2.2	Формула полной вероятности	21
2.3	Формула Байеса	22
2.4	Схема Бернулли	23
3	Одномерные случайные величины	25
3.1	Определение случайной величины	25
3.2	Функция распределения случайной величины	26
3.3	Дискретные случайные величины	27
3.4	Непрерывные случайные величины	27
3.5	Функции от случайной величины	29
4	Числовые характеристики случайных величин	32
4.1	Математическое ожидание случайной величины	32
4.2	Дисперсия. Моменты высших порядков	34
4.3	Квантиль	35
5	Основные законы распределения случайных величин	36
5.1	Биномиальное распределение	36
5.2	Распределение Пуассона	37
5.3	Геометрическое распределение	38
5.4	Равномерное распределение	39
5.5	Экспоненциальное распределение	39
5.6	Нормальное распределение	40
6	Случайные векторы	43
6.1	Функция распределения случайного вектора	43
6.2	Двумерные случайные векторы	44
6.3	Дискретные двумерные случайные векторы	45
6.4	Непрерывные случайные векторы	46
6.5	Независимые случайные величины	47
7	Функции от случайных величин. Свойства математического ожидания и дисперсии	50
7.1	Скалярные функции от случайного векторного аргумента	50
7.2	Математическое ожидание функции от случайного вектора	52
7.3	Свойства математического ожидания	52
7.4	Свойства дисперсии	53
7.5	Формула свертки	54
8	Ковариация и коэффициент корреляции случайных величин	56
9	Многомерное нормальное распределение	60

10	Предельные теоремы теории вероятностей	64
11	Основные понятия выборочной теории	72
12	Точечные оценки	78
12.1	Метод моментов	80
12.2	Метод максимального правдоподобия	81
13	Интервальное оценивание	83
13.1	Некоторые важные вероятностные распределения	83
13.2	Определение доверительного интервала	86
13.3	Доверительные интервалы для параметров нормальной случайной величины	86
	Доверительный интервал для математического ожидания нормальной случайной величины с известной дисперсией	86
	Доверительный интервал для математического ожидания нормальной случайной величины с неизвестной дисперсией	87
	Доверительные интервалы для дисперсии и среднеквадратического отклонения при неизвестном математическом ожидании	88
13.4	Доверительные интервалы для параметров при наблюдении двух нормальных случайных величин	89
	Доверительный интервал для разности математических ожиданий нормальных случайных величин с известными дисперсиями	89
	Доверительный интервал для разности математических ожиданий нормальных случайных величин с неизвестными, но равными дисперсиями	90
13.5	Приближенные доверительные интервалы	91
	Приближенный доверительный интервал для вероятности успеха в схеме Бернулли	92
	Приближенный доверительный интервал для параметра распределения Пуассона	92
14	Проверка гипотез. Параметрические модели	93
14.1	Основные понятия	93
14.2	Проверка гипотез о математическом ожидании нормальной случайной величины с известной дисперсией	95
14.3	Проверка гипотез о математическом ожидании нормальной случайной величины с неизвестной дисперсией	97
14.4	Проверка гипотез о дисперсии нормальной случайной величины при неизвестном математическом ожидании	99
14.5	Проверка гипотез о равенстве математических ожиданий нормальных случайных величин с известными дисперсиями	101
14.6	Проверка гипотез о равенстве математических ожиданий нормальных случайных величин с неизвестными, но равными дисперсиями	101
14.7	Проверка гипотез о равенстве дисперсий нормальных случайных величин с неизвестными математическими ожиданиями	102
15	Проверка непараметрических гипотез	104
16	Метод наименьших квадратов	110

Лекция 1

Случайные события. Вероятность

1.1 Случайные события

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

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

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

В дальнейшем пространство элементарных исходов будем обозначать прописной буквой Ω , а сами элементарные исходы — строчной буквой ω , снабженной, при необходимости, индексами. То, что элемент ω принадлежит Ω , записывают в виде $\omega \in \Omega$, а тот факт, что множество Ω состоит из элементов $\omega_1, \omega_2, \dots, \omega_n, \dots$, и только из них, записывают в виде

$$\Omega = \{\omega_1, \omega_2, \dots; \omega_n, \dots\}$$

или в виде

$$\Omega = \{\omega_i, \quad i = 1, 2, \dots, n, \dots\}.$$

В частности, Ω может содержать конечное число элементарных исходов.

Рассмотрим примеры, поясняющие понятие пространства элементарных исходов.

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

$$\Omega = \{\Gamma, \Pi\}, \quad \Omega = \{\omega_\Gamma, \omega_\Pi\}$$

или

$$\Omega = \{\omega_1, \omega_2\}.$$

При двукратном подбрасывании монеты (или однократном подбрасывании двух монет) пространство элементарных исходов будет, очевидно, содержать 4 элемента, т.е.

$$\Omega = \{\omega_{\Gamma\Gamma}, \omega_{\Gamma\Pi}, \omega_{\Pi\Gamma}, \omega_{\Pi\Pi}\},$$

где $\omega_{\Gamma\Gamma}$ — появление “герба” и при первом, и при втором подбрасываниях, и т.д.

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

$$\Omega = \{\omega_i, \quad i = \overline{1,6}\}.$$

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.

$$\Omega = \{0, 1, \dots, n, \dots\}.$$

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

$$\Omega = \{(x; y) : -\infty < x < +\infty, -\infty < y < +\infty\}.$$

События, действия над ними

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

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

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

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

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

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

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

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

Пример 1.5. В примере 1.2 было показано, что при однократном бросании игральной кости

$$\Omega = \{\omega_i, \quad i = \overline{1,6}\},$$

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

$$A = \{\omega_2; \omega_4; \omega_6\}, \quad B = \{\omega_1; \omega_3; \omega_5\} \quad \text{и} \quad C = \{\omega_3; \omega_6\}.$$

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

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

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

Невозможное событие будем обозначать символом \emptyset .

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

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

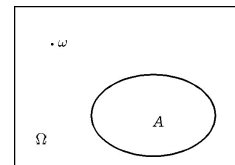


Рис 1.1

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

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

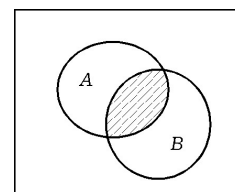


Рис 1.2

Пересечение событий A и B записывают следующим образом:

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

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*.

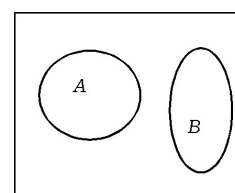


Рис 1.3

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 .

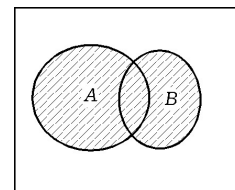


Рис 1.4

Объединение событий A и B записывают в виде

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

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

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

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

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

состоит из элементарных исходов, принадлежащих хотя бы одному из событий A_n , $n \in \mathbb{N}$. В частности, **события** A_1, A_2, \dots, A_n называют **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

$$A_1 A_2 \dots A_n = \emptyset.$$

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

$$C = A \setminus B.$$

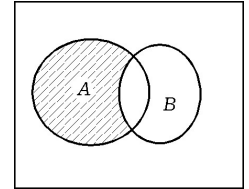


Рис 1.5

Определение 1.8. **Дополнением события** A (обычно обозначают \bar{A}) называют событие, происходящее тогда и только тогда, когда не происходит событие A . Другими словами, $\bar{A} = \Omega \setminus A$. Событие \bar{A} называют также **событием, противоположным** событию A .

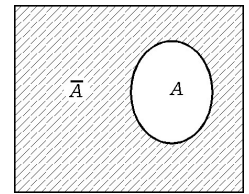


Рис 1.6

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

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

эквивалентна формуле

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

Следует отметить, что все действия над событиями можно получить с помощью только двух действий — объединения и дополнения (или пересечения и дополнения). Основанием для этого утверждения служат законы де Моргана, а также соотношение

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

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

Определение 1.9. 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 .

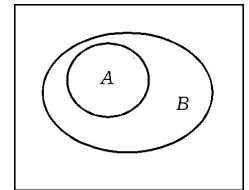


Рис 1.7.

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

Пример 1.7. Рассмотрим техническое устройство (ТУ), состоящее из m элементов. В теории надежности принято говорить, что элементы соединены последовательно, если ТУ прекращает функционировать при отказе любого элемента, и соединены параллельно, если прекращение функционирования ТУ наступает только при отказе всех m элементов. Условное изображение параллельного и последовательного соединений представлено на рис. 1.8 и 1.9 соответственно.

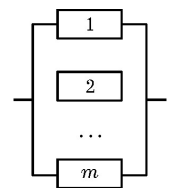


Рис 1.8

Обозначим A событие, означающее отказ ТУ, а A_i — событие, означающее отказ i -го элемента ($i = \overline{1, m}$). Тогда события A и A_i связаны соотношениями

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

для последовательного соединения и параллельного соединения соответственно.

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

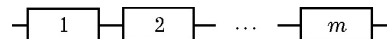


Рис 1.9

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

1. Коммутативность суммы и произведения:

$$A \cup B = B \cup A, \quad AB = BA.$$

2. Ассоциативность суммы и произведения:

$$A \cup B \cup C = A \cup (B \cup C), \quad (AB)C = A(BC).$$

3. Дистрибутивность относительно сложения:

$$(A \cup B)C = AC \cup BC.$$

4. Дистрибутивность относительно умножения (новое свойство, не выполняющееся для чисел):

$$AB \cup C = (A \cup C)(B \cup C).$$

5. Включение A в B , т.е. $A \subset B$, влечет за собой включение \bar{B} в \bar{A} , т.е.

$$\bar{A} \supset \bar{B}.$$

6. Совпадение двойного дополнения с исходным событием:

$$\bar{\bar{A}} = A.$$

7. Совпадение суммы и произведения одинаковых событий с самим событием

$$A \cup A = AA = A.$$

8. *Законы де Моргана:*

$$\overline{A \cup B} = \bar{A} \bar{B}, \quad \overline{AB} = \bar{A} \cup \bar{B}.$$

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

$$\overline{A_1 \cup A_2 \cup \dots \cup A_n} = \bar{A}_1 \bar{A}_2 \dots \bar{A}_n$$

$$\overline{\bar{A}_1 \bar{A}_2 \dots \bar{A}_n} = A_1 \cup A_2 \cup \dots \cup A_n.$$

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 \mathcal{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 \bar{A} и \bar{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 1.10. *Сигма-алгеброй (σ -алгеброй)* событий \mathcal{A} назовем непустую систему подмножеств пространства элементарных исходов Ω , удовлетворяющую следующим двум условиям.

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

Поскольку $\Omega = A \cup \bar{A}$ и $\emptyset = \bar{\Omega}$, то *достоверное событие* Ω и *невозможное событие* \emptyset принадлежат \mathcal{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 1.8. Пусть опыт состоит в подбрасывании один раз тетраэдра, каждая грань которого помечена одним из чисел 1, 2, 3 и 4.

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

Поскольку в рассматриваемом опыте может происходить одно из следующих событий:

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

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

1.2 Вероятность

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

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

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

Классическое определение вероятности

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

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

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

Определение 1.11. *Вероятностью события A* называют отношение числа N_A благоприятствующих событию A элементарных исходов к общему числу N равновозможных элементарных исходов, т.е.

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

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

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

Пример 1.9. Из урны, содержащей $k = 10$ белых и $l = 20$ черных шаров (шары отличаются лишь цветом), наугад вынимают один шар. Требуется найти вероятность $P(A)$ события A , заключающегося в том, что из урны извлечен белый шар.

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

Используя классическое определение вероятности события, докажем следующие свойства.

Свойство 1.1. Для любого события A вероятность удовлетворяет неравенству $P(A) \geq 0$.

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

Свойство 1.2. Для достоверного события Ω (которое содержит все N элементарных исходов) $P(\Omega) = 1$.

Свойство 1.3. Если события A и B несовместны ($AB = \emptyset$), то $P(A+B) = P(A) + P(B)$.

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

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

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

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

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

Недостаток классического определения заключается в том, что оно применимо только к пространствам элементарных исходов, состоящим из конечного числа 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 1.12. *Вероятностью события A называют число $P(A)$, равное отношению меры множества A к мере множества Ω :*

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

где $\mu(A)$ — мера множества A .

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

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

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

Пример 1.10. 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. 1.10). 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

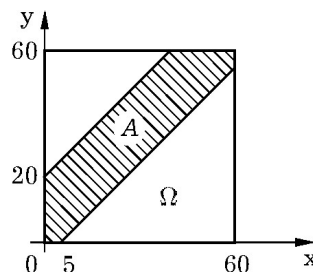


Рис 1.10

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

Область A элементарных исходов, благоприятствующих этому событию, на рис. 1.10 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 1.12, находим $P(A) = \frac{S_A}{S_\Omega} = \frac{1287,5}{3600} \approx 0,36$.

Статистическое определение вероятности

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

Определение 1.13. Пусть произведено n повторений опыта, причем в n_A из них появилось событие A . Число $r_A = n_A/n$ 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 1.14. **Вероятностью события A** называют (эмпирический) предел $P(A)$, к которому стремится частота r_A события A при неограниченном увеличении числа n опытов.

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

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

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

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

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

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

Собственно говоря, задача определения связи вероятности с частотой не потеряла актуальности и в наши дни, когда в теории вероятностей повсеместно используется аксиоматическое определение вероятностей Колмогорова (см. определение 1.15). Это привело к появлению и широкому внедрению в практику обширного раздела теории вероятностей — математической статистики.

Аксиоматическое определение вероятности

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

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

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

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

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

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

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

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

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

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

Определение 1.15. Пусть каждому событию A (т.е. подмножеству A пространства элементарных исходов Ω , принадлежащему σ -алгебре \mathcal{A}) поставлено в соответствие число $P(A)$. Числовую функцию P (заданную на σ -алгебре \mathcal{A}) называют **вероятностью (или вероятностной мерой)**, если она удовлетворяет следующим аксиомам:

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

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

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

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

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

Иногда вместо аксиомы 3 удобно использовать две другие аксиомы.

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

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

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

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

Замечание 1.3. является конечным или счетным множеством, то каждому элементарному исходу $\omega_i \in \Omega$, $i = 1, 2, \dots$, можно поставить в соответствие число $P(\omega_i) = p_i \geq 0$ так, что

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

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

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

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

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

Теорема 1.1. Вероятность удовлетворяет следующим свойствам.

1. Вероятность противоположного события

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

2. Вероятность невозможного события

$$P(\emptyset) = 0.$$

3. Если $A \subset B$, то

$$P(A) \leq P(B)$$

(“большему” событию соответствует большая вероятность).

4. Вероятность заключена между 0 и 1:

$$0 \leq P(A) \leq 1.$$

5. Вероятность объединения двух событий

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

6. Вероятность объединения любого конечного числа событий

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

ДОКАЗАТЕЛЬСТВО. Поскольку

$$\Omega = A + \bar{A},$$

то, согласно расширенной аксиоме сложения,

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

откуда с учетом аксиомы нормированности получаем утверждение 1.

Утверждение 2 вытекает из равенства

$$A = A + \emptyset$$

и расширенной аксиомы сложения.

Пусть $A \subset B$. Тогда

$$B = A + (B \setminus A).$$

В соответствии с расширенной аксиомой сложения

$$P(B) = P(A) + P(B \setminus A).$$

Отсюда и из аксиомы неотрицательности приходим к утверждению 3.

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

Поскольку

$$A \cup B = A + (B \setminus A), \quad B = (B \setminus A) + AB,$$

то, используя расширенную аксиому сложения, находим

$$P(A \cup B) = P(A) + P(B \setminus A)$$

и

$$P(B) = P(B \setminus A) + P(AB).$$

Подставляя в первое из последних двух равенств вероятность $P(B \setminus A)$, выраженную из второго равенства, приходим к утверждению 5.

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

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

Для четырех и более событий это утверждение проверьте самостоятельно.

Замечание 1.4. 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 1.11. Опыт состоит в двукратном подбрасывании симметричной монеты. Найдем вероятность события A , означающего появление “герба” хотя бы один раз. Обозначим A_i появление “герба” при i -м подбрасывании, $i = 1, 2$. Ясно, что

$$A = A_1 \cup A_2,$$

и в соответствии с классической схемой вероятности

$$P(A_1) = P(A_2) = \frac{1}{2}.$$

Если не учитывать, что A_1 и A_2 — совместные события, то можно получить “результат”

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

противоречащий здравому смыслу, поскольку ясно, что *событие* A не является *достоверным*. Применяя теорему сложения для двух совместных событий и учитывая равенство

$$P(A_1 A_2) = \frac{1}{4},$$

находим

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

Определение 1.16. Тройку (Ω, \mathcal{A}, P) , состоящую из пространства элементарных исходов Ω , с σ -алгеброй событий \mathcal{A} и определенной на \mathcal{A} вероятности P , называют *вероятностным пространством*.

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

Лекция 2

Условная вероятность. Формула полной вероятности. Формула Байеса. Схема Бернулли

2.1 Условная вероятность

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

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

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

Определение условной вероятности

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

$$P(A|B) = \frac{N_{AB}}{N_B}.$$

Если теперь поделить числитель и знаменатель полученного выражения на общее число N элементарных исходов, то придем к формуле

$$P(A|B) = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)}.$$

На основании изложенного выше можно дать следующее определение.

Определение 2.1. *Условной вероятностью* события A при условии (наступлении) события B называют отношение вероятности *пересечения событий* A и B к вероятности события B :

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

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

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

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

Теорема 2.1. *Условная вероятность $P(A|B)$ обладает всеми свойствами безусловной вероятности $P(A)$.*

ДОКАЗАТЕЛЬСТВО. Для доказательства достаточно показать, что условная вероятность $P(A|B)$ удовлетворяет аксиомам 1, 2 и 3 (см. определение 1.15).

Из определения 2.1 следует, что условная вероятность, удовлетворяет аксиоме неотрицательности, так как числитель дроби является неотрицательным числом, а знаменатель — положительным числом.

Поскольку $\Omega B = B$, то

$$P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1,$$

т. е. условная вероятность удовлетворяет аксиоме нормированности.

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

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

и

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

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

Смысл теоремы 2.1 заключается в том, что условная вероятность представляет собой безусловную вероятность, заданную на новом *пространстве* Ω_1 *элементарных исходов*, совпадающем с событием B .

Пример 2.1. Рассмотрим опыт с однократным бросанием игральной кости, но не обычной, а с раскрашенными гранями: грани с цифрами 1, 3 и 6 окрашены красным, а грани с цифрами 2, 4 и 5 — белым цветом. Введем события: A_1 — выпадение нечетного числа очков; A_2 — выпадение четного числа очков; B — появление грани красного цвета. Интуитивно ясно, что если произошло событие B , то условная вероятность события A_1 больше, чем условная вероятность события A_2 , поскольку на красных гранях нечетных чисел в два раза больше, чем четных. Заметим, что безусловные вероятности событий A_1 и A_2 при этом одинаковы и равны, очевидно, $1/2$.

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

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

Следовательно, в силу определения 2.1 условной вероятности имеем

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

что подтверждает наше предположение.

Геометрическая интерпретация условной вероятности

При практическом вычислении условной вероятности события A при условии, что событие B произошло, часто удобно трактовать условную вероятность как безусловную, но заданную не на исходном пространстве Ω элементарных исходов, а на новом пространстве $\Omega_1 = B$ элементарных исходов. Действительно, используя *геометрическое определение вероятности*, получаем для безусловной и условной вероятностей события A (на рис. 2.1 заштрихованная область соответствует событию AB):

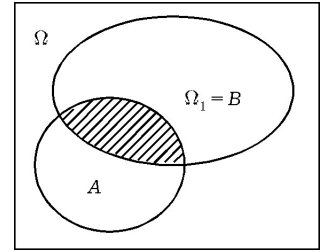


Рис 2.1

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

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

Пример 2.2. Из урны, в которой $a = 7$ белых и $b = 3$ черных шаров, наугад без возвращения извлекают два шара. Пусть событие A_1 состоит в том, что первый извлеченный из урны шар является белым, а A_2 — белым является второй шар. Требуется найти $P(A_2|A_1)$.

Приведем решение этой задачи двумя способами.

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

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

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

Формула умножения вероятностей

При решении различных задач вероятностного характера часто интересующее нас событие A можно достаточно просто выразить через некоторые события A_1, A_2, \dots, A_n с помощью операций объединения или пересечения. Если $A = A_1A_2 \dots A_n$, то для нахождения вероятности $P(A)$ события A обычно удобно использовать следующую теорему.

Теорема 2.2 (теорема умножения вероятностей). Пусть событие $A = A_1A_2 \dots A_n$ (т. е. A — пересечение событий A_1, A_2, \dots, A_n) и $P(A) > 0$. Тогда справедливо равенство

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

называемое *формулой умножения вероятностей*.

ДОКАЗАТЕЛЬСТВО. Поскольку $P(A) = P(A_1A_2 \dots A_n) > 0$, а $A_1A_2 \dots A_k \supseteq A_1A_2 \dots A_n$, то и $P(A_1A_2 \dots A_k) > 0$, $k = \overline{1, n-1}$. Учитывая это неравенство, согласно определению 2.1 условной вероятности, имеем

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

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

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

Аналогично находим

$$P(A_1A_2 \dots A_{n-1}) = P(A_1A_2 \dots A_{n-2})P(A_{n-1}|A_1A_2 \dots A_{n-2}).$$

Тогда

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

Продолжая эту процедуру, получаем формулу умножения вероятностей.

Пример 2.3. 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

$$P(A) = P(A_1 A_2 A_3) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 A_2).$$

Согласно классическому определению 1.11 вероятности, имеем $P(A_1) = \frac{1}{7}$.

Если событие A_1 произошло, то на шести оставшихся карточках буква “О” встречается два раза, поэтому условная вероятность

$$P(A_2 | A_1) = \frac{2}{6} = \frac{1}{3}.$$

Аналогично определяем

$$P(A_3 | A_1 A_2) = \frac{1}{5}.$$

Окончательно получаем

$$P(A) = P(A_1 A_2 A_3) = \frac{1}{7} \cdot \frac{1}{3} \cdot \frac{1}{5} = \frac{1}{105} \approx 0,0095.$$

Независимые и зависимые события

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

Определение 2.2. События A и B , имеющие ненулевую вероятность, называют **независимыми**, если условная вероятность A при условии B совпадает с безусловной вероятностью A или если условная вероятность B при условии A совпадает с безусловной вероятностью B , т.е.

$$P(A|B) = P(A) \tag{2.1}$$

или

$$P(B|A) = P(B), \tag{2.2}$$

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

Теорема 2.3. События A и B , имеющие ненулевую вероятность, являются независимыми тогда и только тогда, когда

$$P(AB) = P(A)P(B). \tag{2.3}$$

ДОКАЗАТЕЛЬСТВО. Пусть выполнено равенство (2.2). Воспользовавшись формулой умножения вероятностей для двух событий, получим

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

К аналогичному выводу приходим и в случае выполнения равенства (2.1), т.е. из условия независимости событий следует (2.3).

Обратно, пусть выполнено равенство (2.3). Тогда, согласно определению 2.1 условной вероятности,

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

т.е. в силу определения 2.2 события A и B независимы.

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

Определение 2.3. События A и B называют независимыми, если выполняется равенство (2.3).

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

Замечание 2.1. Из теоремы 2.3 следует, что если в определении 2.2 независимости выполняется одно из равенств (2.1) или (2.2), то выполняется автоматически и другое, т.е. в определении 2.2 достаточно потребовать выполнения любого одного из них.

Пример 2.4. Из колоды карт, содержащей $n = 36$ карт, наугад извлекают одну карту. Обозначим через A событие, соответствующее тому, что извлеченная карта будет пиковой масти, а B — событие, соответствующее появлению “дамы”. Определим, являются ли зависимыми события A и B .

После вычислений получаем

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

т.е. выполняется равенство (2.1), и поэтому события A и B независимы. Отметим, что в соответствии с замечанием 2.1, имеет место и равенство (2.2)

$$P(A|B) = \frac{1/36}{4/36} = \frac{1}{4} = P(A).$$

Следовательно, события A и B независимы. #

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

Теорема 2.4. Если события A и B независимы, то независимыми также являются пары событий \bar{A} и B , A и \bar{B} , \bar{A} и \bar{B} , если вероятности соответствующих событий ненулевые.

ДОКАЗАТЕЛЬСТВО. В силу теоремы 2.1 и независимости событий A и B имеем: $P(\bar{A}|B) = 1 - P(A|B) = 1 - P(A) = P(\bar{A})$, что означает независимость событий \bar{A} и B . Независимость остальных пар событий можно доказать аналогично.

Определение 2.4. События A_1, A_2, \dots, A_n называют **независимыми в совокупности**, если вероятность пересечения любых двух различных событий равна произведению вероятностей этих событий; вероятность пересечения любых трех событий равна произведению их вероятностей; ...; вероятность пересечения всех событий равна произведению их вероятностей.

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

Теорема 2.5. Если события A_1, A_2, \dots, A_n независимы в совокупности, то и события $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ независимы в совокупности. #

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

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

Замечание 2.2. В силу определения независимости событий в совокупности *формула умножения вероятностей* для независимых в совокупности событий имеет вид $P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n)$. #

Из независимости событий с ненулевыми вероятностями в совокупности, согласно теореме 2.3, следует их попарная независимость. Однако из попарной независимости, вообще говоря, независимость в совокупности не следует, что демонстрирует следующий пример.

Пример 2.5. Опыт состоит в однократном подбрасывании тетраэдра, грани которого “про- нумерованы” следующим образом: на трех гранях стоят цифры 1, 2 и 3 соответственно (одна цифра на каждой из них), а на четвертой присутствуют все цифры 1, 2 и 3.

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

Согласно классическому определению вероятности, получаем

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

Аналогично

$$P(A_i|A_j) = \frac{1}{2}$$

при любых $i, j = \overline{1,3}$, $i \neq j$, т.е. события A_1, A_2 и A_3 являются попарно независимыми. Однако, например,

$$P(A_1|A_2 A_3) = \frac{P(A_1 A_2 A_3)}{P(A_2 A_3)} = \frac{1/4}{1/4} = 1 \neq P(A_1),$$

т. е. события A_1, A_2 и A_3 зависимы в совокупности. #

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

Запишем формулу для вероятности *объединения независимых событий*. Пусть

$$A = A_1 \cup \dots \cup A_n.$$

Тогда в соответствии с *законом де Моргана*

$$\bar{A} = \bar{A}_1 \dots \bar{A}_n.$$

Если события A_1, \dots, A_n независимые, то, согласно теореме 2.5, события $\bar{A}_1, \dots, \bar{A}_n$ также независимые и, значит,

$$P(\bar{A}) = P(\bar{A}_1) \dots P(\bar{A}_n).$$

Отсюда окончательно получаем *формулу для вероятности объединения независимых событий*:

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= 1 - P(\overline{A_1 \cup \dots \cup A_n}) = 1 - P(\bar{A}_1 \cap \dots \cap \bar{A}_n) = \\ &= 1 - P(\bar{A}_1) \dots P(\bar{A}_n) = 1 - [1 - P(A_1)] \dots [1 - P(A_n)]. \end{aligned}$$

Замечание 2.3 (о связи между совместными и зависимыми событиями). Между понятиями “несовместные” и “независимые” события имеется следующая связь:

1) если A и B — несовместные события (и $P(A) \neq 0$, и $P(B) \neq 0$), то они обязательно зависимые;

2) если A и B — совместные события, то они могут быть и зависимыми и независимыми;

3) если A и B — зависимые события, то они могут быть и совместными и несовместными. #

2.2 Формула полной вероятности

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

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

Определение 2.5. События H_1, H_2, \dots, H_n удовлетворяющие условиям 1 и 2, называют *гипотезами*.

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

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

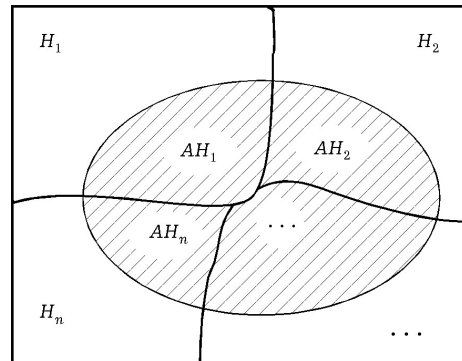


Рис 2.2

Теорема 2.6. Пусть для некоторого события A и гипотез H_1, \dots, H_n известны $P(H_1), \dots, P(H_n)$, которые положительны, и $P(A|H_1), \dots, P(A|H_n)$. Тогда безусловную вероятность $P(A)$ определяют по формуле

$$P(A) = P(H_1)P(A|H_1) + \dots + P(H_n)P(A|H_n), \quad (2.4)$$

которую называют *формулой полной вероятности*.

ДОКАЗАТЕЛЬСТВО. Представим событие A в виде

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

(на рис. 2.2 область, соответствующая событию A , заштрихована). С учетом того, что события AH_i , $i = \overline{1, n}$, несовместны, имеем

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

В соответствии с *формулой умножения вероятностей* получаем

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

Поэтому

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

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

Пример 2.6. Путник должен попасть из пункта B в пункт A в соответствии со схемой дорог изображенной на рис. 2.3. Выбор любой дороги в любом пункте равновозможен. Найдём вероятность события A — достижения путником намеченной цели.

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

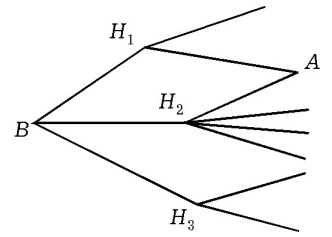


Рис 2.3

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

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

Согласно формуле 2.4 полной вероятности, получаем

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

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

2.3 Формула Байеса

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

Теорема 2.7. Пусть для некоторого события А, $P(A) > 0$, и гипотез H_1, \dots, H_n известны $P(H_1), \dots, P(H_n)$ ($P(H_i) > 0, i = \overline{1, n}$) и $P(A|H_1), \dots, P(A|H_n)$. Тогда условная вероятность $P(H_i|A)$, $i = \overline{1, n}$, гипотезы H_i при условии события А определяется формулой Байеса

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

ДОКАЗАТЕЛЬСТВО. Согласно определению 2.1 условной вероятности, $P(H_i|A) = \frac{P(AH_i)}{P(A)}$. Выражая теперь по формуле умножения вероятностей $P(AH_i)$ через $P(A|H_i)$ и $P(H_i)$, получаем $P(AH_i) = P(H_i)P(A|H_i)$. Поэтому $P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}$.

Подставляя вместо вероятности $P(A)$ ее значение, вычисленное в соответствии с формулой (2.4) полной вероятности, приходим к утверждению теоремы.

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

Пример 2.7. 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. Естественнo ввести следующие гипотезы: H_1 — имеет место заболевание 1; H_2 — имеет место заболевание 2. Из условий задачи ясно, что априорные вероятности гипотез равны: $P(H_1) = 0,4$ и $P(H_2) = 0,6$, а условные вероятности события A при наличии гипотез H_1 и H_2 равны 0,9 и 0,2 соответственно. Используя формулу Байеса, находим

$$P(H_i|A) = \frac{0,4 \cdot 0,9}{0,4 \cdot 0,9 + 0,6 \cdot 0,2} = 0,75.$$

Итак, врач с большей уверенностью признает наличие заболевания 1.

2.4 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 2.6. *Схемой Бернулли (или последовательностью независимых одинаковых испытаний, или биномиальной схемой испытаний) называют последовательность испытаний, удовлетворяющую следующим условиям:*

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

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

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

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

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

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

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

Теорема 2.8. *Вероятность $P_n(k)$ того, что в n испытаниях по схеме Бернулли произойдет ровно k успехов, определяется формулой Бернулли*

$$P_n(k) = C_n^k p^k q^{n-k}, \quad k = \overline{0, n}. \quad (2.5)$$

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

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

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

Так как A_k есть объединение (сумма) всех указанных элементарных исходов, то окончательно получаем для вероятности $P(A_k) = P_n(k)$ формулу (2.5).

Формулу (2.5) называют также **биномиальной**, так как ее правая часть представляет собой $(k+1)$ -й член формулы бинома Ньютона.

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

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

Из формулы Бернулли вытекают два следствия.

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

$$P\{k_1 \leq k \leq k_2\} = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}. \quad (2.6)$$

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

$$P\{k \geq 1\} = 1 - q^n.$$

Пример 2.8. Монету (симметричную) подбрасывают $n = 10$ раз. Определим вероятность выпадения “герба”: а) ровно пять раз; б) не более пяти раз; в) хотя бы один раз.

В соответствии с формулой (2.5) Бернулли имеем:

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

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

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

Пример 2.9. Вероятность выигрыша на один лотерейный билет равна 0,01. Определим, сколько билетов нужно купить, чтобы вероятность хотя бы одного выигрыша в лотерее была не менее заданного значения $P_3 = 0,9$.

Пусть куплено n билетов. Предположим, что общее число билетов, разыгрывающихся в лотерее велико (во много раз больше купленных билетов). При этом можно считать, что каждый билет выигрывает независимо от остальных с вероятностью $p = 0,01$. Тогда вероятность получить k выигрышных билетов можно определить, используя формулу Бернулли. В частности, согласно (2.8), имеем при $q = 1 - p$:

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

откуда получаем

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

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

Лекция 3

Одномерные случайные величины

3.1 Определение случайной величины

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

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

Рассмотрим примеры.

Пример 3.1. В опыте с однократным бросанием игральной кости случайной величиной является число ξ выпавших очков. Множество возможных значений случайной величины ξ имеет вид

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

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

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

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

Пример 3.2. Монету подбрасывают до первого появления “герба”. В этом опыте можно ввести, например, такие случайные величины: ξ — число бросаний до первого появления “герба” с множеством возможных значений $\{1; 2; 3; \dots\}$ и η — число “цифр”, выпавших до первого появления “герба”, с множеством возможных значений $\{0; 1; 2; \dots\}$ (ясно, что $\xi = \eta + 1$). В данном опыте пространство элементарных исходов Ω можно отождествить с множеством

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

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

Пример 3.3. На плоский экран падает частица. Будем считать, что нам известна вероятность попадания частицы в любое (измеримое, т.е. имеющее площадь) множество на экране. Случайными величинами в данном случае будут, например, расстояние ξ от центра экрана до точки падения, квадрат этого расстояния $\eta = \xi^2$, угол ζ в полярной системе координат и т.д. #

Определение 3.1. Случайной величиной ξ , называют любую функцию из Ω в \mathbb{R} , для которой множество $\{\omega \in \Omega : \xi(\omega) < x\}$ — является событием (т.е. принадлежит σ -алгебре событий \mathcal{A} для любого $x \in \mathbb{R}$).

3.2 Функция распределения случайной величины

Определение 3.1 гарантирует, что при любом x неравенство $\xi < x$ есть событие и, следовательно, имеет смысл говорить о его вероятности.

Определение 3.2. *Функцией распределения (вероятностей)* случайной величины ξ называют функцию $F(x)$, значение которой в точке x равно вероятности события $\{\xi < x\}$, т.е. события, состоящего из тех и только тех элементарных исходов ω , для которых $\xi(\omega) < x$:

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

Знание функции распределения вероятностей случайной величины ξ позволяет вычислять события вида $\{\xi < x\}$, $x \in \mathbb{R}$, что вполне достаточно для практических нужд.

