

**Министерство образования и науки Российской Федерации
Федеральное государственное автономное
образовательное учреждение высшего образования
«Санкт-Петербургский политехнический университет»**

Г. Н. Солопченко

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Учебное пособие

**Издание третье
переработанное и дополненное**

**Санкт-Петербург
Издательство Политехнического университета
2015**

ОГЛАВЛЕНИЕ

1. Теория вероятностей.....	7
1.1. Предмет теории вероятностей.....	7
1.2. Случайные события и вероятность.....	7
1.2.1. Случайные события, виды случайных событий, основные понятия и определения.....	7
1.2.2. Вероятность событий.....	10
1.2.3. Условная вероятность.....	13
1.2.3.1. Формула полной вероятности.....	15
1.2.3.2. Формула Байеса.....	15
1.2.4. Последовательность независимых испытаний (схема Бернулли).....	17
1.3. Дискретные случайные величины.....	18
1.3.1. Распределение вероятностей дискретных случайных величин	18
1.3.2. Биномиальное распределение вероятностей.....	20
1.3.3. Числовые характеристики	22
1.3.4. Моменты функций от случайных величин.....	25
1.3.5. Производящая функция моментов	26
1.3.6. Теорема Пуассона.....	28
1.3.7. Локальная теорема Муавра – Лапласа.....	29
1.4. Двумерные дискретные случайные величины.....	30
1.4.1. Распределение вероятностей.....	30
1.4.2. Числовые характеристики.....	32
1.4.3. Линейное преобразование случайного вектора. Числовые характеристики.....	36
1.4.4. Производящая функция моментов двумерного случайного вектора.....	38
1.5. Энтропия и информация (по Шеннону).....	39
1.6. Непрерывные случайные величины.....	43
1.6.1. Функция распределения и плотность распределения вероятностей.....	44
1.6.2. Числовые характеристики	46
1.6.3. Дифференциальная (относительная) энтропия.....	52
1.6.4. Характеристическая функция непрерывной случайной величины.....	54
1.6.5. Линейные функции непрерывных случайных величин.....	56

1.6.6. Примеры плотности распределения непрерывных случайных величин...	57
1.6.6.1. Случайная величина – интервал времени между импульсами в простейшем потоке.....	57
1.6.6.2. Равномерная плотность распределения.....	61
1.6.6.3. Плотность распределения Arcsin.....	62
1.6.6.4. Нормальная плотность распределения (Гаусса).....	63
1.6.6.5. Плотность распределения Лапласа.....	67
1.6.6.6. Плотность распределения Коши.....	69
1.6.6.7. Экспоненциальное семейство распределений.....	70
1.6.7. Функции от непрерывных случайных величин.....	71
1.6.8. Неравенство Чебышева.....	76
1.7. Двумерные непрерывные случайные величины (двумерные случайные векторы)	78
1.7.1. Функции распределения и плотности распределения.....	78
1.7.2. Числовые характеристики.....	81
1.7.3. Ковариационная матрица случайного вектора.....	85
1.7.4. Линейные функции случайных векторов.....	86
1.7.5. Характеристическая функция двумерного случайного вектора.....	88
1.7.6. Плотность распределения суммы двух независимых случайных величин.....	89
1.7.7. Двумерная нормальная плотность распределения.....	90
1.8. Многомерный случайный вектор	95
2. Математическая статистика.....	98
2.1. Задачи математической статистики.....	98
2.2. Кондиционирование результатов экспериментов.....	99
2.3. Точечное оценивание.....	102
2.3.1. Оценивание квантилей.....	102
2.3.2. Точечное оценивание моментов.....	104
2.3.2.1. Оценивание моментов по выборочной функции распределения.....	105
2.3.2.2. Оценивание моментов по выборочной плотности распределения (по гистограмме).....	106
2.3.3. Оценивание параметров законов распределения.....	108
2.3.4. Свойства точечных оценок.....	109
2.3.4.1. Свойства оценок математического ожидания	110
2.3.4.2. Свойства оценок дисперсии.....	113
2.3.4.3. Свойства оценки математического ожидания случайного вектора....	120
2.3.4.4. Оценка ковариационной матрицы случайного вектора.....	121
2.3.5. Метод максимального правдоподобия.....	122
2.3.6. Метод минимума χ^2	126

2.3.7. Оценивание коэффициентов аппроксимирующих полиномов.....	128
2.3.7.1. Формулировка задачи.....	128
2.3.7.2. Измерения однократные.....	131
2.3.7.3. Плотность распределения величины R^2	135
2.3.7.4. Практически важные замечания.....	136
2.3.7.5. Измерения многократные, характеристики погрешностей известны..	138
2.3.7.6. Измерения многократные, характеристики погрешностей неизвестны.....	140
2.3.7.7. Особенности вычислений при реализации МНК и ОМНК.....	143
2.3.7.8. Основные принципы планирования эксперимента, выполняемого с целью полиномиальной аппроксимации.....	146
2.3.7.9. Расширение класса аппроксимирующих полиномов.....	146
2.4. Интервальное оценивание.....	147
2.4.1. Постановка задачи.....	147
2.4.2. Доверительный интервал для математического ожидания. Дисперсия генеральной совокупности известна.....	148
2.4.3. Доверительный интервал для математического ожидания. Дисперсия генеральной совокупности неизвестна.....	149
2.4.4. Доверительный интервал для дисперсии.....	151
2.4.5. Доверительный интервал для вероятности.....	152
2.4.6. Доверительные интервалы для интерквартильного промежутка.....	156
2.4.6.1. Параметрические толерантные пределы.....	156
2.4.6.2. Непараметрические толерантные пределы.....	158
2.4.7. Бутстреп – оценивание.....	165
2.5. Статистические методы проверки гипотез.....	167
2.5.1. Постановка задачи и общие принципы.....	167
2.5.2. Критическая область и критическое значение.....	168
2.5.3. Простые гипотезы.....	170
2.5.4. Сложные гипотезы.....	172
2.5.5. Проверка гипотезы о виде плотности распределения.....	174
2.5.5.1. Критерий “хи - квадрат”.....	174
2.5.5.2. Критерий Колмогорова – Смирнова.....	179
2.5.5.3. Критерий ω^2 Мизеса.....	182
2.5.6. Проверка гипотез при полиномиальной аппроксимации.....	184
2.5.6.1. Критерий Кочрена проверки гипотезы о равенстве дисперсий.....	184
2.5.6.2. Проверка гипотезы о степени аппроксимирующего полинома, характеристики погрешностей измерений известны.....	185
2.5.6.3. Проверка гипотезы о степени аппроксимирующего полинома, характеристики погрешностей измерений неизвестны.....	190

2.6. Последовательная полиномиальная аппроксимация с проверкой гипотезы о степени полинома.....	193
2.7. Проверка сложных гипотез о числовых характеристиках случайных величин с контролем вероятностей ошибок первого и второго рода.....	196
2.7.1. Основные положения.....	196
2.7.2. Проверка сложной гипотезы о математическом ожидании.....	197
2.7.3. Проверка сложной гипотезы о дисперсии.....	203
2.7.4. Проверка сложной гипотезы об интерквартильном промежутке.....	206
2.7.5. Проверка сложной гипотезы о числовых характеристиках случайных величин с применением доверительных интервалов, построенных методом бутстреп.....	210
2.7.6. Проверка сложной гипотезы о вероятности методом статистического последовательного анализа.....	211
Библиографический список.....	214

1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1.1. ПРЕДМЕТ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Теория вероятностей занимается установлением макрозакономерностей, которым подчиняются массовые однородные случайные события. Теория вероятностей не стремится предсказать единичное событие.

Однородные события – события, которые происходят при осуществлении одних и тех же условий **S** и подчиняются определенным макрозакономерностям независимо от природы событий.

Условия **S** необходимо подробно и тщательно описывать в каждом конкретном случае при постановке задачи исследования случайных событий.

1.2. СЛУЧАЙНЫЕ СОБЫТИЯ И ВЕРОЯТНОСТЬ

1.2.1. Случайные события, виды случайных событий, основные понятия и определения

Событие называется случайным, если в результате испытаний при осуществлении некоторой совокупности условий **S** оно может произойти или не произойти. Под “испытанием” может пониматься проявление какого-либо природного явления или спланированный исследователем эксперимент над рукотворным или природным объектом.

Элементарный исход – результат одного испытания в условиях **S**.

Характеризационное свойство (признак) элементарных исходов: элементарные исходы взаимно исключают друг друга, и в результате каждого испытания может произойти только один из элементарных исходов. Обозначение элементарного исхода ω_i .

Примеры элементарных исходов:

результат бросания монеты на идеальную плоскость;

результат бросания игральной кости и выпадение на верхней грани какого-либо числа;

результат одновременного бросания нескольких игральных костей и выпадение на верхних гранях всех костей определенной комбинации цифр.

Все элементарные исходы, возможные при условиях \mathbf{S} , образуют пространство элементарных исходов Ω : $\omega_i \in \Omega, i = 1, 2, \dots$. Каждый элементарный исход влечет за собой появление какого-либо события. В общем случае событие \mathbf{A} может произойти при появлении элементарных исходов, принадлежащих некоторому подмножеству $\Omega_{\mathbf{A}}$ пространства Ω , $\Omega_{\mathbf{A}} \subset \Omega$.

Пусть в целях некоторого исследования сформулированы условия \mathbf{S} и события \mathbf{A}_1 и \mathbf{A}_2 , которые могут произойти в результате испытаний при появлении элементарных исходов, принадлежащих подмножествам $\Omega_{\mathbf{A}_1}$ и $\Omega_{\mathbf{A}_2}$, $\Omega_{\mathbf{A}_1} \subset \Omega$, $\Omega_{\mathbf{A}_2} \subset \Omega$. Запишем это сопоставление событий и элементарных исходов в виде

$$\mathbf{A}_1 : (\omega \in \Omega_{\mathbf{A}_1}), \quad \mathbf{A}_2 : (\omega \in \Omega_{\mathbf{A}_2}).$$

Пусть в этих же условиях определено событие \mathbf{B} следующим образом: “Событие \mathbf{B} происходит или при осуществлении события \mathbf{A}_1 , или при осуществлении события \mathbf{A}_2 ”. При такой формулировке говорят, что событие \mathbf{B} является объединением событий \mathbf{A}_1 и \mathbf{A}_2 и записывают: $\mathbf{B} = \mathbf{A}_1 \cup \mathbf{A}_2$. В этом случае подмножество элементарных исходов, влекущих за собой событие \mathbf{B} , есть объединение подмножеств $\Omega_{\mathbf{A}_1}$ и $\Omega_{\mathbf{A}_2}$:

$$\mathbf{B} : (\omega \in \Omega_{\mathbf{B}}), \text{ где } \Omega_{\mathbf{B}} = \Omega_{\mathbf{A}_1} \cup \Omega_{\mathbf{A}_2}.$$

Если в этих же условиях принято, что событие \mathbf{B} происходит, когда осуществляются события \mathbf{A}_1 , и \mathbf{A}_2 , то говорят, что событие \mathbf{B} есть пере-

сечение событий A_1 и A_2 , и записывают этот факт в виде $B = A_1 \cap A_2$, причем и в этом случае $B : (\omega \in \Omega_B)$, где $\Omega_B = \Omega_{A_1} \cap \Omega_{A_2}$.

В дальнейшем для упрощения обозначений подмножество элементарных исходов Ω_A будем считать событием A , так же его обозначать, и вместо записи $A : (\omega \in \Omega_A, \Omega_A \subset \Omega)$ или $\Omega_A \subset \Omega$ будем писать $A \subset \Omega$. Такое отождествление удобно использовать также для наглядного представления событий (см. рис. 1 – 3).

Обычно пространство Ω представляют внутренностью прямоугольника, каждая точка которой есть представление элементарного события. В этом случае совокупность элементарных событий Ω_A или Ω_B образует некоторые замкнутые фигуры внутри прямоугольника. Поскольку эти элементарные события отождествляются с порождаемыми ими событиями A и B взаимно однозначно, то в целях упрощения эти замкнутые фигуры обозначаются A и B соответственно.

Случайные события могут образовывать *классы событий*, о чем пойдет речь ниже в разд. 1.2.2.

Виды случайных событий:

достоверное событие $T : (\omega \in \Omega_T, \Omega_T = \Omega)$ – событие, которое непременно происходит при появлении любого элементарного исхода в условиях S ;

невозможное событие \emptyset – событие, которое не может произойти ни при одном элементарном исходе из пространства Ω при условиях S ,

события A и B *несовместны*, если появление одного из них исключает появление другого из них, для несовместных событий можно записать: $\Omega_A \subset \Omega, \Omega_B \subset \Omega, \Omega_A \cap \Omega_B = \emptyset$ – пустое множество, или в упрощенных обозначениях: $A \subset \Omega, B \subset \Omega, A \cap B = \emptyset$;

события A и B *противоположны*, если они несовместны и $A \cup B = T$, в этом случае пользуются обозначениями $B = \bar{A}$ или $A = \bar{B}$;

события A_1, A_2, \dots, A_n образуют *полную группу попарно несовместных событий*, если при условиях S осуществляется только одно из этих событий и

$$A_1 \cup A_2 \cup \dots \cup A_n = T, \quad A_i \cap A_j = \emptyset, \quad i \neq j.$$

1.2.2. Вероятность событий

Вероятность (осуществления) события – числовая характеристика возможности случайного события при условиях **S**. Если $\omega \in \Omega_A$, то вероятность события **A** есть вероятностная мера множества Ω_A , обозначается $P(\mathbf{A})$.

Приведем математические определения вероятности события.

Классическое определение:

$P(\mathbf{A})$ – отношение количества случаев, благоприятствующих появлению события **A**, к общему числу испытаний.

Частотное определение:

$$P(\mathbf{A}) = \lim_{n \rightarrow \infty} \frac{m}{n}$$

где n – общее число выполненных испытаний, m – количество случаев появления события **A** при этих испытаниях.

Современная аксиоматика (А. Н. Колмогоров):

$P(\mathbf{A})$ – неотрицательная монотонная счетно-аддитивная мера возможности случайного события, такая, что $P(\mathbf{T}) = 1$.

Пояснения к этой аксиоматике:

неотрицательность: $P(\mathbf{A}) \geq 0$;

монотонность: если $\mathbf{A} \subset \mathbf{B}$ то $P(\mathbf{A}) \leq P(\mathbf{B})$, то есть, если при наступлении события **A** обязательно наступает событие **B**, но обратное необязательно, то $P(\mathbf{A}) \leq P(\mathbf{B})$;

счетная аддитивность: если условия **S** определены, события \mathbf{A}_i попарно несовместны, то есть $\mathbf{A}_i \cap \mathbf{A}_j = \emptyset, i \neq j$, то

$$P(\mathbf{A}_1 \cup \mathbf{A}_2 \cup \dots \cup \mathbf{A}_n) = P\left(\bigcup_{i=1}^n \mathbf{A}_i\right) = \sum_{i=1}^n P(\mathbf{A}_i).$$

В современной теории вероятностная мера определяется на классах событий. Классы событий образуются таким образом, чтобы они давали возможность определить вероятностную меру вначале на простейших событиях, а затем распространить ее на события любой сложности. Для этого класс событий должен содержать в себе не только сходящиеся в этом клас-

се последовательности событий, но также их пределы. Обозначим класс событий \mathfrak{R} .