Теорема 3.1. *Функция распределения удовлетворяет следующим свойствам.*

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

ДОКАЗАТЕЛЬСТВО. При доказательстве будем использовать свойства вероятностей событий, доказанные в теореме 1.1.

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

Если $x < y$, то событие $\{\xi < x\}$ включено в событие $\{\xi < y\}$ и, согласно свойству 3,

$$\mathbf{P}\{\xi < x\} \leq \mathbf{P}\{\xi < y\},$$

т.е. в соответствии с определением 3.2 выполнено утверждение 2.

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

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

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

Замечание 3.1. Можно показать, что любая неубывающая непрерывная слева функция $F(x)$, удовлетворяющая условиям $F(-\infty) = 0$ и $F(+\infty) = 1$, является функцией распределения некоторой случайной величины ξ . #

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

Определение 3.3. Для произвольной случайной величины ξ отображение, которое ставит в соответствие множествам $B \in \mathbb{R}$ вероятность события $\{\xi \in B\}$, т.е. число $\mathbf{P}\{\xi \in B\}$, называют *законом распределения вероятностей*, или *распределением (вероятностей)* случайной величины ξ .

Замечание 3.2. Можно показать, что в определении 3.3 в качестве множеств B из \mathbb{R} достаточно взять всевозможные полуинтервалы $(-\infty, x]$, $x \in \mathbb{R}$. Отсюда следует, закон распределения случайной величины однозначно определяется ее функцией распределения.

3.3 Дискретные случайные величины

Определение 3.4. Случайную величину ξ называют *дискретной*, если множество ее возможных значений конечно или счетно.

Распределение дискретной случайной величины удобно описывать с помощью ряда распределения.

Определение 3.5. *Рядом распределения (вероятностей) дискретной случайной величины* ξ называют таблицу (табл. 3.1), состоящую из двух строк: в верхней строке перечислены все возможные значения случайной величины, а в нижней — *вероятности* $p_n = P\{\xi = x_n\}$, $n = 1, 2, \dots$ того, что случайная величина примет эти значения.

Для проверки правильности составления табл. 3.1 рекомендуется просуммировать вероятности p_n , $n = 1, 2, \dots$. В силу *аксиомы нормированности* эта сумма должна быть равна единице: $\sum_{n=1}^{\infty} p_n = 1$.

ξ	x_1	x_2	\dots	x_n	\dots
P	p_1	p_2	\dots	p_n	\dots

Таблица 3.1.

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 $x_1, x_2, \dots, x_n, \dots$ расположены в порядке возрастания. Тогда для всех $x \leq x_1$ событие $\{\xi < x\}$ является невозможным и поэтому в соответствии с определением 3.2 $F(x) = 0$. Если $x_1 < x \leq x_2$, то событие $\{\xi < x\}$ состоит из тех и только тех элементарных исходов ω , для которых $\xi(\omega) = x_1$, и, следовательно, $F(x) = p_1$.

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

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

а следовательно, $F(x) = p_1 + p_2$ и т.д. Аналогично, при $x_n < x \leq x_{n+1}$ имеем $\{\xi < x\} = \bigcup_{i=1}^n \{\xi = x_i\}$ и $F(x) = p_1 + \dots + p_n$.

Таким образом, функция распределения дискретной случайной величины является кусочно-постоянной функцией, принимающей на промежутке $(-\infty, x_1]$ значение 0, а на промежутках $(x_n, x_{n+1}]$ — значение $p_1 + \dots + p_n$, $n = 1, 2, \dots$.

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 3.1) описывают формулой $P\{\xi = i\} = \frac{1}{6}, i = \overline{1, 6}$.

3.4 Непрерывные случайные величины

Определение 3.6. *Непрерывной* называют *случайную величину* ξ , *функцию распределения* которой $F(x)$ можно представить в виде

$$F(x) = \int_{-\infty}^x f(y) dy. \quad (3.1)$$

Функцию $f(x)$ называют *плотностью распределения (вероятностей)* случайной величины ξ .

Предполагают, что несобственный интеграл в представлении (3.1) сходится.

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

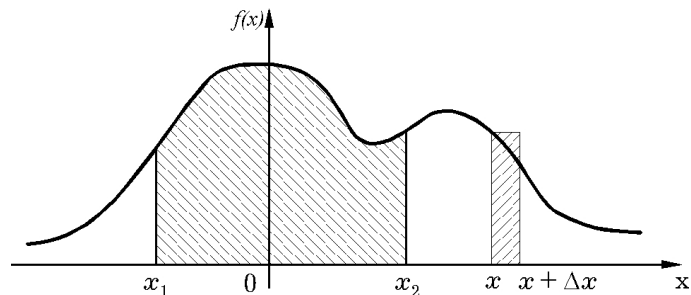


Рис 3.1

$$f(x) = F'(x), \quad (3.2)$$

что следует из свойств интеграла с переменным верхним пределом. Только такие случайные величины мы и будем рассматривать в дальнейшем. На рис. 3.1 изображен типичный вид плотности распределения.

Теорема 3.2. Плотность распределения обладает следующими свойствами.

1) $f(x) \geq 0$.

2) $P\{a \leq \xi < b\} = P\{\xi \in [a, b)\} = \int_a^b f(x) dx$.

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

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

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

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

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

$$F(b) - F(a) = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = \int_a^b f(x) dx,$$

что и доказывает утверждение 2.

В частности, если $a = -\infty$, $b = +\infty$, то событие $\{-\infty < \xi < \infty\}$ является достоверным, и поэтому справедливо утверждение 3.

Согласно свойству 4 (см. теорему 3.1),

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

Если Δx “малó” (см. рис. 3.1), то имеем

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

что и доказывает утверждение 4.

Наконец, поскольку в силу определения 3.6 функция распределения случайной величины есть несобственный интеграл от плотности, то она является непрерывной, что приводит нас к утверждению 5.

Замечание 3.3. В силу свойства 2 плотности распределения вероятность попадания непрерывной случайной величины в промежуток $[x, y)$ численно равна площади криволинейной трапеции, заштрихованной на рис. 3.1.

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

В соответствии со свойством 4 вероятность попадания случайной величины ξ в некоторый “малый” промежуток $(x, x + \Delta x)$ практически пропорциональна Δx с коэффициентом пропорциональности, равным значению плотности распределения в точке x . Поэтому

выражение $f(x)\Delta x$ или $f(x)dx$ называют иногда **элементом вероятности**. Можно также сказать, что непрерывная случайная величина реализует *геометрическую схему* с коэффициентом пропорциональности $f(x)$, но только в “малой” *окрестности* точки x .

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

Замечание 3.4. Можно показать, что любая неотрицательная функция $f(x)$, удовлетворяющая условию $\int_{-\infty}^{\infty} f(x) dx = 1$, является плотностью распределения вероятностей некоторой случайной величины ξ . #

3.5 Функции от случайной величины

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

Определение 3.7. Случайную величину η , которая каждому *элементарному исходу* ω ставит в соответствие число $\eta(\omega) = \varphi(\xi(\omega))$, называют **функцией $\varphi(\xi)$ (скалярной) от скалярной случайной величины ξ** .

Функция $\eta = \varphi(\xi)$ от *дискретной случайной величины* также является дискретной случайной величиной, поскольку она не может принимать больше значений, чем случайная величина ξ . Очевидно, что если случайная величина ξ имеет *ряд распределения*, представленный в табл. 3.2, то ряд распределения случайной величины $\eta = \varphi(\xi)$ определяется табл. 3.3.

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

Таблица 3.2.

При этом, если в верхней строке табл. 3.3 появляются одинаковые значения $\varphi(x_i)$, соответствующие столбцы нужно объединить в один, приписав им суммарную *вероятность*.

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

Таблица 3.3.

Функция $\eta = \varphi(\xi)$ от *непрерывной случайной величины* ξ может быть как непрерывной, так и дискретной (если, например, множество значений функции $\varphi(\xi)$ конечное или счетное).

Теорема 3.3. Пусть случайная величина ξ имеет плотность $f_\xi(x)$. Пусть функция $y = \varphi(x)$ является монотонной и дифференцируемой функцией. Обозначим $x = \psi(y)$ функцию, обратную к $y = \varphi(x)$. Тогда плотность случайной величины $\eta = \varphi(\xi)$ есть

$$f_\eta(y) = f_\xi(\psi(y)) |\psi'(y)|. \quad (3.3)$$

ДОКАЗАТЕЛЬСТВО. Обозначим через F_ξ и F_η функции распределения случайных величин ξ и η соответственно. Если функция $\varphi(x)$ является монотонной, то событие $\{\varphi(\xi(\omega)) < y\}$ эквивалентно событию $\{\xi(\omega) < \psi(y)\}$ (в случае возрастающей функции $\varphi(x)$) или событию $\{\xi(\omega) > \psi(y)\}$ (в случае убывающей $\varphi(x)$). Значит, для возрастающей функции $\varphi(x)$

$$\mathbf{P}\{\varphi(\xi) < y\} = \mathbf{P}\{\xi < \psi(y)\},$$

для убывающей $\varphi(x)$

$$\mathbf{P}\{\varphi(\xi) < y\} = \mathbf{P}\{\xi > \psi(y)\}.$$

Поскольку

$$F_\eta(y) = \mathbf{P}\{\eta < y\},$$

а

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

то окончательно получаем:

для возрастающей функции $\varphi(x)$

$$F_{\eta}(y) = F_{\xi}(\psi(y)); \quad (3.4)$$

для убывающей функции $\varphi(x)$

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

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

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

в случае убывающей функции $\eta(x)$

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

Оба эти случая можно записать в виде (3.3). #

Если $y = \varphi(x)$ — функция немонотонная, то обратная функция неоднозначна. Обозначим через $\psi_1(y), \dots, \psi_k(y)$, все значения (при данном y) обратной функции. Другими словами $\psi_i(y)$, $i = \overline{1, k}$ — все прообразы точки y при отображении $y = \varphi(x)$.

Теорема 3.4. Пусть случайная величина ξ имеет плотность $f_{\xi}(x)$. Пусть функция $y = \varphi(x)$ является кусочно монотонной функцией. Обозначим $x_i = \psi_i(y)$, $i = \overline{1, k}$, прообразы точки y при отображении $y = \varphi(x)$. Если функции $\psi_i(y)$, $i = \overline{1, k}$, дифференцируемы, то плотность случайной величины $\eta = \varphi(\xi)$ есть

$$f_{\eta}(y) = \sum_{i=1}^k f_{\xi}(\psi_i(y)) |\psi'_i(y)|. \quad (3.5)$$

Пример 3.4. Пусть случайная величина ξ имеет непрерывную функцию распределения $F_{\xi}(x)$, которая является возрастающей функцией. Рассмотрим случайную величину $\eta = F_{\xi}(\xi)$. Она принимает значение только на промежутке $[0, 1]$. Обратная функция для функции $y = F_{\xi}(x)$ есть

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

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

$$F_{\xi}(\psi(y)) = y,$$

и, следовательно, в соответствии с формулой (3.4) имеем для $y \in [0, 1]$:

$$F_{\eta}(y) = F_{\xi}(F_{\xi}^{-1}(y)) = y.$$

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

$$F_{\eta}(y) = \mathbf{P}\{\eta < y\} = 0.$$

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

$$F_{\eta}(y) = \mathbf{P}\{\eta < y\} = 1.$$

Итак,

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

Таким образом, случайная величина η имеет равномерное распределение на отрезке $[0, 1]$ (см. определение равномерного распределения на с. 39).

Переходя к обратной функции, видим, что случайная величина

$$\xi = F_{\xi}^{-1}(\eta) \quad (3.6)$$

имеет функцию распределения $F_{\xi}(x)$, если случайная величина η имеет равномерное распределение на отрезке $[0, 1]$.

Полученный результат широко применяют при моделировании случайных величин с заданной функцией распределения $F_{\xi}(x)$. Действительно, если нужно смоделировать такую случайную величину, то достаточно иметь датчик случайных чисел η , распределенных равномерно на отрезке $[0, 1]$, и каждое такое число преобразовать по формуле (3.6).

Например, пусть нужно смоделировать реализацию случайной величины ξ с *экспоненциальной функцией* распределения (см. определение экспоненциального распределения на с. 39)

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

при известном параметре λ . Тогда, учитывая, что

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

реализацию ξ можно получить по формуле

$$\xi = -\frac{1}{\lambda} \ln(1 - \eta),$$

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

Пример 3.5. Пусть случайная величина ξ имеет *стандартное нормальное распределение* (см. определение нормального распределения на с. 40), т.е.

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

Найдем распределение случайной величины $\eta = \xi^2$. Воспользуемся формулой (3.5). В данном случае $\varphi(x) = x^2$. Заметим, что функция $\varphi(x) = x^2$ принимает только неотрицательные значения и, следовательно, при $y < 0$ уравнение $\varphi(x) = y$ не имеет решений. Поэтому случайная величина $\eta = \eta(\xi) = \xi^2$ не может принимать отрицательные значения и, следовательно функция распределения $F_{\eta}(y) = 0$ при $y < 0$. Отсюда дифференцированием получаем $f_{\eta}(y) = F'_{\eta}(y) = 0$, $y < 0$.

Далее, при $y > 0$ уравнение $y = x^2$ имеет два решения: $x = \psi_1(y) = -\sqrt{y}$ и $x = \psi_2(y) = \sqrt{y}$, причем первое решение принадлежит интервалу $(-\infty, 0)$ убывания функции $\varphi(x) = x^2$, а второе — интервалу $(0, +\infty)$ возрастания этой функции. Подставляя $\psi_1(y)$, $\psi_2(y)$ и *плотность стандартного нормального распределения* в формулу (3.5), получаем

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

Лекция 4

Числовые характеристики случайных величин

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

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

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

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

4.1 Математическое ожидание случайной величины

Определение 4.1. *Математическим ожиданием (средним значением) $M\xi$ дискретной случайной величины ξ называют сумму произведений значений x_n случайной величины и вероятностей $p_n = P\{\xi = x_n\}$, с которыми случайная величина принимает эти значения:*

$$M\xi = \sum_{n=1}^{\infty} x_n p_n.$$

При этом предполагается, что

$$\sum_{n=1}^{\infty} |x_n| p_n < +\infty,$$

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

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

$$x_{\text{ц}} = \frac{\sum_{n=1}^{\infty} x_n p_n}{\sum_{n=1}^{\infty} p_n} = \frac{\sum_{n=1}^{\infty} x_n p_n}{1} = \sum_{n=1}^{\infty} x_n p_n,$$

совпадающую с математическим ожиданием $M\xi$ случайной величины ξ .

Пример 4.1. Пусть ξ — число выпавших очков при подбрасывании игральной кости. Так как $p_n = P\{\xi = n\} = \frac{1}{6}$, $n = \overline{1,6}$, то

$$M\xi = \sum_{n=1}^6 n \frac{1}{6} = 7/2 = 3,5.$$

Определение 4.2. Математическим ожиданием (средним значением) $M\xi$ непрерывной случайной величины называют интеграл

$$M\xi = \int_{-\infty}^{+\infty} xf(x) dx.$$

При этом предполагается, что

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

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

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

Пример 4.2. Пусть случайная величина ξ имеет *распределение Коши* с параметрами $m \in \mathbb{R}$ и $\gamma > 0$, т. е. плотность случайной величины ξ равна

$$f(x) = \frac{\gamma}{\pi((x-m)^2 + \gamma^2)}.$$

Тогда

$$\int_{-\infty}^{+\infty} \frac{\gamma|x| dx}{\pi((x-m)^2 + \gamma^2)} = +\infty,$$

поскольку подынтегральная функция эквивалентна $1/(\pi x)$ при $x \rightarrow +\infty$. Поэтому математическое ожидание случайной величины ξ не существует.

Математическое ожидание функции от случайной величины.

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

Рассмотрим сначала *дискретную случайную величину* ξ , принимающую значения x_1, \dots, x_n, \dots с вероятностями $p_n = P\{\xi = x_n\}$, $n = 1, 2, \dots$. Тогда случайная величина $\eta = \varphi(\xi)$ принимает значения $\varphi(x_n)$ с вероятностями $p_n = P\{\xi = x_n\}$, $n = 1, 2, \dots$ (см. табл. 3.3), и ее математическое ожидание определяется формулой

$$M\eta = M\varphi(\xi) = \sum_{n=1}^{\infty} \varphi(x_n)p_n, \quad (4.1)$$

при этом требуется выполнение условия

$$\sum_{n=1}^{\infty} |\varphi(x_n)|p_n < +\infty.$$

Для *непрерывной случайной величины* ξ , имеющей *плотность распределения* $f(x)$, математическое ожидание случайной величины $\eta = \varphi(\xi)$ можно найти, используя аналогичную (4.1) формулу

$$M\eta = M\varphi(\xi) = \int_{-\infty}^{+\infty} \varphi(x)f(x) dx, \quad (4.2)$$

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

4.2 Дисперсия. Моменты высших порядков

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

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

Определение 4.3. *Дисперсией* $D\xi$ случайной величины ξ называют математическое ожидание квадрата отклонения случайной величины ξ от ее среднего значения, т. е. $D\xi = M(\xi - M\xi)^2$.

Используя формулы (4.1) и (4.2), в которых положено $\eta(x) = (x - M\xi)^2$, легко написать расчетные формулы для дисперсий дискретной и непрерывной случайных величин. А именно, если дискретная случайная величина ξ принимает значения x_1, \dots, x_n, \dots с вероятностями $p_n = P\{\xi = x_n\}$, $n = 1, 2, \dots$, то

$$D\xi = \sum_{n=1}^{\infty} (x_n - M\xi)^2 p_n, \quad (4.3)$$

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

$$D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 f(x) dx.$$

Пример 4.3. Пусть ξ — число выпавших очков при подбрасывании игральной кости. Так как $M\xi = 3,5$, то

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

Замечание 4.1. Из определения непосредственно следует, что дисперсия любой случайной величины является неотрицательным числом.

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 4.4. *Начальным моментом k -го порядка μ_k (k -м начальным моментом)* случайной величины ξ называют число $M\xi^k$ — математическое ожидание k -й степени случайной величины ξ .

Определение 4.5. *Центральным моментом k -го порядка ν_k (k -м центральным моментом)* случайной величины ξ называют число $M((\xi - M\xi)^k)$ — математическое ожидание k -й степени случайной величины $\xi - M\xi$.

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

$$\mu_k = \sum_{n=1}^{\infty} x_n^k p_n, \quad \nu_k = \sum_{n=1}^{\infty} (x_n - M\xi)^k p_n,$$

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

$$\mu_k = \int_{-\infty}^{+\infty} x^k f(x) dx, \quad \nu_k = \int_{-\infty}^{+\infty} (x - M\xi)^k f(x) dx.$$

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

4.3 Квантиль

Определение 4.6. Пусть ξ — случайная величина с непрерывной функцией распределения $F(x)$. **Квантилью уровня p** , или **p -квантилью**, $0 < p < 1$, случайной величины ξ (распределения случайной величины ξ) называют число Q_p , удовлетворяющее равенству

$$F(Q_p) = p. \quad (4.4)$$

Замечание 4.2. Учитывая связь (3.1) между функцией распределения $F(x)$ и плотностью распределения $f(x)$ случайной величины ξ , уравнение для определения p -квантили можно записать в виде

$$\int_{-\infty}^{Q_p} f(x) dx = p, \quad (4.5)$$

а если Q_a и Q_b — квантили уровней a и b соответственно, то

$$\int_{Q_a}^{Q_b} f(x) dx = \int_{-\infty}^{Q_b} f(x) dx - \int_{-\infty}^{Q_a} f(x) dx = b - a. \quad \# \quad (4.6)$$

Если $F(x)$ непрерывная функция, то равенство (4.6) определяет квантиль однозначно. Если же функция распределения $F(x)$ случайной величины ξ является разрывной, то для некоторых p уравнение $F(x) = p$ не имеет решения относительно x . В этом случае существует определение квантили более общее, чем определение 4.6.

Определение 4.7. Пусть ξ — случайная величина с (произвольной) функцией распределения $F(x)$. **Квантилью уровня p** , или **p -квантилью**, $0 < p < 1$, случайной величины ξ (распределения случайной величины ξ) называют максимальное значение x , при котором $F(x)$ не больше p , т.е.

$$Q_p = \max\{x : F(x) \leq p\}. \quad \#$$

Нетрудно заметить, что определение 4.6 является частным случаем определения 4.7.

Определение 4.8. Квантиль уровня $1/2$ называют **медианой**.

Пример 4.4. Найдём p -квантиль и медиану экспоненциального распределения. В этом случае Q_p представляет собой решение уравнения

$$1 - e^{-\lambda Q_p} = p.$$

Поэтому

$$Q_p = -\frac{\ln(1-p)}{\lambda}.$$

Медиана экспоненциального распределения равна $\frac{\ln 2}{\lambda}$.

Пример 4.5. Пусть случайная величина ξ представляет собой число успехов в одном испытании по схеме Бернулли с вероятностью успеха p , $q = 1 - p$. Тогда, как видно из графика $F(x)$ функции распределения случайной величины ξ (см. рис. 4.1)

$$Q_\alpha = \begin{cases} 0, & \alpha < q, \\ q, & \alpha \geq q. \end{cases}$$

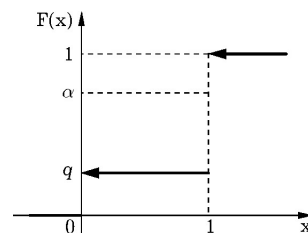


Рис 4.1.

Замечание 4.3. Таблицы квантилей основных вероятностных распределений имеются в специальных справочниках и в большинстве учебников по теории вероятностей и математической статистике. Кроме того, их можно вычислить с помощью различных математических и статистических пакетов, в частности, Matlab, Maple, Excel. Для операционной системы Android квантили вычисляет, например, программа «Probability Distributions» из Google Play.

Лекция 5

Основные законы распределения случайных величин

5.1 Биномиальное распределение

Определение 5.1. Дискретную случайную величину 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

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

или, что тоже самое, рядом распределения, представленным в таблице 5.1.

Говорят также, что эта случайная величина *распределена по биномиальному закону*, или имеет *биномиальное распределение*.

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

Таблица 5.1.

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

$$\sum_{k=0}^m C_m^k p^k q^{m-k} = (p+q)^m, \quad m = 1, 2, \dots, \quad (5.1)$$

известному как бином Ньютона,

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

Замечание 5.1. Биномиальная случайная величина — число успехов в n испытаниях Бернулли с вероятностью успеха p и неудачи $q = 1 - p$ (см. определение 2.6).

Определение 5.2. 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 словами, бернуллиевская случайная величина — биномиальная случайная величина с параметрами $p \in (0,1)$ и $n = 1$.

Пример 5.1. Найдём математическое ожидание и дисперсию бернуллиевской случайной величины ξ с параметром p . Воспользовавшись определениями 5.2 и 4.1, получим

$$M\xi = 0 \cdot (1 - p) + 1 \cdot p = p, \quad (5.2)$$

а согласно формуле (4.3)

$$\begin{aligned} D\xi &= (0 - M\xi)^2 \cdot (1 - p) + (1 - M\xi)^2 \cdot p = \\ &= (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p^2 q + q^2 \cdot p = pq. \end{aligned} \quad (5.3)$$

Пример 5.2. Найдем математическое ожидание биномиальной случайной величины ξ с параметрами n и p . Дважды воспользовавшись равенством (5.1), получим

$$\begin{aligned} M\xi &= \sum_{k=0}^n k C_n^k p^k q^{n-k} = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} = \\ &= \sum_{k=1}^n np \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} = np \sum_{j=0}^{n-1} C_{n-1}^j p^j q^{n-1-j} = np f(p+q)^{n-1} = np. \end{aligned}$$

Более легкий способ вычисления математического ожидания основан на утверждении 3 теоремы 7.1, 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 $\xi = \xi_1 + \dots + \xi_n$, где случайное число ξ_k успехов в k -ом испытании имеет распределение Бернулли, $k = \overline{0, n}$. Поэтому согласно (5.2)

$$M\xi = M\xi_1 + \dots + M\xi_n = \underbrace{p + p + \dots + p}_{n \text{ слагаемых}} = np.$$

Пример 5.3. Пусть ξ — число успехов в n испытаниях по схеме Бернулли. Дисперсию ξ можно подсчитать по формуле (4.3), непосредственно воспользовавшись определением 4.3 дисперсии. Однако мы поступим другим образом. Для этого представим ξ в виде суммы $\xi = \xi_1 + \dots + \xi_n$, где ξ_k — число успехов в k -ом испытании. Дисперсия каждого слагаемого равна (см. (5.3)) pq . Позднее в теореме 7.2 будет доказано, что дисперсия суммы независимых случайных величин равна сумме дисперсий каждого слагаемого. Учитывая, что случайные величины ξ_k являются независимыми, получаем

$$D\xi = D\xi_1 + \dots + D\xi_n = npq.$$

5.2 Распределение Пуассона

Определение 5.3. Дискретную случайную величину назовем *пуассоновской* с параметром $\lambda > 0$, если она принимает значения $0, 1, 2, \dots, n, \dots$ с вероятностями, заданными формулой

$$P\{\xi = n\} = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, \dots,$$

или, что тоже самое, рядом распределения, представленным в таблице 5.2. Говорят также, что она *распределена по пуассоновскому закону*, или имеет *пуассоновское распределение*.

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

Таблица 5.2.

Проверка корректности определения распределения Пуассона дает:

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

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 5.4. Пусть случайная величина ξ имеет *распределение Пуассона*. Тогда

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

Пример 5.5. Найдем дисперсию случайной величины ξ , распределенной по закону Пуассона. Математическое ожидание $M\xi = \lambda$ было найдено в примере 5.4. Определим второй момент:

$$\begin{aligned} M\xi^2 &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} = \lambda \sum_{n=1}^{\infty} n \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} = \left| k = n-1 \right| = \\ &= \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \left(\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \right) = \lambda (M\xi + 1) = \lambda^2 + \lambda. \end{aligned}$$

Воспользовавшись свойством 3 дисперсии из теоремы 7.2, получим

$$D\xi = \lambda^2 + \lambda - \lambda^2 = \lambda,$$

Таким образом, дисперсия ξ , так же как и математическое ожидание, совпадает с параметром λ .

5.3 Геометрическое распределение

Определение 5.4. Дискретную случайную величину назовем *геометрической* с параметром $p \in (0,1)$, если она принимает значения $1, 2, \dots$ с вероятностями, заданными формулой

$$P\{\xi = n\} = pq^{n-1}, \quad q = 1 - p, \quad n = 1, 2, \dots,$$

или, что тоже самое, рядом распределения, представленным в таблице 5.3. Говорят также, что она *распределена по геометрическому закону*, или имеет *геометрическое распределение*.

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

Таблица 5.3.

Покажем, как геометрическое распределение возникает в схеме Бернулли с вероятностью успеха p . Пусть ξ — число испытаний до наступления первого успеха включительно, или, другими словами, номер испытания, в котором наступил первый успех. Тогда ξ — дискретная случайная величина, принимающая значения $1, 2, \dots, n, \dots$. Определим вероятность события $\{\xi = n\}$. Очевидно, что $\xi = 1$, если в первом же испытании произойдет успех. Поэтому $P\{\xi = 1\} = p$. Далее, $\xi = 2$ в том случае, когда в первом испытании произошла неудача, а во втором — успех. Но вероятность такого события (см. теорему 2.8), равна qp , т.е. $P\{\xi = 2\} = qp$. Аналогично $\xi = 3$, если в первых двух испытаниях произошли неудачи, а в третьем — успех, и, значит, $P\{\xi = 3\} = qq p$. Продолжая эту процедуру, получаем $P\{\xi = n\} = q^{n-1}p$. Таким образом, случайная величина ξ имеет ряд распределения, представленный в таблице 5.3.

Правильность составления таблицы вытекает из равенства

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

Пример 5.6. Найдем математическое ожидание геометрической случайной величины ξ :

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

Можно показать, что

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

5.4 Равномерное распределение

Определение 5.5. Случайную величину назовем *равномерно распределенной* на отрезке $[a, b]$ (или на интервале (a, b)), если ее плотность распределения $f(x)$ и функция распределения $F(x)$ имеют вид

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

Говорят также, что она *распределена по равномерному закону* или имеет *равномерное распределение*. Числа a и b называют параметрами распределения.

Графики плотности распределения $f(x)$ и функции распределения $F(x)$ приведены на рис. 5.1. Вероятность попадания равномерно распределенной случайной величины в интервал (x, y) , лежащий внутри отрезка $[a, b]$, равна

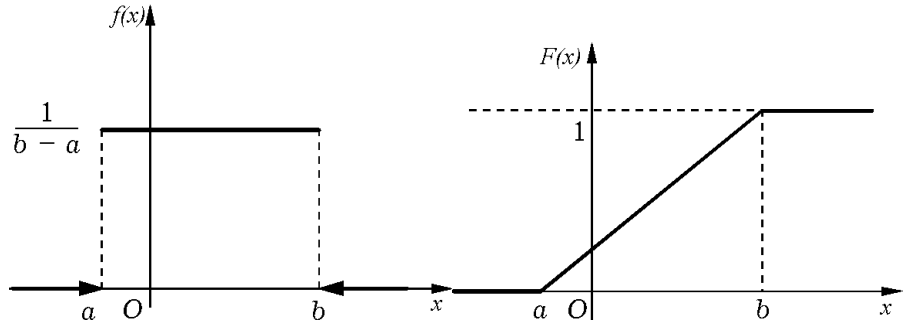


Рис 5.1

$$F(y) - F(x) = (y - x)/(b - a),$$

т.е. пропорциональна длине этого интервала. Таким образом, равномерное распределение реализует *схему геометрической* вероятности при бросании точки на отрезок $[a, b]$.

Пример 5.7. 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

$$M\xi = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_{x=a}^{x=b} = \frac{b+a}{2}. \quad (5.4)$$

Как и следовало ожидать, $M\xi$ совпадает с серединой отрезка $[a, b]$.

Пример 5.8. Дисперсия *равномерно распределенной* на отрезке $[a, b]$ случайной величины ξ определяется формулой

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

5.5 Экспоненциальное распределение

Определение 5.6. Случайную величину назовем *экспоненциальной* (или *показательной*) с параметром $\lambda > 0$, если ее плотность распределения $f(x)$ и функция распределения $F(x)$ имеет вид

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

Говорят также, что она *распределена по экспоненциальному* (или *показательному*) *закону*, или имеет *экспоненциальное* (*показательное*) *распределение*.

Графики плотности распределения и функции распределения экспоненциальной случайной величины приведены на рис. 5.2.

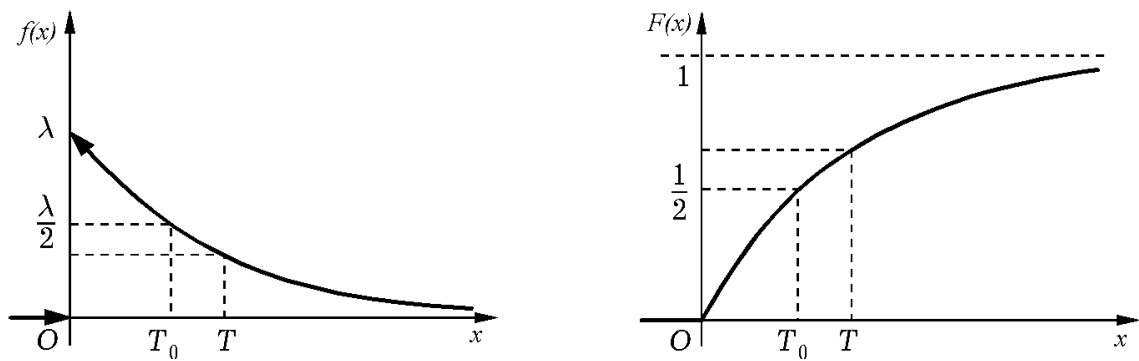


Рис 5.2

Пример 5.9. Найдем математическое ожидание и дисперсию экспоненциально распределенной случайной величины ξ . Применяя формулу интегрирования по частям, получим

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

$$D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 f(x) dx = \int_0^{+\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}.$$

5.6 Нормальное распределение

Определение 5.7. Случайную величину назовем **нормальной** (или **гауссовской**) с параметрами $m \in \mathbb{R}$ и σ^2 ($\sigma > 0$), если ее плотность распределения имеет вид

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

Говорят также, что она **распределена по нормальному** (или **гауссовскому**) **закону**, или имеет **нормальное** (или **гауссовское**) **распределение**.

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

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

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

нормальной случайной величины для различных значений m и σ приведены на рис. 5.3.

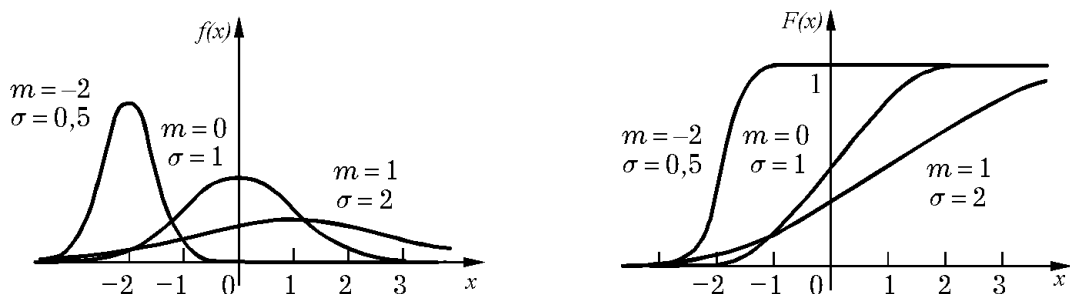


Рис 5.3

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

стандартным и его функцию распределения обозначают $\Phi(x)$, а плотность распределения — $\varphi(x)$.

Таким образом,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx. \quad (5.5)$$

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, в частности, с помощью программы «Probability Distributions» из Google Play.

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

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

называемой функцией Лапласа.

Теорема 5.1. *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 4.6 и формулой (4.4) 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 и выражает через $\Phi(x)$ вероятность попадания нормальной случайной величины в интервал (a, b) .

Теорема 5.2. Пусть ξ — нормальная случайная величина с параметрами m и σ .

Тогда

$$\mathbf{P}\{a < \xi < b\} = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) = \Phi_0\left(\frac{b-m}{\sigma}\right) - \Phi_0\left(\frac{a-m}{\sigma}\right), \quad (5.6)$$

$$\mathbf{M}\xi = m, \quad \mathbf{D}\xi = \sigma^2.$$

ДОКАЗАТЕЛЬСТВО. В соответствии со свойством 2 плотности распределения (см. теорему 3.2) вероятность попадания нормальной случайной величины ξ с параметрами m и σ в интервал (a, b) задается формулой

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

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

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

Поэтому

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

Поскольку $\Phi(x) = \Phi_0(x) + 1/2$, то $\Phi(y) - \Phi(x) = \Phi_0(y) - \Phi_0(x)$ для любых x и y . Таким образом, окончательно получаем 5.6.