Если в условиях \mathbf{S} события $A_i, i = 1, 2, \dots$ принадлежат классу событий \mathfrak{R} , и счетное объединение и счетное пересечение этих событий также принадлежит этому классу, то есть

$$\bigcup_{i=1}^n A_i \in \mathfrak{R} \quad \text{и} \quad \bigcap_{i=1}^n A_i \in \mathfrak{R},$$

то этот класс событий называется *алгеброй событий*.

Если в этих же условиях $A_i \in \mathfrak{R}$, $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{R}$ и $\bigcap_{i=1}^{\infty} A_i \in \mathfrak{R}$, то такой класс событий называется *сигма-алгеброй* (σ -алгеброй).

Пример и з п л а н и м е т р и и. Класс всех многоугольников образует алгебру многоугольников, поскольку пересечение счетного количества многоугольников есть также многоугольник. Дополнение этого класса бесконечными пересечениями и объединениями многоугольников, что мы делали для определения меры (площади) круга, привело к образованию сигма-алгебры, и это позволило с помощью предельного перехода распространить меру (площадь) на геометрические фигуры, которые не являются многоугольниками, но суть пределы, к которым стремятся бесконечные сходящиеся последовательности многоугольников при их объединении и пересечении.

В итоге в соответствии с современной аксиоматикой теории вероятностей говорят, что случайные события и вероятностная мера этих событий определены тройкой: $(\Omega, \mathfrak{R}, P)$, где фигурируют введенные обозначения пространства элементарных событий, сигма-алгебры событий и вероятностной меры на них [4].

Для иллюстрации введенных понятий и свойств вероятностной меры, как было указано ранее, прибегают к геометрическому представлению событий в виде замкнутых фигур внутри прямоугольника, который представляет собой пространство элементарных событий Ω . Вероятностная мера событий отождествляется с площадью соответствующих фигур с учетом того, что вероятностная мера всего прямоугольника равна единице.

Это представление событий и вероятностей показано на рис. 1, который иллюстрирует свойства монотонности (рис. 1, а) и счетной аддитивности (рис. 1, б) вероятностной меры.

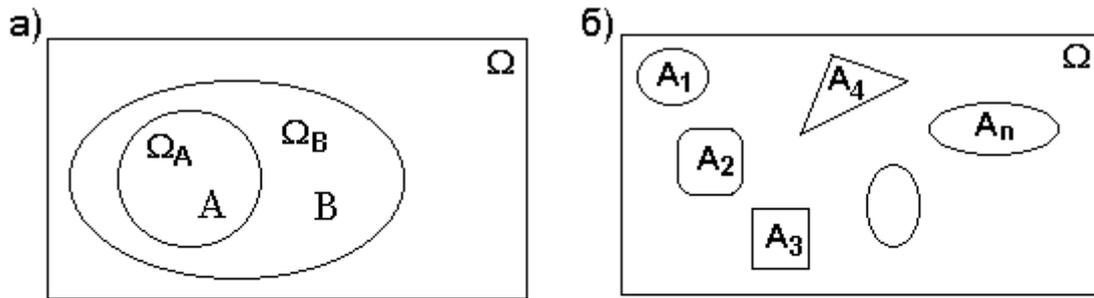


Рис. 1. Геометрическое представление монотонности и счетной аддитивности вероятностной меры

Из аксиоматики Колмогорова следует:

а) $0 \leq P(A) \leq 1$;

б) $T = T \cup \emptyset$, откуда следует: $P(T) = 1 = P(T) + P(\emptyset) = 1 + P(\emptyset)$,

и, наконец, $P(\emptyset) = 0$;

в) $A \subset \Omega$, $B \subset \Omega$, A и B противоположны, то есть $A \cap B = \emptyset$, $A \cup B = T$; тогда $P(A \cup B) = P(A) + P(B) = P(T) = 1$, откуда $P(A) = 1 - P(B)$, то

есть

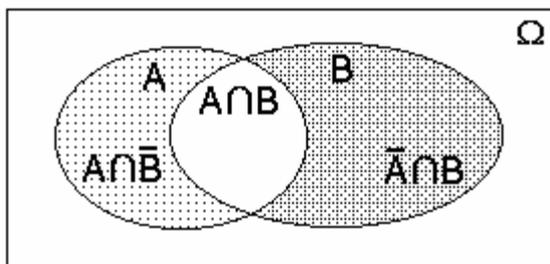


Рис. 2. К выводу формулы для вероятности объединения пересекающихся событий

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

$$P(B) = 1 - P(\bar{B});$$

г) пусть $A \subset \Omega$, $B \subset \Omega$ и события пересекаются, то есть $A \cap B \neq \emptyset$.

В этом случае вероятность объединения событий не равна сумме вероятностей. Для вывода формулы представим объединение пересекающихся событий A и B в

виде объединения трех непересекающихся событий (рис. 2): $A \cup B = (A \cap \bar{B}) \cup (\bar{A} \cap B) \cup (A \cap B)$.

Точно так же представим события A и B :

$$\mathbf{A} = (\mathbf{A} \cap \bar{\mathbf{B}}) \cup (\mathbf{A} \cap \mathbf{B}), \mathbf{B} = (\bar{\mathbf{A}} \cap \mathbf{B}) \cup (\mathbf{A} \cap \mathbf{B}).$$

К этим выражениям можно применить аксиому счетной аддитивности вероятностной меры:

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

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

откуда следует, что $P(\mathbf{A} \cap \bar{\mathbf{B}}) = P(\mathbf{A}) - P(\mathbf{A} \cap \mathbf{B})$ и $P(\bar{\mathbf{A}} \cap \mathbf{B}) = P(\mathbf{B}) - P(\mathbf{A} \cap \mathbf{B})$.

Подставляя эти два выражения в первое, окончательно получим

$$\begin{aligned} P(\mathbf{A} \cup \mathbf{B}) &= P(\mathbf{A}) - P(\mathbf{A} \cap \mathbf{B}) + P(\mathbf{B}) - P(\mathbf{A} \cap \mathbf{B}) + P(\mathbf{A} \cap \mathbf{B}) = \\ &= P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \cap \mathbf{B}). \end{aligned}$$

Очевидно, что в случаях, когда события несовместны, то есть не пересекаются,

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

а в общем случае всегда $P(\mathbf{A} \cup \mathbf{B}) \leq P(\mathbf{A}) + P(\mathbf{B})$.

1.2.3. Условная вероятность

Пусть, как всегда, $\mathbf{A} \subset \Omega$, $\mathbf{B} \subset \Omega$ и сформулированы условия \mathbf{S} .

Условная вероятность есть вероятность осуществления одного из событий при условии, что другое событие состоялось. Обозначение условных вероятностей:

$P(\mathbf{B}/\mathbf{A})$ – условная вероятность события \mathbf{B} при условии осуществления события \mathbf{A} ;

$P(\mathbf{A}/\mathbf{B})$ – условная вероятность события \mathbf{A} при условии осуществления события \mathbf{B} .

Найдем условную вероятность, например, $P(\mathbf{B}/\mathbf{A})$ с помощью классического определения вероятности по разд. 1.2.2. Для этого необходимо знать общее количество n предполагаемых испытаний, в результате которых могут осуществиться события \mathbf{A} , \mathbf{B} , $\mathbf{A} \cap \mathbf{B}$.

Общее количество исходов, при которых возможна реализация условного события \mathbf{B}/\mathbf{A} , определяется исключительно числом появления события \mathbf{A} , поскольку если оно не осуществится, то не осуществится и условное событие \mathbf{B}/\mathbf{A} . Пусть m – число исходов, благоприятствующих появлению события \mathbf{A} . Понятно, что в общем случае $m \leq n$.

В результате предполагаемых испытаний событие **В** может появиться и без появления события **А**, но условное событие **В/А** осуществляется только при совместном появлении событий **А** и **В**. В связи с этим число случаев, благоприятствующих появлению события **А ∩ В**, равно количеству случаев, благоприятствующих появлению события **В/А** или события **А/В**. Обозначим это количество *s*.

Тогда в соответствии с классическим определением вероятности

$$P(\mathbf{B}/\mathbf{A}) = \frac{\text{количество исходов, благоприятствующих событию } \mathbf{A} \cap \mathbf{B}}{\text{количество исходов, благоприятствующих событию } \mathbf{A}},$$

то есть

$$P(\mathbf{B} / \mathbf{A}) = \frac{s}{m} = \frac{s/n}{m/n} = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{A})}.$$

Таким образом, условная вероятность $P(\mathbf{B}/\mathbf{A})$ определяется, как

$$P(\mathbf{B} / \mathbf{A}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{A})}.$$

Точно так же

$$P(\mathbf{A} / \mathbf{B}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{B})}.$$

Из полученных выражений следует, что

$$P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{B} / \mathbf{A})P(\mathbf{A}) = P(\mathbf{A} / \mathbf{B})P(\mathbf{B}).$$

События **А** и **В** независимы, когда $P(\mathbf{B}/\mathbf{A})=P(\mathbf{B})$ и $P(\mathbf{A}/\mathbf{B})=P(\mathbf{A})$.

Отсюда следует формулировка *признака независимости* случайных событий:

события А и В независимы тогда и только тогда, когда

$$P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{A})P(\mathbf{B}).$$

В самом деле, при таком соотношении

$$P(\mathbf{B} / \mathbf{A}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{A})} = \frac{P(\mathbf{A})P(\mathbf{B})}{P(\mathbf{A})} = P(\mathbf{B}).$$

Если **А** и **В** связаны взаимно однозначно, то $m=s$, а потому

$$P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{A}) = P(\mathbf{B}), \quad P(\mathbf{B}/\mathbf{A}) = 1, \quad P(\mathbf{A}/\mathbf{B}) = 1.$$

1.2.3.1. Формула полной вероятности

Пусть при условиях S в Ω возможно событие A . Кроме того в Ω определена полная группа попарно непересекающихся событий $H_i, i=1, 2, \dots, n$, то есть

$$H_i \in \Omega, H_i \cap H_j = \emptyset, \bigcup_{i=1}^n H_i = T.$$

Данная ситуация представлена на рис. 3. Из предъявленных соотношений следует, что

$$A = \bigcup_{i=1}^n (A \cap H_i).$$

Воспользовавшись аксиомой Колмогорова о счетной аддитивности вероятностной меры и математическим определением условных вероятностей, получим:

$$P(A) = P\left(\bigcup_{i=1}^n (A \cap H_i)\right) = \sum_{i=1}^n P(A \cap H_i) = \sum_{i=1}^n P(A)P(H_i/A) = \sum_{i=1}^n P(H_i)P(A/H_i).$$

Таким образом получена *формула полной вероятности*:

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

в которой события H_i именуются *гипотезами*.

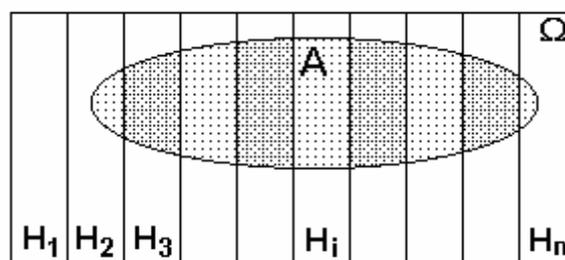


Рис. 3. К выводу формулы полной вероятности

1.2.3.2. Формула Байеса

Воспользовавшись формулой условной вероятности и формулой полной вероятности, получим ценную для многочисленных приложений *формулу Байеса*:

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

Формула Байеса эффективно используется при исследованиях природных явлений, при исследованиях и испытаниях рукотворных объектов в условиях неопределенности математической модели исследуемых объектов и действия мешающих случайных воздействий. В этих условиях события, происходящие при исследованиях, неоднозначно связаны со свойствами и параметрами объектов.

Пусть \mathbf{H}_i – гипотезы (предположения) исследователя о свойствах или параметрах исследуемого объекта, рукотворного или природного. Эти гипотезы могут иметь одинаковый или различный приоритет, который выражается путем задания значений вероятностей $P(\mathbf{H}_i)$. Эти вероятности в данной ситуации суть *априорные вероятности* гипотез \mathbf{H}_i .

В результате эксперимента или исследования событие \mathbf{A} происходит с той или иной вероятностью. Это событие исследователь фиксирует, и по нему он должен вынести суждение об оправданности того или иного априорного предположения (гипотезы). В силу действия случайных факторов и неопределенности математической модели объекта однозначные причинно-следственные связи между предположениями (гипотезами) исследователя и результатами испытаний размыты. После выполнения эксперимента (испытания) фиксируется событие \mathbf{A} . В этой ситуации можно оценить условные вероятности $P(\mathbf{A} / \mathbf{H}_i)$ возможности реализации события \mathbf{A} при справедливости каждой из гипотез. Таким образом после эксперимента правая часть формулы Байеса может быть рассчитана, и формула Байеса дает возможность оценить *апостериорную вероятность* той или иной гипотезы при условии, что результатом эксперимента оказалось событие \mathbf{A} .

Естественно принять в качестве наиболее правдоподобного то предположение (гипотезу), апостериорная вероятность которого окажется наибольшей. Такое *правило принятия решения*, которое основано на применении формулы Байеса, называется *байесовским*. Этой же формулой порожден *принцип максимума апостериорной вероятности*, который часто и эффективно используется в теории и практике систем автоматического регулирования, при математической обработке результатов измерений, при идентификации объектов.

События $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_k$ попарно несовместны, поэтому в соответствии с аксиомой о счетной аддитивности

$$P_n(m) = P(\mathbf{B}_1 \cup \mathbf{B}_2 \cup \dots \cup \mathbf{B}_k) = \sum_{i=1}^k P(\mathbf{B}_i),$$

но $P(\mathbf{B}_i) = p^m q^{n-m}$ независимо от i , поэтому вероятность того, что при n независимых испытаниях событие \mathbf{A} появится ровно m раз, равна

$$P_n(m) = kP(\mathbf{B}_i) = C_n^m p^m q^{n-m} = C_n^m p^m (1-p)^{n-m}.$$

1.3. ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Случайная величина – функция случайного события, $\xi = f(\omega)$. Область определения функции – пространство элементарных исходов Ω , область значений – все вещественные числа.

Дискретная случайная величина – случайная величина, которая может принимать только счетное множество значений.