Далее

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

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

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

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

По определению дисперсии случайной величины ξ

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

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

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

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

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

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

Лекция 6

Случайные векторы

6.1 Функция распределения случайного вектора

Определение 6.1. Совокупность случайных величин $\xi_1 = \xi_1(\omega), \dots, \xi_n = \xi_n(\omega)$, заданных на одном и том же вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$, называют *многомерной (n-мерной) случайной величиной*, или *n-мерным случайным вектором*. При этом сами случайные величины $\xi_1, \xi_2, \dots, \xi_n$ называют *координатами случайного вектора*.

Пример 6.1. 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 6.2. При испытании прибора на надежность совокупность внешних воздействий в некоторый момент времени можно описать случайным вектором $(\xi, \eta, \zeta, \dots)$. 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*. В частности, закон распределения одномерной случайной величины можно задать с помощью *функции распределения*. То же можно сказать и по отношению к n-мерному случайному вектору. Отметим, что в дальнейшем для *пересечения событий* $\{\xi_1 < x_1\}, \dots, \{\xi_n < x_n\}$ вместо записи

$$\{\xi_1 < x_1\} \cap \dots \cap \{\xi_n < x_n\}$$

будем использовать запись

$$\{\xi_1 < x_1, \dots, \xi_n < x_n\}.$$

Определение 6.2. *Функцией распределения (вероятностей) (n-мерного) случайного вектора* (ξ_1, \dots, ξ_n) называют функцию $F(x_1, \dots, x_n)$, значение которой в точке $(x_1, \dots, x_n) \in \mathbb{R}^n$ равно вероятности совместного осуществления (пересечения) событий $\{\xi_1 < x_1\}, \dots, \{\xi_n < x_n\}$, т.е.

$$F(x_1, \dots, x_n) = \mathbf{P}\{\xi_1 < x_1, \dots, \xi_n < x_n\}.$$

Функцию распределения $F(x_1, \dots, x_n)$ называют также *совместной n-мерной функцией распределения* случайных величин ξ_1, \dots, ξ_n .

6.2 Двумерные случайные векторы

Далее для простоты ограничимся рассмотрением двумерных случайных векторов.

Значение двумерной функции распределения $F(x,y)$ случайного вектора (ξ, η) в точке (a,b) , согласно определению 6.2, представляет собой не что иное, как вероятность попадания точки с координатами (ξ, η) в квадрант с вершиной в точке (a,b) , т.е. в множество $\{(x,y) \in \mathbb{R}^2 : x < a, y < b\}$.

Обозначим через $F_\xi(x)$ и $F_\eta(y)$ функции распределения координат ξ и η случайного вектора (ξ, η) . Справедлива следующая теорема.

Теорема 6.1. Двумерная функция распределения удовлетворяет следующим свойствам.

1. $0 \leq F(x,y) \leq 1$.
2. $F(x,y)$ — неубывающая функция по каждому из аргументов x и y .
3. $F(-\infty, y) = F(x, -\infty) = 0$.
4. $F(+\infty, +\infty) = 1$.
5. $P\{a \leq \xi < b, c \leq \eta < d\} = F(b,d) - F(b,c) - F(a,d) + F(a,c)$.
6. $F(x,y)$ — непрерывная слева в любой точке $(x,y) \in \mathbb{R}^2$ по каждому из аргументов x и y функции.
7. $F(x, +\infty) = F_\xi(x)$, $F(+\infty, y) = F_\eta(y)$.

ДОКАЗАТЕЛЬСТВО. Утверждения 1 и 2 доказываются точно так же, как и в одномерном случае (см. теорему 3.1).

События $\{\xi < -\infty\}$ и $\{\eta < -\infty\}$ являются невозможными, а пересечение невозможного события с любым событием, как известно, также невозможное событие, вероятность которого равна нулю. Отсюда с учетом определения 6.2 вытекает утверждение 3.

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

Чтобы найти вероятность попадания двумерной случайной величины (ξ, η) в прямоугольник $\{a \leq x < b, c \leq y < d\}$ (на рис. 6.1 заштрихован), сначала определим вероятность попадания в полуполосу $\{x < a, c \leq y < d\}$ (отмечена двойной штриховкой). Но эта вероятность представляет собой вероятность попадания в квадрант $\{x < a, y < d\}$ за вычетом вероятности попадания в квадрант $\{x < a, y < c\}$, т.е.

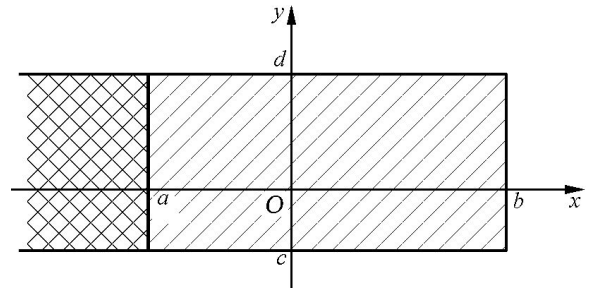


Рис 6.1

$$P\{\xi < a, c \leq \eta < d\} = F(a,d) - F(a,c).$$

Аналогично,

$$P\{\xi < b, c \leq \eta < d\} = F(b,d) - F(b,c).$$

Теперь осталось заметить, что вероятность попадания в прямоугольник $\{a \leq x < b, c \leq y < d\}$ совпадает с вероятностью попадания в полуполосу $\{x < b, c \leq y < d\}$ из которой вычитается вероятность попадания в полуполосу $\{x < a, c \leq y < d\}$, откуда следует утверждение 5.

Подобно одномерному случаю доказывается и утверждение 6.

Наконец, событие $\{\eta < +\infty\}$ является достоверным, поэтому

$$\{\xi < x\} \cap \{\eta < +\infty\} = \{\xi < x\}.$$

Аналогично

$$\{\xi < +\infty\} \cap \{\eta < y\} = \{\eta < y\}.$$

Отсюда приходим к утверждению 7, которое устанавливает естественную связь между двумерной функцией распределения F случайного вектора (ξ, η) и функциями F_ξ и F_η , которые называют *одномерными* (говорят также *частными*, или *маргинальными*) *функциями распределения* случайных величин ξ и η .

6.3 Дискретные двумерные случайные векторы

Определение 6.3. (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 $x_1, \dots, x_i, \dots, \eta$ — значения y_1, \dots, y_j, \dots , а координаты двумерного случайного вектора (ξ, η) — пары значений (x_i, y_j) , $i = 1, 2, \dots, j = 1, 2, \dots$.

ξ	η					P_ξ
	y_1	y_2	\dots	y_j	\dots	
x_1	p_{11}	p_{12}	\dots	p_{1j}	\dots	$p_{1\bullet}$
x_2	p_{21}	p_{22}	\dots	p_{2j}	\dots	$p_{2\bullet}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_i	p_{i1}	p_{i2}	\dots	p_{ij}	\dots	$p_{i\bullet}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
P_η	$p_{\bullet 1}$	$p_{\bullet 2}$	\dots	$p_{\bullet j}$	\dots	

Таблица 6.1.

Такое перечисление удобно представить в виде таблицы 6.1. В этой таблице в верхней строке перечислены все возможные значения y_1, \dots, y_j, \dots случайной величины η , а в левом столбце — значения x_1, \dots, x_i, \dots случайной величины ξ . На пересечении столбца “ y_j ” со строкой “ x_i ” находится вероятность $p_{ij} = P\{\xi = x_i, \eta = y_j\}$ совместного осуществления событий $\{\xi = x_i\}$ и $\{\eta = y_j\}$.

В последнем столбце “ P_ξ ” указывается распределение случайной величины ξ , а в нижней строке “ P_η ” — распределение случайной величины η . А именно, в таблице 6.1 по определению $p_{i\bullet} = P\{\xi = x_i\}$ и $p_{\bullet j} = P\{\eta = y_j\}$ для всех $i = 1, 2, \dots, j = 1, 2, \dots$. Поскольку события $\{\xi = x_i, \eta = y_j\}$, $i = 1, 2, \dots, j = 1, 2, \dots$ несовместны, то

$$p_{i\bullet} = \sum_{n=1}^n p_{ij}, \quad i = 1, 2, \dots, \quad p_{\bullet j} = \sum_{i=1}^m p_{ij}, \quad j = 1, 2, \dots$$

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

Используя таблицу 6.1, нетрудно определить *совместную функцию распределения* $F(x, y)$. Ясно, что для этого необходимо просуммировать p_{ij} по всем тем значениям m и n , для которых $x_i < x, y_j < y$, т.е.

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

Пример 6.3. В соответствии со схемой Бернулли (см. определение 2.6) с вероятностью успеха p и вероятностью неудачи $q = 1 - p$ проводятся два испытания. Выпишем распределение двумерного случайного вектора (ξ, η) , где ξ — число успехов в первом испытании, а η — число успехов во втором испытании. 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 qq . Поэтому

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

и на пересечении столбца “0” со строкой “0” нужно записать q^2 (табл. 6.2). Далее, $\xi = 1$ и $\eta = 0$, если в первом испытании произошел успех, а во втором — неудача, и, значит,

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

Аналогично заполняем второй столбец:

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

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

Таблица 6.2.

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

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

а на пересечении столбца “ P_ξ ” и строки “1” —

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

Построим теперь совместную функцию распределения случайных величин ξ и η :

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

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

Далее, если $0 < x \leq 1$ и $0 < y \leq 1$, то событие $\{\xi < x, \eta < y\}$ эквивалентно событию $\{\xi = 0, \eta = 0\}$, которое, как видно из табл. 6.2, происходит с вероятностью q^2 , и $F(x, y) = q^2$.

Если же $0 < x \leq 1$, а $y > 1$, то событие $\{\xi < x, \eta < y\}$ совпадает с объединением *непесекающихся 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 $x > 1$ и $0 < y \leq 1$.

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

6.4 Непрерывные случайные векторы

Определение 6.4. *Двумерный случайный вектор* (ξ, η) называют **непрерывным**, если ее функцию распределения $F(x, y) = P\{\xi < x, \eta < y\}$ можно представить в виде сходящегося несобственного интеграла:

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

Функцию $f(x, y)$ называют **плотностью распределения вероятностей** 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:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{\partial^2 F(x, y)}{\partial y \partial x}. \quad (6.1)$$

Теорема 6.2. *Двумерная плотность распределения обладает следующими свойствами.*

1. $f(x, y) \geq 0$.

$$2. P\{a < \xi < b, c < \eta < d\} = \int_a^b dx \int_c^d f(x, y) dy.$$

$$3. \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1.$$

$$4. P\{x < \xi < x + \Delta x, y < \eta < y + \Delta y\} \approx f(x, y) \Delta x \Delta y.$$

$$5. P\{\xi = x, \eta = y\} = 0.$$

$$6. P\{(\xi, \eta) \in D\} = \iint_D f(x, y) dx dy.$$

$$7. f_{\xi}(x) = \int_{-\infty}^{+\infty} f(x,y) dy.$$

$$8. f_{\eta}(y) = \int_{-\infty}^{+\infty} f(x,y) dx.$$

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 6.4 двумерной плотности распределения вытекает:

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

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

откуда, дифференцируя интегралы по переменному верхнему пределу и учитывая формулу (3.2), получаем утверждения 7 и 8 для *одномерных плотностей распределения* $f_{\xi}(x)$ и $f_{\eta}(y)$ случайных величин ξ и η .

Пример 6.4. Рассмотрим двумерную случайную величину, плотность распределения которой имеет вид

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

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

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy = \iint_{x^2 + y^2 \leq R^2} A dx dy = \pi A R^2 = 1,$$

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

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

Нетрудно найти одномерную плотность распределения случайной величины ξ :

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

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

6.5 Независимые случайные величины

Определение 6.5. Рассмотрим случайный вектор (ξ, η) . *Случайные величины* ξ и η назовем *независимыми*, если случайные события $\{\xi < x\}$ и $\{\eta < y\}$ являются независимыми для всевозможных действительных x и y . В противном случае *случайные величины* назовем *зависимыми*. #

Можно показать, что независимыми являются и события $\{x_1 \leq \xi < x_2\}$ и $\{y_1 \leq \eta < y_2\}$ для всевозможных действительных x_1, x_2, y_1, y_2 .

Теорема 6.3. Обозначим через $F(x,y)$, $F_\xi(x)$ и $F_\eta(y)$ соответственно функции распределения (ξ,η) , ξ и η . Для того чтобы 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. По определению независимость случайных величин ξ и η равносильна равенству

$$\mathbf{P}\{\xi < x, \eta < y\} = \mathbf{P}\{\xi < x\}\mathbf{P}\{\eta < y\},$$

что равносильно утверждению теоремы, поскольку

$$F(x,y) = \mathbf{P}\{\xi < x, \eta < y\}, \quad F_\xi(x) = \mathbf{P}\{\xi < x\}, \quad F_\eta(y) = \mathbf{P}\{\eta < y\}.$$

Теорема 6.4. Обозначим через $f(x,y)$, $f_\xi(x)$ и $f_\eta(y)$ соответственно плотности распределения (ξ,η) , ξ и η . 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 5.5, $F(x,y) = F_\xi(x)F_\eta(y)$. С учетом формул (6.1) и (3.2) 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 6.4 двумерной плотности распределения и определением 3.6.

$$F(x,y) = \int_{-\infty}^x \int_{-\infty}^y f(s,t) ds dt = \left(\int_{-\infty}^x f_\xi(s) ds \right) \left(\int_{-\infty}^y f_\eta(t) dt \right) = F_\xi(x)F_\eta(y).$$

Справедлива следующая теорема.

Теорема 6.5. Если случайные величины ξ и η независимыми, а функции $\varphi(t)$ и $\psi(t)$ кусочно-непрерывны, то случайные величины $\varphi(\xi)$ и $\psi(\eta)$ также независимы.

Пример 6.5. а. 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 $f_\eta(x)$. Значит, случайные величины ξ и η являются независимыми.

б. Также нетрудно показать, что случайные величины ξ и η из примера 6.4 являются зависимыми.

Теорема 6.6. Дискретные случайные величины ξ и η являются независимыми тогда и только тогда, когда для всех возможных значений x_i и y_j

$$p_{ij} = \mathbf{P}\{\xi = x_i, \eta = y_j\} = \mathbf{P}\{\xi = x_i\}\mathbf{P}\{\eta = y_j\} = p_{i\bullet}p_{\bullet j}. \quad \#$$

Пример 6.6. В схеме Бернулли с двумя испытаниями (см. пример 6.3)

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

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 6.6. Пусть (ξ_1, \dots, ξ_n) — случайный вектор с функцией распределения $F(x_1, \dots, x_n)$. Обозначим через $F_{\xi_i}(x_i)$ функцию распределения координаты ξ_i , $i = 1, \dots, n$. Случайные величины ξ_1, \dots, ξ_n называют **независимыми в совокупности**, если

$$F(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n).$$

Замечание 6.1. Теоремы 6.4 и 6.5 распространяются на любое число случайных величин.

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

Пример 6.7. Свяжем с бросанием тетраэдра из примера 2.5 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,

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

и

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

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

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

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

Лекция 7

Функции от случайных величин. Свойства математического ожидания и дисперсии.

7.1 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 $(\Omega, \mathcal{A}, \mathbf{P})$ двумерный случайный вектор (ξ, η) и числовую функцию $z = \varphi(x, y)$ числовых аргументов x и y .

Определение 7.1. Случайную величину

$$\zeta = \varphi(\xi, \eta) = \varphi(\xi(\omega), \eta(\omega))$$

называют *функцией (скалярной) от двумерной случайной величины (двумерного случайного вектора)* (ξ, η) .

Ясно, что функция $\zeta = \varphi(\xi, \eta)$ от двумерной *дискретной случайной величины* (ξ, η) является дискретной случайной величиной, принимающей значения $\varphi(x_i, y_j)$ с *вероятностью*

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

где x_i и y_j — значения случайных величин ξ и η соответственно.

Чтобы построить *ряд распределения дискретной случайной величины* $\zeta = \varphi(\xi, \eta)$, необходимо, во-первых, не учитывать все те значения $\varphi(x_i, y_j)$, вероятность принять которые случайной величине ζ равна нулю, а во-вторых, объединить в один столбец все одинаковые значения $\varphi(x_i, y_j)$ случайной величины ζ , приписав этому столбцу суммарную вероятность.

Пример 7.1. Пусть ζ — случайная величина, равная суммарному числу успехов в двух испытаниях *по схеме Бернулли*, ξ — число успехов в первом испытании, а η — число успехов во втором испытании. Тогда

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

Поскольку ξ и η могут принимать только значения 0 или 1, то случайная величина ζ 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. 7.1, см. также при-

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

Таблица 7.1.

мер 6.3 и табл. 6.2).

Заметим, что двум средним столбцам соответствует одно и то же значение 1 случайной величины ζ , и их необходимо объединить. Окончательно получаем ряд распределения случайной величины ζ , представленный в табл. 7.2. Как и следовало ожидать, суммарное число успехов ζ в двух испытаниях имеет *биномиальное распределение*. #

ζ	0	1	2
P_ζ	q^2	$2pq$	p^2

Таблица 7.2.

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

$$F_\zeta(z) = \iint_{\varphi(x,y) < z} f(x,y) dx dy, \quad (7.1)$$

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

Поясним геометрически вывод формулы (7.1). Пусть поверхность, определенная функцией $z = \varphi(x, y)$, имеет вид “чаши” (см. рис. 7.1) и z — произвольное значение случайной величины $\zeta = \varphi(\xi, \eta)$. 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, η) ведет к реализации события $\{\zeta < z\}$. Поскольку $\zeta = \varphi(\xi, \eta)$, то

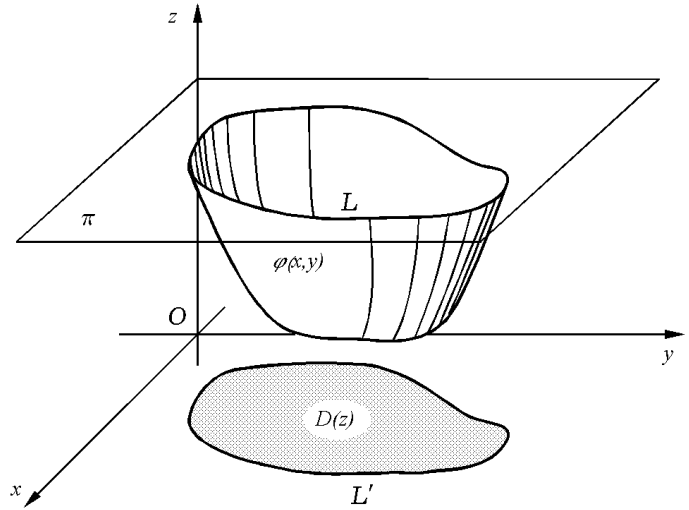


Рис 7.1

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

События $\{\zeta < z\}$ и $\{(\xi, \eta) \in D(z)\}$ совпадают, и в соответствии со свойством 6 двумерной плотности распределения

$$P\{\zeta < z\} = P\{(\xi, \eta) \in D(z)\} = \iint_{\varphi(x,y) < z} f(x,y) dx dy.$$

Учитывая равенство $P\{\zeta < z\} = F_\zeta(z)$, приходим к формуле (7.1).

Пример 7.2. Пусть (ξ, η) — двумерный случайный вектор, имеющий *стандартное двумерное нормальное распределение* (см. определение многомерного нормального распределения на с. 60). Найдем распределение случайной величины $\zeta = \sqrt{\xi^2 + \eta^2}$. В этом случае $\varphi(x, y) = \sqrt{x^2 + y^2}$. Очевидно, что

$$F_\zeta(z) = \begin{cases} 0, & z \leq 0; \\ \iint_{\sqrt{x^2+y^2} < z} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dx dy, & z > 0. \end{cases}$$

Переходя к полярным координатам ρ и φ , имеем

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

Это распределение известно как *распределение Релея*.

7.2 Математическое ожидание функции от случайного вектора

Математическое ожидание $M\zeta$ функции $\zeta = \varphi(\xi, \eta)$ от дискретного случайного вектора (ξ, η) можно найти, воспользовавшись формулой

$$M\zeta = M\varphi(\xi, \eta) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varphi(x_i, y_j) p_{ij}, \quad (7.2)$$

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

Математическое ожидание $M\zeta$ функции $\zeta = \varphi(\xi, \eta)$ от непрерывного случайного вектора (ξ, η) можно найти по формуле

$$M\zeta = M\varphi(\xi, \eta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) f(x, y) dx dy, \quad (7.3)$$

где $f(x, y)$ — плотность распределения случайного вектора (ξ, η) (совместная плотность распределения случайных величин ξ и η).

7.3 Свойства математического ожидания

Докажем теперь теорему о свойствах математического ожидания.

Теорема 7.1. Математическое ожидание удовлетворяет следующим свойствам.

1. Если случайная величина ξ принимает всего одно значение C с вероятностью единица (т.е., по сути дела, не является случайной величиной), то $MC = C$.

2. $M(a\xi + b) = aM\xi + b$, где a, b — постоянные.

3. $M(\xi + \eta) = M\xi + M\eta$.

4. $M(\xi\eta) = M\xi \cdot M\eta$ для независимых случайных величин ξ и η .

ДОКАЗАТЕЛЬСТВО. Если случайная величина ξ принимает всего одно значение C с вероятностью единица, то $MC = C \cdot 1 = C$, откуда следует утверждение 1.

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

Найдем математическое ожидание случайной величины $\eta = a\xi + b$ ($\eta = \varphi(x) = ax + b$):

$$M\eta = M(a\xi + b) = \int_{-\infty}^{+\infty} (ax + b) f_{\xi}(x) dx = a \int_{-\infty}^{+\infty} x f_{\xi}(x) dx + b \int_{-\infty}^{+\infty} f_{\xi}(x) dx = aM\xi + b \cdot 1,$$

т.е. приходим к утверждению 2.

Пусть теперь $\zeta = \xi + \eta$ ($z = \varphi(x, y) = x + y$). Тогда

$$\begin{aligned} M\zeta = M(\xi + \eta) &= \sum_{i,j} (x_i + y_j) p_{ij} = \sum_{i,j} x_i p_{ij} + \sum_{i,j} y_j p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \\ &= \sum_i x_i p_{i\bullet} + \sum_j y_j p_{\bullet j} = M\xi + M\eta, \end{aligned}$$

и, значит, утверждение 3 доказано.

Наконец, если ξ и η независимые случайные величины, то для математического ожидания их произведения $\zeta = \xi\eta$ (воспользовавшись формулой 7.3 и теоремой 6.4) имеем:

$$\begin{aligned} M\zeta = M(\xi\eta) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{\xi}(x) f_{\eta}(y) dx dy = \\ &= \left(\int_{-\infty}^{+\infty} x f_{\xi}(x) dx \right) \left(\int_{-\infty}^{+\infty} y f_{\eta}(y) dy \right) = M\xi M\eta. \end{aligned}$$

Замечание 7.1. Свойства 3 и 4 допускают обобщение на n случайных величин:

$$M(\xi_1 + \xi_2 + \dots + \xi_n) = M\xi_1 + M\xi_2 + \dots + M\xi_n,$$

а если $\xi_1, \xi_2, \dots, \xi_n$ независимы в совокупности, то

$$M(\xi_1 \cdot \xi_2 \dots \xi_n) = M\xi_1 \cdot M\xi_2 \dots M\xi_n.$$

Пример 7.3. Представим число успехов ξ в n испытаниях по схеме Бернулли в виде $\xi = \xi_1 + \dots + \xi_n$, где ξ_i — число успехов в i -м испытании. Нетрудно видеть, что $M\xi_i = 0 \cdot q + 1 \cdot p = p$. Значит, в силу свойства 3 $M\xi = M\xi_1 + \dots + M\xi_n = np$, что совпадает с результатами примера 5.2, но получено с минимальными вычислениями. #

Определение 7.2. Вектор $(M\xi_1, \dots, M\xi_n)$ называют **вектором математических ожиданий (средних значений)**

7.4 Свойства дисперсии

Теорема 7.2. Дисперсия удовлетворяет следующим свойствам.

1. Если случайная величина ξ принимает всего одно значение C с вероятностью единица, то $DC = 0$.

2. $D(a\xi + b) = a^2 D\xi$.

3. $D\xi = M\xi^2 - (M\xi)^2$.

4. $D(\xi + \eta) = D\xi + D\eta$ для независимых случайных величин ξ и η .

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

Определим дисперсию случайной величины $\eta = a\xi + b$. Используя свойство 2 математического ожидания, имеем

$$\begin{aligned} D\eta &= M(\eta - M\eta)^2 = M(a\xi + b - M(a\xi + b))^2 = M(a\xi + b - aM\xi - b)^2 = \\ &= M(a(\xi - M\xi))^2 = M(a^2(\xi - M\xi)^2) = a^2 M(\xi - M\xi)^2. \end{aligned}$$

Поэтому справедливо утверждение 2.

Далее, согласно свойствам 2 и 3 математического ожидания, получаем

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

т.е. приходим к утверждению 3.

Наконец, пусть ξ и η — независимые случайные величины. Тогда, используя свойство 3 дисперсии и свойство 4 теоремы 7.1, получим

$$\begin{aligned} D(\xi + \eta) &= M(\xi + \eta)^2 - (M(\xi + \eta))^2 = M(\xi^2 - 2\xi\eta + \eta^2) - (M\xi^2 - 2M(\xi\eta) + M\eta^2) = \\ &= (M(\xi^2) - (M\xi)^2) - 2(M(\xi\eta) - (M\xi)(M\eta)) + (M(\eta^2) - (M\eta)^2) = D\xi + D\eta, \end{aligned}$$

поскольку $M(\xi\eta) = (M\xi)(M\eta)$. Значит, имеет место утверждение 4.

Замечание 7.2. Можно показать, что справедливо и свойство, обратное свойству 1, а следовательно, имеет место утверждение: дисперсия случайной величины ξ равна нулю тогда и только тогда, когда ξ с вероятностью 1 принимает всего одно значение.

Замечание 7.3. Очевидно, что свойство 4 справедливо для суммы не только двух, но и любого числа n попарно независимых случайных величин

$$D(\xi_1 + \dots + \xi_n) = D\xi_1 + \dots + D\xi_n. \quad \#$$

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

7.5 Формула свертки

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

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

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

$$\zeta = \xi + \eta.$$

Тогда $\zeta = \varphi(\xi, \eta)$, где

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

и, согласно формуле (7.1), находим:

$$\begin{aligned} F_{\zeta}(z) &= \iint_{x+y < z} f(x, y) dx dy = \iint_{x+y < z} f_{\xi}(x)f_{\eta}(y) dx dy = \\ &= \int_{-\infty}^{+\infty} f_{\xi}(x) dx \int_{-\infty}^{z-x} f_{\eta}(y) dy = \int_{-\infty}^{+\infty} F_{\eta}(z-x)f_{\xi}(x) dx. \end{aligned}$$

Дифференцируя последнюю формулу по z под знаком интеграла, получаем (с учетом переобозначения $x = x$) выражение для *плотности $f_{\zeta}(z)$ распределения суммы ξ и η* :

$$f_{\zeta}(z) = \int_{-\infty}^{+\infty} f_{\eta}(z-x)f_{\xi}(x) dx. \quad (7.4)$$

В этом случае говорят, что плотность распределения $f_{\zeta}(z)$ случайной величины ζ является **сверткой (композицией) плотностей распределения** $f_{\xi}(x)$ и $f_{\eta}(x)$ слагаемых ξ и η или что закон распределения суммы двух независимых случайных величин является **сверткой (композицией) законов распределения** слагаемых. Соотношение (7.4) условно записывают в виде $f_{\zeta} = f_{\eta} * f_{\xi}$. Формулу (7.4) называют **формулой свертки** для плотностей распределения случайных величин ξ и η .

Пример 7.4. Пусть ξ и η — независимые нормальные случайные величины с математическими ожиданиями m_1 и m_2 и дисперсиями σ_1^2 и σ_2^2 . Найдём плотность распределения суммы $\zeta = \xi + \eta$. Воспользовавшись формулой свертки, имеем

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

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

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

Таким образом, случайная величина ζ также нормальна и имеет математическое ожидание $m_1 + m_2$ и дисперсию $\sigma_1^2 + \sigma_2^2$. Другими словами, композиция плотностей нормальных законов распределения является плотностью нормального закона распределения.

Замечание 7.4. Выведенное в примере 7.4 свойство справедливо для суммы любого конечного числа слагаемых, распределенных по нормальному закону. А именно, если ξ_1, \dots, ξ_n — независимые нормальные случайные величины с математическими ожиданиями m_1, \dots, m_n и дисперсиями $\sigma_1^2, \dots, \sigma_n^2$, то их сумма $\xi_1 + \dots + \xi_n$ является нормальной случайной величиной с математическим ожиданием $m_1 + \dots + m_n$ и дисперсией $\sigma_1^2 + \dots + \sigma_n^2$.

Лекция 8

Ковариация и коэффициент корреляции случайных величин

Пусть (ξ, η) — двумерный случайный вектор.

Определение 8.1. Ковариацией (корреляционным моментом) $\text{cov}(\xi, \eta)$ случайных величин ξ и η называют математическое ожидание произведения случайных величин $\overset{o}{\xi} = \xi - M\xi$ и $\overset{o}{\eta} = \eta - M\eta$:

$$\text{cov}(\xi, \eta) = M(\overset{o}{\xi} \overset{o}{\eta}) = M((\xi - M\xi)(\eta - M\eta)).$$

Запишем формулы, определяющие ковариацию.

Согласно (7.2) для дискретных случайных величин ξ и η

$$\text{cov}(\xi, \eta) = \sum_{i,j} (x_i - M\xi)(y_j - M\eta) p_{ij},$$

а согласно (7.3) для непрерывных случайных величин ξ и η

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

Заметим, что введение понятия ковариации позволяет записать выражение для дисперсии суммы случайных величин и к уже имеющимся свойствам дисперсии из теоремы 7.2 добавить еще одно (свойство 5)

$$D(\xi + \eta) = D\xi + D\eta + 2\text{cov}(\xi, \eta), \quad (8.1)$$

справедливое для произвольных, а не только независимых случайных величин ξ и η . Действительно,

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

Свойство 5 дисперсии допускает обобщение на произвольное число слагаемых:

$$D(\xi_1 + \dots + \xi_n) = D\xi_1 + \dots + D\xi_n + 2 \sum_{1 \leq i < j \leq n} \text{cov}(\xi_i, \xi_j).$$

Следующая теорема устанавливает основные свойства ковариации.

Теорема 8.1. Ковариация имеет следующие свойства

1. $\text{cov}(\xi, \xi) = D\xi$.
2. $\text{cov}(\xi, \eta) = 0$ для независимых случайных величин ξ и η
3. Если $\eta_i = a_i \xi_i + b_i$, $i = 1, 2$, то $\text{cov}(\eta_1, \eta_2) = a_1 a_2 \text{cov}(\xi_1, \xi_2)$.

4. $-\sqrt{D\xi D\eta} \leq \text{cov}(\xi, \eta) \leq \sqrt{D\xi D\eta}$, причем

$$|\text{cov}(\xi, \eta)| = \sqrt{D\xi D\eta} \quad (8.2)$$

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

$$\eta = a\xi - b. \quad (8.3)$$

$$5. \text{cov}(\xi, \eta) = M(\xi\eta) - M\xi M\eta.$$

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

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

$$\begin{aligned} \text{cov}(\eta_1, \eta_2) &= M((\eta_1 - M\eta_1)(\eta_2 - M\eta_2)) = \\ &= M((a_1\xi_1 + b_1 - a_1M\xi_1 - b_1)(a_2\xi_2 + b_2 - a_2M\xi_2 - b_2)) = \\ &= M(a_1a_2(\xi_1 - M\xi_1)\xi_2 - M\xi_2)) = a_1a_2\text{cov}(\xi_1, \xi_2). \end{aligned}$$

Поэтому справедливо утверждение 3.

Рассмотрим дисперсию случайной величины $\eta_x = x\xi - \eta$, где x — произвольное число. В силу свойств дисперсии и свойства 3 ковариации

$$D\eta_x = D(x\xi) + 2\text{cov}(x\xi, -\eta) + D(-\eta) = x^2D\xi - 2x\text{cov}(\xi, \eta) + D\eta.$$

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

$$D = (2\text{cov}(\xi, \eta))^2 - 4D\xi D\eta \quad (8.4)$$

квадратного трехчлена $D\eta_x$ является неположительным, т.е. $|\text{cov}(\xi, \eta)| \leq \sqrt{D\xi D\eta}$.

Далее, пусть выполнено равенство (8.2). Значит, дискриминант (8.4) равен нулю, и поэтому уравнение $D\eta_x = 0$ имеет решение, которое обозначим a . Тогда в соответствии с замечанием 7.2 случайная величина $\eta_a = a\xi - \eta$ принимает всего одно значение (допустим, b), и, следовательно, $\eta = a\xi - b$, т.е. из (8.2) вытекает (8.3).

Наоборот, пусть выполнено (8.3). Тогда $\eta_a = a\xi - \eta = b$ и в соответствии со свойством 1 дисперсии выполнено $D\eta_a = 0$, а значит, квадратный трехчлен $D\eta_x$ имеет один действительный корень. Т.к. этот же трехчлен неотрицательный, то он имеет действительный корень лишь в случае, когда его дискриминант в (8.4) равен нулю, откуда следует (8.2). Таким образом, из (8.3) вытекает (8.2). Утверждение 4 полностью доказано.

Наконец, свойство 5, которое часто бывает полезным при численном подсчете ковариации, вытекает из свойства 4 теоремы 7.2 и равенства (8.1).

Если случайные величины ξ и η являются независимыми, то согласно свойству 5 этой теоремы и свойству 4 теоремы 7.1

$$\text{cov}(\xi, \eta) = M(\xi\eta) - M\xi M\eta = 0,$$

откуда приходим к утверждению 2.

Замечание 8.1. Если случайные величины связаны линейной зависимостью $\eta = a\xi - b$, то в соответствии со свойствами 3 и 1 $\text{cov}(\xi, \eta) = a\text{cov}(\xi, \xi) = aD\xi$. Поэтому знак ковариации совпадает со знаком коэффициента a и равенство (8.2) допускает следующее уточнение:

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

Замечание 8.2. Из свойств дисперсии и ковариации можно получить еще одно полезное при расчетах свойство дисперсии

$$D(a\xi + b\eta + c) = a^2D\xi + b^2D\eta + 2ab\text{cov}(\xi, \eta).$$

Докажите это свойство самостоятельно.

Замечание 8.3. Как следует из свойства 2 теоремы 8.1, ковариация независимых случайных величин равна нулю. Однако обратное, вообще говоря, неверно. Существуют зависимые и даже функционально зависимые случайные величины, ковариация которых равна нулю, что демонстрирует следующий пример.

Пример 8.1. Пусть случайная величина ξ имеет *равномерное* в интервале $(-1, 1)$ *распределение*, а случайная величина η связана со случайной величиной ξ функциональной зависимостью $\eta = \xi^2$. Покажем, что $\text{cov}(\xi, \eta) = 0$, несмотря на функциональную зависимость ξ и η .

Действительно, учитывая равенство $M\xi = 0$ (см. (5.4)) и свойство 5 ковариации, имеем

$$\text{cov}(\xi, \eta) = M(\xi \cdot \eta) - M\xi \cdot M\eta = M\xi^3 = \int_{-\infty}^{+\infty} x^3 f(x) dx = \frac{1}{2} \int_{-1}^1 x^3 dx = 0.$$

Определение 8.2. Случайные величины ξ и η называют *некоррелированными*, если их ковариация равна нулю, т.е. $\text{cov}(\xi, \eta) = 0$.

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

Рассмотрим теперь n -мерный случайный вектор $\xi = (\xi_1, \dots, \xi_n)$.

Определение 8.3. Матрицей ковариаций (ковариационной матрицей) случайного вектора ξ называют матрицу $\Sigma = (\sigma_{ij}) = (\text{cov}(\xi_i, \xi_j))$, состоящую из ковариаций случайных величин ξ_i и ξ_j .

Пример 8.2. Рассмотрим двумерную случайную величину (ξ, η) , распределенную по *нормальному закону* (см. лекцию 9). Тогда

$$\begin{aligned} \text{cov}(\xi, \eta) &= \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{(x-m_1)(y-m_2)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\left(\frac{(x-m_1)^2}{2\sigma_1^2(1-\rho^2)} - \frac{2\rho(x-m_1)(y-m_2)}{2\sigma_1\sigma_2(1-\rho^2)} + \frac{(y-m_2)^2}{2\sigma_2^2(1-\rho^2)} \right) \right] dx dy. \end{aligned}$$

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

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

Внутренний интеграл равен ρu . Поэтому

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

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

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

Определение 8.4. Коэффициентом корреляции случайных величин ξ и η называют число $\rho = \rho(\xi, \eta)$, определяемое равенством (предполагается, что $D\xi > 0$ и $D\eta > 0$)

$$\rho = \frac{\text{cov}(\xi, \eta)}{\sqrt{D\xi \cdot D\eta}}.$$

Теорема 8.2. Коэффициент корреляции имеет следующие свойства.

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

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

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

По аналогии с ковариационной матрицей для случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ можно ввести корреляционную матрицу.

Определение 8.5. Корреляционной (нормированной ковариационной) матрицей случайного вектора ξ называют матрицу $P = (\rho_{ij}) = (\rho(\xi_i, \xi_j))$, состоящую из коэффициентов корреляций случайных величин ξ_i и ξ_j .

Лекция 9

Многомерное нормальное распределение

Нормальное распределение одномерной случайной величины рассматривалось в лекции 5. Сейчас обратимся к многомерному случаю. При этом сначала введем *двумерное нормальное распределение случайного вектора* $\zeta = (\xi, \eta)$, а затем обобщим полученные результаты на случайный вектор произвольной размерности $n > 2$.

Пусть координаты ξ и η случайного вектора $\zeta = (\xi, \eta)$ являются случайными величинами, распределенными по нормальному закону, т.е. имеют *плотности распределения*

$$f_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2}\right),$$
$$f_{\eta}(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-m_2)^2}{2\sigma_2^2}\right).$$

Напомним, что параметры m_i и $\sigma_i > 0$, $i = \overline{1,2}$, этих распределений совпадают с математическими ожиданиями и средними квадратическими отклонениями случайных величин ξ и η .

Если ξ и η являются независимыми случайными величинами, то, согласно теореме 6.4, плотность $f(x,y)$ случайного вектора $\zeta = (\xi, \eta)$ имеет вид

$$f(x,y) = \frac{1}{(\sqrt{2\pi})^2 \sigma_1 \sigma_2} \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2} - \frac{(y-m_2)^2}{2\sigma_2^2}\right).$$

В общем случае вектор $\zeta = (\xi, \eta)$ имеет (*невыврожденное*) *двумерное нормальное распределение*, если его плотность распределения определяется формулой

$$f(x,y) = \frac{1}{(\sqrt{2\pi})^2 \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x-m_1, y-m_2)}, \quad (9.1)$$

где функция

$$Q(s,t) = \frac{1}{1-\rho^2} \left(\frac{s^2}{\sigma_1^2} - \frac{2\rho st}{\sigma_1 \sigma_2} + \frac{t^2}{\sigma_2^2} \right), \quad (9.2)$$

есть положительно определенная квадратичная форма (т.е. $Q(s,t) > 0$ для любых $(s,t) \in \mathbb{R}$, $(s,t) \neq (0,0)$).

Двумерное нормальное распределение зависит от пяти параметров: $m_1, m_2, \sigma_1, \sigma_2, \rho$. Можно показать, что $m_1 = M\xi$, $m_2 = M\eta$, $\sigma_1^2 = D\xi$, $\sigma_2^2 = D\eta$, ρ — коэффициент корреляции случайных величин ξ и η .

Последние три параметра запишем для дальнейшего обобщения на случай $n > 2$ в виде *матрицы ковариаций (ковариационной матрицы)* Σ вектора $\zeta = (\xi, \eta)$:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

где $\sigma_{ii} = \sigma_i^2$, $i = \overline{1,2}$, а $\sigma_{12} = \sigma_{21} = \rho \sigma_1 \sigma_2$.

Если ввести матрицу $\tilde{\Sigma}$, обратную матрице Σ , т.е. $\tilde{\Sigma} = \Sigma^{-1}$ и вектор $v = (v_1, v_2)$, то квадратичную форму (9.2) можно записать в матричной форме в виде

$$Q(v) = v \tilde{\Sigma} v^T, \quad (9.3)$$

где знак “ T ” означает транспонирование. Действительно, если учитывать, что

$$\tilde{\Sigma} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix},$$

то, подставляя $\tilde{\Sigma}$ в (9.3), приходим к выражению (9.2).

Далее, если заметить, что

$$\sigma_1 \sigma_2 \sqrt{1 - \rho^2} = \sqrt{\det \Sigma},$$

где $\det \Sigma$ — определитель матрицы Σ , то выражение (9.1), обозначив $m = (m_1, m_2)$, можно записать в виде

$$f(u) = \frac{1}{(\sqrt{2\pi})^2 (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(u-m)\tilde{\Sigma}(u-m)^T}, \quad u \in \mathbb{R}^2.$$

Теперь можно записать плотность (невырожденного) нормального распределения для случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ произвольной размерности $n > 2$.

Определение 9.1. Случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ назовем ***n -мерным нормальным вектором*** или случайным вектором, имеющим нормальное распределение с вектором математических ожиданий $m = (m_1, \dots, m_n)$ и ковариационной матрицей $\Sigma = (\sigma_{ij})$, $i, j = \overline{1, n}$, если его плотность имеет вид

$$f(x) = \frac{1}{(\sqrt{2\pi})^n (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-m)\tilde{\Sigma}(x-m)^T}, \quad x \in \mathbb{R}^n,$$

где $\tilde{\Sigma}$ — матрица, обратная к матрице Σ .

Если матрица Σ (а значит, и матрица $\tilde{\Sigma} = \Sigma^{-1}$) совпадает с единичной матрицей I , а вектор $m = (0, \dots, 0)$, то

$$f(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)}.$$

Такую плотность по аналогии с одномерным случаем (см. с. 41) называют ***плотностью стандартного многомерного (n -мерного) нормального распределения***.

Справедлива следующая теорема.

Теорема 9.1. Случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ имеет нормальное распределение тогда и только тогда, когда для любых действительных чисел c_1, \dots, c_n случайная величина

$$c_1 \xi_1 + \dots + c_n \xi_n$$

является нормальной.

ДОКАЗАТЕЛЬСТВО. Без доказательства.

Эта теорема позволяет дать эквивалентное определение многомерного случайного вектора.

Определение 9.2. Случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ называется ***n -мерным нормальным случайным вектором***, если для любых действительных чисел c_1, \dots, c_n случайная величина

$$c_1 \xi_1 + \dots + c_n \xi_n$$

является нормальной.

Рассмотрим основные свойства многомерного нормального распределения.

1. ***Закон распределения*** каждой из координат случайного вектора ξ , имеющего n -мерное нормальное распределение с вектором средних $m = (m_1, \dots, m_n)$ и матрицей ковариаций $\Sigma = \sigma_{ij}$, является нормальным с параметрами m_i и σ_i .

ДОКАЗАТЕЛЬСТВО. Докажем это утверждение для случая $n = 2$ (общий случай требует более громоздких преобразований).

Найдем плотность распределения $f_{\xi}(x)$, если $f(x, y)$ определяется формулами (9.1) и (9.2). Воспользовавшись свойством 7 (см. теорему 6.2) двумерной плотности распределения, имеем

$$f_{\xi}(x) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\tilde{Q}(x,y)} dy,$$

где

$$\tilde{Q}(x, y) = -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-m_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-m_1}{\sigma_1} \right) \left(\frac{y-m_2}{\sigma_2} \right) + \left(\frac{y-m_2}{\sigma_2} \right)^2 \right].$$

Делая замену

$$t = \frac{\frac{y-m_2}{\sigma_2} - \frac{\rho(x-m_1)}{\sigma_1}}{\sqrt{1-\rho^2}},$$

после преобразований получаем

$$f_{\xi}(x) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1} e^{-\frac{t^2}{2} - \frac{(x-m_1)^2}{2\sigma_1^2}} dt.$$

Поскольку

$$\int_{-\infty}^{+\infty} e^{-t^2/2} dt = \sqrt{2\pi},$$

приходим к окончательному ответу

$$f_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-m_1)^2/2\sigma_1^2},$$

что и доказывает требуемое утверждение.

Аналогично можно показать, что

$$f_{\eta}(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}.$$

2. Если ковариационная матрица Σ случайного вектора ξ , распределенного по нормальному закону (невырожденному), является диагональной, то координаты вектора ξ_1, \dots, ξ_n являются независимыми случайными величинами.

ДОКАЗАТЕЛЬСТВО. Действительно, матрица $\tilde{\Sigma} = \Sigma^{-1}$ также является диагональной и имеет вид

$$\tilde{\Sigma} = \begin{pmatrix} \sigma_1^{-2} & 0 & \dots & 0 \\ 0 & \sigma_2^{-2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^{-2} \end{pmatrix},$$

и, следовательно, формула (9.1) для совместной (n -мерной) плотности распределения имеет вид

$$f(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n \sigma_1 \dots \sigma_n} e^{-\left[\frac{(x_1-m_1)^2}{2\sigma_1^2} + \dots + \frac{(x_n-m_n)^2}{2\sigma_n^2} \right]} = f_{\xi_1}(x_1) \dots f_{\xi_n}(x_n),$$

т.е. случайные величины ξ_1, \dots, ξ_n являются независимыми (см. замечание 6.1).

Заметим, что если $\sigma_{ij} = 0$ для некоторых i и j или, что то же самое, коэффициент корреляции $\rho_{ij} = 0$, то говорят, что *случайные величины* ξ_i и ξ_j являются *некоррелированными*.

Таким образом, из некоррелированности координат случайного вектора, распределенного по нормальному закону, следует (в силу теоремы 6.4) их независимость. Поскольку

независимые случайные величины являются некоррелированными, то для нормально распределенных случайных векторов некоррелированность координат равносильна их независимости.

3. Если вектор $\xi = (\xi_1, \dots, \xi_n)$ имеет нормальный закон распределения с вектором средних $m = (m_1, \dots, m_n)$ и матрицей ковариаций Σ , то вектор $\xi' = (\xi_1, \dots, \xi_{n-1})$ также распределен по нормальному закону с вектором средних $m' = (m_1, \dots, m_{n-1})$ и матрицей ковариаций Σ' , полученной из матрицы Σ вычеркиванием последних строки и столбца.

ДОКАЗАТЕЛЬСТВО. Это свойство доказывается так же, как и свойство 1, но в силу громоздкости вывода оно здесь не приводится.

Из свойства 3 методом математической индукции можно показать, что любой набор координат n -мерного случайного вектора $\xi = (\xi_1, \dots, \xi_n)$, распределенного по нормальному закону, снова имеет нормальное распределение. В частности, двумерный случайный вектор (ξ, η) распределен по нормальному закону с вектором средних (m_1, m_2) и матрицей ковариаций $\Sigma' = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$.

Лекция 10

Предельные теоремы теории вероятностей

С самого начала изучения курса теории вероятностей мы говорили о том, что практическое применение методов этой математической дисциплины основывается на законе предельного постоянства *частоты события*, установленном эмпирически. Согласно этому закону, если один и тот же опыт повторяется многократно, то частота появления конкретного случайного события теряет свойства случайности и приближается к некоторому пределу, который в соответствии со *статистическим определением вероятности* (см. определение 1.14 на с. 11 лекции 1) и называют *вероятностью*.

Однако для того чтобы теория согласовывалась с практикой, при *аксиоматическом определении вероятности*, которое мы использовали, этот закон предельного постоянства частоты должен быть обоснован теоретически. Иначе говоря, он должен быть сформулирован и доказан в виде одной или нескольких теорем. В теории вероятностей теоремы такого типа обычно называют различными формами *закона больших чисел*. В этой лекции мы докажем некоторые формы этого закона, которые, в частности, поясняют смысл *математического ожидания случайной величины*, и то, почему его называют также *средним значением*.

Далее доказывается простейший вариант *центральной предельной теоремы*, уточняющей закон больших чисел. Центральная предельная теорема, в свою очередь, объясняет то широкое распространение, которое получило на практике *нормальное распределение*.

Сначала введем некоторые типы сходимости случайных величин. Пусть

$$\xi_1, \xi_2, \dots, \xi_n, \dots$$

представляет собой последовательность *случайных величин*, заданных на одном и том же *вероятностном пространстве*. Опишем типы сходимости этой последовательности к некоторой случайной величине ξ .

Казалось бы, наиболее разумно понимать сходимость последовательности случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ следующим образом. Для каждого *элементарного события* (исхода) $\omega \in \Omega$ последовательность $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega), \dots$ представляет собой обычную числовую последовательность, и можно определить сходимость последовательности случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ к случайной величине ξ как сходимость числовых последовательностей $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega), \dots$ к числу $\xi(\omega)$ при всех $\omega \in \Omega$.

К сожалению, такая сходимость (ее называют *сходимостью всюду*) редко встречается на практике.

Определение 10.1. Если последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ случайных величин удовлетворяет условию

$$P \left\{ \lim_{n \rightarrow \infty} \xi_n = \xi \right\} = 1,$$

то говорят о *сходимости* этой последовательности к случайной величине ξ *с вероятностью 1* или *почти наверное* и обозначают

$$\xi_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi.$$

Другими словами множество $A = \{\omega \in \Omega : \lim_{n \rightarrow \infty} \xi_n(\omega) \neq \xi(\omega)\}$ тех ω из Ω , для которых числовая (неслучайная) последовательность $\xi_n(\omega)$ не стремится к $\xi(\omega)$, настолько мало, что $P(A) = 0$. #

Иногда и сходимости с вероятностью 1 не существует, но удастся установить следующий более слабый вид сходимости, называемый сходимостью по вероятности.

Определение 10.2. Если последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ случайных величин для любого $\varepsilon > 0$ удовлетворяет условию

$$\lim_{n \rightarrow \infty} P\{|\xi_n - \xi| < \varepsilon\} = 1,$$

или равносильному условию

$$\lim_{n \rightarrow \infty} P\{|\xi_n - \xi| > \varepsilon\} = 0,$$

то говорят о *сходимости* этой последовательности к случайной величине ξ *по вероятности*. Сходимость к ξ по вероятности записывается в виде

$$\xi_n \xrightarrow[n \rightarrow \infty]{P} \xi. \quad \#$$

Сходимость по вероятности означает, что вероятность того, что ξ_n отклоняется от ξ больше, чем на любое наперед заданное число ε , стремится к нулю с возрастанием n .

Можно показать, что из сходимости с вероятностью 1 следует сходимость по вероятности. Обратное, вообще говоря, неверно: существуют последовательности, сходящиеся по вероятности, но не сходящиеся почти наверное.

Существуют также примеры последовательностей, сходящихся по вероятности, но не сходящихся ни в одной точке, т.е. ни при каком ω . Это происходит потому что множества $A_n = \{\omega \in \Omega : |\xi_n(\omega) - \xi(\omega)| < \varepsilon\}$ при разных n состоят, вообще говоря, из разных ω , поскольку некоторые из ω с ростом n могут мигрировать, выпадая из A_n при одних n и снова возвращаясь при других n . Другими словами, не следует думать, что если какое-то ω попадет в A_N при некотором N , то это ω будет принадлежать A_n при всех $n > N$.

Неравенства Чебышёва. Закон больших чисел

Прежде чем приступить к рассмотрению закона больших чисел, докажем два неравенства Чебышёва. Заметим, что неравенства Чебышёва представляют и самостоятельный интерес, поскольку в современной теории вероятностей широко используются неравенства такого типа.

Теорема 10.1. Для каждой неотрицательной случайной величины ξ , имеющей математическое ожидание $M\xi$, при любом $\varepsilon > 0$ справедливо соотношение

$$P\{\xi \geq \varepsilon\} \leq \frac{M\xi}{\varepsilon},$$

называемое *первым неравенством Чебышёва*.

ДОКАЗАТЕЛЬСТВО. Доказательство проведем для непрерывной случайной величины ξ с плотностью распределения $f(x)$ Поскольку случайная величина ξ является неотрицательной, то

$$M\xi = \int_0^{+\infty} xf(x) dx.$$

Так как подынтегральное выражение неотрицательное, то при уменьшении области интегрирования интеграл может только уменьшиться. Поэтому

$$M\xi = \int_0^{\varepsilon} xf(x) dx + \int_{\varepsilon}^{+\infty} xf(x) dx \geq \int_{\varepsilon}^{+\infty} xf(x) dx.$$

Заменяя в подынтегральном выражении сомножитель x на ε , имеем

$$\int_{\varepsilon}^{+\infty} xf(x) dx \geq \varepsilon \int_{\varepsilon}^{+\infty} f(x) dx.$$

Остается заметить, что последний интеграл (равный площади области, заштрихованной на 10.1) представляет собой вероятность события $\xi \geq \varepsilon$, и, значит,

$$M\xi \geq \varepsilon P\{\xi \geq \varepsilon\},$$

откуда и вытекает первое неравенство Чебышёва. Аналогично первое неравенство Чебышёва доказывается и для дискретной случайной величины, при этом нужно только заменить интеграл суммой. #

Ясно, что применять первое неравенство Чебышёва имеет смысл только тогда, когда $\varepsilon > M\xi$; в противном случае оно даёт тривиальную оценку.

Пример 10.1. Пусть ξ — время опоздания студента на лекцию, причём известно, что $M\xi = 1$ мин. Воспользовавшись первым неравенством Чебышёва, оценим вероятность $P\{\xi \geq 5\}$ того, что студент опоздает не менее, чем на 5 мин.

Имеем

$$P\{\xi \geq 5\} \leq \frac{M\xi}{5} = 0,2.$$

Таким образом, искомая вероятность не более 0,2, т.е. в среднем из каждых пяти студентов опаздывает, по крайней мере, на 5 мин не более чем один студент. #

Рассмотрим теперь случайную величину ξ , имеющую дисперсию $D\xi = \sigma^2$. Мы уже говорили, что дисперсия является показателем разброса ξ вокруг математического ожидания $M\xi$. Однако с точки зрения исследователя разброс естественнее характеризовать вероятностью $P\{|\xi - M\xi| \geq \varepsilon\}$ отклонения случайной величины ξ от $M\xi$ на величину, большую некоторого заданного ε . Следующее неравенство позволяет оценить эту вероятность с помощью дисперсии σ^2 .

Теорема 10.2. Для каждой случайной величины ξ , имеющей дисперсию $D\xi$, при любом $\varepsilon > 0$ справедливо *второе неравенство Чебышёва*

$$P\{|\xi - M\xi| \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}.$$

ДОКАЗАТЕЛЬСТВО. Для доказательства воспользуемся утверждением первого неравенства Чебышёва. Применяя к случайной величине $\eta = (\xi - M\xi)^2$ это неравенство, в котором ε заменено на ε^2 , получаем

$$P\{|\xi - M\xi| \geq \varepsilon\} = P\{(\xi - M\xi)^2 \geq \varepsilon^2\} = P\{\eta \geq \varepsilon^2\} \leq \frac{M\eta}{\varepsilon^2} = \frac{D\xi}{\varepsilon^2},$$

что и доказывает второе неравенство Чебышёва. #

Геометрический смысл второго неравенства Чебышёва понятен из рис. 10.2. Второе неравенство Чебышёва имеет содержательный смысл лишь при $\varepsilon > \sigma$.

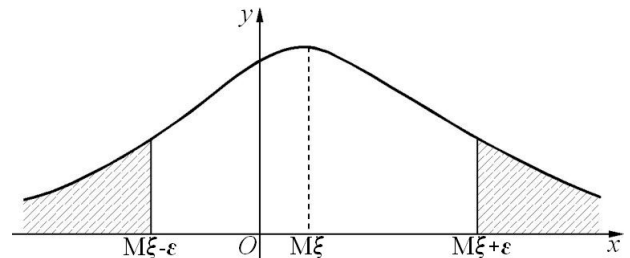


Рис 10.2

Пример 10.2. Пусть в условиях предыдущего примера известно дополнительно, что $\sigma = \sqrt{D\xi} = 1$. Оценим минимальное значение x_0 , при котором вероятность опоздания студента на время не менее x_0 не превышает заданного значения $P_3 = 0,1$.

Для решения поставленной задачи воспользуемся вторым неравенством Чебышёва. Тогда

$$P_3 \leq P\{\xi \geq x_0\} = P\{\xi - M\xi \geq x_0 - M\xi\} \leq P\{|\xi - M\xi| \geq x_0 - M\xi\} \leq \frac{\sigma^2}{(x_0 - M\xi)^2}.$$

Значит, $(x_0 - M\xi)^2 \leq \frac{\sigma^2}{P_3}$ и $x_0 \leq M\xi + \sqrt{\frac{\sigma^2}{P_3}}$. Подставляя конкретные значения, имеем $x_0 \leq 1 + \sqrt{\frac{1}{0,1}} \approx 4,16$. Таким образом, вероятность опоздания студента на время более 4,16 мин не более 0,1. #

Сравнивая полученный результат с результатом примера 10.1, видим, что дополнительная информация о дисперсии времени опоздания позволяет дать более точную оценку искомой вероятности.

Рассмотрим некоторые формы закона больших чисел.

Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность случайных величин, имеющих математические ожидания $m_i = M\xi_i$.

Определение 10.3. Последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ случайных величин удовлетворяет **закону больших чисел (слабому)**, если для любого $\varepsilon > 0$

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n m_i \right| \geq \varepsilon \right\} \xrightarrow{n \rightarrow \infty} 0,$$

т.е.

$$\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n m_i \xrightarrow[n \rightarrow \infty]{P} 0. \quad \#$$

Иными словами, выполнение закона больших чисел отражает предельную устойчивость средних арифметических случайных величин. А именно, среднее арифметическое большого числа случайных величин практически перестает быть случайным, почти не отличаясь от неслучайного среднего арифметического своих математических ожиданий.

Очевидно, что последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ удовлетворяет закону больших чисел тогда и только тогда, когда среднее арифметическое случайных величин $\xi_1 - m_1, \xi_2 - m_2, \dots, \xi_n - m_n$ сходится по вероятности к нулю при $n \rightarrow \infty$.

Теорема 10.3 (Закон больших чисел в форме Чебышёва). Пусть последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимых случайных величин такова, что существуют $M\xi_i = m_i$ и $D\xi_i$, причем дисперсии $D\xi_i$ ограничены в совокупности (т.е. $D\xi_i \leq C$ для некоторой постоянной C), $i = 1, 2, \dots$

Тогда

$$\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n m_i \xrightarrow[n \rightarrow \infty]{P} 0,$$

то есть для последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$ выполнен закон больших чисел. При этом говорят также, что к последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$ случайных величин применим **закон больших чисел в форме Чебышёва**.

ДОКАЗАТЕЛЬСТВО. Теорема является следствием второго неравенства Чебышёва. Действительно, в силу свойств математического ожидания и дисперсии (см. теоремы 7.1 и 7.2)

$$M \left(\frac{1}{n} \sum_{i=1}^n \xi_i \right) = \frac{1}{n} \sum_{i=1}^n m_i, \quad D \left(\frac{1}{n} \sum_{i=1}^n \xi_i \right) = \frac{1}{n^2} \sum_{i=1}^n D\xi_i \leq \frac{Cn}{n^2} = \frac{C}{n}.$$

Применяя теперь второе неравенство Чебышёва к случайной величине $\frac{1}{n} \sum_{i=1}^n \xi_i$, получаем для любого $\varepsilon > 0$

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n m_i \right| \geq \varepsilon \right\} \leq \frac{C}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Таким образом, мы показали, что для последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$ выполняется закон больших чисел.

Определение 10.4. Случайные величины $\xi_1, \xi_2, \dots, \xi_n, \dots$ назовем **одинаково распределенными**, если все они имеют одну и ту же функцию распределения.

Замечание 10.1. У одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ математические ожидания и дисперсии (если таковые существуют) совпадают: $M\xi_i = m$, $D\xi_i = \sigma^2$, $i = 1, 2, \dots$, где m и σ — некоторые действительные числа, причем $\sigma > 0$. Например, если $\xi_1, \xi_2, \dots, \xi_n, \dots$ — непрерывные случайные величины с одинаковой функцией распределения $F(x)$, то $M\xi_i = \int_{-\infty}^{\infty} xF'(x)dx$ не зависит от i . #

Примером последовательности одинаково распределенных случайных величин может служить количество ξ_i выпавших шестерок при i -ом подбрасывании одной игральной кости (в этом случае $P\{\xi_i = 0\} = 5/6$ и $P\{\xi_i = 1\} = 1/6$ для всех i) или ошибка ε_i i -го измерения некоторой физической величины в одних и тех же условиях (один и тот же измерительный прибор, один и тот же лаборант-измеритель и т.д.).

Следствие 10.1. Если в условиях теоремы 10.3 случайные величины ξ_i , $i = 1, 2, \dots$, являются также одинаково распределенными с общим математическим ожиданием m , то последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ случайных величин удовлетворяет закону больших чисел в следующей форме:

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow[n \rightarrow \infty]{P} m.$$

ДОКАЗАТЕЛЬСТВО. Доказательство проведите самостоятельно.

Замечание 10.2. Следствие 10.1 обосновывает широко распространенную практику измерения величин, согласно которой проводится несколько измерений, а затем измеряемая величина оценивается средним арифметическим этих измерений.

А именно, пусть m — неслучайная величина, подлежащая измерению. Обозначим через $\xi_1, \xi_2, \dots, \xi_n$ результаты n измерений, при этом ошибки измерений $\delta_i = \xi_i - m$, $i = 1, \dots, n$ разумно считать случайными. Предположим, что у измерительных ошибок отсутствует систематическая составляющая, т.е. $M\delta_i = 0$, $i = 1, \dots, n$. Предположим также, что все измерения проведены одним и тем же прибором, ошибки измерений которого не выходят из некоего измерительного диапазона, т.е. дисперсия ошибок $D\delta_i = d$, $i = 1, \dots, n$ существует и одинакова. И наконец, предположим, что на результат каждого измерения не влияют результаты предыдущих измерений, т.е. случайные величины $\delta_1, \delta_2, \dots, \delta_n$, а следовательно, и величины $\xi_1, \xi_2, \dots, \xi_n$ являются независимыми. Тогда

$$M\xi_i = M(m + \delta_i) = m + M\delta_i = m + 0 = m,$$

и, согласно закону больших чисел $m \approx \frac{1}{n} \sum_{i=1}^n \xi_i$.

Следствие 10.2. Пусть проводится n испытаний по схеме Бернулли и Y_n — общее число успехов в n испытаниях. Тогда наблюдаемая частота успехов (доля успехов)

$$\hat{p} = \frac{Y_n}{n}$$

сходится по вероятности к вероятности p успеха в одном испытании

$$\hat{p} \xrightarrow[n \rightarrow \infty]{P} p.$$

ДОКАЗАТЕЛЬСТВО. Обозначим ξ_i число успехов в i -м испытании Бернулли (ξ_i принимают значения 0 или 1 с вероятностями $q = 1 - p$ и p соответственно). Тогда $Y_n = \sum_{i=1}^n \xi_i$ и частоту успехов в n испытаниях можно определить в виде $\hat{p} = \frac{1}{n} \sum_{i=1}^n \xi_i$, причем $M\xi_i = p$ и $D\xi_i = pq$. Значит, выполняются все условия следствия 10.1, из которого вытекает утверждение теоремы.

Следствие 10.2 называют также *теоремой Бернулли*, или *законом больших чисел в форме Бернулли*, который является частным случаем закона больших чисел в форме Чебышёва.

Теорему 10.3 и следствие 10.2 можно усилить, заменив сходимость по вероятности на более сильный вид сходимости с вероятностью 1. Приведем здесь без доказательства закон больших чисел в форме Колмогорова.

Теорема 10.4 (закон больших чисел в форме Колмогорова). Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ последовательность независимых одинаково распределенных случайных величин с $M\xi_i = m$ и $M|\xi_i| < \infty, i = 1, 2, \dots$

Тогда

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow[n \rightarrow \infty]{n.н.} m. \quad \#$$

Это утверждение называют усиленным законом больших чисел, поскольку сходимость с вероятностью 1 является более сильным видом сходимости по сравнению со сходимостью по вероятности в законе больших чисел в форме Чебышёва. В связи с этим закон больших чисел в форме Чебышёва называют слабым законом больших чисел.

Теорема 10.4 позволяет усилить следствие 10.2, заменив в его утверждении сходимость по вероятности на сходимость почти наверное.

Замечание 10.3. Следствие 10.2 является «прибором», с помощью которого можно измерять вероятность любого события A . Для оценивания $P(A)$ нужно воспроизвести случайный эксперимент n раз, подсчитать число Y_n экспериментов, в которых событие A наблюдалось, и затем аппроксимировать $P(A)$ частотой Y_n/n появления A в n экспериментах: $P(A) \approx Y_n/n$. Чем больше n , тем точнее будет оценка.

Пример 10.3. Пусть дана последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимых случайных величин, причем ряд распределения случайной величины ξ_n представлен в табл. 10.1. Покажем, что к этой последовательности применим закон больших чисел в форме Чебышёва. Для этого вычислим дисперсию $D\xi_n$. Имеем

ξ_n	$-5n$	0	$5n$
P	$\frac{1}{2n^2}$	$1 - \frac{1}{n^2}$	$\frac{1}{2n^2}$

Таблица 10.1.

$$M\xi_n = (-5n) \cdot \frac{1}{2n^2} + 0 \cdot \left(1 - \frac{1}{n^2}\right) + 5n \cdot \frac{1}{2n^2} = 0,$$

$$D\xi_n = M\xi_n^2 - (M\xi_n)^2 = M\xi_n^2 = (-5n)^2 \frac{1}{2n^2} + 0^2 \left(1 - \frac{1}{n^2}\right) + (5n)^2 \frac{1}{2n^2} = 25.$$

Итак, дисперсии $D\xi_n$ ограничены в совокупности (например, числом 25), и к последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$ применим закон больших чисел в форме Чебышёва.

Центральная предельная теорема

Рассмотрим последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимых одинаково распределенных случайных величин, имеющих математическое ожидание $M\xi_n = m$. Предположим также, что существует дисперсия $D\xi_n = \sigma^2$. Закон больших чисел (слабый) для этой последовательности можно представить в следующей форме:

$$\frac{1}{n} \sum_{i=1}^n (\xi_i - M\xi_i) = \frac{1}{n} (S_n - nm) = \bar{\xi} - m \xrightarrow[n \rightarrow \infty]{P} 0,$$

где $S_n = \sum_{i=1}^n \xi_i$ — суммарное значение первых n случайных величин последовательности, а сходимость можно понимать как в смысле сходимости по вероятности (слабый закон больших чисел), так и в смысле сходимости с вероятностью 1 (усиленный закон больших чисел).

Однако сразу возникает вопрос: поскольку случайные величины ξ_n имеют не только математическое ожидание, но и дисперсию, то нельзя ли доказать более “тонкую” предельную теорему, позволяющую точнее описать предельное поведение распределений величин $S_n - nm$? Такая теорема существует, ее называют центральной предельной теоремой.

Теорема 10.5 (центральная предельная теорема). Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых одинаково распределенных случайных величин, $M\xi_n = m, D\xi_n = \sigma^2, S_n = \sum_{i=1}^n \xi_i$. Тогда

$$P \left\{ \frac{S_n - nm}{\sqrt{n}\sigma^2} < x \right\} \xrightarrow[n \rightarrow \infty]{} \Phi(x),$$

где $\Phi(x)$ — функция стандартного нормального распределения.

Ценность центральной предельной теоремы заключается в следующем.

Во-первых, 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.

Во-вторых, 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 $\mathbf{P}\left\{\frac{S_n - nm}{\sqrt{n}\sigma^2} < x\right\}$ 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 10.5 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.

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

Следствие 10.3 (интегральная теорема Муавра — Лапласа). Обозначим через Y_n суммарное число успехов в n испытаниях по схеме Бернулли с вероятностью успеха p и вероятностью неудачи $q = 1 - p$. Тогда с ростом n последовательность функций распределения случайных величин $(Y_n - np)/\sqrt{npq}$ сходится к функции стандартного нормального распределения, т.е.

$$\mathbf{P}\left\{\frac{Y_n - np}{\sqrt{npq}} < x\right\} \xrightarrow{n \rightarrow \infty} \Phi(x).$$

ДОКАЗАТЕЛЬСТВО. Пусть ξ_i — число успехов в i -м испытании. Тогда $\mathbf{M}\xi_i = p$, $\mathbf{D}\xi_i = pq$. Представляя Y_n в виде $Y_n = \xi_1 + \dots + \xi_n$ и используя центральную предельную теорему, приходим к утверждению следствия.

Пример 10.4. Для определения неслучайной величины m (например, скорости движения некоторого объекта) делают n измерений ξ_1, \dots, ξ_n этой величины, причем i -е измерение проводят с погрешностью ξ_i , т.е. $\xi_i = m + \xi_i$. Предположим, что погрешности измерений являются независимыми и одинаково распределенными случайными величинами с математическим ожиданием $\mathbf{M}\xi_i = 0$ (отсутствуют систематические погрешности наблюдений) и дисперсией $\mathbf{D}\xi_i = \sigma^2$.

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

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

$$\begin{aligned} \mathbf{P}\{|\bar{\xi} - m| < \varepsilon\} &= \mathbf{P}\left\{|\zeta| < \frac{\varepsilon\sqrt{n}}{\sigma}\right\} = \mathbf{P}\left\{-\frac{\varepsilon\sqrt{n}}{\sigma} < \zeta < \frac{\varepsilon\sqrt{n}}{\sigma}\right\} \approx \\ &\approx \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 = 2\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right). \end{aligned}$$

Пример 10.5. Согласно замечанию 10.3 частота $\hat{p} = Y_n/n$ появления события A в n экспериментах стремится к вероятности $p = \mathbf{P}(A)$ этого события: $p \approx \hat{p}$.