1.3.1. Распределение вероятностей дискретных случайных величин

Пусть $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n$ – полная группа непересекающихся событий \mathbf{B}_i , которые могут возникать при выполнении некоторых испытаний в условиях \mathbf{S} . $P(\mathbf{B}_1), P(\mathbf{B}_2), \dots, P(\mathbf{B}_n)$ – вероятности событий \mathbf{B}_i .

Пусть также каждому из событий \mathbf{B}_i по какому-либо правилу поставлено во взаимно однозначное соответствие вещественное число x_i . Так мы определили вещественную функцию случайных событий.

При выполнении испытаний эта функция будет принимать значения x_1, x_2, \dots, x_n с вероятностями

$$P(x_1) = P(\mathbf{B}_1) = p_1, P(x_2) = P(\mathbf{B}_2) = p_2, \dots, P(x_n) = P(\mathbf{B}_n) = p_n.$$

Таким образом определена дискретная случайная величина ξ набором значений, которые она может принимать, и набором вероятностей, с которыми она может принимать эти значения:

$$\xi : \begin{cases} x_1, x_2, \dots, x_n \\ p_1, p_2, \dots, p_n \end{cases}.$$

Векторная запись этого определения экономнее: $\xi : (\mathbf{x}, \mathbf{p})$, где \mathbf{x} – вектор значений x_1, x_2, \dots, x_n , \mathbf{p} – вектор вероятностей p_1, p_2, \dots, p_n .

Совокупность значений (\mathbf{x}, \mathbf{p}) представляет собой наиболее полное описание дискретной случайной величины и называется *распределением вероятностей*.

Поскольку случайные события \mathbf{B}_i образуют полную группу непересекающихся событий,

$$\sum_{i=1}^n P(\mathbf{B}_i) = \sum_{i=1}^n p_i = 1.$$

Наряду с распределением вероятностей в теории вероятностей используется еще одна полная характеристика случайных величин, а именно, *функция распределения вероятностей*:

$$F(x) = P(\xi \leq x).$$

На рис. 4 приведен пример графического представления распределения вероятностей и функции распределения вероятностей дискретной случайной величины. Распределение вероятностей дискретной случайной величины изображается решетчатой функцией, высота каждого вертикального отрезка пропорциональна вероятности, с которой случайная величина принимает соответствующее значение.

Как следует из математического определения, функция распределения является неубывающей ступенчатой функцией. Каждый скачок этой функции происходит при значениях, которые может принимать случайная величина, а высота этих скачков пропорциональна соответствующим вероятностям.

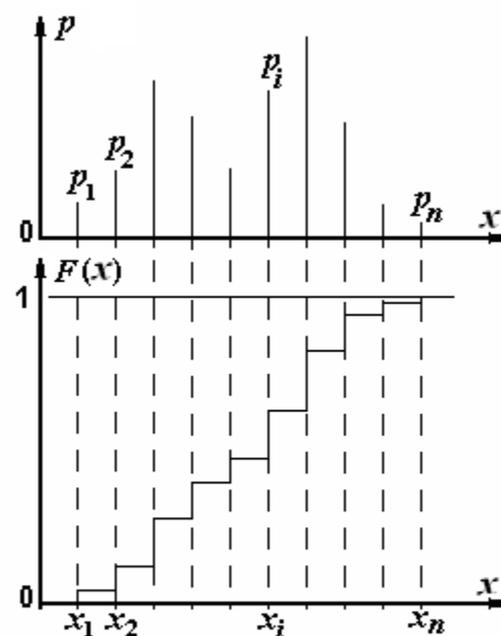


Рис. 4. Распределение и функция распределения вероятностей

Примером дискретной случайной величины может служить величина, порожденная бросанием шестигранной игральной кости. Случайным событием является положение кости на плоскости, которое она принимает в результате ее бросания на эту плоскость. Число вариантов этих положений – шесть, и если кость выполнена в виде идеального куба, то эти положения равновероятны. Правило, в соответствии с которым каждому такому случайному событию сопоставлено число, реализовано путем нанесения на грани кости целых чисел от одного до шести. Таким образом определена случайная величина

$$\xi: \begin{cases} 1, 2, 3, 4, 5, 6 \\ \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \end{cases}.$$

Поскольку все вероятности одинаковы, распределение вероятностей подобного вида называется *равномерным распределением*, а соответствующая случайная величина – *равномерно распределенной случайной величиной*.

Графическое представление такого распределения – решетчатая функция с отрезками равной высоты. Высота ступенек функции распределения одинакова.

1.3.2. Биномиальное распределение вероятностей

Схема Бернулли порождает дискретную случайную величину μ , а именно количество случаев появления события **A** в последовательности n независимых испытаний. Множество значений, которые может принимать эта случайная величина: $m = 0, 1, 2, 3, \dots, n$. Вероятности, с которыми случайная величина принимает эти значения: $P_n(m) = C_n^m p^m (1-p)^{n-m}$. Для того чтобы убедиться, является ли набор этих вероятностей распределением, просуммируем их с использованием формулы бинома Ньютона:

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

Таким образом, случайная величина, порожденная испытаниями по схеме Бернулли, задана своими значениями и распределением вероятностей:

$$\mu: \begin{cases} 0 & 1 & 2 & \dots & i & \dots \\ p_n(0) & p_n(1) & p_n(2) & \dots & p_n(i) & \dots \end{cases}$$

Это распределение называется *биномиальным распределением*.

Для того чтобы представить вид этого распределения, исследуем его на максимальное значение. Для этого вычислим отношение двух соседствующих значений вероятностей:

$$\frac{P_n(m+1)}{P_n(m)} = \frac{n! m! (n-m)!}{(m+1)! (n-m-1)! n!} \frac{p^{m+1} (1-p)^{n-m-1}}{p^m (1-p)^{n-m}} = \frac{n-m}{m+1} \frac{p}{1-p}.$$

Выясним интервал возрастания вероятности, то есть интервал, в котором

$$\frac{n-m}{m+1} \frac{p}{1-p} > 1.$$

В этом неравенстве все сомножители, делимые и делители больше нуля, поэтому его решение не вызовет у читателя затруднений. В результате получим, что интервал монотонного возрастания вероятности простирается от $m = 0$ до $m = \text{Ent}[np+p-1]$, где $\text{Ent}[\bullet]$ – целая часть числа.

Точно так же отыскивается область монотонного убывания:

$$\frac{n-m}{m+1} \frac{p}{1-p} < 1,$$

откуда получаем, что монотонное убывание вероятности начинается от значения $m = \text{Ent}[np+p]$.

Итак, мы нашли, что наибольшее значение вероятность $P_n(m)$ в схеме Бернулли принимает при значении m , ближайшем к np . А это в свою очередь означает, что ближайшее целое к np есть наиболее вероятное значение количества появления события **A** в схеме Бернулли.

1.3.3. Числовые характеристики

Распределение вероятностей случайной величины – это максимально полная ее характеристика, но в то же время очень громоздкая. Кроме того вид распределения реальных случайных величин, которые порождаются техническими причинами в измерительных системах, системах автоматизации, системах передачи и приема информации, чувствителен к неконтролируемым случайным факторам. Поэтому в технических приложениях чаще всего используются не распределения вероятностей, а числовые характеристики расположения случайной величины на числовой оси, разброса ее значений и других свойств.

В качестве таких характеристик применяются *моменты распределения вероятностей*. Это название вызывает естественное желание применить механическую модель, которой мы будем пользоваться при изложении нижеследующего материала.

Механическая система, моделирующая распределение вероятностей, может быть представлена в виде жесткого невесомого стержня (числовая ось), на которую нанизаны точечные массы, расположенные в точках x_1, x_2, \dots, x_n . Значения вероятностей p_1, p_2, \dots, p_n моделируются, как значения массы (или веса) этих точек. Таким образом эти точечные массы придают описанной механической системе вращающий момент относительно начала координат.

В качестве характеристики расположения всех значений случайной величины на оси естественно использовать координату центра тяжести описанной механической системы. Как известно из теоретической механики, координата центра тяжести системы определяется из условия равенства моментов: вращающего момента, создаваемого точечными массами, и противодействующего момента, который должна создавать равнодействующая сила, приложенная к центру тяжести.

Итак, пусть p_1, p_2, \dots, p_n – веса точечных масс, то есть силы, создающие вращающий момент относительно начала координат. Тогда отрезки $[0, x_1], [0, x_2], \dots, [0, x_n]$ – плечи. Равнодействующая сила равна сумме сил, создаваемых точечными массами. Если $x_{ц.т.}$ – координата центра тяжести,

то отрезок $[0, x_{ц.т}]$ – плечо равнодействующей силы. Уравнение моментов имеет вид

$$\sum_{i=1}^n p_i x_i = x_{ц.т} \sum_{i=1}^n p_i, \quad \sum_{i=1}^n p_i = 1.$$

Отсюда следует, что

$$x_{ц.т} = \mu_1[\xi] = \sum_{i=1}^n p_i x_i,$$

где $\mu_1[\xi]$ – первая из характеристик распределения вероятностей и самой случайной величины, которая имеет два равноправных названия: *математическое ожидание* случайной величины ξ и *первый начальный момент* случайной величины ξ . Для обозначения математического ожидания используют также равноправное обозначение $M[\xi]$.

Рассеяние, разбросанность по числовой оси построенной нами механической системы – аналога распределения вероятностей – не зависит от расположения центра тяжести и характеризуется моментом инерции системы относительно ее центра тяжести. Напомним, что момент инерции точки относительно оси вращения пропорционален массе точки и квадрату ее расстояния от оси. Применительно к распределению вероятностей эта характеристика называется *вторым центральным моментом распределения* (случайной величины) или *дисперсией случайной величины*, вычисляется по очевидной формуле для моментов инерции и имеет несколько равноправных обозначений:

$$\mu_2^0[\xi] = D[\xi] = \sigma_\xi^2 = \sum_{i=1}^n p_i (x_i - \mu[\xi])^2,$$

где σ_ξ – *среднеквадратическое значение (отклонение)* случайной величины.

Понятно, что если величина не случайна, а потому в любых обстоятельствах может принимать только одно значение с вероятностью 1, то ее дисперсия будет равна нулю.

Кружок над обозначением момента означает, что этот момент вычисляется относительно центра тяжести.

В теории вероятностей, в отличие от механики, широко используются начальные и центральные моменты более высокого порядка:

$$\text{начальные моменты порядка } k : \quad \mu_k [\xi] = \sum_{i=1}^n p_i x_i^k ;$$

$$\text{центральные моменты порядка } k : \quad \overset{\circ}{\mu}_k [\xi] = \sum_{i=1}^n p_i (x_i - \mu_1 [\xi])^k .$$

Найдем несколько полезных соотношений. Вначале вычислим первый центральный момент:

$$\overset{\circ}{\mu}_1 [\xi] = \sum_{i=1}^n p_i (x_i - \mu_1 [\xi]) = \sum_{i=1}^n p_i x_i - \mu_1 [\xi] \sum_{i=1}^n p_i = \mu_1 [\xi] - \mu_1 [\xi] = 0 ,$$

чего и следовало ожидать.

Найдем теперь соотношение между вторыми моментами: начальным и центральным.

$$\begin{aligned} D[\xi] = \overset{\circ}{\mu}_2 [\xi] &= M[(\xi - \mu_1 [\xi])^2] = \sum_{i=1}^n p_i (x_i - \mu_1 [\xi])^2 = \\ &= \sum_{i=1}^n p_i x_i^2 + \sum_{i=1}^n p_i \mu_1^2 [\xi] - 2 \sum_{i=1}^n p_i x_i \mu_1 [\xi] = \\ &= \mu_2 [\xi] + \mu_1^2 [\xi] \sum_{i=1}^n p_i - 2 \mu_1 [\xi] \sum_{i=1}^n p_i x_i . \end{aligned}$$

Поскольку $\sum_{i=1}^n p_i = 1$ и $\sum_{i=1}^n p_i x_i = \mu_1 [\xi]$, окончательно получим

$$\overset{\circ}{\mu}_2 [\xi] = D[\xi] = \mu_2 [\xi] - \mu_1^2 [\xi] .$$

Размерность моментов k -го порядка есть k -ая степень размерности значений, которые принимает случайная величина. Например, размерность дисперсии есть квадрат размерности значений случайной величины.

1.3.4. Моменты функций от случайных величин

Очевидно, что любая функция от случайной величины есть также случайная величина. Пусть определена взаимно однозначная функция случайной величины $\eta = f(\xi)$. Значения, которые принимают случайные величины ξ и η , связаны функцией $y = f(x)$. Если случайная величина ξ задана в виде

$$\xi : \begin{cases} x_1, x_2, \dots, x_n \\ p_1, p_2, \dots, p_n \end{cases},$$

то случайная величина η будет представлена своими значениями и их вероятностями следующим образом:

$$\eta : \begin{cases} f(x_1), f(x_2), \dots, f(x_n) \\ p_1, p_2, \dots, p_n \end{cases} \quad \text{или} \quad \eta : \begin{cases} y_1, y_2, \dots, y_n \\ p_1, p_2, \dots, p_n \end{cases},$$

где $y_1 = f(x_1), y_2 = f(x_2), \dots, y_n = f(x_n)$.

Моменты случайной величины η записываются очевидным образом:

начальные
$$\mu_k[\eta] = \sum_{i=1}^n p_i y_i^k = \sum_{i=1}^n p_i f^k(x_i);$$

центральные

$$\mu_k^o[\eta] = \sum_{i=1}^n p_i (y_i - \mu_1[\eta])^k = \sum_{i=1}^n p_i (f(x_i) - M[f(\xi)])^k.$$

Рассмотрим полезный для приложений частный случай линейной функции случайной величины, а именно функцию $\eta = a\xi + b$. Ее распределение:

$$\eta : \begin{cases} ax_1 + b, ax_2 + b, \dots, ax_n + b \\ p_1, p_2, \dots, p_n \end{cases},$$

поэтому
$$\mu_1[\eta] = \sum_{i=1}^n p_i (ax_i + b) = a \sum_{i=1}^n p_i x_i + b \sum_{i=1}^n p_i = a\mu_1[\xi] + b.$$

Полученное выражение показывает, что математическое ожидание линейной функции от случайной величины есть функция от математического ожидания этой величины:

$$M[a\xi + b] = aM[\xi] + b.$$

Найдем второй центральный момент, то есть дисперсию этой функции:

$$\overset{0}{\mu}_2[\eta] = D[\eta] = \sum_{i=1}^n p_i (ax_i + b - aM[\xi] - b)^2 = a^2 \sum_{i=1}^n p_i (x_i - \mu_1[\xi])^2 = a^2 \overset{0}{\mu}_2[\xi].$$

В терминах и в обозначении дисперсий это соотношение имеет вид:

$$D[a\xi + b] = a^2 D[\xi].$$

Как и следовало ожидать, из этого выражения следует, что смещение значений случайной величины не влияет на ее дисперсию.

Пусть $\eta = \xi^2$. Тогда $M[\eta] = \mu_1[\eta] = \sum_{i=1}^n p_i x_i^2 = \mu_2[\xi]$.

Пусть $\eta = (\xi - M[\xi])^2$. Тогда

$$M[\eta] = M[(\xi - M[\xi])^2] = \sum_{i=1}^n p_i (x_i - M[\xi])^2 = D[\xi].$$

1.3.5. Производящая функция моментов

Производящей функцией моментов случайной величины ξ называется математическое ожидание функции $\eta = \exp(v\xi)$, где v – аргумент производящей функции моментов:

$$g_\xi(v) = M[\exp(v\xi)] = \sum_{i=1}^n p_i \exp(vx_i).$$

Производящая функция моментов обладает рядом полезных свойств.

1) $g_\xi(0) = \sum_{i=1}^n p_i \exp(0) = \sum_{i=1}^n p_i = 1$;

2) Первая производная от $g_\xi(v)$ по аргументу v :

$$\frac{dg_\xi(v)}{dv} = \sum_{i=1}^n p_i x_i \exp(vx_i), \text{ при } v = 0 \text{ получим } \frac{dg_\xi(v)}{dv} = \sum_{i=1}^n p_i x_i = \mu_1[\xi];$$

3) Вторая производная от $g_\xi(v)$ по аргументу v :

$$\frac{d^2 g_\xi(v)}{d^2 v} = \sum_{i=1}^n p_i x_i^2 \exp(vx_i), \text{ при } v = 0 \text{ получим}$$

$$\frac{d^2 g_{\xi}(v)}{d^2 v} = \sum_{i=1}^n p_i x_i^2 = \mu_2[\xi];$$

4) k -я производная от $g_{\xi}(v)$ по аргументу v :

$$\frac{d^k g_{\xi}(v)}{d^k v} = \sum_{i=1}^n p_i x_i^k \exp(vx_i), \text{ при } v = 0 \text{ получим}$$

$$\frac{d^k g_{\xi}(v)}{d^k v} = \sum_{i=1}^n p_i x_i^k = \mu_k[\xi].$$

Таким образом, чтобы получить значение k -го начального момента, достаточно продифференцировать производящую функцию моментов k раз по v и подставить в полученную производную $v = 0$.

П р и м е р. Написать производящую функцию моментов для биномиального распределения и вычислить математическое ожидание и дисперсию случайной величины, распределенной по биномиальному закону.

$$g_{\xi}(v) = \sum_{m=0}^n P_n(m) \exp(vm) = \sum_{m=0}^n C_n^m p^m e^{vm} q^{n-m} = \sum_{m=0}^n C_n^m (pe^v)^m q^{n-m} = (pe^v + q)^n.$$

Первая производная от $g_{\xi}(v)$ по v

$$\left[(pe^v + q)^n \right]'_v = npe^v (pe^v + q)^{n-1}.$$

Вторая производная от $g_{\xi}(v)$ по v

$$\left[(pe^v + q)^n \right]''_v = npe^v (pe^v + q)^{n-1} + np^2 e^{2v} (n-1) (pe^v + q)^{n-2}.$$

Вычислим эти производные при $v = 0$:

$$\mu_1[m] = M[m] = np, \quad \mu_2[m] = D[m] + M^2[m] = np + n^2 p^2 - np^2.$$

Окончательно получим: $M[m] = np$, $D[m] = np(1-p) = npq$.

Сопоставляя полученное выражение для математического ожидания числа появления события **A** в испытаниях по схеме Бернулли с наиболее вероятным значением этого числа, видим, что они совпадают.

З а м е ч а н и е о сходимости распределений вероятности и производящих функций моментов.

Пусть имеется последовательность распределений вероятностей дискретной случайной величины $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n, \dots$

Пусть $g_i(v)$ – производящие функции соответствующих распределений \mathbf{p}_i из этой последовательности, которые также образуют последовательность $g_1(v), g_2(v), \dots, g_n(v)$

Если последовательность $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n, \dots$ сходится, имеет предел и пределом этой последовательности является распределение \mathbf{p}_{\lim} , то последовательность $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n, \dots$ также сходится, имеет предел, и ее пределом является производящая функция моментов предельного распределения μ . Справедливо и обратное утверждение.

Обратим внимание на то, что конструкция производящей функции моментов близка конструкции обратного дискретного преобразования Фурье, отсюда вытекают полезные свойства производящих функций моментов и близость их свойств свойствам дискретного преобразования Фурье.

1.3.6. Теорема Пуассона

Проанализируем асимптотическое поведение вероятности появления m событий в схеме Бернулли при $n \rightarrow \infty$, $np = \text{const} = \alpha$. Цель – упрощение вычислений, трудоемкость которых сильно возрастает с ростом n .

Задача состоит в нахождении предела последовательности:

$$\lim_{\substack{n \rightarrow \infty \\ np = \alpha}} P_n(m) = \lim_{\substack{n \rightarrow \infty \\ np = \alpha}} C_n^m p^m (1-p)^{n-m}.$$

Из равенства $np = \alpha$ следует, что $p = \frac{\alpha}{n}$. Кроме того,

$$\begin{aligned} C_n^m p^m (1-p)^{n-m} &= \frac{n!}{m!(n-m)!} \left(\frac{\alpha}{n}\right)^m \left(1 - \frac{\alpha}{n}\right)^{n-m} = \\ &= \frac{\alpha^m}{m!} \frac{n(n-1)(n-2)\dots(n-m+1)}{n^m} \left(1 - \frac{\alpha}{n}\right)^{-m} \left(1 - \frac{\alpha}{n}\right)^n = \\ &= \left(\frac{\alpha^m}{m!}\right) \cdot 1 \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) \left(1 - \frac{\alpha}{n}\right)^{-m} \left(1 - \frac{\alpha}{n}\right)^n. \end{aligned}$$

В полученном выражении первый сомножитель не содержит n . Предел последнего сомножителя при $n \rightarrow \infty$ равен. Пределы остальных сомножителей при $n \rightarrow \infty$ равны 1. В результате получаем асимптотическое представление вероятностей из схемы Бернулли, или, что то же самое, биномиального распределения, в виде

$$\lim_{\substack{n \rightarrow \infty \\ np = \alpha}} P_n(m) = \lim_{\substack{n \rightarrow \infty \\ np = \alpha}} C_n^m p^m (1-p)^{n-m} = \frac{\alpha^m}{m!} e^{-\alpha}.$$

Этот результат получен Пуассоном и успешно применяется для расчета вероятностей редких событий (при $n \rightarrow \infty$ вероятность p стремится к 0) при массовых явлениях (испытаниях, опытах).

Полученные предельные значения вероятностей образуют в совокупности распределение вероятностей случайной величины. В самом деле,

$$\sum_{m=0}^{\infty} \frac{\alpha^m}{m!} e^{-\alpha} = e^{-\alpha} \sum_{m=0}^{\infty} \frac{\alpha^m}{m!} = e^{-\alpha} e^{\alpha} = 1.$$

Это распределение называется распределением Пуассона. Математическое ожидание и дисперсия этой случайной величины равны

$$M[m] = np = \alpha, \quad D[m] = npq \approx np = \alpha,$$

поскольку $q = 1 - p \approx 1$ в силу малости p .

Производящая функция распределения Пуассона:

$$g(v) = \sum_{m=0}^{\infty} e^{vm} \frac{\alpha^m}{m!} e^{-\alpha} = e^{-\alpha} \sum_{m=0}^{\infty} \frac{(\alpha e^v)^m}{m!} = e^{-\alpha} e^{\alpha e^v} = \exp[-\alpha(1 - e^v)].$$

1.3.7. Локальная теорема Муавра-Лапласа

В отличие от теоремы Пуассона теорема Муавра-Лапласа посвящена установлению асимптотики для вероятностей событий по схеме Бернулли при $n \rightarrow \infty$ и при $p = \text{const}$.

Здесь без вывода и доказательства приводится результат, полученный Муавром и Лапласом.

Напомним, что в разд. 1.3.5 были получены следующие выражения для математического ожидания и дисперсии случайной величины: числа появления события **A** при n испытаниях по схеме Бернулли

$$M[m] = np, \quad D[m] = npq = np(1-p),$$

где p – вероятность появления события A при одном испытании.

В соответствии с локальной теоремой Муавра-Лапласа значения вероятностей $P_n(m) = C_n^m p^m q^{n-m}$ при $n \rightarrow \infty$ и $p = \text{const}$ аппроксимируются функцией

$$P_n(m) \approx \frac{1}{\sqrt{2\pi npq}} \exp\left[-\frac{(m - np)^2}{2npq}\right].$$

Эта функция симметрична и имеет максимум при $m = np$.

1.4. ДВУМЕРНЫЕ ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

1.4.1. Распределение вероятностей

Будем рассматривать двумерную случайную величину как двумерный случайный вектор

$$\zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix}.$$

Компонентами вектора ζ являются дискретные случайные величины ξ и η , которые могут принимать значения x_1, x_2, \dots, x_n и y_1, y_2, \dots, y_m соответственно. Реализации вектора ζ будем обозначать вектором \mathbf{z} .

При каждом испытании компоненты вектора ζ могут принимать значения (x_i, y_j) с вероятностью совместного осуществления двух событий $P(\xi = x_i \cap \eta = y_j)$. В дальнейшем будем пользоваться упрощенными обозначениями этой вероятности в виде $p(x_i, y_j)$ или p_{ij} .

Кроме того, заметим, что события $(\xi = x_i)$ образуют полную группу попарно несовместных событий. То же самое можно утверждать и о событиях $(\eta = y_j)$.

Представим *совместное распределение вероятностей* $p(x_i, y_j)$ в виде таблицы 1.

События $(\xi = x_i) \cap (\eta = y_1)$, $(\xi = x_i) \cap (\eta = y_2)$, ..., $(\xi = x_i) \cap (\eta = y_m)$ не пересекаются, поэтому в силу счетной аддитивности вероятностной меры и в соответствии с формулой полной вероятности

$$\sum_{j=1}^m P(x_i, y_j) = \sum_{j=1}^m P(x_i / y_j) P(y_j) = P(x_i).$$

Таблица 1

Совместное распределение вероятностей дискретного случайного вектора

$\xi \backslash \eta$	y_1	y_2	\dots	y_j	\dots	y_m	Маргинальное распределение случайной величины ξ
x_1	p_{11}	p_{12}	\dots	p_{1j}	\dots	p_{1m}	$P(\xi = x_1) = p_{\xi 1}$
x_2	p_{21}	p_{22}	\dots	p_{2j}	\dots	p_{2m}	$P(\xi = x_2) = p_{\xi 2}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_i	p_{i1}	p_{i2}	\dots	p_{ij}	\dots	p_{im}	$P(\xi = x_i) = p_{\xi i}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_n	p_{n1}	p_{n2}	\dots	p_{nj}	\dots	p_{nm}	$P(\xi = x_n) = p_{\xi n}$

По этой же причине

$$\sum_{i=1}^n P(x_i, y_j) = \sum_{i=1}^n P(y_j / x_i) P(x_i) = P(y_j).$$

Обе полученные вероятности есть не что иное, как *частные (маргинальные)* распределения вероятностей каждой компоненту случайного вектора ζ . Эти вероятности снабжаются специальным индексом, отражающим принадлежность каждого частного распределения к той или иной компоненте:

$$P_{\xi}(x_i) = p_{\xi i} = \sum_{j=1}^m P(x_i / y_j) P(y_j) = \sum_{j=1}^m P(x_i, y_j),$$

$$P_{\eta}(y_j) = p_{\eta j} = \sum_{i=1}^n P(y_j / x_i) P(x_i) = \sum_{i=1}^n P(x_i, y_j).$$

Сумма всех вероятностей

$$\sum_{i=1}^n \sum_{j=1}^m p_{ij} = \sum_{i=1}^n p(x_i) = \sum_{i=1}^n p_{\xi i} = \sum_{j=1}^m p(y_j) = \sum_{j=1}^m p_{\eta j} = 1.$$

Мы получили *маргинальные (частные) распределения* случайных компонент ξ и η (см. также разд. 1.3.1):

$$\xi: \begin{cases} x_1, x_2, \dots, x_n \\ p_{\xi 1}, p_{\xi 2}, \dots, p_{\xi n} \end{cases}; \quad \eta: \begin{cases} y_1, y_2, \dots, y_m \\ p_{\eta 1}, p_{\eta 2}, \dots, p_{\eta m} \end{cases}.$$

Признак независимости случайных компонент вектора ζ : случайные компоненты ξ и η вектора ζ независимы **тогда и только тогда**, когда их совместное распределение вероятностей может быть представлено как произведение маргинальных (частных) распределений (см. также разд. 1.2.3):

$$p(x_i, y_j) = p_{\xi}(x_i)p_{\eta}(y_j) = p_{\xi i}p_{\eta j}.$$

1.4.2. Числовые характеристики

Числовые характеристики, а именно моменты отдельных составляющих вектора ζ определяются через маргинальные (частные) распределения точно так же, как для одномерной (скалярной) дискретной случайной величины:

начальные моменты k -го порядка

$$\mu_k(\xi) = \sum_{i=1}^n x_i^k p(x_i) = \sum_{i=1}^n \sum_{j=1}^m x_i^k p(x_i, y_j), \quad \mu_k(\eta) = \sum_{j=1}^m y_j^k p(y_j) = \sum_{i=1}^n \sum_{j=1}^m y_j^k p(x_i, y_j),$$

в частности, математические ожидания

$$M[\xi] = \mu_1(\xi) = \sum_{i=1}^n x_i p(x_i) = \sum_{i=1}^n \sum_{j=1}^m x_i p(x_i, y_j),$$

$$M[\eta] = \mu_1(\eta) = \sum_{j=1}^m y_j p(y_j) = \sum_{i=1}^n \sum_{j=1}^m y_j p(x_i, y_j);$$

центральные моменты k -го порядка

$$\overset{\circ}{\mu}_k(\xi) = \sum_{i=1}^n (x_i - M[\xi])^k p(x_i) = \sum_{i=1}^n \sum_{j=1}^m (x_i - M[\xi])^k p(x_i, y_j),$$

$$\overset{\circ}{\mu}_k(\eta) = \sum_{j=1}^m (y_j - M[\eta])^k p(y_j) = \sum_{i=1}^n \sum_{j=1}^m (y_j - M[\eta])^k p(x_i, y_j),$$

в частности, дисперсии

$$D[\xi] = \sigma_{\xi}^2 = \overset{\circ}{\mu}_2(\xi) = \sum_{i=1}^n (x_i - M[\xi])^2 p(x_i) = \sum_{i=1}^n \sum_{j=1}^m (x_i - M[\xi])^2 p(x_i, y_j),$$

$$D[\eta] = \sigma_{\eta}^2 = \overset{\circ}{\mu}_k(\eta) = \sum_{j=1}^m (y_j - M[\eta])^2 p(y_j) = \sum_{i=1}^n \sum_{j=1}^m (y_j - M[\eta])^2 p(x_i, y_j).$$

Для составляющих случайного вектора определены *смешанные моменты*:

начальные моменты порядка k, r

$$\mu_{k,r}(\xi, \eta) = \sum_{i=1}^n \sum_{j=1}^m x_i^k y_j^r p(x_i, y_j);$$

центральные моменты порядка k, r

$$\overset{\circ}{\mu}_{k,r}(\xi, \eta) = \sum_{i=1}^n \sum_{j=1}^m (x_i - M[\xi])^k (y_j - M[\eta])^r p(x_i, y_j).$$

Особое значение для дальнейшего имеет центральный смешанный момент порядка (1, 1), который называется *корреляционным моментом* или *ковариацией*:

$$\begin{aligned} \overset{\circ}{\mu}_{1,1}(\xi, \eta) &= \text{cov}(\xi, \eta) = M[(\xi - M[\xi])(\eta - M[\eta])] = \\ &= \sum_{i=1}^n \sum_{j=1}^m (x_i - M[\xi])(y_j - M[\eta]) p(x_i, y_j). \end{aligned}$$

Для того чтобы установить соотношение между центральным и начальным смешанными моментами раскроем скобки в последнем выражении и выполним несложные преобразования:

$$\begin{aligned} \overset{\circ}{\mu}_{1,1}(\xi, \eta) &= \text{cov}(\xi, \eta) = \sum_{i=1}^n \sum_{j=1}^m (x_i y_j - M[\xi] y_j - M[\eta] x_i + M[\eta] M[\xi]) p(x_i, y_j) = \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j p(x_i, y_j) - M[\eta] \sum_{i=1}^n \sum_{j=1}^m x_i p(x_i, y_j) - M[\xi] \sum_{i=1}^n \sum_{j=1}^m y_j p(x_i, y_j) + M[\xi] M[\eta]. \end{aligned}$$

Окончательно получим $\overset{\circ}{\mu}_{1,1}(\xi, \eta) = \mu_{1,1}(\xi, \eta) - M[\xi] M[\eta]$.