Оценим вероятность того, что \hat{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 $\hat{p} = \frac{1}{n} \sum_{i=1}^n \xi_i$, $\mathbf{M}\xi_i = p$, $\mathbf{D}\xi_i = pq$. Поэтому из примера 10.4 вытекает, что

$$\mathbf{P}\{|\hat{p} - p| < \varepsilon\} \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) - 1 = 2\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right).$$

Лекция 11

Основные понятия выборочной теории

Начнем изучение математической статистики с обсуждения следующих вопросов: что такое математическая статистика, какие задачи она решает, какова связь между математической статистикой и теорией вероятностей и в чем состоит различие между этими дисциплинами.

Для этого сначала рассмотрим примеры статистических задач, а потом перейдем к обобщениям и строгим формулировкам.

Пример 11.1. Пусть m — неслучайная величина, подлежащая измерению, например m — период колебания маятника. Результатом измерения величины m является неслучайное число x . Предположим, что измерительный прибор измеряет m со случайной ошибкой. В силу этого число x можно интерпретировать как значение (реализацию, наблюдение) $X(\omega)$ некоторой случайной величины X , которое соответствует исходу ω случайного измерительного эксперимента. Повторив этот эксперимент, мы получим, вообще говоря, другое, отличное от x , неслучайное число. Заметим, что функция распределения F случайной величины X полностью определяется измерительным экспериментом, в том числе измерительным прибором.

Пусть эксперимент по измерению величины m повторяется в неизменном виде (при одном и том же комплексе условий) n раз и обозначим через x_1, x_2, \dots, x_n результаты этих измерений. Разумно считать, что вектор (x_1, x_2, \dots, x_n) является реализацией случайного вектора (X_1, X_2, \dots, X_n) , где каждая из координат X_i в силу неизменности комплекса условий, в которых повторяется эксперимент, имеет одну и ту же функцию распределения, совпадающую с функцией распределения F случайной величины X . Предположим также, что на результат каждого измерения не влияют результаты предыдущих измерений. В этой ситуации случайные величины X_1, X_2, \dots, X_n логично полагать независимыми.

Основной задачей эксперимента является оценивание величины m по результатам наблюдений x_1, x_2, \dots, x_n и определение точности этого оценивания. Другой важной задачей является проверка различных гипотез о m , например о том, что $m = m_0$, где m_0 некоторая известная величина (в примере с маятником m_0 может быть периодом колебания, необходимым для правильного функционирования маятника).

Пример 11.2. Пусть p — вероятность наступления некоторого события A . Для оценивания p проводится n испытаний Бернулли по наблюдению за A . Результатом этого эксперимента является вектор (x_1, x_2, \dots, x_n) , где $x_i = 1$, если событие A наблюдалось в i -м эксперименте и $x_i = 0$ в противном случае. Здесь также можно считать, что вектор (x_1, x_2, \dots, x_n) является реализацией случайного вектора (X_1, X_2, \dots, X_n) , где каждая из координат X_i имеет распределение Бернулли (см. определение 5.2) с параметром p , т.е. $X_i = 1$ с вероятностью p и $X_i = 0$ с вероятностью $1 - p$, $i = 1, \dots, n$. В схеме Бернулли испытания являются независимыми, поэтому случайные величины X_1, X_2, \dots, X_n будут независимыми. Здесь как и в примере ?? одной из задач является оценивание величины p по результатам наблюдений x_1, x_2, \dots, x_n , а также определение точности этого оценивания. Другая распространенная задача заключается в проверке различных гипотез о p , например о том, что $p = p_0$, где p_0 некоторая известная величина (например, можно проверить гипотезу о том, что вероятность выпадения герба при подбрасывании монеты равна $p_0 = 1/2$, т.е. монета симметричная). #

Обобщая эти примеры, сформулируем принципиальную задачу математической статистики следующим образом. Есть случайная величина X , функция распределения F которой

частично или полностью неизвестна. Нужно по наблюдениям этой случайной величины получить всю возможную информацию о F , т.е. о вероятностных свойствах X .

Как видно из этих примеров, математическая статистика разрабатывает методы анализа результатов наблюдений случайной величины и определения по этим наблюдениям характеристик наблюдаемой случайной величины, а также вероятностей различных событий, связанных с этой случайной величиной. Другими словами, основными задачами математической статистики являются разработка методов нахождения оценок и исследования точности их приближения к оцениваемым характеристикам и разработка методов проверки гипотез об этих характеристиках.

Для этих целей математическая статистика использует математический аппарат теории вероятностей и в этом смысле ее можно считать частью теории вероятностей.

С другой стороны, задачи, решаемые математической статистикой, в каком-то смысле противоположны задачам теории вероятностей. В типичной задаче по теории вероятностей по известным вероятностям простых событий нужно определить вероятности сложных событий. Например, по известной вероятности $1/2$ выпадения герба при подбрасывании монеты нужно вычислить вероятность появления трех гербов в пяти подбрасываниях. Но откуда известно, что вероятность выпадения герба равна $1/2$, ведь возможно монета несимметрична. Математическая статистика как раз и разрабатывает методы, которые позволяют по наблюдениям оценить вероятность выпадения герба и проверить гипотезу о симметричности монеты.

В задачах, изучаемых в математической статистике, предполагается, что эксперимент по наблюдению над случайной величиной хотя бы теоретически может быть повторен сколько угодно раз в одних и тех же условиях. Под словами “в одних и тех же условиях” понимается, что распределение случайной величины X_i , $i = 1, 2, \dots, n$ не зависит от номера испытания i и совпадает с распределением X . В этом случае принято говорить о **независимых повторных экспериментах (испытаниях)** или о **независимых повторных наблюдениях** над случайной величиной X .

Выборка

Определение 11.1. Совокупность независимых случайных величин X_1, \dots, X_n , каждая из которых имеет то же распределение, что и случайная величина X , будем называть **выборкой** и записывать $\vec{X}_n = (X_1, \dots, X_n)$ (иногда просто X_1, \dots, X_n). При этом число n называют **объемом выборки**, а случайные величины X_i — **элементами выборки**. Любое возможное значение $\vec{x}_n = (x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$ выборки \vec{X}_n будем называть **реализацией выборки** \vec{X}_n . #

Замечание 11.1. Таким образом, выборка является математической моделью независимых наблюдений случайной величины, проводимых в одинаковых условиях. Поэтому в дальнейшем, говоря о независимых наблюдениях X_1, \dots, X_n случайной величины X будем иметь ввиду выборку, подразумевая, что случайные величины X_1, \dots, X_n не только независимы, но и одинаково распределены, имея ту же функцию распределения, что и X . #

Реализация выборки (если это не приводит к недоразумениям) для удобства нередко отождествляется с самой выборкой, подобно тому, как, например, символ $\sin(x)$ в зависимости от контекста может обозначать и функцию “синус” и число, равное значению этой функции в точке x .

Реализацию $\vec{x}_n = \vec{X}_n(\omega)$ выборки \vec{X}_n можно интерпретировать как совокупность чисел x_1, \dots, x_n , полученных в результате проведения n повторных независимых наблюдений над случайной величиной X .

Для того чтобы подчеркнуть, что речь идет о наблюдении именно случайной величины X , а не какой-нибудь другой, обычно говорят, что \vec{X}_n — выборка из закона распределения X или из распределения X . Заметим, что в соответствии с определением 3.2 закона распределения вероятностей случайной величины и замечанием 3.2, его можно отождествлять с ее функцией распределения. Поэтому часто используется и другая терминология. Так, если $F(x)$ (или $f(x)$) — функция распределения (или плотность) случайной величины X , говорят, что \vec{X}_n — случайная выборка из $F(x)$ (или из $f(x)$).

Вариационный ряд

Одним из самых простых преобразований статистических данных является их упорядочивание по величине. Пусть X_1, \dots, X_n — случайная выборка из распределения X , а x_1, \dots, x_n — реализация этой выборки. Обозначим через $x_{(1)}, x_{(2)}, \dots, x_{(i)}, \dots, x_{(n)}$ величины x_1, \dots, x_n , расположенные в неубывающем порядке:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(n)}. \quad (11.1)$$

В частности, $x_{(1)}$ — наименьшее, $x_{(n)}$ — наибольшее из чисел x_1, \dots, x_n .

Обозначим через $X_{(i)}$ случайную величину, которая при каждой реализации x_1, \dots, x_n выборки X_1, \dots, X_n принимает значение, равное $x_{(i)}$, $i = \overline{1, n}$.

Определение 11.2. Последовательность случайных величин $X_{(1)}, X_{(2)}, \dots, X_{(i)}, \dots, X_{(n)}$ называют **вариационным рядом выборки**. При этом $X_{(i)}$ называют i -м **членом вариационного ряда выборки**, $i = \overline{1, n}$.

Последовательность чисел $x_{(1)}, x_{(2)}, \dots, x_{(i)}, \dots, x_{(n)}$, удовлетворяющих условию (11.1), называют **реализацией вариационного ряда выборки**, число $x_{(i)} = X_{(i)}(\omega)$ называют **реализацией i -го члена вариационного ряда**, $i = \overline{1, n}$.

Так же как и в случае с выборкой, если это не приводит к недоразумениям, реализация вариационного ряда выборки для удобства часто отождествляется с самим вариационным рядом выборки.

Среди наблюдений x_1, \dots, x_n могут быть одинаковые. Так бывает либо когда наблюдаемая случайная величина X — дискретная, либо когда X — непрерывная, но ее значения при измерениях округляют.

Пусть среди наблюдений x_1, \dots, x_n есть только r различных ($r < n$). Обозначим через $z_{(1)}, \dots, z_{(r)}$ эти r различных значений, расположенных в порядке возрастания, а через n_k — число повторений значения $z_{(k)}$ среди x_1, \dots, x_n , $k = \overline{1, r}$. Очевидно, что $\sum_{k=1}^r n_k = n$.

Определение 11.3. **Статистическим рядом** выборки называют таблицу 11.1, где в первой строке расположены элементы $z_{(1)}, \dots, z_{(r)}$, а во второй — числа их повторений n_1, \dots, n_r . Отношение n_k/n называют частотой значения $z_{(k)}$, $k = \overline{1, r}$.

$z_{(1)}$	$z_{(2)}$	\dots	$z_{(r)}$
n_1	n_2	\dots	n_r

Таблица 11.1.

Пример 11.3. результате опыта получения выборка объема $n = 30$: 217, 218, 220, 219, 220, 221, 219, 220, 221, 217, 218, 219, 220, 218, 217, 220, 219, 221, 221, 220, 219, 222, 223, 220, 216, 220, 219, 220, 218, 222. Построим статистический ряд для данной выборки. Наименьшее значение в выборке $z_{(1)} = 216$, наибольшее — $z_{(8)} = 223$. Подсчитываем частоту n_k , $k = \overline{1, 8}$, каждого из восьми различных значений в выборке и строим таблицу 11.2. #

$z_{(k)}$	216	217	218	219	220	221	222	223
n_k	1	3	4	6	9	4	2	1

Таблица 11.2.

Выборочная функция распределения

Пусть X_1, \dots, X_n независимые наблюдения случайной величины X с функцией распределения $F(x)$. Обозначим через $n(x; \vec{X}_n)$ случайную величину, которая для каждого $x \in \mathbb{R}$ и каждой реализации $\vec{x}_n = (x_1, \dots, x_n)$ выборки $\vec{X}_n = (X_1, \dots, X_n)$ принимает значение $n(x; \vec{x}_n)$, равное числу элементов среди x_1, \dots, x_n , меньших x .

Отметим, что $n(x; \vec{X}_n)$ при фиксированном x является биномиальной случайной величиной с вероятностью “успеха” $p = P\{X < x\} = F(x)$, а если зафиксировать ω , то $n(x; \vec{X}_n(\omega))$ будет неслучайной неубывающей кусочно-постоянной (ступенчатой) функцией на \mathbb{R} , принимающей значения из множества $0, 1, \dots, n$.

Определение 11.4. Функцию

$$\hat{F}(x; \vec{X}_n) = \frac{n(x; \vec{X}_n)}{n}, \quad x \in \mathbb{R},$$

будем называть **выборочной функцией распределения**.

Согласно определению 11.4 при любом фиксированном x функция $\hat{F}(x; \vec{X}_n)$ есть случайная величина, которая принимает одно из значений $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, \frac{n}{n} = 1$ и $P\{\hat{F}(x; \vec{X}_n) = i/n\} = C_n^i p^i (1-p)^{n-i}$, где $p = F(x)$.

Теорема 11.1. Для любого фиксированного x последовательность случайных величин $\{\hat{F}(x; \vec{X}_n)\}$ сходится по вероятности при $n \rightarrow \infty$ к значению $F(x)$ функции распределения случайной величины X в точке x .

ДОКАЗАТЕЛЬСТВО. При любом фиксированном x выборочная функция распределения $\hat{F}(x; \vec{X}_n)$ есть частота события $\{X < x\}$. В соответствии с законом больших чисел в форме Бернулли (см. следствие 10.2), частота при $n \rightarrow \infty$ сходится по вероятности к вероятности события $\{X < x\}$. Следовательно,

$$\hat{F}(x; \vec{X}_n) \xrightarrow[n \rightarrow \infty]{P} P\{X < x\} = F(x).$$

Теорема доказана.

Замечание 11.2. Теорема 11.1 показывает принципиальную возможность сколь угодно точного восстановления теоретической функции распределения случайной величины по ее наблюдениям. Отсюда следует принципиальная возможность определения по наблюдениям случайной величины всех ее вероятностных характеристик, в частности, математического ожидания, дисперсии, вероятностей различных событий, связанных со этой случайной величиной. #

Если все выборочные значения x_1, \dots, x_n различны, то функцию $\hat{F}(x; \vec{x}_n)$, обозначив для краткости $F_n(x)$ можно записать в виде

$$F_n(x) = \begin{cases} 0, & x \leq x_{(1)}; \\ \frac{i}{n}, & x_{(i)} < x \leq x_{(i+1)}, \quad i = \overline{1, n-1}; \\ 1, & x > x_{(n)}, \end{cases}$$

т.е. в каждой точке $x_{(i)}$ функция $F_n(x)$ имеет скачок величиной $1/n$. График функции $F_n(x)$ изображен на рис. 11.1.

При больших объемах выборки n ($n > 50$) иногда производят группирование исходных данных следующим образом. Промежуток $J = [x_{(1)}, x_{(n)}]$, содержащий все выборочные значения, разбивают на m промежутков J_1, \dots, J_m , как правило, одинаковой длины Δ и таких, что каждый из них, кроме последнего, содержит левую границу, а последний содержит обе границы. Затем подсчитывают число n_i элементов выборки, попавших в i -й промежуток J_i , $i = \overline{1, m}$, $n = n_1 + \dots + n_m$, а результаты представляют в виде следующей таблицы 11.3, которую называют **интервальным статистическим рядом**.

Иногда в верхней строке таблицы 11.3 указывают не интервал, а его середину \tilde{x}_i , а в нижней строке вместо n_i записывают частоту n_i/n . Число промежутков m , на которые разбивают промежуток J , содержащий все выборочные значения, выбирают в зависимости от объема выборки n . Для ориентировочной оценки величины m можно пользоваться следующей приближенной формулой

$$m \approx \log_2 n + 1.$$

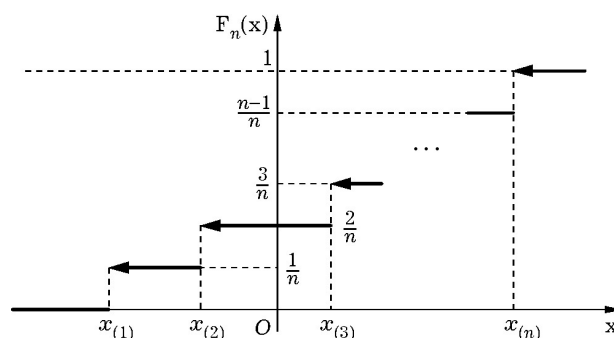


Рис 11.1.

J_1	J_2	\dots	J_m
n_1	n_2	\dots	n_m

Таблица 11.3.

Определение 11.5. График функции

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i; \\ 0, & x \notin J, \end{cases} \quad (11.2)$$

представляющий собой кусочно постоянную функцию называют **гистограммой** (см. рис. 11.2).

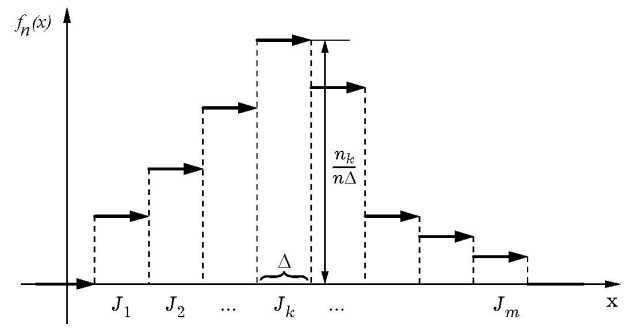


Рис 11.2

Часто гистограммой называют диаграмму, составленную из прямоугольников с основанием Δ и высотами $n_i/(n\Delta)$, $i = \overline{1, m}$. Нетрудно увидеть, что суммарная площадь всех прямоугольников, образующих такую диаграмму, равна 1, так как

$$\sum_{i=1}^m \frac{n_i}{n\Delta} \Delta = \frac{1}{n} \sum_{i=1}^m n_i = 1.$$

Площадь каждого прямоугольника n_i/n есть частота попадания элементов выборки в соответствующий интервал J_i статистического ряда.

Рассмотрим случайную величину $n_i(\vec{X}_n)/n$, которая для каждой реализации \vec{x}_n случайной выборки \vec{X}_n равна частоте n_i/n . В соответствии с законом больших чисел в форме Бернулли (см. следствие 10.2) случайная величина $n_i(\vec{X}_n)/n$ при $n \rightarrow \infty$ будет сходиться по вероятности к вероятности попадания случайной величины X в промежуток J_i , $i = \overline{1, m}$, т.е.

$$\frac{n_i(\vec{X}_n)}{n} \xrightarrow[n \rightarrow \infty]{P} P\{X \in J_i\} = \int_{J_i} f(x) dx,$$

где $f(x)$ — плотность распределения случайной величины X . Если длина Δ промежутков достаточно мала и объем выборки n велик, то с вероятностью, близкой к 1, можно утверждать, что

$$\frac{n_i}{n} \approx f(\tilde{x}_i)\Delta,$$

или

$$\frac{n_i}{n\Delta} \approx f(\tilde{x}_i),$$

где \tilde{x}_i — середина промежутка J_i , $i = \overline{1, m}$. Таким образом, при большом объеме выборки n и достаточно малом Δ с вероятностью, близкой к 1, можно считать, что $f_n(x) \approx f(x)$. Иными словами, функция $f_n(x)$ является статистическим аналогом плотности распределения $f(x)$, наблюдаемой в эксперименте случайной величины X , а гистограмма выглядит приблизительно как график плотности случайной величины X .

Пусть \vec{X}_n — независимые наблюдения случайной величины X с функцией распределения $F(x)$ (и плотностью распределения $f(x)$, если X — непрерывная случайная величина).

Определение 11.6. Случайную величину

$$\hat{\mu}_k(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i^k$$

называют **выборочным начальным моментом k -го порядка**. В частности, выборочный начальный момент первого порядка $\hat{\mu}_1(\vec{X}_n)$ называют **выборочным средним** и обозначают \bar{X} (реализацию $\hat{\mu}_1(\vec{x}_n)$ случайной величины \bar{X} будем обозначать \bar{x}). Таким образом,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Определение 11.7. Случайную величину

$$\hat{v}_k(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

называют **выборочным центральным моментом k -го порядка**. В частности, выборочный центральный момент 2-го порядка $\hat{v}_2(\vec{X}_n)$ называют **выборочной дисперсией** и обозначают $\hat{\sigma}^2(\vec{X}_n)$. Таким образом,

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Можно показать, что

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2.$$

Определение 11.8. Случайную величину $\hat{\sigma}(\vec{X}_n) = \sqrt{\hat{\sigma}^2(\vec{X}_n)}$ называют **выборочным средне-неквадратическим отклонением**.

Пусть $(X_1, Y_1), \dots, (X_n, Y_n)$ — независимые наблюдения случайного вектора (X, Y) . Случайную величину

$$\hat{K}(\vec{X}_n, \vec{Y}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

называют **выборочной ковариацией**.

Определение 11.9. Случайную величину

$$\hat{\rho}(\vec{X}_n, \vec{Y}_n) = \frac{\hat{K}(\vec{X}_n, \vec{Y}_n)}{\hat{\sigma}(\vec{X}_n) \hat{\sigma}(\vec{Y}_n)},$$

где

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \hat{\sigma}^2(\vec{Y}_n) = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

называют **выборочным коэффициентом корреляции**.

Значения выборочных моментов (неслучайные числа) будем для краткости называть теми же терминами, что и сами моменты (случайные величины).

Замечание 11.3. Опираясь на закон больших чисел, можно доказать, что при увеличении объема выборки n перечисленные выше выборочные характеристики сходятся по вероятности к соответствующим теоретическим моментам. В частности, при $n \rightarrow \infty$

$$\bar{X} \xrightarrow[n \rightarrow \infty]{P} MX, \quad \hat{\sigma}^2(\vec{X}_n) \xrightarrow[n \rightarrow \infty]{P} DX, \quad \hat{K}(\vec{X}_n, \vec{Y}_n) \xrightarrow[n \rightarrow \infty]{P} \text{cov}(X, Y), \quad \hat{\rho}(\vec{X}_n, \vec{Y}_n) \xrightarrow[n \rightarrow \infty]{P} \rho(X, Y).$$

Поэтому, если n велико, то

$$MX \approx \bar{x}, \quad DX \approx \hat{\sigma}^2(\vec{x}_n), \quad \text{cov}(X, Y) \approx \hat{K}(\vec{x}_n, \vec{x}_n), \quad \rho(X, Y) \approx \hat{\rho}(\vec{x}_n, \vec{x}_n).$$

Лекция 12

Точечные оценки

Одной из задач математической статистики является *оценка* неизвестных параметров, от которых зависит распределение случайной величины или случайного вектора.

Предположим, что функция распределения случайной величины X принадлежит множеству $\{F(x; \theta) : \theta \in \Theta\}$, где вид $F(x; \theta)$ функции распределения задан, а вектор параметров $\theta = (\theta_1, \dots, \theta_r)$ неизвестен. О параметре θ известно только то, что он принадлежит некоторому известному множеству $\Theta \in \mathbb{R}^r$. В этом случае говорят, что функция распределения $F(x; \theta)$ известна с точностью до параметра θ , а модель $\{F(x; \theta) : \theta \in \Theta\}$ называют *параметрической моделью*. Отметим, что параметр θ неслучаен, и что все вероятностные характеристики случайной величины X , в частности математическое ожидание и дисперсия, вообще говоря, зависят от параметра θ .

Требуется найти оценку параметра θ или некоторой функции от θ (например, математического ожидания, дисперсии) по независимым наблюдениям $\vec{X}_n = (X_1, \dots, X_n)$ случайной величины X .

Пример 12.1. Предположим, что о случайной величине X известно только то, что она является нормальной, но ее параметры (см. определение 5.7 и теорему 5.2) $\theta_1 = \mathbf{M}X$ и $\theta_2 = \mathbf{D}X$ неизвестны. Другими словами, известно, что функция распределения случайной величины X имеет вид

$$F(x; \theta) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(t - \theta_1)^2}{2\theta_2}\right) dt,$$

где параметр $\theta = (\theta_1, \theta_2)$ принадлежит множеству $\Theta = \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 \in \mathbb{R}, \theta_2 > 0\}$. Нужно оценить параметр θ по результатам наблюдений случайной величины X . #

В математической статистике существует традиция, согласно которой оценки параметра обозначаются тем же символом, что и параметр, но с добавлением символа $\hat{\cdot}$, \sim , $*$ или какого-нибудь аналогичного (обычно указанных трех символов бывает достаточно). Например, оценка параметра θ обычно обозначается $\hat{\theta}$, $\tilde{\theta}$ или θ^* , а оценка какого-нибудь другого параметра, например α , обозначается $\hat{\alpha}$, $\tilde{\alpha}$ или α^* и т.д. В дальнейшем будем следовать этой традиции.

Определение 12.1. *Точечной оценкой* параметра $\theta \in \Theta$ назовем любую функцию от наблюдений $\hat{\theta}(\vec{X}_n)$. #

Как видно из этого определения, точечную оценку построить легко, достаточно взять любую функцию от наблюдений, например

$$\hat{\theta}(\vec{X}_n) = \bar{X}, \quad \tilde{\theta}(\vec{X}_n) = \max(X_1, \dots, X_n), \quad \theta^* = X_1$$

и т.д. Труднее построить хорошую оценку, т.е. такую, которая принимала бы значения, близкие к θ . Как следует из теоремы 11.1 о сходимости выборочной функции распределения к теоретической, принципиально должны существовать оценки, которые с ростом n , т.е. с ростом информации о X будут принимать значения все ближе и ближе к θ . Это свойство оценки формализуется в следующем определении.

Определение 12.2. Оценку $\hat{\theta}(\vec{X}_n)$ параметра $\theta \in \Theta$ называют *состоятельной*, если с ростом объема выборки n она сходится по вероятности к оцениваемому параметру θ , т.е.

$$\hat{\theta}(\vec{X}_n) \xrightarrow[n \rightarrow \infty]{P} \theta.$$

#

Как следует из этого определения, состоятельность — важнейшее свойство любой оценки, несостоятельные оценки, по большому счету, никому не нужны.

Определение 12.3. Оценку $\hat{\theta}(\vec{X}_n)$ параметра $\theta \in \Theta$ называют *несмещенной*, если ее математическое ожидание совпадает с θ , т.е. $M\hat{\theta}(\vec{X}_n) = \theta$ для любого n . #

Несмещенность — желательное, но не обязательное свойство. Оно показывает, что оценка как случайная величина в среднем совпадает с оцениваемым параметром, т.е. систематического отклонения $M\hat{\theta}(\vec{X}_n)$ от θ нет. Большинство распространенных оценок не являются несмещенными, однако обладают свойством *асимптотической несмещенности*, которое означает, что смещение стремится к нулю с ростом числа наблюдений, т.е. $M\hat{\theta}(\vec{X}_n) - \theta \rightarrow 0$ при $n \rightarrow \infty$.

Замечание 12.1. Определения 12.2 и 12.3 иллюстрируют удобство взгляда на выборку \vec{X}_n как на случайный объект. В результате такого подхода оценка $\hat{\theta}(\vec{X}_n)$ является случайной величиной и становятся осмысленными такие важные понятия как состоятельность и несмещенность.

Теорема 12.1. Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения случайной величины X (выбока) и существует DX .

Тогда выборочное среднее $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ является несмещенной и состоятельной оценкой математического ожидания случайной величины X .

ДОКАЗАТЕЛЬСТВО. По определению выборки случайные величины $X_i, i = \overline{1, n}$ являются независимыми случайными величинами и имеют то же самое распределение, что и случайная величина X . Следовательно, $MX_i = MX$ и $DX_i = DX, i = \overline{1, n}$.

В силу свойств математического ожидания имеем

$$M\bar{X} = M\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n MX_i = \frac{1}{n} \sum_{i=1}^n MX = \frac{1}{n} nMX = MX,$$

что и доказывает несмещенность оценки \bar{X} .

Далее, поскольку последовательность X_1, \dots, X_n состоит из независимых одинаково распределенных случайных величин с конечной дисперсией, то в силу закона больших чисел в форме Чебышева для любого $\varepsilon > 0$

$$P\{|\bar{X} - MX| < \varepsilon\} \rightarrow 1, \quad n \rightarrow \infty,$$

т.е. оценка \bar{X} сходится по вероятности к оцениваемому параметру, а это и означает ее состоятельность. #

Пример 12.2. Можно показать, что случайная величина

$$S^2(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

является несмещенной и состоятельной оценкой дисперсии DX произвольной случайной величины X . Ее называют *исправленной выборочной дисперсией*.

Замечание 12.2. Можно доказать, что выборочные начальные и центральные моменты являются состоятельными оценками соответствующих теоретических моментов (см. определения 4.4 и 4.5 начальных и центральных теоретических моментов), если только теоретические моменты существуют. Однако эти оценки, кроме \bar{X} , являются смещенными.

Пример 12.3. Пусть X_1, \dots, X_n — независимые наблюдения нормальной случайной величины X с неизвестным средним значением θ и известной дисперсией σ^2 .

Оценка $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) = X_1$ является несмещенной для θ , ибо $\mathbf{M}X_1 = \mathbf{M}X = \theta$, но не является состоятельной, так как, во-первых, X_1 не зависит от объема выборки и, следовательно, ее распределение не меняется с ростом n , а во-вторых, из теорем 5.1 и 5.2 вытекает, что

$$\mathbf{P}\{|X_1 - \theta| < \varepsilon\} = \frac{2}{\sigma\sqrt{2\pi}} \int_0^\varepsilon e^{-\frac{t^2}{2\sigma^2}} dt \neq 1.$$

12.1 Метод моментов

Метод моментов был предложен английским статистиком К. Пирсоном и является одним из первых общих методов оценивания. Он состоит в следующем.

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ независимые наблюдения случайной величины X , функция распределения которой $F(x; \theta)$ известна с точностью до вектора параметров $\theta = (\theta_1, \dots, \theta_r)$. Требуется найти *оценку* параметра θ по случайной выборке \vec{X}_n .

Обозначим (см. определения 4.4 и 4.5) через $\mu_k(\theta) = \mathbf{M}(X^k)$, $v_k(\theta) = \mathbf{M}((X - \mathbf{M}X)^k)$ соответственно начальный и центральный моменты порядка k , $k = 1, 2, \dots$, которые в рассматриваемой нами параметрической модели также будут зависеть от параметра θ .

Определение 12.4. *Оценкой $\hat{\theta}(\vec{X}_n) = (\hat{\theta}_1(\vec{X}_n), \dots, \hat{\theta}_r(\vec{X}_n))$ параметра θ методом моментов* назовем решение любой системы из r уравнений вида

$$\begin{cases} \hat{\mu}_{i_\alpha}(\vec{X}_n) = \mu_{i_\alpha}(\theta), & \alpha = \overline{1, k}, \\ \hat{v}_{j_\beta}(\vec{X}_n) = v_{j_\beta}(\theta), & \beta = \overline{1, l}, \end{cases}, \quad k + l = r, \quad (12.1)$$

относительно неизвестных $\theta_1, \dots, \theta_r$. Индексы i_α и j_β выбирают таким образом, чтобы эта система уравнений решалась как можно проще (обычно так бывает, если i_α и j_β невелики, например 1, 2 и т.д.). #

Можно показать, что при условии непрерывной зависимости решения этой системы от $\hat{\mu}_{i_\alpha}$ и \hat{v}_{j_β} , оценка $\hat{\theta}(\vec{X}_n)$, полученная методом моментов, является состоятельной. При этом уравнения (12.1) во многих случаях просты и их решение не вызывает больших вычислительных сложностей.

Понятно, что метод моментов не применим, когда моменты случайной величины нужного порядка не существуют. Например, если X имеет *распределения Коши* (см. пример 4.2), то не существует даже начальный момент первого порядка — математическое ожидание $\mathbf{M}X$.

Пример 12.4. Пусть (X_1, \dots, X_n) независимые наблюдения случайной величины X , равномерно распределенной в интервале (a, b) с неизвестными параметрами a и b .

Найдем методом моментов точечные оценки этих параметров.

Известно (см. определение 5.5 и примеры 5.7 и 5.8), что для равномерно распределенной случайной величины X

$$\mathbf{M}X = \frac{a+b}{2}, \quad \mathbf{D}X = \frac{(b-a)^2}{12}.$$

Выборочное среднее \bar{X} и выборочная дисперсия $\hat{\sigma}^2(\vec{X}_n)$ вычисляются по формулам

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Составляем систему двух уравнений

$$\begin{cases} \frac{a+b}{2} = \bar{X}, \\ \frac{(b-a)^2}{12} = \hat{\sigma}^2(\vec{X}_n). \end{cases}$$

Решая систему, получаем

$$b = \bar{X} + \sqrt{3} \hat{\sigma}(\vec{X}_n), \quad a = \bar{X} - \sqrt{3} \hat{\sigma}(\vec{X}_n).$$

Окончательно имеем

$$\hat{b}(\vec{X}_n) = \bar{X} + \sqrt{3} \hat{\sigma}(\vec{X}_n), \quad \hat{a}(\vec{X}_n) = \bar{X} - \sqrt{3} \hat{\sigma}(\vec{X}_n).$$

12.2 Метод максимального правдоподобия

Одним из наиболее универсальных методов оценивания параметров является **метод максимального правдоподобия** (предложенный Р. Фишером), суть которого состоит в следующем.

Пусть \vec{X}_n — независимые наблюдения случайной величины X . Если X непрерывна, что через $P(x; \theta)$ обозначим ее плотность $f(x; \theta)$, а если X дискретна, то по определению положим $P(x; \theta) = \mathbf{P}\{X = x\}$ для любого $x \in \mathbb{R}$:

$$P(x; \theta) = \begin{cases} f(x; \theta), & \text{если } X \text{ непрерывная случайная величина;} \\ \mathbf{P}\{X = x\}, & \text{если } X \text{ дискретная случайная величина.} \end{cases}$$

Определение 12.5. Функцию $L(\vec{X}_n; \theta)$, определенную равенством

$$L(\vec{X}_n; \theta) = \prod_{i=1}^n P(X_i; \theta), \quad (12.2)$$

и рассматриваемую как функцию от θ при фиксированных X_1, \dots, X_n , называют **функцией правдоподобия**, а функцию

$$l(\vec{X}_n; \theta) = \ln L(\vec{X}_n; \theta) = \sum_{i=1}^n \ln P(X_i; \theta) \quad (12.3)$$

— **логарифмической функцией правдоподобия**.

Определение 12.6. **Оценкой максимального правдоподобия** $\hat{\theta}(\vec{X}_n)$ параметра θ называют точку максимума функции правдоподобия (12.2). Другими словами для любой выборки \vec{X}_n случайная величина $\hat{\theta}(\vec{X}_n)$ удовлетворяют условию

$$L(\vec{X}_n; \hat{\theta}(\vec{X}_n)) = \max_{\theta \in \Theta} L(\vec{X}_n; \theta). \quad \#$$

Если функция $L(\vec{X}_n; \theta)$ дифференцируема как функция аргумента θ при любом \vec{X}_n и максимум $L(\vec{X}_n; \theta)$ достигается во внутренней точке из Θ , то оценка максимального правдоподобия в случае скалярного параметра $\theta \in \mathbb{R}$ удовлетворяет уравнению (необходимому условию экстремума)

$$\frac{dL(\vec{X}_n; \theta)}{d\theta} = 0, \quad (12.4)$$

а в случае векторного параметра $\theta \in \mathbb{R}^r$ — системе уравнений

$$\frac{\partial L(\vec{X}_n; \theta)}{\partial \theta_k} = 0, \quad k = \overline{1, r}. \quad (12.5)$$

Из-за того, что \ln является монотонно возрастающей функцией, функции $L(\vec{X}_n; \theta)$ и $l(\vec{X}_n; \theta)$ достигают максимума в одной и той же точке. Но дифференцировать произведение в правой части (12.2) намного труднее, чем дифференцировать сумму в правой части (12.3). Поэтому уравнение (12.4) обычно заменяется уравнением

$$\frac{dl(\vec{X}_n; \theta)}{d\theta} = 0,$$

а система (12.5) — системой

$$\frac{\partial l(\vec{X}_n; \theta)}{\partial \theta_k} = 0, \quad k = \overline{1, r}. \quad (12.6)$$

Уравнения (12.4)–(12.6) называют **уравнениями правдоподобия**. Для наиболее важных семейств распределений $f(x; \theta)$ уравнение правдоподобия имеет единственное решение $\hat{\theta} = \hat{\theta}_1, \dots, \hat{\theta}_r$. Во многих случаях уравнения (12.4)–(12.6), являющиеся, как правило, нелинейными, приходится решать численными методами.