Если ξ и η независимы, то

$$\begin{aligned} \overset{\circ}{\mu}_{1,1}(\xi, \eta) &= \sum_{i=1}^n \sum_{j=1}^m (x_i - M[\xi])(y_j - M[\eta]) p(x_i) p(y_j) = \\ &= \sum_{i=1}^n (x_i - M[\xi]) p(x_i) \sum_{j=1}^m (y_j - M[\eta]) p(y_j) = M[(\xi - M[\xi])] M[(\eta - M[\eta])]. \end{aligned}$$

Но, как было установлено в разд. 1.3.3, $M[\xi - M[\xi]] = 0$ и $M[\eta - M[\eta]] = 0$, поэтому центральный смешанный момент $\overset{\circ}{\mu}_{1,1}(\xi, \eta)$ независимых случайных величин равен нулю. Однако из того, что $\overset{\circ}{\mu}_{1,1}(\xi, \eta) = 0$, независимость случайных величин ξ и η , вообще говоря, не следует. О случайных величинах, корреляционный момент которых равен нулю, говорят, что они *не коррелированы*. Для оценки степени коррелированности случайных величин в приложениях удобнее использовать безразмерный коэффициент корреляции $\rho(\xi, \eta)$. Его значение не зависит от масштаба, в котором выражены значения случайных величин:

$$\rho(\xi, \eta) = \frac{\overset{\circ}{\mu}_{1,1}(\xi, \eta)}{\sigma_{\xi} \sigma_{\eta}}.$$

С целью определения диапазона значений коэффициента корреляции рассмотрим крайний случай взаимно однозначной зависимости между ξ и η , а именно допустим, что $\eta = a\xi + b$.

Из предположенной линейной зависимости следует (см. также разд. 1.3.4):

$$M[\eta] = a M[\xi] + b, \quad \sigma_{\eta}^2 = a^2 \sigma_{\xi}^2, \quad \sigma_{\eta} = |a| \sigma_{\xi}$$

$$\overset{\circ}{\mu}_{1,1}(\xi, \eta) = \sum_{i=1}^n \sum_{j=1}^m (x_i - M[\xi])(ax_j + b - aM[\xi] - b)p(x_i, y_j).$$

После простых преобразований получим

$$\overset{\circ}{\mu}_{1,1}(\xi, \eta) = a \sum_{i=1}^n (x_i - M[\xi])^2 p(x_i) = a \sigma_{\xi}^2,$$

$$\rho(\xi, \eta) = \frac{\overset{\circ}{\mu}_{1,1}(\xi, \eta)}{\sigma_{\xi} \sigma_{\eta}} = \frac{a \sigma_{\xi}^2}{|a| \sigma_{\xi}^2} = 1 \cdot \text{sign}(a).$$

Таким образом мы установили, что коэффициент корреляции не превышает единицу по абсолютной величине: $-1 \leq \rho(\xi, \eta) \leq 1$.

Математическое ожидание случайного вектора – вектор, составляющие (компоненты) которого суть математические ожидания соответствующих компонент:

$$\mathbf{M}[\zeta] = \begin{pmatrix} M[\xi] \\ M[\eta] \end{pmatrix}.$$

Дисперсии компонент случайного вектора ζ и их ковариации объединяют в ковариационную матрицу следующим образом:

$$\Sigma_{\zeta} = \begin{pmatrix} \sigma_{\xi}^2 & \text{cov}(\xi, \eta) \\ \text{cov}(\xi, \eta) & \sigma_{\eta}^2 \end{pmatrix}.$$

В теории вероятностей часто используется корреляционная матрица, которая получается из ковариационной матрицы путем деления ее элементов на произведение среднеквадратических значений:

$$\mathbf{C}_{\zeta} = \begin{pmatrix} \frac{\sigma_{\xi}^2}{\sigma_{\xi}\sigma_{\xi}} & \frac{\text{cov}(\xi, \eta)}{\sigma_{\xi}\sigma_{\eta}} \\ \frac{\text{cov}(\xi, \eta)}{\sigma_{\eta}\sigma_{\xi}} & \frac{\sigma_{\eta}^2}{\sigma_{\eta}\sigma_{\eta}} \end{pmatrix} = \begin{pmatrix} 1 & \rho(\xi, \eta) \\ \rho(\xi, \eta) & 1 \end{pmatrix}.$$

Эти матрицы симметричны и неотрицательно определены. Если компоненты случайного вектора независимы или хотя бы некоррелированы, матрицы Σ_{ζ} и \mathbf{C}_{ζ} диагональны.

Математическое определение ковариационной матрицы:

$$\Sigma_{\zeta} = M \left[(\zeta - M[\zeta])(\zeta - M[\zeta])^T \right],$$

где T – символ транспонирования.

Раскроем это выражение.

$$\begin{aligned} \Sigma_{\zeta} &= M \left[\begin{pmatrix} \xi - M[\xi] \\ \eta - M[\eta] \end{pmatrix} (\xi - M[\xi], \eta - M[\eta]) \right] = \\ &= M \begin{bmatrix} (\xi - M[\xi])^2 & (\xi - M[\xi])(\eta - M[\eta]) \\ (\xi - M[\xi])(\eta - M[\eta]) & (\eta - M[\eta])^2 \end{bmatrix}. \end{aligned}$$

Математическое ожидание случайной матрицы есть матрица, элементы которой суть математические ожидания:

$$\Sigma_{\zeta} = \begin{pmatrix} M[(\xi - M[\xi])^2] & M[(\xi - M[\xi])(\eta - M[\eta])] \\ M[(\xi - M[\xi])(\eta - M[\eta])] & M[(\eta - M[\eta])^2] \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma_{\xi}^2 & \text{cov}(\xi, \eta) \\ \text{cov}(\xi, \eta) & \sigma_{\eta}^2 \end{pmatrix},$$

с чем мы уже ознакомились в настоящем разделе.

1.4.3. Линейное преобразование случайного вектора.

Числовые характеристики

Общий вид линейного преобразования случайного вектора есть умножение его на матрицу и добавление произвольного неслучайного вектора:

$$\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\zeta} + \mathbf{b}.$$

Раскроем это преобразование:

$$\boldsymbol{\psi} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \mathbf{b} = \begin{pmatrix} a_{11}\xi + a_{12}\eta \\ a_{21}\xi + a_{22}\eta \end{pmatrix} + \mathbf{b}.$$

Вычислим вначале математическое ожидание от первой составляющей первого из векторов:

$$\begin{aligned} M[a_{11}\xi + a_{12}\eta] &= \sum_{i=1}^n \sum_{j=1}^m (a_{11}x_i + a_{12}y_j) p(x_i, y_j) = \\ &= \sum_{i=1}^n \sum_{j=1}^m a_{11}x_i p(x_i, y_j) + \sum_{i=1}^n \sum_{j=1}^m a_{12}y_j p(x_i, y_j) = a_{11}M[\xi] + a_{12}M[\eta]. \end{aligned}$$

Точно так же

$$M[a_{21}\xi + a_{22}\eta] = a_{21}M[\xi] + a_{22}M[\eta].$$

Поэтому

$$M[\boldsymbol{\psi}] = \begin{pmatrix} a_{11}M[\xi] + a_{12}M[\eta] \\ a_{21}M[\xi] + a_{22}M[\eta] \end{pmatrix} + \mathbf{b} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} M[\xi] \\ M[\eta] \end{pmatrix} + \mathbf{b} = \mathbf{A}M[\boldsymbol{\zeta}] + \mathbf{b}.$$

В соответствии с математическим определением ковариационной матрицы

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{\psi}} &= M[(\boldsymbol{\psi} - M[\boldsymbol{\psi}])(\boldsymbol{\psi} - M[\boldsymbol{\psi}])^T] = \\ &= M[(\mathbf{A}\boldsymbol{\zeta} + \mathbf{b} - \mathbf{A}M[\boldsymbol{\zeta}] - \mathbf{b})(\mathbf{A}\boldsymbol{\zeta} + \mathbf{b} - \mathbf{A}M[\boldsymbol{\zeta}] - \mathbf{b})^T] = M[\mathbf{A}(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])^T \mathbf{A}^T] = \\ &= \mathbf{A}M[(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])^T] \mathbf{A}^T = \mathbf{A}\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\mathbf{A}^T. \end{aligned}$$

Таким образом, если случайный вектор ζ претерпевает преобразование $\psi = \mathbf{A}\zeta + \mathbf{b}$, то математическое ожидание и ковариационная матрица результата такого преобразования вычисляются по формулам

$$M[\mathbf{A}\zeta + \mathbf{b}] = \mathbf{A}M[\zeta] + \mathbf{b}, \quad \Sigma_{\psi} = \mathbf{A}\Sigma_{\zeta}\mathbf{A}^T.$$

Мы снова убеждаемся в том, что ковариационная матрица не зависит от смещения, которое задается вектором \mathbf{b} .

Рассмотрим в качестве примера важный частный случай. Пусть матрица \mathbf{A} имеет вид $\mathbf{A} = \begin{pmatrix} a & d \end{pmatrix}$, а вектор $\mathbf{b} = 0$. Тогда ψ – скаляр ($\psi = \mathbf{A}\zeta = a\xi + d\eta$), и ковариационная матрица Σ_{ψ} также вырождается в скаляр, а именно в дисперсию, которую будем обозначать σ_{ψ}^2 . Найдем математическое ожидание и дисперсию случайной величины ψ , пользуясь полученными формулами.

$$\begin{aligned} M[\psi] &= \mathbf{A}M[\zeta] = \begin{pmatrix} a & d \end{pmatrix} \begin{pmatrix} M[\xi] \\ M[\eta] \end{pmatrix} = aM[\xi] + dM[\eta], \\ \Sigma_{\psi} &= \sigma_{\psi}^2 = \begin{pmatrix} a & d \end{pmatrix} \begin{pmatrix} \sigma_{\xi}^2 & \text{cov}(\xi, \eta) \\ \text{cov}(\xi, \eta) & \sigma_{\eta}^2 \end{pmatrix} \begin{pmatrix} a \\ d \end{pmatrix} = \\ &= \begin{pmatrix} a\sigma_{\xi}^2 + d\text{cov}(\xi, \eta) & a\text{cov}(\xi, \eta) + d\sigma_{\eta}^2 \end{pmatrix} \begin{pmatrix} a \\ d \end{pmatrix}. \end{aligned}$$

Перемножив эти два вектора, окончательно получим

$$\sigma_{\psi}^2 = a^2\sigma_{\xi}^2 + d^2\sigma_{\eta}^2 + 2ad\text{cov}(\xi, \eta) = a^2\sigma_{\xi}^2 + d^2\sigma_{\eta}^2 + 2ad\sigma_{\xi}\sigma_{\eta}\rho(\xi, \eta).$$

Частные случаи:

случайные величины ξ и η независимы или хотя бы не коррелированы, тогда $\sigma_{\psi}^2 = a^2\sigma_{\xi}^2 + d^2\sigma_{\eta}^2$;

коэффициенты $a = b = 1$, то есть случайная величина ψ есть сумма двух некоррелированных случайных величин ξ и η , тогда $\sigma_{\psi}^2 = \sigma_{\xi}^2 + \sigma_{\eta}^2$, то есть дисперсия суммы некоррелированных случайных величин равна сумме дисперсий слагаемых.

1.4.4. Производящая функция моментов двумерного случайного вектора

Математическое определение производящей функции моментов

$$g_{\xi, \eta}(v_1, v_2) = M \left[e^{v_1 \xi} e^{v_2 \eta} \right] = \sum_{i=1}^n \sum_{j=1}^m e^{v_1 x_i} e^{v_2 y_j} p(x_i, y_j).$$

Свойства этой функции аналогичны свойствам производящей функции моментов одномерной случайной величины, которые были установлены в разд. 1.3.5:

$$\frac{\partial^k}{\partial v_1^k} g_{\xi, \eta}(v_1, v_2) = \sum_{i=1}^n \sum_{j=1}^m x_i^k e^{v_1 x_i} e^{v_2 y_j} p(x_i, y_j) \Big|_{v_1=0, v_2=0} = \sum_{i=1}^n \sum_{j=1}^m x_i^k p(x_i, y_j) = \mu_k(\xi),$$

$$\frac{\partial^k}{\partial v_2^k} g_{\xi, \eta}(v_1, v_2) = \sum_{i=1}^n \sum_{j=1}^m y_j^k e^{v_1 x_i} e^{v_2 y_j} p(x_i, y_j) \Big|_{v_1=0, v_2=0} = \sum_{i=1}^n \sum_{j=1}^m y_j^k p(x_i, y_j) = \mu_k(\eta).$$

Дополнительно к этим свойствам производящие функции моментов двумерных случайных величин позволяют находить смешанные начальные моменты:

$$\begin{aligned} & \frac{\partial^{k+r}}{\partial v_1^k \partial v_2^r} g_{\xi, \eta}(v_1, v_2) = \\ & = \sum_{i=1}^n \sum_{j=1}^m x_i^k y_j^r e^{v_1 x_i} e^{v_2 y_j} p(x_i, y_j) \Big|_{v_1=0, v_2=0} = \sum_{i=1}^n \sum_{j=1}^m x_i^k y_j^r p(x_i, y_j) = \mu_{k,r}(\xi, \eta). \end{aligned}$$

Найдем производящую функцию моментов линейной комбинации двух независимых случайных величин $\psi = a\xi + d\eta$:

$$g_{\psi}(v) = M \left[e^{v\psi} \right] = M \left[e^{av\xi + dv\eta} \right] = \sum_{i=1}^n \sum_{j=1}^m e^{avx_i} e^{dvy_j} p(x_i, y_j).$$

В силу независимости ξ и η совместное распределение $p(x_i, y_j) = p(x_i)p(y_j)$, поэтому

$$g_{\psi}(v) = \sum_{i=1}^n e^{avx_i} p(x_i) \sum_{j=1}^m e^{dvy_j} p(y_j) = g_{\xi}(av) g_{\eta}(dv).$$

При $a = d = 1$ получим удобное выражение $g_{\xi+\eta}(v) = g_{\xi}(v) g_{\eta}(v)$, из которого следует, что производящая функция моментов суммы двух неза-

висимых случайных величин равна произведению производящих функций моментов слагаемых.

Пример. Пусть в одинаковых условиях выполнены две серии независимых испытаний по схеме Бернулли (см. разд. 1.2.4). Количество испытаний в первой серии равно n , во второй серии – k . Производящие функции моментов (см. разд. 1.3.5): $g_1(v) = (pe^v + q)^n$ и $g_2(v) = (pe^v + q)^k$. Найти производящую функцию моментов для случайной величины – суммы количества появления события **A** в этих двух сериях испытаний.

Искомая функция есть произведение:

$$g(v) = (pe^v + q)^n (pe^v + q)^k = (pe^v + q)^{n+k}.$$

Оказывается, что полученная функция – это производящая функция моментов биномиального распределения, соответствующего схеме Бернулли с количеством испытаний, равным $n + k$.

Этот результат свидетельствует о том, что сумма случайных величин, подчиняющихся биномиальному распределению, есть случайная величина, распределенная также по биномиальному закону. Это свойство случайных величин и их распределений называется *безграничной делимостью*. Приведем точную формулировку обнаруженного свойства.

Распределение вероятностей случайной величины безгранично делимо тогда и только тогда, когда эта случайная величина может быть представлена суммой любого количества независимых случайных величин, подчиненных тому же распределению вероятностей.

1.5. ЭНТРОПИЯ И ИНФОРМАЦИЯ (ПО ШЕННОНУ)

К. Шеннон заложил основы теории информации в интересах научной поддержки систем передачи кодированной информации по каналам связи.

Пусть **X** – ансамбль возможных сообщений, которые могут быть переданы по каналам связи. Априорные вероятности, с которыми на передающей стороне может появиться i -ое сообщение, обозначим $p_i = p(x_i)$.

В качестве характеристики априорной неопределенности генерации того или иного сообщения на передающей стороне Шеннон предложил применить следующий функционал энтропии:

$$H(\mathbf{X}) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i).$$

В этом функционале используется логарифм по основанию 2 в связи с тем, что в цифровых каналах связи информация обычно представляется в двоичных кодах и энтропия измеряется в *битах*. Заметим, что единица измерения энтропии – бит отличается от одноименной единицы измерения объема памяти и длины кодовых слов, которая численно равна числу двоичных разрядов слова или количеству разрядов элементарных ячеек памяти компьютера. Так, если ансамбль \mathbf{X} состоит из четырех восьмибитовых кодовых слов, вероятности передачи которых одинаковы и равны $1/4$, то энтропия этого ансамбля равна 2 битам. Объем памяти, необходимой для хранения этого ансамбля восьмибитовых кодовых слов, равен 32 бит.

Энтропия является характеристикой неопределенности состояния ансамбля \mathbf{X} в отличие от характеристики возможности наступления того или иного события, то есть вероятности. Энтропия не зависит от значений, которые может принимать тот или иной элемент ансамбля и от способа его представления. В частности, элементами ансамбля могут быть словесные описания или фотографические изображения. Энтропия характеризует степень хаотичности ансамбля и принимает максимальное значение, когда вероятности $p(x_i)$ одинаковы.

Инструментарий теории информации, выдвинутой К. Шенноном, эффективно используется в теории и практике передачи информации и кодирования. В этих применениях ансамбль \mathbf{X} – ансамбль возможных сообщений. Основные результаты теории информации относятся к передаче информации в двоичном коде и к каналам, ориентированным на передачу двоичных кодов. Передача такой информации осуществляется двоичными кодовыми словами, и при каждой такой передаче искажение отдельного символа в передаваемом слове заключается в том, что вместо 1 получатель принимает 0 или вместо 0 получатель принимает 1. При таких условиях канал передачи двоичной информации характеризуется вероятностью ис-

кажения символа. Схематическое представление простейшего канала связи, а именно двоичного симметричного канала, представлено на рис 5.

Симметричным этот канал называется потому, что вероятности искажения символов 0 и 1 одинаковы и равны q , как это показано на рисунке. Вероятность p – вероятность неискаженной передачи символа.

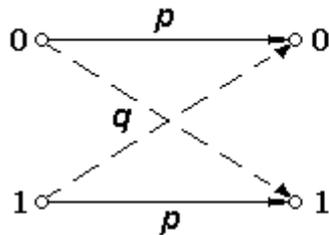


Рис. 5. Двоичный симметричный канал

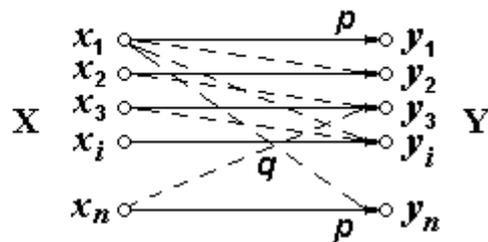


Рис. 6. Передача сообщений по каналу с искажениями

Поскольку каждое сообщение $x_i \in \mathbf{X}$ передается кодовым словом, которое может состоять из нескольких символов 0 или 1, возможностей искажения слова при передаче по такому каналу гораздо больше (см. рис. 6, где принимаемые сообщения обозначены $y_1, y_2, \dots, y_i, \dots \in \mathbf{Y}$). Вероятность искажения слова в общем случае отличается от вероятности искажения отдельного символа.

Понятно, что при передаче по каналам связи сообщения (слова) $x_i \in \mathbf{X}$ и при получении сообщения $y_j \in \mathbf{Y}$ энтропия ансамбля \mathbf{X} , несомненно, уменьшится и будет исчисляться условной энтропией на сообщении

$$H(\mathbf{X} / y_j) = - \sum_{i=1}^n p(x_i / y_j) \log_2 p(x_i / y_j).$$

Для того чтобы характеризовать всю систему приема-передачи, используют среднюю условную энтропию:

$$H(\mathbf{X}/\mathbf{Y}) = \sum_{j=1}^n p(y_j) H(\mathbf{X}/y_j) = - \sum_{j=1}^n \sum_{i=1}^n p(y_j) p(x_i / y_j) \log_2 p(x_i / y_j) =$$

$$\begin{aligned}
&= -\sum_{j=1}^n \sum_{i=1}^n p(x_i, y_j) \log \left(\frac{p(x_i, y_j)}{p(y_j)} \right) = \\
&= -\sum_{j=1}^n \sum_{i=1}^n p(x_i, y_j) \log_2 p(x_i, y_j) + \sum_{j=1}^n \sum_{i=1}^n p(x_i, y_j) \log_2 p(y_j) = \\
&= -\sum_{j=1}^n \sum_{i=1}^n p(x_i, y_j) \log_2 p(x_i, y_j) + \sum_{j=1}^n p(y_j) \log_2 p(y_j).
\end{aligned}$$

Окончательно получим

$$H(\mathbf{X}/\mathbf{Y}) = H(\mathbf{X}, \mathbf{Y}) - H(\mathbf{Y}),$$

где $H(\mathbf{X}, \mathbf{Y})$ – совместная энтропия двух ансамблей, которая, как следует из полученного выражения, равна

$$H(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}/\mathbf{Y}) + H(\mathbf{Y}).$$

Точно так же

$$H(\mathbf{Y}/\mathbf{X}) = H(\mathbf{X}, \mathbf{Y}) - H(\mathbf{X}), \quad H(\mathbf{X}, \mathbf{Y}) = H(\mathbf{Y}/\mathbf{X}) + H(\mathbf{X}).$$

Количество информации об x_i , полученное в одном сообщении y_j :

$$I(x_i, y_j) = \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}.$$

Среднее количество информации об \mathbf{X} , полученное в сообщении y_j

$$\begin{aligned}
I(\mathbf{X}, y_j) &= \sum_{i=1}^n p(x_i / y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} = \sum_{i=1}^n p(x_i / y_j) \log \frac{p(x_i / y_j)}{p(x_i)} = \\
&= -\sum_{i=1}^n p(x_i / y_j) \log p(x_i) - H(\mathbf{X} / y_j).
\end{aligned}$$

Количество информации измеряется в тех же единицах, что и энтропия, – в битах.

Средняя взаимная информация, содержащаяся в \mathbf{Y} об \mathbf{X} или в \mathbf{X} об \mathbf{Y} :

$$I(\mathbf{X}, \mathbf{Y}) = \sum_{j=1}^n p(y_j) I(\mathbf{X}, y_j) = H(\mathbf{X}) - H(\mathbf{X}/\mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X}, \mathbf{Y}),$$

то есть это количество информации численно равно количеству неопределенности, устраненной при получении одного сообщения.

Рассмотрим частные случаи.

1. Ансамбль детерминирован, неопределенность отсутствует, поскольку этот ансамбль содержит лишь одну компоненту, то есть

$$p(x_k) = 1, \quad p(x_i) = 0, i \neq k.$$

В этом случае $H(\mathbf{X}) = 0$, $I(\mathbf{X}, \mathbf{Y}) = 0$. Это означает, что сообщение об известном факте информации не несет.

2. Сообщения x_i, y_j независимы, то есть

$$p(x_i, y_j) = p(x_i)p(y_j), \quad p(x_i / y_j) = p(x_i), \quad p(y_j / x_i) = p(y_j).$$

Тогда

$$H(\mathbf{X}/\mathbf{Y}) = -\sum_{j=1}^n \sum_{i=1}^n p(x_i) \cdot p(y_j) \log p(x_i) = -\sum_{i=1}^n p(x_i) \log p(x_i) = H(\mathbf{X}),$$

откуда следует, что $I(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}/\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}) = 0$, чего и следовало ожидать, ибо независимость x_i и y_j означает отсутствие связи, а потому в \mathbf{Y} не может содержаться информация об \mathbf{X} и наоборот.

3. Связь между элементами \mathbf{X} и \mathbf{Y} взаимно однозначна, то есть

$$p(x_i, y_i) = p(x_i) = p(y_i), \quad p(x_i, y_j) = 0, i \neq j, \quad p(x_i / y_i) = 1.$$

В этом случае

$$H(\mathbf{X}/\mathbf{Y}) = -\sum_{j=1}^n \sum_{i=1}^n p(x_i, y_j) \log p(x_i / y_j) = 0.$$

Тогда $I(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{Y}, \mathbf{X}) = H(\mathbf{X})$. Это значит, что в рассмотренном случае мы имеем дело с идеальным неискажающим каналом связи, и потому количество информации максимально. В общем случае

$$0 \leq I(\mathbf{X}, \mathbf{Y}) \leq H(\mathbf{X}).$$

1.6. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Непрерывная случайная величина – функция случайных событий, принимающая значения на полуоткрытых интервалах.

В соответствии с современной аксиоматикой для определения непрерывной случайной величины устанавливают:

пространство элементарных событий Ω ;

сигма-алгебру полуоткрытых интервалов \mathfrak{R} ;

вероятностную меру на \mathfrak{R} .

Тогда говорят, что непрерывная случайная величина задана тройкой $(\Omega, \mathfrak{R}, P)$.

Такое задание непрерывной случайной величины позволяет применять вероятностную меру точки, поскольку сигма-алгебра полуоткрытых интервалов, по определению, содержит точку как результат более чем счетного пересечения этих интервалов, то есть как предел сходящейся последовательности полуоткрытых интервалов.

1.6.1. Функция распределения и плотность распределения вероятностей

Пусть случайная величина задана тройкой $(\Omega, \mathfrak{R}, P)$. Будем обозначать случайные величины греческими буквами, а значения, которые они принимают – соответствующими латинскими. Например, если случайная величина обозначена буквой ξ , то значения, которые она принимает, будем обозначать буквой x .

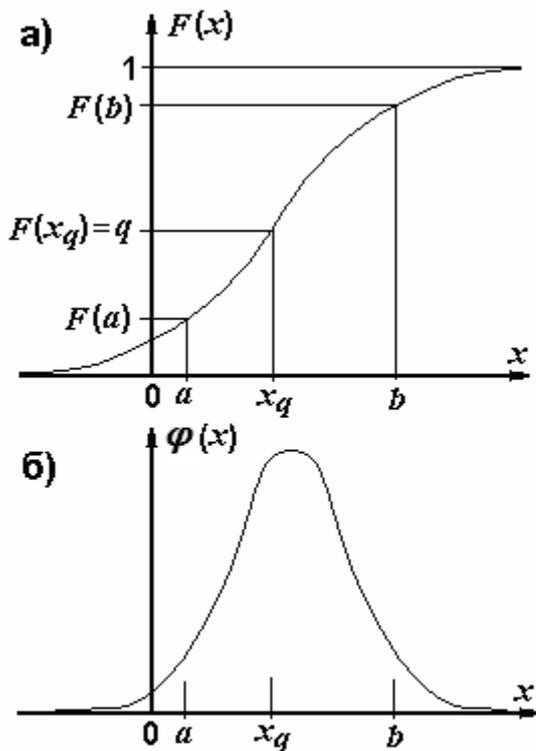


Рис. 7. Функция распределения и плотность распределения

Математическое определение (интегральной) функции распределения непрерывной случайной величины (обычно слово “интегральной” из этого термина опускают):

$$F(x) = P(\xi \leq x).$$

Понятно, что эта функция неотрицательна и не убывает на всей оси. Область значений функции распределения – $[0, 1]$. Пример графика функции распределения непрерывной случайной величины показан на рис. 7 а.