Пример 12.5. Пусть \vec{X}_n — случайная выборка из $\mathcal{N}(\theta_1, \theta_2)$. Методом максимального правдоподобия найдем оценку вектора параметров $\theta = (\theta_1; \theta_2)$.

В этом случае плотность равна

$$f(x; \theta) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x - \theta_1)^2}{2\theta_2}\right),$$

поэтому

$$P(x; \theta) = f(x; \theta) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x - \theta_1)^2}{2\theta_2}\right),$$

функция правдоподобия есть

$$L(\vec{X}_n; \theta) = \prod_{i=1}^n P(X_i; \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(X_i - \theta_1)^2}{2\theta_2}\right) = \frac{1}{(\sqrt{2\pi\theta_2})^n} \exp\left(-\frac{1}{2\theta_2} \sum_{i=1}^n (X_i - \theta_1)^2\right)$$

а логарифмическая функция правдоподобия имеет вид

$$l(\vec{X}_n; \theta) = -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \theta_2 - \frac{1}{2\theta_2} \sum_{i=1}^n (X_i - \theta_1)^2.$$

Поскольку число неизвестных параметров $r = 2$, система уравнений правдоподобия (12.6) будет состоять из двух уравнений:

$$\begin{cases} \frac{\partial}{\partial \theta_1} l(\vec{X}_n; \theta) = \frac{1}{\theta_2} \sum_{i=1}^n (X_i - \theta_1) = 0, \\ \frac{\partial}{\partial \theta_2} l(\vec{X}_n; \theta) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^n (X_i - \theta_1)^2 = 0. \end{cases}$$

Первое уравнение системы равносильно уравнению

$$\sum_{i=1}^n (X_i - \theta_1) = 0,$$

откуда $\theta_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$. Подставляя $\theta_1 = \bar{X}$ во второе уравнение, получим

$$\theta_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Таким образом,

$$\hat{\theta}_1(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\theta}_2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Следовательно, оценками максимального правдоподобия для математического ожидания $\mathbf{M}X = \theta_1$ и дисперсии $\mathbf{D}X = \theta_2$ нормальной случайной величины являются соответственно выборочное среднее

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

и выборочная дисперсия

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad \#$$

Лекция 13

Интервальное оценивание

13.1 Некоторые важные вероятностные распределения

Для построения интервальных оценок нам понадобятся три важных распределения вероятностей: распределение Стьюдента, χ^2 -распределение и распределение Фишера. В их определение входят гамма и бета функции, поэтому начнем с определения этих функций.

Определение 13.1. Определим гамма-функцию $\Gamma(p)$ и бета-функцию $B(x,y)$ по формулам

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt, \quad z > 0,$$

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad x > 0, y > 0. \quad \#$$

Составить представление об этих функциях помогут следующие соотношения. Справедливы рекуррентные формулы

$$\Gamma(z+1) = z\Gamma(z), \quad B(x,y) = B(x+1,y) + B(x,y+1).$$

Если n и k целые неотрицательные числа, то

$$\Gamma(n+1) = n!, \quad C_n^k = \frac{1}{(n+1)B(n-k+1, k+1)}.$$

Определение 13.2. Случайную величину с плотностью

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

называют случайной величиной, имеющей **распределение χ^2 (хи-квадрат)** или **χ^2 -распределение** с n степенями свободы. $\#$

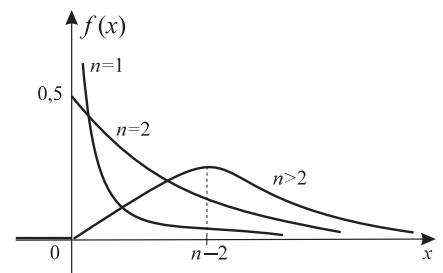


Рис 13.1

График функции $f(x)$ (рис. 13.1) асимметричен и начиная с $n > 2$ имеет один максимум в точке $x = n - 2$.

Определение 13.3. Случайную величину с плотностью

$$f(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}, \quad -\infty < x < \infty, \quad (13.1)$$

называют случайной величиной, имеющей **распределение Стьюдента** с n степенями свободы. $\#$

График плотности $f(x)$ (рис. 13.2) симметричен при всех $n = 1, 2, \dots$ относительно оси ординат.

Обозначим через $t_p(n)$ квантиль уровня p распределения Стюдента с n степенями свободы (см. определение квантили 4.6).

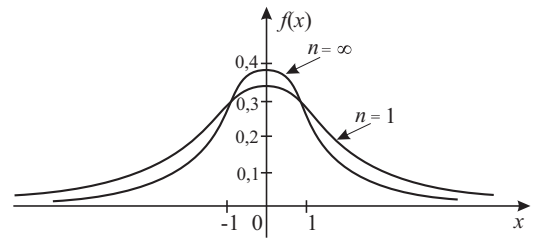


Рис 13.2

Теорема 13.1. Квантиль $t_p(n)$ уровня p распределения Стюдента с n степенями свободы удовлетворяет соотношению

$$t_p(n) = -t_{1-p}(n). \quad (13.2)$$

ДОКАЗАТЕЛЬСТВО. Согласно определению квантили 4.6 и формуле (4.5) в замечании 4.2 для любого $p \in (0, 1)$

$$\int_{-\infty}^{t_p(n)} f(x) dx = p,$$

где $f(x)$ — плотность распределения Стюдента (13.1). Далее, т.к. эта функция четная ($f(-x) = f(x) \forall x \in \mathbb{R}$), то воспользовавшись хорошо известными свойствами интеграла и свойством плотности $\int_{-\infty}^{+\infty} f(y) dy = 1$ (свойство 3 теоремы 3.2), получим

$$\int_{-\infty}^{-t_p(n)} f(x) dx = \left| \begin{matrix} y = -x \\ dy = -dx \end{matrix} \right| = - \int_{+\infty}^{t_p(n)} f(-y) dy = \int_{t_p(n)}^{+\infty} f(y) dy = \int_{-\infty}^{+\infty} f(y) dy - \int_{-\infty}^{t_p(n)} f(y) dy = 1 - p,$$

т.е. $t_{1-p}(n) = -t_p(n)$. #

Можно показать (опираясь, в частности, на второй замечательный предел), что при $n \rightarrow \infty$ плотность распределения Стюдента (13.1) стремится к плотности распределения вероятности $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ стандартной нормальной случайной величины с функцией распределения $\Phi(x)$ вида (5.5) (распределения $\mathcal{N}(0,1)$) в каждой точке $x \in \mathbb{R}$, а квантиль распределения Стюдента уровня p с n степенями свободы стремится к квантили уровня p стандартного нормального распределения.

Определение 13.4. Случайную величину с плотностью

$$f(x) = \begin{cases} \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{\left(1 + \frac{mx}{n}\right)^{\frac{m+n}{2}}}, & x \geq 0; \\ 0, & x < 0, \end{cases}$$

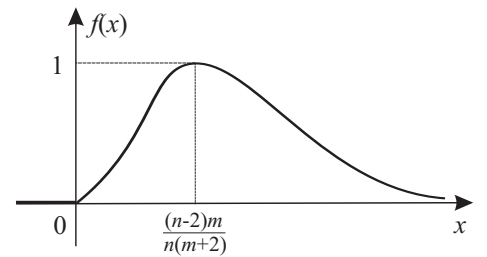


Рис 13.3

назовем случайной величиной, имеющей **распределение Фишера** (говорят также **F-распределение**, **распределение Снедекора**) с числом степеней свободы m и n . #

График функции $f(x)$ (см. рис. 13.3) асимметричен и при $n > 2$ имеет максимум в точке $\frac{(n-2)m}{(m+2)n}$, близкой к единице при больших значениях m и n .

Для квантилей $F_p(m, n)$ распределения Фишера уровня p с m и n степенями свободы имеет место соотношение

$$F_p(m, n) = \frac{1}{F_{1-p}(n, m)}.$$

По поводу вычисления квантилей см. замечание 4.3.

Справедливы следующие теоремы.

Теорема 13.2. Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{N}(\mu, \sigma^2)$, т.е. X_1, \dots, X_n — независимые одинаково распределенные нормальные случайные величины с математическим ожиданием μ и дисперсией σ^2 .

Тогда:

- 1) выборочное среднее $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ и исправленная выборочная дисперсия $S^2(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ являются независимыми случайными величинами;
- 2) случайная величина $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$ имеет стандартное нормальное распределение, т.е. является нормальной с нулевым математическим ожиданием и единичной дисперсией;
- 3) случайная величина $\frac{\sqrt{n}(\bar{X} - \mu)}{S(\vec{X}_n)}$ имеет распределение Стьюдента с $n - 1$ степенью свободы;
- 4) случайная величина $\frac{(n-1)S^2(\vec{X}_n)}{\sigma^2}$ имеет χ^2 -распределение с $n - 1$ степенью свободы.

Замечание 13.1. Случайные величины \bar{X} и $S^2(\vec{X}_n)$ определяются через одни и те же случайные величины X_1, \dots, X_n . Более того, в выражение для $S^2(\vec{X}_n)$ входит \bar{X} . Тем удивительнее, что \bar{X} и $S^2(\vec{X}_n)$ независимы.

Теорема 13.3. Пусть $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$ — две независимые выборки из распределений $\mathcal{N}(\mu_1, \sigma_1^2)$ и $\mathcal{N}(\mu_2, \sigma_2^2)$ соответственно,

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad S^2(\vec{X}_m) = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2,$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad S^2(\vec{Y}_n) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

$$S(\vec{X}_m, \vec{Y}_n) = \sqrt{\frac{(m-1)S^2(\vec{X}_m) + (n-1)S^2(\vec{Y}_n)}{m+n-2}}.$$

Тогда:

- 1) случайная величина

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

имеет стандартное нормальное распределение;

- 2) случайная величина $\frac{S^2(\vec{X}_m)}{S^2(\vec{Y}_n)}$ имеет распределение Фишера с числом степеней свободы $m - 1$ и $n - 1$;

- 3) если $\sigma_1 = \sigma_2$, то случайная величина

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

имеет распределение Стьюдента с $m + n - 2$ степенями свободы.

Замечание 13.2. Случайные величины, определенные в утверждениях 2)–4) теоремы 13.2 и в утверждениях 1)–3) теоремы 13.3 выражаются через X_1, \dots, X_n и формально зависят от них, в частности от их математического ожидания μ и дисперсии σ^2 . Ценность этих теорем в том, что определенные в них случайные величины имеют плотности распределения вероятностей (см. определения 13.2–13.4), которые не зависят от μ и σ^2 , а зависят только от количества наблюдений m и n . Следовательно, от μ и σ^2 не будут зависеть и квантили этих случайных величин.

13.2 Определение доверительного интервала

При оценивании неизвестных параметров наряду с рассмотренными выше точечными оценками используются также интервальные оценки. В отличие от точечной оценки интервальная оценка позволяет получить не только оценку параметра, но и охарактеризовать (в некотором вероятностном смысле) точность этого оценивания. Для простоты будем предполагать оцениваемый параметр одномерным.

Определение 13.5. Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения случайной величины или случайного вектора X с функцией распределения $F(x; \theta)$, зависящей от скалярного параметра θ , значение которого неизвестно. Предположим, что для параметра θ построен интервал $(\underline{\theta}(\vec{X}_n), \bar{\theta}(\vec{X}_n))$, где $\underline{\theta}(\vec{X}_n)$ и $\bar{\theta}(\vec{X}_n)$ являются такими функциями случайной выборки \vec{X}_n , что для некоторого $\gamma \in (0, 1)$ выполняется равенство

$$\mathbf{P} \left\{ \underline{\theta}(\vec{X}_n) < \theta < \bar{\theta}(\vec{X}_n) \right\} = \gamma. \quad (13.3)$$

В этом случае интервал $(\underline{\theta}(\vec{X}_n), \bar{\theta}(\vec{X}_n))$ называют **доверительным интервалом** для параметра θ с **коэффициентом доверия** γ или **γ -доверительным интервалом**, а случайные величины $\underline{\theta}(\vec{X}_n)$ и $\bar{\theta}(\vec{X}_n)$ соответственно **нижней** и **верхней границами** доверительного интервала. #

Доверительный интервал $(\underline{\theta}(\vec{X}_n), \bar{\theta}(\vec{X}_n))$ представляет собой интервал со случайными границами, который с заданной вероятностью γ накрывает неизвестное неслучайное истинное значение параметра θ .

Наряду с термином “коэффициент доверия” широко используют также термины **доверительная вероятность** и **уровень доверия**. На практике коэффициент доверия γ чаще всего выбирают близким к 1, например, равным 0,9, 0,95 или 0,99.

В некоторых ситуациях (например, при рассмотрении дискретных случайных величин) вместо равенства (13.3) удастся обеспечить лишь неравенство

$$\mathbf{P} \left\{ \underline{\theta}(\vec{X}_n) < \theta < \bar{\theta}(\vec{X}_n) \right\} \geq \gamma,$$

т.е. построить доверительный интервал для параметра θ с коэффициентом доверия, не меньшим γ .

13.3 Доверительные интервалы для параметров нормальной случайной величины

Модели, описанные в этом разделе, возникают, например, в случае измерения некоторой величины μ прибором, характеристики которого известны, записаны в его техническом паспорте и заключаются в том, что ошибки измерения $(X_1 - \mu, \dots, X_n - \mu)$ являются независимыми, нормальными и с нулевой систематической составляющей. Разброс ошибок может быть как известным, так и неизвестным (дисперсия как раз и описывает разброс).

Нормальность ошибок обосновывается обычно двумя способами. Во-первых, ошибка может быть суммой нескольких ошибок, независимых друг от друга. В этом случае нормальность суммарной ошибки вытекает из центральной предельной теоремы. Во-вторых, гипотезу о нормальности ошибок можно проверить статистическими методами, например, при помощи критерия согласия К. Пирсона, который изложен в последующих лекциях.

Если измерительный прибор тестировался на большом числе измерений $\vec{X}_n = (X_1, \dots, X_n)$, то выборочную дисперсию $S^2(\vec{X}_n)$ можно отождествить с истинной дисперсией наблюдений, поскольку $S^2(\vec{X}_n)$ является ее состоятельной оценкой (см. пример 12.2).

Доверительный интервал для математического ожидания нормальной случайной величины с известной дисперсией

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения нормальной случайной величины X с неизвестным математическим ожиданием μ и известной дисперсией σ^2 . Тогда по теоре-

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \quad (13.4)$$

имеет стандартное нормальное распределение, т.е. является нормальной с математическим ожиданием 0 и дисперсией 1. Доказать это несложно. Действительно, согласно замечанию 7.4 случайная величина $X_1 + \dots + X_n$ является нормальной, а согласно свойствам 2 и 3 теоремы 7.1 и свойствам 2 и 4 теоремы 7.2 случайная величина (13.4) имеет математическое ожидание 0 и дисперсию 1.

Далее, обозначим через u_p квантиль уровня p стандартного нормального распределения. Согласно определению 4.6 $\Phi(u_p) = p$, а по свойству 4 теоремы 5.1 $-u_p = u_{1-p}$ для любого $p \in (0, 1)$. Поэтому для любого $\gamma \in (0, 1)$

$$\begin{aligned} \mathbf{P} \left\{ -u_{\frac{1+\gamma}{2}} < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u_{\frac{1+\gamma}{2}} \right\} &= \Phi(u_{\frac{1+\gamma}{2}}) - \Phi(-u_{\frac{1+\gamma}{2}}) = \\ &= \Phi(u_{\frac{1+\gamma}{2}}) - (1 - \Phi(u_{\frac{1+\gamma}{2}})) = 2\Phi(u_{\frac{1+\gamma}{2}}) - 1 = 2\frac{1+\gamma}{2} - 1 = \gamma. \end{aligned} \quad (13.5)$$

Умножая все части двойного неравенства в (13.5) на $-\frac{\sigma}{\sqrt{n}}$ (при этом все знаки в неравенстве изменятся на противоположные), а затем прибавляя ко всем частям получившегося неравенства \bar{X} , получим

$$\begin{aligned} \mathbf{P} \left\{ -u_{\frac{1+\gamma}{2}} < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u_{\frac{1+\gamma}{2}} \right\} &= \mathbf{P} \left\{ u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} > \mu - \bar{X} > -\frac{\sigma}{\sqrt{n}} u_{\frac{1+\gamma}{2}} \right\} = \\ &= \mathbf{P} \left\{ \bar{X} + u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right\} = \mathbf{P} \left\{ \bar{X} - u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right\} = \gamma, \end{aligned}$$

т.е. случайный интервал

$$\left(\bar{X} - u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right) \quad (13.6)$$

накрывает неизвестное математическое ожидание μ с необходимой доверительной вероятностью γ . Таким образом случайные величины $\underline{\theta}(\vec{X}_n)$ и $\bar{\theta}(\vec{X}_n)$ из (13.3) в данном случае обозначаемые соответственно $\underline{\mu}(\vec{X}_n)$ и $\bar{\mu}(\vec{X}_n)$ имеют вид

$$\underline{\mu}(\vec{X}_n) = \bar{X} - u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{\mu}(\vec{X}_n) = \bar{X} + u_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}.$$

Замечание 13.3. Можно показать, что доверительный интервал, задаваемый формулой (13.6), имеет наименьшую длину среди всех доверительных интервалов уровня доверия γ для математического ожидания нормальной случайной величины с известной дисперсией и в этом смысле является наилучшим. #

Видно, что длина доверительного интервала (13.6), характеризующего точность $\bar{X} - \mu$ оценивания, равна $2u_{\frac{1+\gamma}{2}} \sigma / \sqrt{n}$ и стремится к нулю при $n \rightarrow \infty$. Другими словами, чем больше наблюдений, тем информативнее и полезнее доверительная оценка. Также из неравенства $u_{\frac{1+\gamma}{2}} \sigma / \sqrt{n} \leq \varepsilon$ можно найти наименьшее количество наблюдений, позволяющих оценить μ с точностью ε (т.е. $|\bar{X} - \mu| \leq \varepsilon$) и надежностью γ . А именно,

$$n \geq \left(\frac{u_{\frac{1+\gamma}{2}} \sigma}{\varepsilon} \right)^2.$$

Доверительный интервал для математического ожидания нормальной случайной величины с неизвестной дисперсией

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения нормальной случайной величины X с неизвестным математическим ожиданием μ и неизвестной дисперсией σ^2 . Тогда по теореме 13.2 случайная величина

$$\frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n}$$

имеет *распределение Стьюдента* с $n - 1$ степенью свободы. Обозначим через $t_p(k)$ квантили уровня p распределения Стьюдента с k степенями свободы, $k = 1, 2, \dots$. Из (13.2) вытекает, что $t_p(k) = -t_{1-p}(k)$. Поэтому с учетом (4.6)

$$\begin{aligned} \mathbf{P} \left\{ -t_{\frac{1+\gamma}{2}}(n-1) < \frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n} < t_{\frac{1+\gamma}{2}}(n-1) \right\} = \\ = \mathbf{P} \left\{ t_{\frac{1-\gamma}{2}}(n-1) < \frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n} < t_{\frac{1+\gamma}{2}}(n-1) \right\} = \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = \gamma. \end{aligned}$$

Умножая все части двойного неравенства

$$-t_{\frac{1+\gamma}{2}}(n-1) < \frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n} < t_{\frac{1+\gamma}{2}}(n-1)$$

на $-\frac{S(\vec{X}_n)}{\sqrt{n}}$, а затем прибавляя \bar{X} , получим, что

$$\begin{aligned} \mathbf{P} \left\{ -t_{\frac{1+\gamma}{2}}(n-1) < \frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n} < t_{\frac{1+\gamma}{2}}(n-1) \right\} = \\ = \mathbf{P} \left\{ \bar{X} - t_{\frac{1+\gamma}{2}}(n-1) \frac{S(\vec{X}_n)}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{1+\gamma}{2}}(n-1) \frac{S(\vec{X}_n)}{\sqrt{n}} \right\} = \gamma. \end{aligned}$$

Отсюда заключаем, что нижняя и верхняя границы доверительного интервала уровня доверия γ для математического ожидания μ нормальной случайной величины с неизвестной дисперсией можно определить по формулам

$$\underline{\mu}(\vec{X}_n) = \bar{X} - t_{\frac{1+\gamma}{2}}(n-1) \frac{S(\vec{X}_n)}{\sqrt{n}}, \quad \bar{\mu}(\vec{X}_n) = \bar{X} + t_{\frac{1+\gamma}{2}}(n-1) \frac{S(\vec{X}_n)}{\sqrt{n}}. \quad (13.7)$$

Замечание 13.4. Можно показать, что доверительный интервал с границами (13.7) имеет наименьшую длину среди всех доверительных интервалов уровня доверия γ для математического ожидания нормальной случайной величины с неизвестной дисперсией.

Доверительные интервалы для дисперсии и среднеквадратического отклонения при неизвестном математическом ожидании

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения нормальной случайной величины X с неизвестным математическим ожиданием μ и неизвестной дисперсией σ^2 . Тогда по теореме 13.2 случайная величина

$$\frac{(n-1)S^2(\vec{X}_n)}{\sigma^2}$$

имеет χ^2 -распределение с $n - 1$ степенью свободы. Поэтому с учетом (4.6)

$$\mathbf{P} \left\{ \chi_{\frac{1-\gamma}{2}}^2(n-1) < \frac{(n-1)S^2(\vec{X}_n)}{\sigma^2} < \chi_{\frac{1+\gamma}{2}}^2(n-1) \right\} = \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = \gamma,$$

где $\chi_{\frac{1-\gamma}{2}}^2(n-1)$ и $\chi_{\frac{1+\gamma}{2}}^2(n-1)$ — квантили уровней $\frac{1-\gamma}{2}$ и $\frac{1+\gamma}{2}$ соответственно χ^2 -распределения с $n - 1$ степенью свободы. Деля все части двойного неравенства на $(n-1)S^2(\vec{X}_n)$, а затем переходя к неравенству для обратных величин, получим

$$\begin{aligned} \mathbf{P} \left\{ \chi_{\frac{1-\gamma}{2}}^2(n-1) < \frac{(n-1)S^2(\vec{X}_n)}{\sigma^2} < \chi_{\frac{1+\gamma}{2}}^2(n-1) \right\} = \mathbf{P} \left\{ \frac{\chi_{\frac{1-\gamma}{2}}^2(n-1)}{(n-1)S^2(\vec{X}_n)} < \frac{1}{\sigma^2} < \frac{\chi_{\frac{1+\gamma}{2}}^2(n-1)}{(n-1)S^2(\vec{X}_n)} \right\} = \\ = \mathbf{P} \left\{ \frac{(n-1)S^2(\vec{X}_n)}{\chi_{\frac{1-\gamma}{2}}^2(n-1)} > \sigma^2 > \frac{(n-1)S^2(\vec{X}_n)}{\chi_{\frac{1+\gamma}{2}}^2(n-1)} \right\} = \mathbf{P} \left\{ \frac{(n-1)S^2(\vec{X}_n)}{\chi_{\frac{1+\gamma}{2}}^2(n-1)} < \sigma^2 < \frac{(n-1)S^2(\vec{X}_n)}{\chi_{\frac{1-\gamma}{2}}^2(n-1)} \right\} = \gamma, \end{aligned}$$

т.е.

$$\left(\frac{(n-1)S^2(\vec{X}_n)}{\chi^2_{\frac{1+\gamma}{2}}(n-1)}, \frac{(n-1)S^2(\vec{X}_n)}{\chi^2_{\frac{1-\gamma}{2}}(n-1)} \right) \quad (13.8)$$

доверительный интервал для дисперсии σ^2 уровня доверия γ .

Извлекая квадратный корень из границ доверительного интервала для дисперсии, получим доверительный интервал уровня доверия γ для среднеквадратического отклонения σ нормальной случайной величины с неизвестным математическим ожиданием μ

$$\left(\sqrt{\frac{(n-1)S^2(\vec{X}_n)}{\chi^2_{\frac{1+\gamma}{2}}(n-1)}}, \sqrt{\frac{(n-1)S^2(\vec{X}_n)}{\chi^2_{\frac{1-\gamma}{2}}(n-1)}} \right).$$

Замечание 13.5. Доверительный интервал (13.8) не является самым коротким среди всех доверительных интервалов уровня доверия γ для дисперсии нормальной случайной величины с неизвестным математическим ожиданием. Но на практике он является наиболее употребительным из-за своей простоты.

13.4 Доверительные интервалы для параметров при наблюдении двух нормальных случайных величин

Среди прикладных задач встречаются такие, при решении которых приходится сравнивать математические ожидания двух случайных величин.

Например, ожидается, что добавление специальных веществ уменьшает жесткость воды. Обозначим жесткость воды до и после добавления специальных веществ через μ_1 и μ_2 . Проводятся независимые измерения жесткости воды до и после добавления специальных веществ по m и n пробам соответственно. Предполагается, что измерения производятся со случайными ошибками, систематическая ошибка отсутствует. Математической моделью этого эксперимента будем считать две независимые выборки $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$, полученные в результате наблюдений двух независимых случайных величин X и Y . Обозначим $\Delta = \mu_1 - \mu_2$. Нас интересует оценка параметра Δ , которая позволит оценить эффект изменения жесткости воды при добавления специальных веществ.

Эту задачу называют задачей сравнения двух математических ожиданий (двух средних). Для такого сравнения полезно иметь доверительный интервал для разности соответствующих математических ожиданий случайных величин X и Y . В случае нормальных случайных величин такой интервал можно легко построить на основании теоремы 13.3.

Доверительный интервал для разности математических ожиданий нормальных случайных величин с известными дисперсиями

Пусть $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$ — две независимые выборки соответственно из распределений $\mathcal{N}(\mu_1, \sigma_1^2)$ и $\mathcal{N}(\mu_2, \sigma_2^2)$ с неизвестными математическими ожиданиями μ_1 и μ_2 и известными дисперсиями σ_1^2 и σ_2^2 .

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Согласно теореме 13.3 случайная величина

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \quad (13.9)$$

является нормальной с математическим ожиданием 0 и дисперсией 1. Действительно, согласно замечанию 7.4 случайные величины $X_1 + \dots + X_m$ и $Y_1 + \dots + Y_n$ и их разность являются нормальными, а согласно свойствам 2 и 3 теоремы 7.1 и свойствам 2 и 4 теоремы 7.2

случайная величина (13.9) имеет математическое ожидание 0 и дисперсию 1. Поэтому

$$\begin{aligned} \mathbf{P} \left\{ -u_{\frac{1+\gamma}{2}} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < u_{\frac{1+\gamma}{2}} \right\} &= \Phi(u_{\frac{1+\gamma}{2}}) - \Phi(-u_{\frac{1+\gamma}{2}}) = \\ &= \Phi(u_{\frac{1+\gamma}{2}}) - (1 - \Phi(u_{\frac{1+\gamma}{2}})) = 2\Phi(u_{\frac{1+\gamma}{2}}) - 1 = 2\frac{1+\gamma}{2} - 1 = \gamma, \end{aligned}$$

где $u_{\frac{1+\gamma}{2}}$ — квантиль уровня $\frac{1+\gamma}{2}$ стандартного нормального распределения. Умножая все части двойного неравенства на $-\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$, а затем прибавляя $\bar{X} - \bar{Y}$, получим

$$\mathbf{P} \left\{ \bar{X} - \bar{Y} - u_{\frac{1+\gamma}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < \mu_1 - \mu_2 < \bar{X} - \bar{Y} + u_{\frac{1+\gamma}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \right\} = \gamma,$$

т.е. случайный интервал

$$\left(\bar{X} - \bar{Y} - u_{\frac{1+\gamma}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}, \quad \bar{X} - \bar{Y} + u_{\frac{1+\gamma}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \right) \quad (13.10)$$

накрывает неизвестную разность математических ожиданий $\mu_1 - \mu_2$ с необходимой доверительной вероятностью γ .

Замечание 13.6. Можно показать, что доверительный интервал (13.10) имеет наименьшую длину среди всех доверительных интервалов уровня доверия γ для разности математических ожиданий двух нормальных случайных величин с известными дисперсиями.

Доверительный интервал для разности математических ожиданий нормальных случайных величин с неизвестными, но равными дисперсиями

Пусть $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$ — две независимые выборки соответственно из распределений $\mathcal{N}(\mu_1, \sigma^2)$ и $\mathcal{N}(\mu_2, \sigma^2)$ с неизвестными математическими ожиданиями μ_1 и μ_2 и неизвестной дисперсией σ^2 . Обозначим

$$\begin{aligned} \bar{X} &= \frac{1}{m} \sum_{i=1}^m X_i, \quad S^2(\vec{X}_m) = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \\ \bar{Y} &= \frac{1}{n} \sum_{i=1}^n Y_i, \quad S^2(\vec{Y}_n) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2, \\ S(\vec{X}_m, \vec{Y}_n) &= \sqrt{\frac{(m-1)S^2(\vec{X}_m) + (n-1)S^2(\vec{Y}_n)}{m+n-2}}. \end{aligned}$$

Согласно теореме 13.3 случайная величина

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

имеет распределение Стьюдента с $m+n-2$ степенями свободы. Поэтому

$$\mathbf{P} \left\{ -t_{\frac{1+\gamma}{2}}(m+n-2) < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}} < t_{\frac{1+\gamma}{2}}(m+n-2) \right\} = \gamma,$$

где $t_{\frac{1+\gamma}{2}}(m+n-2)$ — квантиль уровня $\frac{1+\gamma}{2}$ распределения Стьюдента с $m+n-2$ степенями свободы. Умножая все части двойного неравенства на

$$-S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}$$

а затем прибавляя $\bar{X} - \bar{Y}$, заключаем, что нижняя $\underline{\theta}(\vec{X}_m, \vec{Y}_n)$ и верхняя $\bar{\theta}(\vec{X}_m, \vec{Y}_n)$ границы интервальной оценки с коэффициентом доверия $\gamma = \gamma$ для разности $\mu_1 - \mu_2$ в случае с неизвестными, но равными дисперсиями можно определить по формулам

$$\begin{cases} \underline{\theta}(\vec{X}_m, \vec{Y}_n) = \bar{X} - \bar{Y} - t_{\frac{1+\gamma}{2}}(m+n-2)S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}, \\ \bar{\theta}(\vec{X}_m, \vec{Y}_n) = \bar{X} - \bar{Y} + t_{\frac{1+\gamma}{2}}(m+n-2)S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}. \end{cases} \quad (13.11)$$

Замечание 13.7. Можно показать, что доверительный интервал (13.11) имеет наименьшую длину среди всех доверительных интервалов уровня доверия γ для разности математических ожиданий двух нормальных случайных величин с неизвестными дисперсиями.

Замечание 13.8. Как можно узнать, что дисперсии случайных величин равны, если они неизвестны? Во-первых, так будет, если обе выборки являются измерениями каких-то величин μ_1 и μ_2 , сделанных одним и тем же измерительным прибором. Во-вторых, существуют статистические методы проверки по наблюдениям $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$ гипотезы о равенстве дисперсий.

Замечание 13.9. Если дисперсии случайных величин неизвестны и не равны, то точный доверительный интервал для разности $\mu_1 - \mu_2$ еще не придуман, известен лишь приближенный доверительный интервал.

13.5 Приближенные доверительные интервалы

Приближенный доверительный интервал для математического ожидания случайной величины

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — случайная выборка объема n из распределения произвольной случайной величины X с неизвестными математическим ожиданием $\mu = MX$ и дисперсией $\sigma^2 = DX$.

В соответствии с центральной предельной теоремой 10.5 функция распределения случайной величины

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

стремится к функции распределения $\Phi(x)$ стандартной нормальной случайной величины $\mathcal{N}(0,1)$. Поэтому

$$\begin{aligned} \mathbf{P} \left\{ -u_{\frac{1+\gamma}{2}} < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u_{\frac{1+\gamma}{2}} \right\} &\approx \Phi(u_{\frac{1+\gamma}{2}}) - \Phi(-u_{\frac{1+\gamma}{2}}) = \\ &= \Phi(u_{\frac{1+\gamma}{2}}) - (1 - \Phi(u_{\frac{1+\gamma}{2}})) = 2\Phi(u_{\frac{1+\gamma}{2}}) - 1 = 2\frac{1+\gamma}{2} - 1 = \gamma, \end{aligned}$$

где $u_{\frac{1+\gamma}{2}}$ — квантиль уровня $\frac{1+\gamma}{2}$ стандартного нормального распределения. Умножая все части двойного неравенства $-u_{\frac{1+\gamma}{2}} < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u_{\frac{1+\gamma}{2}}$ на $-\frac{\sigma}{\sqrt{n}}$, а затем прибавляя \bar{X} , получим

$$\mathbf{P} \left\{ \bar{X} - \frac{\sigma}{\sqrt{n}} u_{\frac{1+\gamma}{2}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} u_{\frac{1+\gamma}{2}} \right\} \approx \gamma.$$

Заменяя неизвестную величину σ ее оценкой, например, случайной величиной

$$S(\vec{X}_n) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

получим приближенный доверительный интервал

$$\left(\bar{X} - u_{\frac{1+\gamma}{2}} \frac{S(\vec{X}_n)}{\sqrt{n}}, \quad \bar{X} + u_{\frac{1+\gamma}{2}} \frac{S(\vec{X}_n)}{\sqrt{n}} \right) \quad (13.12)$$

для математического ожидания μ с доверительной вероятностью γ .

Приближенный доверительный интервал для вероятности успеха в схеме Бернулли

Пусть X_i число успехов в i -м, $i = \overline{1, n}$, испытании по схеме Бернулли с неизвестной вероятностью успеха p , $q = 1 - p$. Тогда согласно примерам 5.2 и 5.3 $MX_i = p$, $DX_i = pq$. Обозначим через $\hat{p} = \bar{X}$ и $\hat{q} = 1 - \hat{p}$ доли успехов и неудач соответственно. Согласно следствию 10.2 из закона больших чисел $\hat{p} \rightarrow p$ при $n \rightarrow \infty$. Поэтому оценкой дисперсии DX_i будет случайная величина $\hat{p}\hat{q}$. Следовательно приближенный доверительный интервал (13.12) будет иметь вид

$$\left(\hat{p} - u_{\frac{1+\gamma}{2}} \frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}}, \quad \hat{p} + u_{\frac{1+\gamma}{2}} \frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}} \right)$$

или

$$\left(\hat{p} - u_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}, \quad \hat{p} + u_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

и являться приближенным доверительным интервалом для вероятности успеха p с доверительной вероятностью γ .