Представим полуоткрытый интервал $(-\infty, b]$, как объединение двух непересекающихся полуоткрытых интервалов :

$$(-\infty, b] = (-\infty, a] \cup (a, b].$$

Тогда в связи с аддитивностью вероятностной меры

$$P(-\infty < \xi \leq b) = P(-\infty < \xi \leq a \cup a < \xi \leq b) = P(-\infty < \xi \leq a) + P(a < \xi \leq b),$$

и, по определению функции распределения,

$$P(-\infty < \xi \leq b) = F(b) = F(a) + P(a < \xi \leq b).$$

Отсюда следует, что $P(a < \xi \leq b) = F(b) - F(a)$.

Механической аналогией, удобной для понимания смысла функции распределения непрерывной случайной величины, может служить ось с переменной плотностью материала, из которого она изготовлена. Тогда механическим аналогом функции распределения может служить зависимость веса левого отрезка оси от координаты разреза. Последнее выражение показывает, что вес отрезка оси, вырезанного из середины, равен разности между весом отрезка $(-\infty, b]$ и отрезка $(-\infty, a]$.

Встав на путь механических аналогий, мы не можем не заинтересоваться такой характеристикой, как удельный вес или плотность материала оси, тем более что плотность изменяется по длине оси, то есть плотность есть функция от x : $\varphi(x)$.

По определению плотности (или удельного веса) бесконечно тонкого стержня,

$$\varphi(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x < \xi \leq x + \Delta x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{dF(x)}{dx}.$$

Функция $\varphi(x)$ называется *плотностью распределения вероятностей* или *законом распределения вероятностей* непрерывной случайной величины. Она представлена на рис. 7, б). Поскольку функция распределения – неубывающая функция, $\varphi(x) \geq 0$.

Для плотности распределения справедливы следующие соотношения:

$$F(x) = \int_{-\infty}^x \varphi(u) du, \quad F(\infty) = \int_{-\infty}^{\infty} \varphi(u) du = 1,$$

где u так же, как x – значения, которые может принимать случайная величина ξ .

Последнее из этих соотношений является условием нормировки плотности распределения вероятностей.

Вероятность того, что в результате испытания случайная величина примет значение из полуоткрытого интервала $(a, b]$, то есть вероятностная

мера этого интервала, вычисляется как площадь криволинейной трапеции, находящейся под кривой плотности распределения над этим интервалом:

$$P(a < \xi \leq b) = F(b) - F(a) = \int_a^b \varphi(x) dx.$$

Таким образом, площадь под всей кривой плотности распределения есть вероятность достоверного события, которое заключается в том, что в результате испытания случайная величина примет какое-либо значение.

Кроме того, из приведенных соотношений следует, что событие $\xi = c$ невозможно, поскольку вероятность того, что в результате испытания непрерывная случайная величина примет конкретное значение, равна нулю:

$$P(\xi = c) = F(c) - F(c) = \int_c^c \varphi(x) dx = 0.$$

Иными словами, это означает, что в условиях, когда определена непрерывная случайная величина, вероятностная мера точки есть нуль.

Отметим, что, хотя точка не является полуоткрытым интервалом, мы имеем право рассматривать вероятностную меру точки, поскольку в классе полуоткрытых интервалов может быть построена бесконечная сходящаяся последовательность, пределом которой является точка, принадлежащая сигма-алгебре \mathfrak{A} .

1.6.2. Числовые характеристики

М о м е н т ы. Моменты непрерывных случайных величин определяются так же, как в разд. 1.3.3 были определены моменты дискретных случайных величин, с той разницей, что здесь вместо суммирования будет применено интегрирование. Обозначения моментов оставим прежними.

Начальные моменты порядка k (или k -е начальные моменты):

$$\mu_k[\xi] = M[\xi^k] = \int_{-\infty}^{\infty} x^k \varphi(x) dx.$$

Центральные моменты порядка k (или k -е центральные моменты):

$$\mu_k^o[\xi] = M[(\xi - M[\xi])^k] = \int_{-\infty}^{\infty} (x - M[\xi])^k \varphi(x) dx.$$

В теории вероятностей и математической статистике особую роль играют несколько моментов от первого до четвертого порядка.

Первый начальный момент $\mu_1[\xi]$ или *математическое ожидание* характеризует положение непрерывной случайной величины на оси. Если прибегнуть к механическим аналогиям, как это было в разд. 1.3.3, математическое ожидание случайной величины – это абсцисса центра тяжести фигуры под кривой плотности распределения.

Эту характеристику находят, как и в разд. 1.3.3, из равенства моментов относительно начала координат: вращающего момента, создаваемого “весом” оси, плотность которой $\varphi(x)$ зависит от координаты, и момента, который должна создать равнодействующая сила, равная “весу” всей оси, чтобы установилось равенство этих моментов:

$$\int_{-\infty}^{\infty} x\varphi(x)dx = \mu_1[\xi] \int_{-\infty}^{\infty} \varphi(x)dx .$$

Поскольку $\int_{-\infty}^{\infty} \varphi(x)dx = 1$, оконча-

тельно получим выражение для абсциссы центра тяжести, то есть математического ожидания:

$$\mu_1[\xi] = M[\xi] = \int_{-\infty}^{\infty} x\varphi(x)dx .$$

Характеристикой разброса значений случайной величины является дисперсия или второй центральный момент (см. рис. 8):

$$D[\xi] = \mu_2^0[\xi] = \sigma_{\xi}^2 = M[(\xi - M[\xi])^2] = \int_{-\infty}^{\infty} (x - M[\xi])^2 \varphi(x)dx ,$$

который в механической интерпретации есть момент инерции стержня с плотностью $\varphi(x)$ относительно центра тяжести (математического ожида-

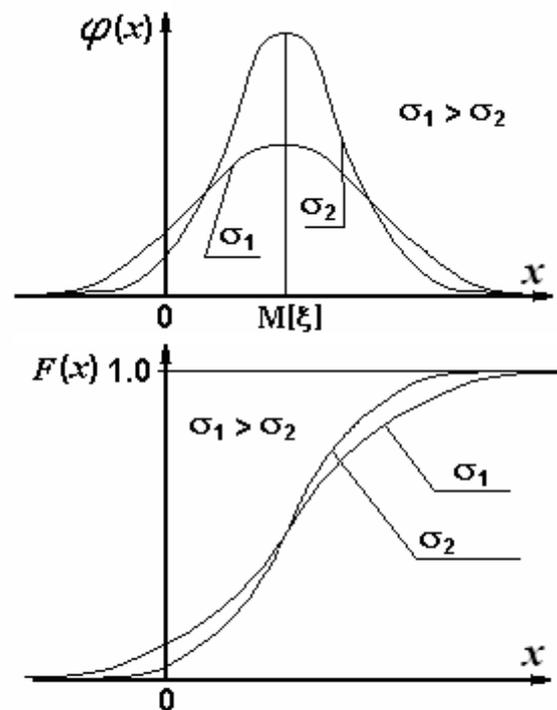


Рис. 8. Ширина кривой плотности распределения при различных значениях дисперсии

ния). С увеличением дисперсии кривая плотности распределения расширяется. Площадь под кривой остается постоянной и равна 1.

В левой части этого выражения приведены все возможные обозначения дисперсии. В частности, корень квадратный из дисперсии

$$\sigma_{\xi} = \sqrt{D[\xi]}$$

имеет самостоятельное применение и называется *среднеквадратическим значением* или *среднеквадратическим отклонением* (сокращенно с.к.о.) случайной величины.

Раскрывая скобки под интегралом в последнем соотношении, найдем полезную связь между вторыми начальным и центральным моментами:

$$\begin{aligned} \overset{\circ}{\mu}_2[\xi] &= \int_{-\infty}^{\infty} (x - M[\xi])^2 \varphi(x) dx = \int_{-\infty}^{\infty} x^2 \varphi(x) dx + \mu_1^2[\xi] \int_{-\infty}^{\infty} \varphi(x) dx - 2\mu_1[\xi] \int_{-\infty}^{\infty} x \varphi(x) dx = \\ &= \mu_2[\xi] + \mu_1^2[\xi] - 2\mu_1^2[\xi]. \end{aligned}$$

Окончательно получим (см. также разд. 1.3.3)

$$\overset{\circ}{\mu}_2[\xi] = \mu_2[\xi] - \mu_1^2[\xi].$$

Данное соотношение часто применяется при компьютерной обработке результатов измерений (испытаний) в темпе получения данных.

Х а р а к т е р и с т и к и ф о р м ы к р и в о й п л о т н о с т и р а с - п р е д е л е н и я. Одной из характеристик формы кривой плотности распределения, а именно, ширины этой кривой является второй центральный момент, то есть дисперсия.

Другими характеристиками формы являются безразмерные величины, производные от третьего и четвертого центральных моментов, то есть *коэффициент асимметрии* As (или асимметрия) и *коэффициент эксцесса* ex (или эксцесс):

$$As = \frac{\overset{\circ}{\mu}_3[\xi]}{\sqrt{\overset{\circ}{\mu}_2^3[\xi]}}, \quad ex = \frac{\overset{\circ}{\mu}_4[\xi]}{\overset{\circ}{\mu}_2^2[\xi]}.$$

Асимметрия – это характеристика степени отклонения от симметричности формы кривой плотности распределения относительно матема-

тического ожидания случайной величины. Асимметрия случайных величин с симметричной формой кривой плотности распределения равна нулю. Как видно из выражения для центральных моментов нечетных порядков, эти моменты равны нулю для всех симметричных плотностей, поскольку в этих случаях подынтегральная функция – нечетная.

Экссесс – это характеристика остроты и степени приближения к оси абсцисс кривой плотности распределения. Экссесс возрастает при заострении вершины кривой плотности распределения или при уменьшении скорости приближения этой кривой к оси абсцисс на бесконечности.

Специфические точки кривых плотности распределения и функции распределения.

Мода – значение абсциссы x_{mod} , при котором кривая плотности распределения имеет максимум. Мода указывает положение высоковероятной области значений случайной величины, то есть области, в которой вероятностная мера интервала фиксированной ширины максимальна. Это свойство становится очевидным, если вспомнить о том, что вероятностная мера интервала – это площадь под кривой плотности распределения над этим интервалом.

Медиана – значение абсциссы x_{med} , при котором фигура под кривой плотности распределения делится на две равновеликие части, площади которых равны по 1/2 каждая, а это значит, что $F(x_{med}) = 0,5$.

Кривые плотности распределения могут иметь более одного максимума, то есть более одной моды. Тогда такие распределения называются *многомодальными*. Из многомодальных распределений чаще всего в тех-

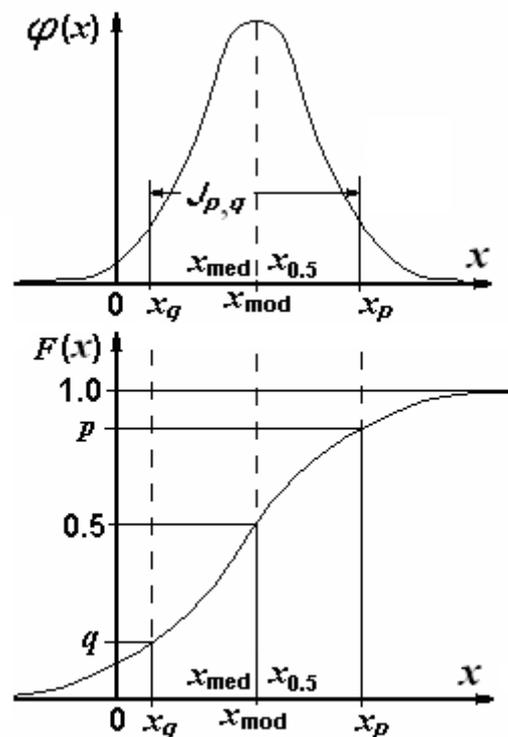


Рис. 9. Характерные значения случайной величины

нических приложениях встречаются *двумодальные распределения*, которые порождены явлениями типа гистерезиса или сухого трения. На рис. 7 – 9 представлены примеры *одномодальных симметричных плотностей* распределения. В частности, из рис. 9 видно, что мода x_{mod} , медиана x_{med} и математическое ожидание таких распределений совпадают:

$$x_{\text{mod}} = x_{\text{med}} = M[\xi].$$

Квантиль – значение абсциссы x_q , которое является решением уравнения (см. рис. 9) $F(x_q) = q$.

Квантиль x_q называется *q-й* или *q·100-процентной квантилью функции распределения* (или плотности распределения, или случайной величины).

В частности, медиана, как следует из ее определения, является 50-процентной квантилью и обозначается, как $x_{0.5}$ или x_{med} .

На рис. 9 показаны квантили: *p·100-процентная квантиль* x_p , *q·100-процентная квантиль* x_q и медиана.

Наиболее употребительные квантили:

уже упоминавшаяся медиана $x_{\text{med}} = x_{0.5}$,

квартиль – 25-процентная квантиль,

дециль – 10-процентная квантиль.

Промежуток между квантилями x_p и x_q , то есть *интерквантильный промежуток* обозначается J_{p-q} и наряду с дисперсией является характеристикой ширины кривой плотности распределения случайной величины.

В технических приложениях интерквантильный промежуток есть характеристика погрешности регулирования, позиционирования или результата измерения. В условиях действия стохастических внешних факторов нормы на эти технические характеристики систем автоматического регулирования, роботов, средств и результатов измерений задаются в виде интервалов с указанием их вероятностной меры, а именно, вероятности того, что указанные погрешности не выйдут за пределы этих норм. Значение ве-

роятности, как правило, задают равным 0.95, а границы интервалов могут быть установлены одним из двух способов: симметрично относительно математического ожидания исследуемой величины или относительно начала координат. Различие между этими вариантами показано на рис. 10. В наших терминах эти интервалы суть не что иное, как интерквантильные промежутки:

$\overset{\circ}{J}_P$ – интерквантильный промежуток, симметричный относительно математического ожидания (см. рис. 10 а),

J_P – интерквантильный промежуток, симметричный относительно начала координат (см. рис. 10 б).

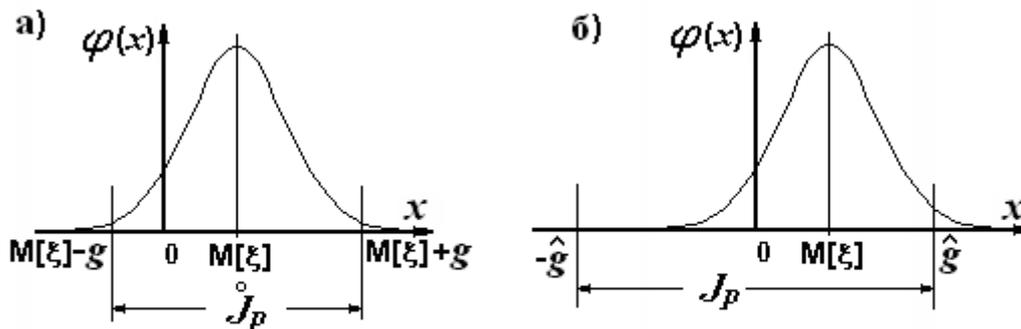


Рис. 10. Варианты задания интерквантильного промежутка

Второй промежуток шире, чем первый. Такая ситуация возникает при назначении норм погрешности технического средства (регулятора, позиционера, средства измерений) при условии, что не устанавливаются отдельные нормы случайной и систематической составляющих этой погрешности.