Приближенный доверительный интервал для параметра распределения Пуассона

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — случайная выборка пуассоновской случайной величины X с параметром λ . Согласно примерам 5.4 и 5.5 $MX = \lambda$, $DX = \lambda$. Следовательно дисперсию DX можно оценить как и математическое ожидание MX случайной величиной \bar{X} . Поэтому приближенный доверительный интервал (13.12) будет иметь вид

$$\left(\bar{X} - u_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{X}}{n}}, \quad \bar{X} + u_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{X}}{n}} \right)$$

и являться приближенным доверительным интервалом для параметра λ с доверительной вероятностью γ .

Лекция 14

Проверка гипотез. Параметрические модели

Рассмотрим пример, иллюстрирующий проблемы, о которых пойдет речь в этой лекции.

Пример 14.1. По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. В результате изменения конструкции двигателя ожидается, что расход топлива уменьшится. Для проверки проводятся испытания 25 случайно отобранных автомобилей с модернизированным двигателем, причем выборочное среднее \bar{X} расходов топлива на 100 км пробега по результатам испытаний составило 9.3 л.

Повлияло ли изменение конструкции двигателя на расход топлива? Другими словами уменьшение среднего расхода топлива следует объяснить улучшением конструкции двигателя или случайными факторами, повлиявшими на расход топлива каждого из 25 автомобилей. Такими случайными факторами могут быть ошибки измерения количества топлива в бензобаке автомобилей, неоднородность качества топлива, погрешности при изготовлении деталей двигателя, изменения погодных и дорожных условий во время испытаний.

Заметим, что в силу этих случайных эффектов \bar{X} не будет в точности совпадать с 10 даже если не вносить никаких изменений в конструкцию двигателя. Если бы было $\bar{X} = 7.3$, то скорее всего расход топлива уменьшился, а если бы было $\bar{X} = 9.99$, то скорее всего нет. Как найти порог C , чтобы при $\bar{X} < C$ можно было принять гипотезу об уменьшении расхода топлива, а при $\bar{X} \geq C$ принять гипотезу об отсутствии уменьшения расхода топлива? При этом можно ошибиться, приняв гипотезу об уменьшении расхода топлива, хотя она на самом деле неверна (просто наблюдения X_1, \dots, X_{25} случайно попались нетипичные), а можно ошибиться, отклонив гипотезу об уменьшении расхода топлива, хотя она на самом деле верна. Как вычислить вероятности этих ошибок и можно ли их уменьшить?

14.1 Основные понятия

Определение 14.1. Пусть X — произвольная одномерная или многомерная случайная величина. *Статистической гипотезой* называют любое утверждение о функции распределения случайной величины X , при этом слово “статистическая” для краткости обычно опускают.

Пример 14.2. Гипотезы будем обозначать буквой H . Приведем примеры гипотез.

- 1) $H_1 : MX = 10$ (т.е. функция распределения $F(x)$ случайной величины X такова, что $\int_{-\infty}^{\infty} xF'(x) dx = 10$);
- 2) $H_2 : DX < 3$;
- 3) $H_3 : X$ — пуассоновская случайная величина;
- 4) $H_4 : MX_1 \neq MX_2$ для вектора $X = (X_1, X_2)$.

Определение 14.2. Пусть функция распределения $F(x; \theta)$ случайной величины X известна с точностью до параметра θ . *Параметрической гипотезой* называют любое утверждение о параметре θ функции распределения $F(x; \theta)$.

Пример 14.3. Пусть $F(x; \theta)$ — функция распределения случайной величины X , $\theta \in \mathbb{R}^r$. Приведем примеры параметрических гипотез.

- 1) $H_1: \theta_2 \neq -3$;
- 2) $H_2: \theta_1 > 5$;
- 3) $H_3: \theta_1 = \theta_2$.

Определение 14.3. *Статистическую гипотезу* называют *простой*, если она однозначно определяет функцию распределения случайной величины. В противном случае гипотезу называют *сложной*.

В частности, параметрическую гипотезу H называют *простой*, если она имеет вид $H: \theta = \theta_0$, где θ_0 — некоторое заданное значение параметра. Статистическую гипотезу называют *сложной*, если она имеет вид $H: \theta \in D$, где D — некоторое множество значений параметра θ , состоящее более чем из одного элемента.

Пример 14.4. Утверждение «случайная величина X является экспоненциальной с математическим ожиданием 5» — это пример простой гипотезы, поскольку она однозначно определяет параметр λ экспоненциального распределения, совпадающий с величиной $1/MX$ (см. определение 5.6 и пример 5.9).

Утверждение «случайная величина X является экспоненциальной» (без указания конкретного значения параметра) — пример сложной гипотезы.

Пример 14.5. Пусть X — нормальная случайная величина с неизвестным математическим ожиданием μ и известной дисперсией σ^2 , а μ_0 и μ_1 — некоторые числа.

Тогда гипотеза $H: \mu = \mu_0$ является простой, а гипотезы $H_1: \mu \geq \mu_0$; $H_2: \mu \leq \mu_0$; $H_3: \mu_0 \leq \mu \leq \mu_1$ — сложными.

Пусть теперь оба параметра μ и σ неизвестны. В этом случае гипотеза $H: \mu = \mu_0$ становится сложной, так как ей соответствует множество значений двумерного вектора $\theta = (\mu, \sigma)$, для которых $\mu = \mu_0$, $0 < \sigma < \infty$. #

Пусть имеется случайная величина X с функцией распределения $F(x)$ и две гипотезы о $F(x)$. Одну из них назовем *основной*, или *нулевой, гипотезой* и обозначим H_0 , а вторую — *альтернативной*, или *конкурирующей, гипотезой* и обозначим H_A .

Гипотезы H_0 и H_A не равнозначны. Гипотеза H_0 обычно свидетельствует о равенстве или о симметрии, гипотеза H_A об отсутствии равенства или симметрии.

Пример 14.6. Примеры основной и альтернативной гипотез.

- 1) H_0 — подбрасывается симметричная игральная кость (идеальный куб сделанный из однородного материала), H_A — подбрасывается несимметричная игральная кость;
- 2) H_0 — вероятность выпадения герба равна $1/2$, H_A — вероятность выпадения герба не равна $1/2$;
- 3) H_0 — вероятности рождения мальчика и девочки совпадают, H_A — вероятности рождения мальчика и девочки не совпадают. #

Пусть имеется случайная выборка \vec{X}_n из X (независимые наблюдения случайной величины X). Задача в теории проверки гипотез ставится следующим образом. На основании анализа наблюдений нужно сделать выбор: принять основную гипотезу H_0 или отклонить ее в пользу альтернативы H_A .

Определение 14.4. *Критерием*, или *статистическим критерием*, проверки гипотез называют правило, по которому по данным выборки \vec{X}_n принимается решение о справедливости либо основной, либо альтернативной гипотезы.

Критерий задают с помощью *критического множества* $W \in \mathbb{R}^n$. Решение принимают следующим образом:

- если выборка \vec{X}_n принадлежит критическому множеству W , то отвергают основную гипотезу H_0 и принимают альтернативную гипотезу H_A ;
- если выборка \vec{X}_n не принадлежит критическому множеству W (т.е. принадлежит дополнению $\bar{W} = \mathbb{R}^n \setminus W$ множества W до \mathbb{R}^n), то отвергают альтернативную гипотезу H_A и принимают основную гипотезу H_0 . Множество \bar{W} называют *доверительным множеством*.

При использовании любого критерия возможны ошибки следующих видов:

- принять гипотезу H_A , когда верна H_0 — **ошибка первого рода**;
- принять гипотезу H_0 , когда верна H_A — **ошибка второго рода**.

Вероятности совершения ошибок первого и второго рода обозначают α и β :

$$\alpha = \mathbf{P}\{\vec{X}_n \in W \mid H_0\}, \quad \beta = \mathbf{P}\{\vec{X}_n \in \bar{W} \mid H_A\}.$$

Здесь $\mathbf{P}\{A \mid H_0\}$ и $\mathbf{P}\{A \mid H_A\}$ — вероятности события A при условии, что справедливы соответственно гипотезы H_0 и H_A .

Вероятность совершения ошибки первого рода α называют также **уровнем значимости критерия**.

Величину $1 - \beta$, равную вероятности отвергнуть основную гипотезу H_0 , когда она неверна, называют **мощностью критерия**. #

Как видно из определения 14.4 критерий проверки гипотез построить легко, достаточно выбрать произвольное подмножество W в \mathbb{R}^n в качестве критического. Трудно построить хороший критерий, т.е. такой, у которого вероятности ошибок α и β малы.

К сожалению при построении большинства критериев попытка уменьшить α приводит к увеличению β и наоборот. Поэтому поступают следующим образом. Фиксируют уровень значимости α критерия (вероятность совершения ошибки первого рода) и среди всех критериев уровня значимости α пытаются найти (построить) наиболее мощный критерий, т.е. критерий с наименьшей вероятностью ошибки второго рода β . В приложениях обычно α равно 0.01, 0.05, 0.1 (в медицине и биологии почти всегда 0.05).

Уровень значимости показывает уровень риска, на который исследователь готов идти, отклоняя H_0 , когда она верна. Например, более решительный исследователь выбирает в качестве α число 0.1, более осторожный — 0.01. Величина α выбирается еще до сбора наблюдений \vec{X}_n .

Фиксация α , а не β связана с неравнозначностью гипотез. Гипотезы формулируются таким образом, что последствия отклонения H_0 , когда она верна, являются более тяжелыми по сравнению с последствиями принятия H_0 , когда она неверна. Например, H_0 — человек болен, H_A — человек здоров. Здесь ошибка первого рода (вероятность которой α) означает, что человек болен, а его считают здоровым и не лечат. В этом случае человек может серьезно подорвать здоровье и даже умереть. Ошибка второго рода (вероятность которой β) означает, что человек здоров, а его считают больным и лечат. В этом случае скорее всего здоровье человека не ухудшится и он просто будет испытывать дискомфорт.

Отметим, что у большинства используемых в науке критериев обе ошибки можно свести к нулю при увеличении количества наблюдений, т.е. α и β одновременно стремятся к нулю при $n \rightarrow \infty$.

14.2 Проверка гипотез о математическом ожидании нормальной случайной величины с известной дисперсией

Пусть X — нормальная случайная величина с неизвестным математическим ожиданием μ и известной дисперсией σ^2 .

Построим критерий уровня значимости α проверки гипотезы

$$H_0 : \mu = \mu_0$$

против альтернативы

$$H_1 : \mu < \mu_0,$$

где μ_0 — некоторое известное число.

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения случайной величины X ,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \xi = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}.$$

Эту модель можно проиллюстрировать примером 14.1. Тогда μ — средний расход топлива после изменения конструкции двигателя, $n = 25$, $\bar{X} = 9.3$, $\mu_0 = 10$, ошибки измерения расхода топлива на 100 км пробега каждым автомобилем считаем независимыми нормальными случайными величинами с нулевым математическим ожиданием и известной дисперсией σ^2 .

Согласно свойствам математического ожидания (см. свойства 2 и 3 теоремы 7.1) случайная величина ξ имеет математическое ожидание $(\mu - \mu_0)\sqrt{n}/\sigma$, в частности $M\xi = 0$ при гипотезе H_0 . Поэтому при гипотезе H_0 типичные значения случайной величины ξ будут вблизи нуля, а при альтернативе H_1 — вблизи $(\mu - \mu_0)\sqrt{n}/\sigma$, т.е. в основном отрицательные. Следовательно разумно отклонить H_0 в пользу H_1 , если $\xi < C$ и принять в противном случае, где C — некоторая постоянная.

По теореме 13.2 случайная величина ξ при гипотезе H_0 является стандартной нормальной величиной. В этом случае уровень значимости α критерия по определению равен

$$P\{\xi < C|H_0\} = P\{\xi < C|\xi \sim \mathcal{N}(0, 1)\} = \Phi(C).$$

Таким образом $\Phi(C) = \alpha$ и $C = u_\alpha$, т.е. C совпадает с квантилью u_α уровня α стандартного нормального распределения.

Критическое W и доверительное \bar{W} множества имеют вид

$$W = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} < u_\alpha \right\},$$

$$\bar{W} = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} \geq u_\alpha \right\},$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Итак, мы построили критерий уровня значимости α проверки гипотезы $H_0 : \mu = \mu_0$ против альтернативы $H_1 : \mu < \mu_0$, в соответствии с которым H_0 отклоняется в пользу H_1 на уровне значимости α , если $\xi < u_\alpha$, где u_α — квантиль уровня α стандартного нормального распределения. В противном случае, т.е. если $\xi \geq u_\alpha$, говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Нельзя говорить, что если $\xi \geq u_\alpha$, то гипотеза H_0 доказана, поскольку естественнонаучную гипотезу в принципе нельзя доказать (а опровергнуть можно, приведя хотя бы один опровергающий пример). Например, гипотеза, в соответствие с которой Вселенная описывается законами Ньютона, около двухсот лет многократно «доказывалась» тем, что построенные в соответствии с законами Ньютона здания и мосты не разрушались, а самолеты не падали. Но любой пример, подтверждающий специальную теорию относительности Эйнштейна, в частности, GPS-навигаторы, уничтожает все эти «доказательства».

Можно доказать, что построенный критерий является наилучшим, т.е. имеет наибольшую мощность (наименьшую ошибку второго рода β) среди всех других критериев проверки $H_0 : \mu = \mu_0$ против $H_1 : \mu < \mu_0$ фиксированного уровня значимости α .

Аналогично строится критерий уровня значимости α проверки гипотезы

$$H_0 : \mu = \mu_0$$

против альтернативы

$$H_2 : \mu > \mu_0.$$

При альтернативе H_2 типичные значения случайной величины ξ будут положительными. Поэтому разумно отклонить H_0 в пользу H_2 , если $\xi > C$ и принять в противном случае, где C — некоторая постоянная. В этом случае уровень значимости критерия по определению равен

$$P\{\xi > C|H_0\} = 1 - P\{\xi < C|H_0\} = 1 - \Phi(C) = \alpha.$$

Следовательно $\Phi(C) = 1 - \alpha$, т.е. $C = u_{1-\alpha}$ — квантиль уровня $1 - \alpha$ стандартного нормального распределения.

Таким образом мы построили критерий уровня значимости α проверки гипотезы $H_0 : \mu = \mu_0$ против альтернативы $H_2 : \mu > \mu_0$, в соответствии с которым H_0 отклоняется в пользу H_2 на уровне значимости α , если $\xi > u_{1-\alpha}$, где $u_{1-\alpha}$ – квантиль уровня $1 - \alpha$ стандартного нормального распределения. В противном случае, т.е. если $\xi \leq u_{1-\alpha}$, говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Теперь построим критерий уровня значимости α проверки гипотезы

$$H_0 : \mu = \mu_0$$

против альтернативы

$$H_3 : \mu \neq \mu_0.$$

При альтернативе H_3 типичные значения случайной величины ξ не будут вблизи нуля. Если $\mu - \mu_0 > 0$, то они будут в основном положительными, а если $\mu - \mu_0 < 0$, то они будут в основном отрицательными. Поэтому разумно с учетом симметрии вариантов $\mu - \mu_0 > 0$ и $\mu - \mu_0 < 0$ отклонить H_0 в пользу H_3 , если $|\xi| > C$ и принять в противном случае, где C – некоторая постоянная. В этом случае уровень значимости критерия по определению равен

$$\begin{aligned} P\{|\xi| > C | H_0\} &= 1 - P\{|\xi| < C | H_0\} = 1 - P\{-C < \xi < C | H_0\} = \\ &= 1 - (\Phi(C) - \Phi(-C)) = 1 - (2\Phi(C) - 1) = 2 - 2\Phi(C) = \alpha. \end{aligned}$$

Следовательно $\Phi(C) = 1 - \alpha/2$, т.е. $C = u_{1-\alpha/2}$ – квантиль уровня $1 - \alpha/2$ стандартного нормального распределения.

Таким образом мы построили критерий уровня значимости α проверки гипотезы $H_0 : \mu = \mu_0$ против альтернативы $H_3 : \mu \neq \mu_0$, в соответствии с которым H_0 отклоняется в пользу H_3 на уровне значимости α , если $|\xi| > u_{1-\alpha/2}$, где $u_{1-\alpha/2}$ – квантиль уровня $1 - \alpha/2$ стандартного нормального распределения. В противном случае, т.е. если $|\xi| \leq u_{1-\alpha/2}$, говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Замечание 14.1. Можно доказать, что все три построенные критерия являются наилучшими, т.е. имеют наибольшую мощность (наименьшую ошибку второго рода β) среди всех критериев фиксированного уровня значимости α .

Альтернативы H_1 и H_2 называются *односторонними*, альтернатива H_3 – *двусторонней*. Несмотря на то, что альтернативы H_1 и H_2 являются частными случаями альтернативы H_3 , они не являются лишними. Дело в том, что уточнение альтернативы путем замены H_3 на H_1 или H_2 приводит к уменьшению вероятности ошибки второго рода β (при фиксированных α и n). Поэтому везде, где только возможно, нужно заменять H_3 на H_1 или H_2 . При этом обоснование такой замены лежит в предметной области исследования и не является математическим. Например, пусть μ_0 – первоначальное (10 лет назад) давление инертного газа в баллоне, а μ – нынешнее давление этого же газа в этом же баллоне после 10 лет хранения. Если исследователь уверен, что по прошествии 10 лет давление может остаться прежним, может уменьшиться из-за рассохшихся от превышения срока годности прокладок в кране баллона, но никак не может увеличиться, то альтернативу $H_3 : \mu \neq \mu_0$ целесообразнее заменить на альтернативу $H_1 : \mu < \mu_0$.

14.3 Проверка гипотез о математическом ожидании нормальной случайной величины с неизвестной дисперсией

Пусть X – нормальная случайная величина с неизвестным математическим ожиданием μ и неизвестной дисперсией σ^2 .

Построим критерий уровня значимости α проверки гипотезы

$$H_0 : \mu = \mu_0$$

против альтернативы

$$H_1 : \mu < \mu_0,$$

где μ_0 — некоторое известное число.

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения случайной величины X . Обозначим

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \tau = \frac{\bar{X} - \mu_0}{S(\vec{X}_n)} \sqrt{n}.$$

По теореме 13.2 случайная величина

$$\frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n}$$

имеет распределение Стьюдента с $n - 1$ степенью свободы, в частности нулевое математическое ожидание, поскольку плотность распределения Стьюдента (13.1) является четной функцией. Поэтому

$$\mathbf{M}\tau = \mathbf{M}\left(\frac{\bar{X} - \mu_0}{S(\vec{X}_n)} \sqrt{n}\right) = \mathbf{M}\left(\frac{\bar{X} - \mu}{S(\vec{X}_n)} \sqrt{n}\right) + \mathbf{M}\left(\frac{\mu - \mu_0}{S(\vec{X}_n)} \sqrt{n}\right) = \mathbf{M}\left(\frac{\mu - \mu_0}{S(\vec{X}_n)} \sqrt{n}\right).$$

Отсюда вытекает, что $\mathbf{M}\tau = 0$ при гипотезе H_0 . Кроме того, поскольку $S(\vec{X}_n) > 0$, то $\mathbf{M}\tau < 0$ при $\mu < \mu_0$ и $\mathbf{M}\tau > 0$ при $\mu > \mu_0$. Следовательно при гипотезе H_0 типичные значения случайной величины τ будут вблизи нуля, а при альтернативе H_1 типичные значения случайной величины τ будут в основном отрицательными. Таким образом разумно отклонить H_0 в пользу H_1 , если $\tau < C$ и принять в противном случае, где C — некоторая постоянная.

По теореме 13.2 случайная величина τ при гипотезе H_0 имеет распределение Стьюдента с $n - 1$ степенью свободы. Следовательно уровень значимости критерия будет равен α , если $C = t_\alpha(n - 1)$, где $t_\alpha(n - 1)$ — квантиль уровня α распределения Стьюдента с $n - 1$ степенью свободы, поскольку

$$\mathbf{P}\{\tau < C | H_0\} = \mathbf{P}\{\tau < t_\alpha(n - 1) | H_0\} = \alpha.$$

Итак, мы построили критерий уровня значимости α проверки гипотезы $H_0 : \mu = \mu_0$ против альтернативы $H_1 : \mu < \mu_0$, в соответствии с которым H_0 отклоняется в пользу H_1 на уровне значимости α , если $\tau < t_\alpha(n - 1)$, где $t_\alpha(n - 1)$ — квантиль уровня α распределения Стьюдента с $n - 1$ степенью свободы. В противном случае, т.е. если $\tau \geq t_\alpha(n - 1)$, говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Аналогично строится критерий уровня значимости α проверки гипотезы

$$H_0 : \mu = \mu_0$$

против альтернативы

$$H_2 : \mu > \mu_0.$$

При альтернативе H_2 типичные значения случайной величины τ будут положительными. Поэтому разумно отклонить H_0 в пользу H_2 , если $\tau > C$ и принять в противном случае, где C — некоторая постоянная. Покажем, что C совпадает с квантилью $t_{1-\alpha}(n - 1)$ уровня $1 - \alpha$ распределения Стьюдента с $n - 1$ степенью свободы. Действительно, согласно определению уровня значимости критерия и определению квантили

$$\mathbf{P}\{\tau > t_{1-\alpha}(n - 1) | H_0\} = 1 - \mathbf{P}\{\tau < t_{1-\alpha}(n - 1) | H_0\} = 1 - (1 - \alpha) = \alpha.$$

Таким образом построен критерий уровня значимости α проверки гипотезы $H_0 : \mu = \mu_0$ против альтернативы $H_2 : \mu > \mu_0$, в соответствии с которым H_0 отклоняется в пользу H_2 на уровне значимости α , если $\tau > t_{1-\alpha}(n - 1)$, где $t_{1-\alpha}(n - 1)$ — квантиль уровня $1 - \alpha$ распределения Стьюдента с $n - 1$ степенью свободы. В противном случае, т.е. если $\tau \leq t_{1-\alpha}(n - 1)$, говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Теперь построим критерий уровня значимости α проверки гипотезы

$$H_0 : \mu = \mu_0$$

против альтернативы

$$H_3: \mu \neq \mu_0.$$

При альтернативе H_3 типичные значения случайной величины τ не будут вблизи нуля. Если $\mu - \mu_0 > 0$, то они будут в основном положительными, а если $\mu - \mu_0 < 0$, то они будут в основном отрицательными. Поэтому разумно с учетом симметрии случаев $\mu - \mu_0 > 0$ и $\mu - \mu_0 < 0$ отклонить H_0 в пользу H_3 , если $|\tau| > C$ и принять в противном случае, где C — некоторая постоянная.

Покажем, что C совпадает с квантилью $t_{1-\alpha/2}(n-1)$ уровня $1 - \alpha/2$ распределения Стьюдента с $n - 1$ степенью свободы. Действительно, согласно определению уровня значимости критерия, определению квантили, свойству $-t_{1-\alpha/2}(n-1) = t_{\alpha/2}(n-1)$ и теореме 13.2

$$\begin{aligned} \mathbf{P}\{|\tau| > t_{1-\alpha/2}(n-1) | H_0\} &= 1 - \mathbf{P}\{|\tau| < t_{1-\alpha/2}(n-1) | H_0\} = \\ &= 1 - \mathbf{P}\{-t_{1-\alpha/2}(n-1) < \tau < t_{1-\alpha/2}(n-1) | H_0\} = \\ &= 1 - \mathbf{P}\{t_{\alpha/2}(n-1) < \tau < t_{1-\alpha/2}(n-1) | H_0\} = 1 - ((1 - \alpha/2) - \alpha/2) = \alpha. \end{aligned}$$

Таким образом мы построили критерий уровня значимости α проверки гипотезы $H_0: \mu = \mu_0$ против альтернативы $H_3: \mu \neq \mu_0$, в соответствии с которым H_0 отклоняется в пользу H_3 на уровне значимости α , если $|\tau| > t_{1-\alpha/2}(n-1)$, где $t_{1-\alpha/2}(n-1)$ — квантиль уровня $1 - \alpha/2$ распределения Стьюдента с $n - 1$ степенью свободы. В противном случае (если $|\tau| \leq t_{1-\alpha/2}(n-1)$) говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Замечание 14.2. Можно доказать, что все три построенные критерия являются наилучшими, т.е. имеют наибольшую мощность (наименьшую ошибку второго рода β) среди всех критериев фиксированного уровня значимости α .

14.4 Проверка гипотез о дисперсии нормальной случайной величины при неизвестном математическом ожидании

Пусть X — нормальная случайная величина с неизвестным математическим ожиданием μ и неизвестной дисперсией σ^2 .

Построим критерий уровня значимости α проверки гипотезы

$$H_0: \sigma^2 = \sigma_0^2$$

против альтернативы

$$H_1: \sigma^2 < \sigma_0^2,$$

где σ_0^2 — некоторое известное число.

Пусть $\vec{X}_n = (X_1, \dots, X_n)$ — независимые наблюдения случайной величины X . Обозначим

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \eta = \frac{(n-1)S^2(\vec{X}_n)}{\sigma_0^2}.$$

Опираясь на закон больших чисел можно показать (это несложно), что $S^2(\vec{X}_n)$ — состоятельная оценка σ^2 , т.е. $S^2(\vec{X}_n) \approx \sigma^2$. Поэтому при альтернативе H_1 случайная величина η принимает преимущественно меньшие значения, чем при гипотезе H_0 . Следовательно разумно отклонить H_0 в пользу H_1 , если $\eta < C$ и принять H_0 , если $\eta \geq C$, где C — некоторая постоянная.

По теореме 13.2 случайная величина

$$\frac{(n-1)S^2(\vec{X}_n)}{\sigma^2}$$

имеет χ^2 -распределение с $n - 1$ степенью свободы. Поэтому случайная величина η при гипотезе H_0 имеет χ^2 -распределение с $n - 1$ степенью свободы.

Следовательно уровень значимости критерия будет равен α , т.е. будет справедливо равенство

$$\mathbf{P}\{\eta < C|H_0\} = \alpha,$$

только если $C = \chi^2_{\alpha}(n - 1)$, где $\chi^2_{\alpha}(n - 1)$ – квантиль уровня α χ^2 -распределения с $n - 1$ степенью свободы.

Итак, мы построили критерий уровня значимости α проверки гипотезы $H_0 : \sigma^2 = \sigma_0^2$ против альтернативы $H_1 : \sigma^2 < \sigma_0^2$, в соответствии с которым H_0 отклоняется в пользу H_1 на уровне значимости α , если $\eta < \chi^2_{\alpha}(n - 1)$, где $\chi^2_{\alpha}(n - 1)$ – квантиль уровня α χ^2 -распределения с $n - 1$ степенью свободы. В противном случае, т.е. если $\eta \geq \chi^2_{\alpha}(n - 1)$, говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Построим теперь критерий уровня значимости α проверки гипотезы

$$H_0 : \sigma^2 = \sigma_0^2$$

против альтернативы

$$H_2 : \sigma^2 > \sigma_0^2,$$

где σ_0^2 – некоторое известное число.

Так как $S^2(\vec{X}_n) \approx \sigma^2$, то при альтернативе H_2 случайная величина η принимает преимущественно бóльшие значения, чем при гипотезе H_0 . Следовательно разумно отклонить H_0 в пользу H_2 , если $\eta > C$ и принять H_0 , если $\eta \leq C$, где C – некоторая постоянная.

Так как случайная величина η при гипотезе H_0 имеет χ^2 -распределение с $n - 1$ степенью свободы, то уровень значимости критерия будет равен α , если $C = \chi^2_{1-\alpha}(n - 1)$, где $\chi^2_{1-\alpha}(n - 1)$ – квантиль уровня $1 - \alpha$ χ^2 -распределения с $n - 1$ степенью свободы. Действительно

$$\mathbf{P}\{\eta > \chi^2_{1-\alpha}(n - 1)|H_0\} = 1 - \mathbf{P}\{\eta < \chi^2_{1-\alpha}(n - 1)|H_0\} = 1 - (1 - \alpha) = \alpha.$$

Итак, мы построили критерий уровня значимости α проверки гипотезы $H_0 : \sigma^2 = \sigma_0^2$ против альтернативы $H_2 : \sigma^2 > \sigma_0^2$, в соответствии с которым H_0 отклоняется в пользу H_2 на уровне значимости α , если $\eta > \chi^2_{1-\alpha}(n - 1)$, где $\chi^2_{1-\alpha}(n - 1)$ – квантиль уровня α χ^2 -распределения с $n - 1$ степенью свободы. В противном случае, т.е. если $\eta \leq \chi^2_{1-\alpha}(n - 1)$, говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Построим теперь критерий уровня значимости α проверки гипотезы

$$H_0 : \sigma^2 = \sigma_0^2$$

против альтернативы

$$H_3 : \sigma^2 \neq \sigma_0^2,$$

где σ_0^2 – некоторое известное число.

Так как $S^2(\vec{X}_n) \approx \sigma^2$, то при альтернативе H_3 случайная величина η принимает либо преимущественно бóльшие, либо преимущественно меньшие значения, чем при гипотезе H_0 . Следовательно разумно принять H_0 , если $\chi^2_{\frac{\alpha}{2}}(n - 1) \leq \eta \leq \chi^2_{1-\frac{\alpha}{2}}(n - 1)$ и отклонить в противном случае. Здесь $\chi^2_{\frac{\alpha}{2}}(n - 1)$ и $\chi^2_{1-\frac{\alpha}{2}}(n - 1)$ – квантили соответственно уровней $\alpha/2$ и $1 - \alpha/2$ χ^2 -распределения с $n - 1$ степенью свободы.

При этом уровень значимости критерия будет равен

$$1 - \mathbf{P}\{\chi^2_{\frac{\alpha}{2}}(n - 1) \leq \eta \leq \chi^2_{1-\frac{\alpha}{2}}(n - 1)\} = 1 - \left(\left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} \right) = \alpha.$$

Итак, мы построили критерий уровня значимости α проверки гипотезы $H_0 : \sigma^2 = \sigma_0^2$ против альтернативы $H_3 : \sigma^2 \neq \sigma_0^2$, в соответствии с которым H_0 принимается на уровне значимости α , если $\chi^2_{\frac{\alpha}{2}}(n - 1) \leq \eta \leq \chi^2_{1-\frac{\alpha}{2}}(n - 1)$ и отклоняется в противном случае.

14.5 Проверка гипотез о равенстве математических ожиданий нормальных случайных величин с известными дисперсиями

Пусть X и Y — нормальные случайные величины с неизвестными математическими ожиданиями μ_1 и μ_2 и известными дисперсиями σ_1^2 и σ_2^2 .

Построим три критерия уровня значимости α проверки гипотезы

$$H_0: \mu_1 = \mu_2$$

против каждой из трех альтернатив

$$H_1: \mu_1 < \mu_2,$$

$$H_2: \mu_1 > \mu_2,$$

$$H_3: \mu_1 \neq \mu_2.$$

Пусть $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$ — две независимые выборки соответственно из распределений $\mathcal{N}(\mu_1, \sigma_1^2)$ и $\mathcal{N}(\mu_2, \sigma_2^2)$,

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \xi = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}.$$

Согласно свойствам математического ожидания (см. свойства 2 и 3 теоремы 7.1) случайная величина ξ имеет математическое ожидание

$$\frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}},$$

в частности $M\xi = 0$ при гипотезе H_0 . По теореме 13.3 случайная величина ξ при гипотезе H_0 является стандартной нормальной величиной. Поэтому, рассуждая так же как в разделе 14.2 на с. 95, получим, что три критерия проверки H_0 соответственно против альтернатив H_0 , H_2 и H_3 имеют следующий вид.

Обозначим через u_α , $u_{1-\alpha}$ и $u_{1-\alpha/2}$ соответственно квантили уровней α , $1 - \alpha$ и $1 - \alpha/2$ стандартного нормального распределения. Гипотеза H_0 отклоняется:

- 1) в пользу H_1 , если $\xi < u_\alpha$;
- 2) в пользу H_2 , если $\xi > u_{1-\alpha}$;
- 3) в пользу H_3 , если $|\xi| > u_{1-\alpha/2}$.

В противном случае говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Замечание 14.3. Можно доказать, что все три построенные критерия являются наилучшими, т.е. имеют наибольшую мощность (наименьшую ошибку второго рода β) среди всех критериев фиксированного уровня значимости α .

14.6 Проверка гипотез о равенстве математических ожиданий нормальных случайных величин с неизвестными, но равными дисперсиями

Пусть X и Y — нормальные случайные величины с неизвестными математическими ожиданиями μ_1 и μ_2 и неизвестными, но равными дисперсиями σ^2 .

Построим три критерия уровня значимости α проверки гипотезы

$$H_0: \mu_1 = \mu_2$$

против каждой из трех альтернатив

$$H_1 : \mu_1 < \mu_2,$$

$$H_2 : \mu_1 > \mu_2,$$

$$H_3 : \mu_1 \neq \mu_2.$$

Пусть $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$ — две независимые выборки соответственно из распределений $\mathcal{N}(\mu_1, \sigma^2)$ и $\mathcal{N}(\mu_2, \sigma^2)$. Обозначим

$$\begin{aligned}\bar{X} &= \frac{1}{m} \sum_{i=1}^m X_i, & S^2(\vec{X}_m) &= \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \\ \bar{Y} &= \frac{1}{n} \sum_{i=1}^n Y_i, & S^2(\vec{Y}_n) &= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2, \\ S(\vec{X}_m, \vec{Y}_n) &= \sqrt{\frac{(m-1)S^2(\vec{X}_m) + (n-1)S^2(\vec{Y}_n)}{m+n-2}}, \\ \tau &= \frac{\bar{X} - \bar{Y}}{S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}}.\end{aligned}$$

По теореме 13.3 случайная величина

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

имеет распределение Стьюдента с $m+n-2$ степенями свободы, в частности нулевое математическое ожидание, поскольку плотность распределения Стьюдента является четной функцией (см. (13.1)). Поэтому

$$\mathbf{M}\tau = \mathbf{M} \left(\frac{\mu_1 - \mu_2}{S(\vec{X}_m, \vec{Y}_n) \sqrt{\frac{1}{m} + \frac{1}{n}}} \right).$$

Далее, согласно теореме 13.3 случайная величина τ при гипотезе H_0 имеет распределение Стьюдента с $m+n-2$ степенями свободы. Поэтому, рассуждая так же как в разделе 14.3 на с. 97, получим, что три критерия проверки H_0 соответственно против альтернатив H_1 , H_2 и H_3 имеют следующий вид.

Обозначим через $t_\alpha(m+n-2)$, $t_{1-\alpha}(m+n-2)$ и $t_{1-\alpha/2}(m+n-2)$ соответственно квантили уровней α , $1-\alpha$ и $1-\alpha/2$ распределения Стьюдента с $m+n-2$ степенями свободы. Гипотеза H_0 отклоняется:

- 1) в пользу H_1 , если $\tau < t_\alpha(m+n-2)$;
- 2) в пользу H_2 , если $\tau > t_{1-\alpha}(m+n-2)$;
- 3) в пользу H_3 , если $|\tau| > t_{1-\alpha/2}(m+n-2)$.

В противном случае говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Замечание 14.4. Можно доказать, что все три построенные критерия являются наилучшими, т.е. имеют наибольшую мощность (наименьшую ошибку второго рода β) среди всех критериев фиксированного уровня значимости α .

14.7 Проверка гипотез о равенстве дисперсий нормальных случайных величин с неизвестными математическими ожиданиями

Пусть X и Y — нормальные случайные величины с неизвестными математическими ожиданиями μ_1 и μ_2 и неизвестными дисперсиями σ_1^2 и σ_2^2 .

$$H_0 : \sigma_1^2 = \sigma_2^2$$

против каждой из трех альтернатив

$$H_1 : \sigma_1^2 < \sigma_2^2,$$

$$H_2 : \sigma_1^2 > \sigma_2^2,$$

$$H_3 : \sigma_1^2 \neq \sigma_2^2.$$

Пусть $\vec{X}_m = (X_1, \dots, X_m)$ и $\vec{Y}_n = (Y_1, \dots, Y_n)$ — две независимые выборки соответственно из распределений $\mathcal{N}(\mu_1, \sigma_1^2)$ и $\mathcal{N}(\mu_2, \sigma_2^2)$. Обозначим

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad S^2(\vec{X}_m) = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2,$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad S^2(\vec{Y}_n) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

$$F = \frac{S^2(\vec{X}_m)}{S^2(\vec{Y}_n)}.$$

При гипотезе H_0 значения случайной величины F скорее всего будут близки к единице, большие и малые значения F неправдоподобны. При альтернативе H_2 более вероятны именно большие значения F , а при альтернативе H_1 — малые. Поэтому разумно отклонять H_0 в пользу H_1 , если $F < C_1$ и принимать в противном случае. Аналогично, следует отклонять H_0 в пользу H_2 , если $F > C_2$ и принимать в противном случае. Также нужно отклонять H_0 в пользу H_3 , если $F < C_3$ или $F > C_4$ и принимать в противном случае $C_3 \leq F \leq C_4$. Здесь C_1, C_2, C_3 и C_4 — некоторые постоянные. Так как по теореме 13.3 случайная величина F имеет распределение Фишера с числом степеней свободы $m-1$ и $n-1$, то нетрудно показать, что постоянные C_1, C_2, C_3 и C_4 будут квантилями распределения Фишера с числом степеней свободы $m-1$ и $n-1$.

Поэтому, рассуждая так же как в и в предыдущих разделах, получим, что три критерия проверки H_0 соответственно против альтернатив H_1, H_2 и H_3 имеют следующий вид.

Обозначим через $f_p(m-1, n-1)$, $p \in (0, 1)$, квантиль уровня p распределения Фишера с числом степеней свободы $m-1$ и $n-1$. Гипотеза H_0 отклоняется:

- 1) в пользу H_1 , если $F < f_\alpha(m-1, n-1)$;
- 2) в пользу H_2 , если $F > f_{1-\alpha}(m-1, n-1)$;
- 3) в пользу H_3 , если $F < f_{\alpha/2}(m-1, n-1)$ или $F > f_{1-\alpha/2}(m-1, n-1)$.

В противном случае говорят, что гипотеза H_0 принимается на уровне значимости α или, что нет основания отклонить H_0 или, что H_0 не противоречит экспериментальным данным \vec{X}_n .

Лекция 15

Проверка непараметрических гипотез

Критерий согласия χ^2 . Простая гипотеза

Пусть наблюдается дискретная случайная величина X , принимающая r различных значений u_1, \dots, u_r с положительными вероятностями p_1, \dots, p_r :

$$\mathbf{P}\{X = u_k\} = p_k, \quad k = \overline{1, r}, \quad \sum_{k=1}^r p_k = 1.$$

Допустим, что в выборке $\vec{x}_n = (x_1, \dots, x_n)$ число u_k встретилось $n_k(\vec{x}_n)$ раз, $k = \overline{1, r}$. Отметим, что $\sum_{k=1}^r n_k(\vec{x}_n) = n$, т.е. случайные величины $n_1(\vec{X}_n), \dots, n_r(\vec{X}_n)$ зависимы.

Теорема 15.1 (теорема Пирсона). *Распределение случайной величины*

$$\sum_{k=1}^r \frac{(n_k(\vec{X}_n) - np_k)^2}{np_k}$$

при $n \rightarrow \infty$ сходится к χ^2 -распределению с $r - 1$ степенями свободы. #

Этой теоремой можно воспользоваться для проверки *простой гипотезы*

$$H_0: p_1 = p_{10}, \dots, p_r = p_{r0}, \quad (15.1)$$

где p_{10}, \dots, p_{r0} — известные величины, против альтернативной гипотезы

$$H_1: \text{существуют такие } k, \text{ что } p_k \neq p_{k0}, k = \overline{1, r}. \quad (15.2)$$

Если истинной является гипотеза H_0 , то по закону больших чисел

$$\frac{n_k(\vec{X}_n)}{n} - p_{k0} \rightarrow 0, \quad k = \overline{1, r},$$

а если верна H_1 , то

$$\frac{n_k(\vec{X}_n)}{n} - p_{k0} \rightarrow p_k - p_{k0} \neq 0, \quad \text{для некоторых } k = \overline{1, r}.$$

Поэтому при H_1 случайная величина

$$\chi^2(\vec{X}_n) = \sum_{k=1}^r \frac{(n_k(\vec{X}_n) - np_{k0})^2}{np_{k0}} = n \sum_{k=1}^r \frac{\left(\frac{n_k(\vec{X}_n)}{n} - p_{k0}\right)^2}{p_{k0}} \quad (15.3)$$

стремится к бесконечности и, следовательно, в эксперименте, как правило, принимает большие значения, чем при H_0 , когда ее распределение стремится к распределению χ^2 с $r - 1$ степенями свободы.

Таким образом, становится естественным следующее определение **критерия согласия** χ^2 (хи-квадрат). Этот критерий при больших n на уровне значимости α отклоняет гипотезу H_0 в пользу альтернативной гипотезы H_1 , если

$$\chi^2(\vec{x}_n) > \chi^2_{1-\alpha}(r-1),$$

где $\chi^2_{1-\alpha}(r-1)$ — квантиль уровня $1 - \alpha$ χ^2 -распределения с $r - 1$ степенями свободы, а $\chi^2(\vec{x}_n)$ — реализация случайной величины (15.3).

Если же

$$\chi^2(\vec{x}_n) \leq \chi^2_{1-\alpha}(r-1),$$

то делается вывод о том, что гипотеза H_0 не противоречит статистическим данным и ее следует принять.

Критерием χ^2 при небольших объемах выборки n пользоваться нельзя. На практике при небольших r необходимо, чтобы выполнялись условия $np_k \geq 10$, $k = \overline{1, r}$, а если r велико ($r \geq 20$), достаточно, чтобы было $np_k \geq 5$, $k = \overline{1, r}$.

Критерий χ^2 можно использовать и тогда, когда случайная величина X непрерывна или дискретна, но принимает счетное множество значений с положительными вероятностями. В этом случае множество M возможных значений X разбивают на r непересекающихся подмножеств M_k , $k = \overline{1, r}$, таким образом, чтобы вероятность p_k , $k = \overline{1, r}$, попадания случайной величины X в k подмножество M_k удовлетворяла условию $np_k \geq 5$ или $np_k \geq 10$, $k = \overline{1, r}$. Если X — непрерывная случайная величина, то в качестве M_k , $k = \overline{1, r}$, обычно берут множества вида

$$(-\infty, s_1), [s_1, s_2), \dots, [s_{r-2}, s_{r-1}), [s_{r-1}, \infty),$$

где $s_1 < s_2 < \dots < s_{r-1}$, $s_k \in \mathbb{R}$, $k = \overline{1, r-1}$.

Определим дискретную случайную величину X' , принимающую значение k тогда и только тогда, когда $X \in M_k$, $k = \overline{1, r}$. В этом случае исходная задача проверки статистических гипотез сводится к проверке основной гипотезы (15.1) при альтернативной гипотезе (15.2), где в случае непрерывной случайной величины X

$$p_{k0} = \int_{M_k} f_0(t) dt \quad -$$

вероятность попадания случайной величины X в множество M_k , а $f_0(t)$ — плотность X при H_0 .

Если X — дискретная случайная величина, имеющая счетное множество возможных значений z_1, z_2, \dots , и $\mathbf{P}\{X = z_j\} = q_j > 0$, $j = 1, 2, \dots$, то вместо проверки гипотезы

$$H_0: q_j = q_{j0}, \quad j = 1, 2, \dots,$$

где q_{j0} , $j = 1, 2, \dots$, — известные числа, при альтернативной гипотезе

$$H_1: \text{существуют такие } j, \text{ что } q_j \neq q_{j0}, \quad j = 1, 2, \dots,$$

проверяют гипотезу (15.1) при альтернативной гипотезе (15.2), где вероятности p_{k0} , $k = \overline{1, r}$, вычисляют по формулам

$$p_{k0} = \sum_{z_j \in M_k} q_{j0}, \quad k = \overline{1, r}.$$

Далее для выборки \vec{x}_n находят число $n_k(\vec{x}_n)$ ее элементов, принадлежащих множеству M_k , $k = \overline{1, r}$. Затем, подставляя \vec{x}_n вместо \vec{X}_n в формулу (15.3), определяют реализацию $\chi^2(\vec{x}_n)$ случайной величины $\chi^2(\vec{X}_n)$. Гипотеза H_0 отклоняется в пользу гипотезы H_1 , если $\chi^2(\vec{x}_n) > \chi^2_{1-\alpha}(r-1)$ и принимается в противном случае.

Недостатком использования критерия χ^2 для случайных величин, принимающих бесконечное множество значений, является некоторая потеря информации при переходе от X к случайной величине X' с конечным числом значений.

Пример 15.1. При 4040 бросаниях монеты французский естествоиспытатель Ж.Л.Л. Бюффон (1707–1788) получил 2048 выпадений “герба” и 1992 выпадений “решки”. Совместимо

ли это с гипотезой о том, что вероятность выпадения “герба” при одном бросании равна $1/2$?

Здесь $n = 4040$, $r = 2$, $n_1(\vec{x}_n) = 2048$, $n_2(\vec{x}_n) = 1992$, $p_{10} = p_{20} = 0,5$, число степеней свободы $r - 1 = 1$, и при $\alpha = 0,05$ находим $\chi^2_{0,95}(1) = 3,841$.

Проверим гипотезу H_0 о том, что вероятности p_1 и p_2 выпадения “герба” и “решки” равны $1/2$. На основании (15.3) получаем

$$\chi^2(\vec{x}_n) = \frac{(2048 - 4040 \cdot 0,5)^2}{4040 \cdot 0,5} + \frac{(1992 - 4040 \cdot 0,5)^2}{4040 \cdot 0,5} = 0,776.$$

Так как $0,776 < 3,841$, то статистические данные не противоречат гипотезе H_0 .

Критерий χ^2 для сложной гипотезы

Пусть функция распределения дискретной случайной величины X , принимающей конечное множество значений u_1, \dots, u_r , зависит от неизвестного d -мерного вектора параметров θ . Тогда вероятность p_k того, что X примет возможное значение u_k , зависит от θ , т.е. $p_k = p_k(\theta)$, $k = \overline{1, r}$. А так как вероятности $p_1(\theta), \dots, p_r(\theta)$ полностью определяют функцию распределения случайной величины X , то в рассматриваемом случае *основная гипотеза* принимает следующий вид:

$$H_0: P\{X = u_k\} = p_k(\theta), \quad k = \overline{1, r}, \quad \theta \in \Theta \subset \mathbb{R}^d.$$

Эту сложную гипотезу можно проверить при помощи модификации критерия χ^2 Пирсона.

Пусть $\hat{\theta}(\vec{x}_n)$ — значение оценки $\hat{\theta}(\vec{X}_n)$ максимального правдоподобия для θ , а $n_k(\vec{x}_n)$ — количество элементов выборки \vec{x}_n , равных u_k , $k = \overline{1, r}$. Оценку $\hat{\theta}(\vec{X}_n)$ получают в результате минимизации логарифма функции правдоподобия

$$L(\vec{X}_n; \theta) = \frac{n!}{n_1! \dots n_r!} \prod_{k=1}^r p_k^{n_k(\vec{x}_n)}(\theta), \quad \sum_{i=1}^r n_i(\vec{X}_n) = n,$$

как решение системы уравнений

$$\sum_{k=1}^r \frac{n_k(\vec{X}_n)}{p_k(\theta)} \frac{\partial p_k(\theta)}{\partial \theta_j} = 0, \quad j = \overline{1, d}.$$

Можно показать, что при некоторых предположениях о гладкости функций $p_k(\theta)$, $k = \overline{1, r}$, распределение случайной величины

$$\chi^2(\vec{X}_n) = \sum_{i=1}^r \frac{(n_i(\vec{X}_n) - np_i(\hat{\theta}(\vec{X}_n)))^2}{np_i(\hat{\theta}(\vec{X}_n))}$$

при $n \rightarrow \infty$ сходится к χ^2 -распределению с $r - d - 1$ степенями свободы.

Если X — непрерывная случайная величина с функцией распределения $F(t)$, то, разбивая множество возможных значений X на конечное число непересекающихся подмножеств и переходя к дискретной случайной величине X' , можно проверить сложную гипотезу

$$H_0: F(t) \in \{F(t; \theta), \theta \in \Theta \subset \mathbb{R}^d\}.$$

Необходимо только помнить, что оценку максимального правдоподобия $\hat{\theta}(\vec{X}_n)$ следует строить не по наблюдениям X_1, \dots, X_n случайной величины X , а по значениям $n_1(\vec{x}'_n), \dots, n_r(\vec{x}'_n)$ случайной величины X' , что, как правило, гораздо труднее. Построение такой оценки для наиболее распространенных параметрических семейств распределений (нормального, экспоненциального, пуассоновского и т.д.) можно найти в специальной литературе (см. Г. Крамер).

Критерий независимости, основанный на выборочном коэффициенте корреляции

Пусть (X_i, Y_i) , $i = \overline{1, n}$ — выборка из распределения нормального случайного вектора (X, Y) . Обозначим

$$\hat{\rho}(\vec{X}_n, \vec{Y}_n) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

выборочный коэффициент корреляции.

При проверке статистической гипотезы $H_0: \rho = 0$ (т.е. гипотезы о том, что нормально распределенные случайные величины независимы) используют статистику

$$t = \frac{\hat{\rho}(\vec{X}_n, \vec{Y}_n) \sqrt{n-2}}{\sqrt{1 - \hat{\rho}^2(\vec{X}_n, \vec{Y}_n)}},$$

которая имеет *распределение Стьюдента* с $n - 2$ степенями свободы. Если окажется, что

$$\frac{|\hat{\rho}| \sqrt{n-2}}{\sqrt{1 - \hat{\rho}^2}} \leq t_{1-\alpha/2}(n-2),$$

то гипотезу H_0 принимают при уровне значимости α , где $t_{1-\alpha/2}(n-2)$ — квантиль уровня $1 - \alpha/2$ распределения Стьюдента с $n - 2$ степенями свободы.

Пример 15.2. Вычислим значение $\hat{\rho}$ для пары случайных величин (ξ, η) , где ξ — рост (в см), а η — масса тела (в кг) наугад выбранного студента-первокурсника. Выборка объема $n = 15$ представлена в таблице 15.1.

Чтобы оценить показатель ρ связи двух случайных величин, сначала найдем выборочные средние этих величин:

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = \frac{2620}{15} = 173,3; \quad \bar{y} = \frac{1}{15} \sum_{i=1}^{15} y_i = \frac{945}{15} = 63,1.$$

Номер наблюдения	Рост, см		Масса тела, кг	
	x_i	$x_i - \bar{x}$	y_i	$y_i - \bar{y}$
1	165	-8,3	72,9	9,8
2	171	-2,3	48,4	-14,7
3	182	8,7	66,3	3,2
4	165	-8,3	64,1	1,0
5	183	9,7	62,7	-0,4
6	180	6,7	76,0	12,9
7	183	9,7	73,8	10,7
8	166	-7,3	50,6	-12,5
9	173	-0,3	52,3	-10,8
10	172	-1,3	56,5	-6,6
11	174	0,7	66,8	3,7
12	170	-3,3	61,6	-1,5
13	164	-9,3	72,8	9,7
14	168	-5,3	52,6	-10,5
15	184	10,7	68,6	5,5
Σ	2600		945	

Таблица 15.1

Затем определяем суммы

$$\sum_{i=1}^{15} (x_i - \bar{x})^2 = 747,33; \quad \sum_{i=1}^{15} (y_i - \bar{y})^2 = 1171,4; \quad \sum_{i=1}^{15} (x_i - \bar{x})(y_i - \bar{y}) = 293,3.$$

Таким образом,

$$\hat{\rho} = \frac{293,3}{\sqrt{747,33 \cdot 1171,4}} = 0,313.$$

Пример 15.3. В примере 15.2 найдено значение точечной оценки $\hat{\rho} = 0,313$. Проверим гипотезу $H_0: \rho = 0$ на уровне значимости $\alpha = 0,1$.

По таблице квантилей распределения Стьюдента находим квантиль $t_{0,95}(13) = 1,77$ и сравниваем со значением

$$\hat{\rho} \frac{\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} = 0,313 \frac{\sqrt{13}}{\sqrt{0,902}} = 1,19.$$

Поскольку $1,19 < 1,77$, то гипотезу $\rho = 0$ принимаем.

Таблицы сопряженности признаков и критерий χ^2

Пусть

$$(\vec{X}_n, \vec{Y}_n) = ((X_1, Y_1), \dots, (X_n, Y_n))$$

независимые наблюдения дискретного случайного вектора (X, Y) , где случайная величина X может принимать значения u_1, \dots, u_r , а случайная величина Y — значения v_1, \dots, v_s . Определим случайную величину $n_{ij}(\vec{X}_n, \vec{Y}_n)$, реализация n_{ij} которой равна количеству элементов выборки $(\vec{x}_n, \vec{y}_n) = ((x_1, y_1), \dots, (x_n, y_n))$, совпадающих с элементом $(u_i; v_j)$, $i = \overline{1, r}$, $j = \overline{1, s}$.

Введем случайные величины $n_i(\vec{X}_n, \vec{Y}_n)$ и $n_j(\vec{X}_n, \vec{Y}_n)$, значения n_i и n_j которых определим по формулам

$$n_i = \sum_{j=1}^s n_{ij}, \quad n_j = \sum_{i=1}^r n_{ij}.$$

При этом n_i — количество элементов выборки (\vec{x}_n, \vec{y}_n) , в которых встретилось значение u_i , а n_j — количество элементов выборки (\vec{x}_n, \vec{y}_n) , в которых встретилось значение v_j . Кроме того, имеют место очевидные равенства

$$\sum_{i=1}^r n_i = \sum_{j=1}^s n_j = \sum_{i=1}^r \sum_{j=1}^s n_{ij} = n.$$

В рассматриваемом случае результаты наблюдений удобно оформлять в виде таблицы 15.2, называемой **таблицей сопряженности признаков**. Пусть далее

$$p_{ij} = \mathbf{P}\{X = u_i, Y = v_j\}, \quad p_i = \mathbf{P}\{X = u_i\}, \quad p_j = \mathbf{P}\{Y = v_j\}, \quad i = \overline{1, r}, \quad j = \overline{1, s}.$$

Дискретные случайные величины X и Y независимы тогда и только тогда, когда

$$\mathbf{P}\{X = u_i, Y = v_j\} = \mathbf{P}\{X = u_i\} \mathbf{P}\{Y = v_j\}, \quad i = \overline{1, r}, \quad j = \overline{1, s}.$$

Поэтому основную гипотезу о независимости дискретных случайных величин X и Y можно представить в следующем виде:

$$H_0: p_{ij} = p_i p_j, \quad i = \overline{1, r}, \quad j = \overline{1, s}. \quad (15.4)$$

При этом, как правило, в качестве альтернативной используют гипотезу

$$H_1: p_{ij} \neq p_i p_j \text{ для некоторых } i = \overline{1, r}, \quad j = \overline{1, s}. \quad (15.5)$$

Для проверки основной гипотезы (15.4) при альтернативной гипотезе (15.5) К. Пирсон предложил использовать статистику $\hat{\chi}^2(\vec{X}_n, \vec{Y}_n)$, называемую **статистикой Фишера — Пирсона**, реализация $\hat{\chi}^2(\vec{x}_n, \vec{y}_n)$ которой определяется формулой

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = n \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_i n_j}{n}\right)^2}{n_i n_j}. \quad (15.6)$$

Из закона больших чисел следует, что при $n \rightarrow \infty$

$$\frac{n_{ij}(\vec{X}_n, \vec{Y}_n)}{n} \rightarrow p_{ij}, \quad \frac{n_i(\vec{X}_n, \vec{Y}_n)}{n} \rightarrow p_i, \quad \frac{n_j(\vec{X}_n, \vec{Y}_n)}{n} \rightarrow p_j, \quad i = \overline{1, r}, \quad j = \overline{1, s}.$$

Поэтому при истинности гипотезы H_0 и больших объемах выборки (\vec{x}_n, \vec{y}_n) должно выполняться приближенное равенство

$$n_{ij} \approx n_i n_j, \quad i = \overline{1, r}, \quad j = \overline{1, s},$$

и, следовательно, значения (15.6) статистики $\hat{\chi}^2(\vec{X}_n, \vec{Y}_n)$ должны быть “не слишком велики”. “Слишком большие” значения должны свидетельствовать о том, что H_0 неверна.

Ответ на вопрос о том, какие значения нужно считать слишком большими, а какие — нет, дает следующая теорема.

Теорема 15.2. Если истинна гипотеза H_0 , то распределение статистики $\hat{\chi}^2(\vec{X}_n, \vec{Y}_n)$ при $n \rightarrow \infty$ сходится к распределению случайной величины, имеющей χ^2 -распределение с числом степеней свободы $k = (r-1)(s-1)$:

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\hat{\chi}^2(\vec{X}_n, \vec{Y}_n) < z\} = \int_0^z \frac{t^{\frac{k}{2}-1}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} e^{-\frac{t}{2}} dt, \quad z > 0. \quad \#$$

В соответствии с теоремой 15.2 критерий независимости χ^2 отклоняет гипотезу H_0 на уровне значимости $1 - \alpha$, если

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) > \chi_{1-\alpha}^2((r-1)(s-1)),$$

где $\chi_{1-\alpha}^2((r-1)(s-1))$ — квантиль уровня значимости $1 - \alpha$ χ^2 -распределения с числом степеней свободы $(r-1)(s-1)$. При этом считается, что критерий χ^2 можно использовать, если $n_i n_j / n \geq 5$.

Правую часть равенства (15.6) можно преобразовать к форме, более удобной для практического использования:

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = n \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_i n_j} - 1 \right).$$

В частном, но очень распространенном случае таблиц сопряженности при $r = s = 2$ формула (15.6) для вычисления $\hat{\chi}^2(\vec{x}_n, \vec{y}_n)$ имеет еще более простой вид:

$$\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 n_2 n_1 n_2}. \quad (15.7)$$

Для таблиц сопряженности при $r = s = 2$, как правило, используют статистику $\tilde{\chi}^2(\vec{X}_n, \vec{Y}_n)$ с реализациями

$$\tilde{\chi}^2(\vec{x}_n, \vec{y}_n) = \frac{(n|n_{11}n_{22} - n_{12}n_{21}| - n/2)^2}{n_1 n_2 n_1 n_2},$$

называемую **статистикой Фишера — Пирсона с поправкой Йейтса на непрерывность**, распределение которой лучше согласуется с χ^2 -распределением.

Пример 15.4. В табл. 15.3 приведены данные о распределении цвета волос на голове и бровей у 46542 человек. Проверим на уровне значимости $\alpha = 0,05$ гипотезу о независимости этих признаков. Здесь $n = 46592$, $r = s = 2$, $n_{11} = 30472$, $n_{12} = 3238$, $n_{21} = 3364$, $n_{22} = 9468$, $n_1 = 33710$, $n_2 = 12832$, $n_1 = 33836$, $n_2 = 12706$, число степеней свободы $(r-1)(s-1) = 1$. Из (15.7) получаем $\hat{\chi}^2(\vec{x}_n, \vec{y}_n) = 19,288$. По таблице квантилей χ^2 -распределения находим $\chi_{0,95}^2(1) = 3,84$. Так как $19,288 > 3,84$, то гипотезу о независимости признаков следует отклонить.

Цвет бровей	Цвет волос на голове		Сумма
	светлые	темные	
Светлые	30472	3238	33710
Темные	3364	9468	12832
Сумма	33836	12706	46542

Таблица 15.3.

Лекция 16

Метод наименьших квадратов

Рассмотрим задачу о подборе функции одного переменного — подборе по неточным наблюдениям (измерениям). Предположим, что переменные y и x_1, \dots, x_p связаны линейным соотношением

$$y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_p x_p,$$

где коэффициенты $\theta = (\theta_1, \dots, \theta_p)$ неизвестны. При некоторых значениях $x_{i1}, x_{i2}, \dots, x_{ip}$, $i = \overline{1, n}$, переменных x_1, \dots, x_p (называемых обычно факторами) были произведены измерения переменной y (называемой откликом) со случайной ошибкой ε_i , так что вместо неслучайных величин

$$y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_p x_{ip}, \quad i = \overline{1, n},$$

наблюдались случайные величины

$$Y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_p x_{ip} + \varepsilon_i, \quad i = \overline{1, n}. \quad (16.1)$$

Возникает задача оценивания неизвестных коэффициентов $\theta = (\theta_1, \dots, \theta_p)$ по наблюдениям $Y = (y_1, y_2, \dots, y_n)^T$ и элементам x_{ij} матрицы X размера $n \times p$.

Основное предположение об ошибках состоит в том, что случайные величины $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ считаются независимыми и $M\varepsilon_i = 0$, т.е. систематических ошибок при измерении отклика нет. Менее важные предположения заключаются в том, что ε_i распределены одинаково и по нормальному закону $N(0, \sigma^2)$. Величина σ обычно считается неизвестной. Она численно выражает неточность (изменчивость) измерений, т.е. масштаб случайных ошибок.

Систему (16.1) можно записать в матричном виде

$$Y = X\theta + \varepsilon. \quad (16.2)$$

Один из способов оценивания коэффициентов $\theta = (\theta_1, \dots, \theta_p)$, называемый методом наименьших квадратов состоит в следующем.

Определение 16.1. Оценкой $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$ параметра $\theta = (\theta_1, \dots, \theta_p)$ по методу наименьших квадратов называется точка минимума функции

$$S(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T (Y - X\theta) = \sum_{i=1}^n (Y_i - \theta_1 x_{i1} - \theta_2 x_{i2} - \dots - \theta_p x_{ip})^2.$$

Теорема 16.1. Предположим, что ранг матрицы X равен p . Тогда оценка наименьших квадратов имеет вид

$$\hat{\theta} = (X^T X)^{-1} X^T Y. \quad (16.3)$$

Теорема 16.2. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ — независимые одинаково распределенные случайные величины с $M\varepsilon_i = 0$ и конечной дисперсией $D\varepsilon_i = \sigma^2$. Тогда оценка наименьших квадратов

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

является несмещенной и состоятельной оценкой параметра $\theta = (\theta_1, \dots, \theta_p)$.

Обозначим

$$S(\theta) = (Y - X\theta)^T(Y - X\theta),$$

(d_1, d_2, \dots, d_p) — диагональные элементы матрицы $(X^T X)^{-1}$.

Теорема 16.3. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ — независимые одинаково распределенные нормальные случайные величины с $M\varepsilon_i = 0$ и конечной дисперсией $D\varepsilon_i = \sigma^2$. Тогда оценка наименьших квадратов

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

является несмещенной, состоятельной оценкой параметра $\theta = (\theta_1, \dots, \theta_p)$ и нормальным случайным вектором с математическим ожиданием $\theta = (\theta_1, \dots, \theta_p)$ и ковариационной матрицей $\sigma^2(X^T X)^{-1}$. Интервальная оценка для θ_j уровня доверия $1 - \alpha$ имеет вид $(\hat{\theta}_j - \Delta, \hat{\theta}_j + \Delta)$, где

$$\Delta = t_{1-\alpha}(n-p) \sqrt{\frac{d_j}{n-p} S(\hat{\theta})},$$

а $t_{1-\alpha}(n-p)$ — квантиль распределения Стьюдента уровня $1 - \alpha$ с $n - p$ степенями свободы.

Рассмотрим теперь задачу оценивания зависимости

$$y = \theta_1 \varphi_1(t) + \theta_2 \varphi_2(t) + \dots + \theta_p \varphi_p(t),$$

считая функции $\varphi_1, \varphi_2, \dots, \varphi_p$ известными, по измерениям $Y = (Y_1, Y_2, \dots, Y_n)$ величины y в неслучайных точках t_1, t_2, \dots, t_n со случайными ошибками $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$:

$$Y_i = \theta_1 \varphi_1(t_i) + \theta_2 \varphi_2(t_i) + \dots + \theta_p \varphi_p(t_i) + \varepsilon_i, \quad i = \overline{1, n}. \quad (16.4)$$

Обозначив

$$x_{ij} = \varphi_j(t_i), \quad i = \overline{1, n}, j = \overline{1, p},$$

сведем модель (16.4) к модели (16.2).

Пример 16.1. В “Основах химии” Д. И. Менделеев приводит следующие данные о количестве у азот-

t_i	0	4	10	15	21	29	36	51	68
y_i	66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

Таблица 16.1.

натриевой соли $NaNO_3$, которое можно растворить в 100 г воды в зависимости от температуры t (см. таб. 16.1). Построим, опираясь на эти экспериментальные данные, наилучшую линейную аппроксимацию этой зависимости. Другими словами, найдем приближенную эмпирическую формулу вида

$$y = \theta_1 + \theta_2 t,$$

описывающую зависимость между рассматриваемыми величинами.

Оценим коэффициенты $\theta = (\theta_1, \theta_2)$ по $n = 9$ наблюдениям $y = (y_1, y_2, \dots, y_n)$ случайных величин (Y_1, Y_2, \dots, Y_n) . Используя формулу 16.3, получим для $p = 2$

$$y^T = (66.7, 71.0, 76.3, 80.6, 85.7, 92.9, 99.4, 113.6, 125.1),$$

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 4 & 10 & 15 & 21 & 29 & 36 & 51 & 68 \end{pmatrix},$$

$$X^T X = \begin{pmatrix} 9 & 234 \\ 234 & 10144 \end{pmatrix},$$

$$(X^T X)^{-1} = \begin{pmatrix} 0.2776 & -0.006404 \\ -0.006404 & 0.0002463 \end{pmatrix},$$

$$\hat{\theta} = (X^T X)^{-1} X^T y = (67.5077941981390, 0.870640394088670).$$

Таким образом, приближение $\hat{y}(t)$ зависимости y от t имеет вид

$$\hat{y}(t) = 67.5077941981390 + 0.870640394088670t.$$

Другими словами, количество y азотнатриевой соли $NaNO_3$, которое можно растворить в 100 г воды в зависимости от температуры t можно приближенно вычислить по формуле

$$y \approx 67.5077941981390 + 0.870640394088670t.$$

На рис. 16.1 построены график наблюдений Менделеева и график прогноза второй степени.

Согласно этому линейному прогнозу в 100 г воды можно растворить сколь угодно большое количество соли. Очевидно, что это не так. Другими словами, линейный прогноз не учитывает эффект насыщения солевого раствора и, поэтому, хорош при изменении температуры в небольшом диапазоне.

Для обнаружения эффекта насыщения, найдем теперь квадратичное приближение зависимости из табл. 16.1. Другими словами, аппроксимируем эту зависимость параболой, т.е. построим по этим данным приближенную эмпирическую формулу вида

$$y = \theta_1 + \theta_2 t + \theta_3 t^2,$$

описывающую зависимость между рассматриваемыми величинами.

Оценим коэффициенты $\theta = (\theta_1, \theta_2, \theta_3)$ по $n = 9$ наблюдениям $y = (y_1, y_2, \dots, y_n)$ случайных величин (Y_1, Y_2, \dots, Y_n) . Используя формулу 16.3, получим для $p = 3$

$$y^T = (66.7, 71.0, 76.3, 80.6, 85.7, 92.9, 99.4, 113.6, 125.1),$$

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 4 & 10 & 15 & 21 & 29 & 36 & 51 & 68 \\ 0 & 16 & 100 & 225 & 441 & 841 & 1296 & 2601 & 4624 \end{pmatrix},$$

$$X^T X = \begin{pmatrix} 9 & 234 & 10144 \\ 234 & 10144 & 531828 \\ 10144 & 531828 & 30788836 \end{pmatrix},$$

$$(X^T X)^{-1} = \begin{pmatrix} 0.4878864808 & -0.0299495315 & 0.0003565864 \\ -0.0299495315 & 0.0028828545 & -0.0000399292 \\ 0.0003565864 & -0.0000399292 & 0.0000006047 \end{pmatrix},$$

$$\hat{\theta} = (X^T X)^{-1} X^T y = (66.7061866865310, 0.960401469741897, -0.00135938723367342).$$

Таким образом, приближение $\hat{y}(t)$ зависимости y от t имеет вид

$$\hat{y}(t) = 66.7061866865310 + 0.960401469741897t - 0.00135938723367342t^2.$$

Другими словами, количество y азотнатриевой соли $NaNO_3$, которое можно растворить в 100 г воды в зависимости от температуры t можно приближенно вычислить по формуле

$$y \approx 66.71 + 0.9604t - 0.0013593t^2.$$

На рис. 16.2 построены график наблюдений Менделеева и график прогноза второй степени. Видно, что коэффициент -0.00135938723367342 при t^2 отрицательный, что свидетельствует о наличии эффекта насыщения при растворении соли.

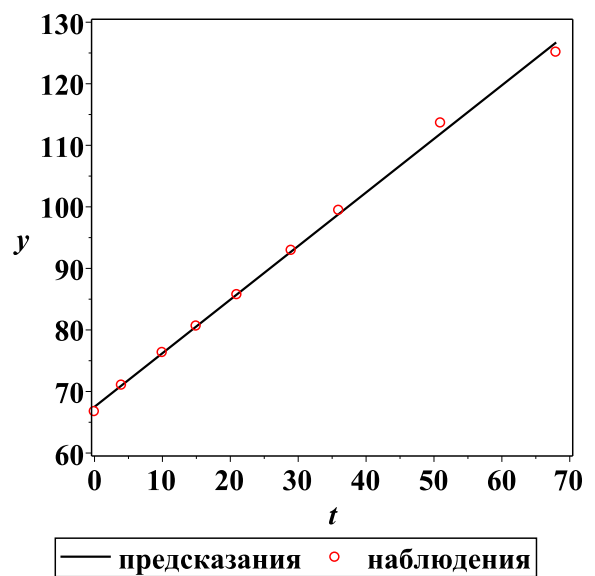


Рис. 16.1