Если же нормы систематической и случайной составляющих погрешности устанавливаются отдельно (на каждую из них) или если систематическая составляющая исключена путем введения поправки, или если систематическая составляющая погрешности отсутствует, то для нормирования случайной составляющей применяется интерквантильный промежуток

$\overset{\circ}{J}_P$, симметричный относительно математического ожидания.

1.6.3. Дифференциальная (относительная) энтропия

Пусть $\varphi(x)$ – плотность распределения непрерывной случайной величины ξ . Воспользуемся выражением для энтропии, введенное выше в разд. 1.5:

$$H(\mathbf{X}) = - \sum_{i=-\infty}^{\infty} p_i \log_2 p_i,$$

где в качестве p_i будем использовать вероятностные меры интервалов одинаковой ширины Δx .

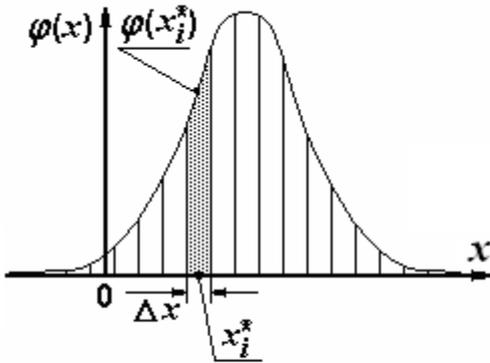


Рис. 11. К определению дифференциальной энтропии

Каждая из этих вероятностных мер численно равна площади полосы шириной Δx , вырезанной из фигуры, ограниченной плотностью распределения и осью абсцисс, как это показано на рис. 11.

По теореме о среднем, эта вероятностная мера равна площади прямоугольника, равновеликого заштрихованной криволинейной трапеции. Основание этого прямоугольника Δx ,

высота его равна $\varphi(x_i^*)$, то есть плотности распределения в некоторой средней точке x_i^* . Тогда

$$p_i = \varphi(x_i^*) \Delta x,$$

$$\begin{aligned} H(\mathbf{X}) &= - \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x \log_2 [\varphi(x_i^*) \Delta x] = \\ &= - \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x \log_2 \varphi(x_i^*) - \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x \log_2 \Delta x = \\ &= - \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x \log_2 \varphi(x_i^*) - (\log_2 \Delta x) \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x = \\ &= - \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x \log_2 \varphi(x_i^*) - \log_2 \Delta x. \end{aligned}$$

Теперь для получения окончательного решения необходимо совершить предельный переход при $\Delta x \rightarrow 0$. Однако из-за того, что $\lim_{\Delta x \rightarrow 0} \log_2 \Delta x = -\infty$, предел $\lim_{\Delta x \rightarrow 0} H(\mathbf{X}) = \infty$. Этого следовало ожидать, поскольку непрерывная случайная величина на любом сколь угодно малом отрезке может принимать бесчисленное множество значений, и в отличие от дискретной величины все ее значения образуют более чем счетное множество. Мера неопределенности такого ансамбля не может быть меньше бесконечности вне зависимости от вида плотности распределения.

Для того чтобы, несмотря на это обстоятельство, получить возможность сопоставления непрерывных случайных величин по значению энтропии, условились отсчитывать энтропию непрерывной случайной величины от некоторого уровня, который принимается за нуль. В качестве такого нулевого отсчета энтропии непрерывной случайной величины используется энтропия другой непрерывной случайной величины, плотность распределения которой равномерна на интервале $(0, 1)$. График этой плотности распределения представлен на рис. 12. Ее аналитическая запись имеет вид

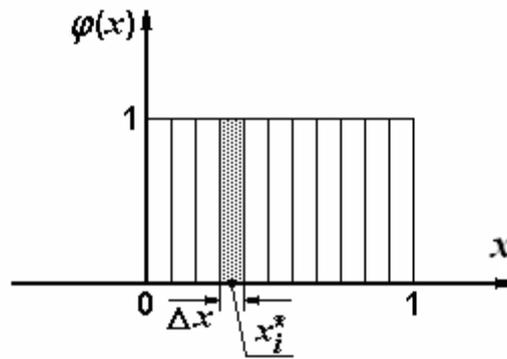


Рис. 12. К расчету энтропии непрерывной случайной величины с равномерной плотностью распределения

Как и ранее, $p_i = \varphi(x_i^*) \Delta x$,
 $H_{\text{равн}}(\mathbf{X}) = -\sum_{i=0}^n \varphi(x_i^*) \Delta x \log_2 [\varphi(x_i^*) \Delta x] = -\sum_{i=0}^n \Delta x \log_2 \Delta x = -n \Delta x \log_2 \Delta x$,

где $n \Delta x = 1$. С учетом этого равенства

$$H_{\text{равн}}(\mathbf{X}) = -\log_2 \Delta x.$$

Энтропия этой случайной величины в пределе при $\Delta x \rightarrow 0$ также стремится к бесконечности, но принимая эту энтропию за начало отсчета, определим *относительную (дифференциальную) энтропию* непрерывной случайной величины, как предел разности:

$$H_\varepsilon(\mathbf{X}) = \lim_{\Delta x \rightarrow 0} (H(\mathbf{X}) - H_{\text{равн}}(\mathbf{X})) = \lim_{\Delta x \rightarrow 0} \left[- \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x \log_2 \varphi(x_i^*) - \log_2 \Delta x + \log_2 \Delta x \right] =$$

$$= - \lim_{\Delta x \rightarrow 0} \sum_{i=-\infty}^{\infty} \varphi(x_i^*) \Delta x \log_2 \varphi(x_i^*) = - \int_{-\infty}^{\infty} \varphi(x) \log_2 \varphi(x) dx.$$

Итак, *относительная, или дифференциальная энтропия* отсчитывается от энтропии случайной величины, распределенной равномерно на интервале $(0, 1]$, поэтому она может быть отрицательной. Выпишем отдельно формулу, которая определяет относительную энтропию:

$$H_\varepsilon(\mathbf{X}) = - \int_{-\infty}^{\infty} \varphi(x) \log_2 \varphi(x) dx.$$

1.6.4. Характеристическая функция непрерывной случайной величины

Характеристической функцией непрерывной случайной величины ξ называется математическое ожидание случайной функции $\eta = \exp(jv\xi)$ (см., например, [3, 4]):

$$g_\xi(v) = M \left[e^{jv\xi} \right] = \int_{-\infty}^{\infty} e^{jv x} \varphi(x) dx.$$

Иными словами, характеристическая функция случайной величины есть интегральное преобразование плотности распределения этой случайной величины. Это преобразование – частный случай применения обратного преобразования Фурье к функциям, обладающим специфическими свойствами, присущими плотности распределения, а именно к неотрицательным функциям, интеграл от которых по всему множеству их определения равен 1. Отсюда следует, что характеристическая функция и плотность распределения связаны взаимно однозначно, то есть

$$\varphi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jvx} g_{\xi}(v) dv.$$

Приведем несколько полезных свойств характеристической функции, первое из которых порождено спецификой плотности распределения случайной величины.

а) При $v = 0$

$$g_{\xi}(0) = \int_{-\infty}^{\infty} \varphi(x) dx = 1,$$

$$\frac{dg_{\xi}(v)}{dv} = \int_{-\infty}^{\infty} jxe^{jvx} \varphi(x) dx = j \int_{-\infty}^{\infty} x \varphi(x) dx = jM[\xi] = j\mu_1[\xi],$$

$$\frac{dg_{\xi}^2(v)}{dv^2} = \int_{-\infty}^{\infty} (jx)^2 e^{jvx} \varphi(x) dx = (j)^2 \int_{-\infty}^{\infty} (x)^2 \varphi(x) dx = (j)^2 \mu_2[\xi],$$

.....

$$\frac{dg_{\xi}^k(v)}{dv^k} = \int_{-\infty}^{\infty} (jx)^k e^{jvx} \varphi(x) dx = (j)^k \int_{-\infty}^{\infty} (x)^k \varphi(x) dx = (j)^k \mu_k[\xi].$$

Эти равенства означают, что для определения *начальных моментов всех порядков* достаточно знать выражение для характеристической функции, продифференцировать его k раз по аргументу v , подставить в полученную производную значение $v = 0$ и разделить результат на $(j)^k$. В частности, если необходимо определить дисперсию, то придется найти первый и второй начальные моменты и затем воспользоваться соотношением, которое было получено в разд. 1.3.3:

$$D[\xi] = \mu_2[\xi] - M^2[\xi].$$

б) Если задана случайная величина η , которая является линейной функцией случайной величины ξ : $\eta = a\xi + b$, то характеристическая функция случайной величины η есть

$$g_{\eta}(v) = M[e^{jv(a\xi+b)}] = \int_{-\infty}^{\infty} e^{jv(ax+b)} \varphi(x) dx = e^{jvb} g_{\xi}(av).$$

Это свойство, в частности, означает, что при простом смещении значений случайной величины по оси абсцисс на величину b характеристическая функция умножается на экспоненту в степени jvb : если $\eta = \xi + b$, то $g_{\eta}(v) = e^{jvb} g_{\xi}(v)$.

в) Пусть имеется последовательность $\varphi_1(x), \varphi_2(x), \dots$ плотностей распределения непрерывных случайных величин. Пусть $g_1(v), g_2(v), \dots$ – характеристические функции этих случайных величин. Если последовательность плотностей распределения сходится и предельная плотность распределения $\varphi(x)$, то последовательность характеристических функций также сходится и имеет предельную функцию $g(v)$, которая есть характеристическая функция случайной величины с предельной плотностью распределения $\varphi(x)$. Иными словами, из сходимости плотностей распределения случайных величин следует сходимость их характеристических функций. Справедливо и обратное утверждение: из сходимости характеристических функций случайных величин следует сходимость их плотностей распределений.

Это свойство очевидным образом следует из взаимно однозначной связи между характеристическими функциями и плотностями распределения.

Необходимые условия того, чтобы некоторая функция $g(v)$ была характеристической функцией:

$g(v)$ непрерывна по v ;

$g(v)$ определена на каждом конечном интервале v ;

$g(0) = 1$;

$|g(v)| \leq 1$.

1.6.5. Линейные функции непрерывных случайных величин

Пусть задана функция случайной величины: $\eta = a\xi + b$. Вычислим математическое ожидание и дисперсию этой функции.

$$M[\eta] = \mu_1[\eta] = \int_{-\infty}^{\infty} (ax + b)\varphi(x)dx = a \int_{-\infty}^{\infty} x\varphi(x)dx + b \int_{-\infty}^{\infty} \varphi(x)dx,$$

откуда получаем:

$$M[\eta] = M[a\xi + b] = aM[\xi] + b,$$

иными словами, математическое ожидание линейной функции есть функция математического ожидания.

$$D[\eta] = M[(ax + b - M[a\xi + b])^2] = M[(ax - aM[\xi])^2] = a^2 M[(x - M[\xi])^2],$$

откуда:

$$D[\eta] = D[a\xi + b] = a^2 D[\xi].$$

Как и следовало ожидать, дисперсия не зависит от сдвига, а множитель выносится за знак дисперсии во второй степени.

1.6.6. Примеры плотности распределения непрерывных случайных величин

1.6.6.1. Случайная величина – интервал времени между импульсами в простейшем потоке

Простейший случайный поток импульсов, заявок на обслуживание и тому подобных последовательностей случайных событий задается следующими свойствами:

поток импульсов (заявок) рассматривается в течение интервала времени, равного t ;

в любом бесконечно малом, но не нулевом промежутке времени Δt , на которые поделен интервал времени $(0, t]$ и который будем называть *элементарным промежутком*, с вероятностью $p > 0$ может появиться только один импульс (одна заявка);

появление импульса (заявки) в каждом из элементарных промежутков времени Δt есть событие, не зависящее от предыстории, то есть от появления импульса (заявки) в предыдущие промежутки времени, это свойство – *отсутствие последействия*, или *отсутствие памяти*.

Такие условия осуществления событий, которые заключаются в случайном появлении импульсов (заявок) в следующие друг за другом элементарные промежутки времени, в точности соответствуют схеме Бернулли последовательности независимых испытаний (см. разд. 1.3.2). Если n – количество элементарных промежутков на интервале $(0, t]$, то $t = n\Delta t$, а np – математическое ожидание количества импульсов (заявок) на этом ин-

тервале (то есть среднее значение их количества). Тогда естественным образом может быть вычислена средняя частота или интенсивность следования импульсов (заявок) в течение интервала $(0, t]$ как отношение среднего их количества за этот интервал к длительности интервала: $\lambda = np/t$. Отсюда, в частности, следует, что $np = \lambda t$. Единица измерения частоты $\lambda - [1/c]$.

Найдем функцию распределения непрерывной случайной величины τ – времени между двумя последовательными импульсами (заявками) в потоке:

$$F(t) = P(\tau \leq t).$$

Неравенство $\tau \leq t$ выполняется, когда за время t появляется один импульс (заявка), или два импульса (заявки), или три импульса (заявки) и т. д. Определить вероятность бесконечного объединения подобных событий затруднительно. Гораздо легче и продуктивнее определить эту вероятность через вероятность противоположного события, а именно, через вероятность того, что за время t не появится ни одна заявка (отказ, импульс), то есть

$$F(t) = P(\tau \leq t) = 1 - P(\tau > t).$$

Напомним, что исходной моделью для модели простейшего потока была схема независимых испытаний Бернулли (см. разд. 1.3.2). Поскольку потоки указанных событий реализуются в непрерывном времени, устремим Δt к нулю (при этом $n \rightarrow \infty$, $np = \lambda t = \text{const}$), и с использованием теоремы Пуассона (см. разд. 1.3.6) найдем вероятность появления m импульсов в n интервалах времени, или, что то же самое, в течение времени t , как предел:

$$P_n(m) = \lim_{\substack{n \rightarrow \infty \\ np = \lambda t = \text{const}}} C_n^m p^m (1-p)^{n-m} = \frac{(\alpha)^m}{m!} e^{-np} = \frac{(\lambda t)^m}{m!} e^{-\lambda t}.$$

Вероятность того, что в течение времени t не появится ни один импульс (или ни одна заявка), то есть $m = 0$, равна $P(\tau > t) = 1 - P(\tau \leq t) = P_n(0) = e^{-\lambda t}$. Отсюда функция распределения интервала времени между двумя последовательными импульсами (заявками) в случайном потоке

$$F(t) = P(\tau \leq t) = 1 - e^{-\lambda t}.$$

Плотность распределения этой случайной величины есть производная от $F(t)$ по t :

$$\varphi(t) = \lambda e^{-\lambda t}.$$

В результате мы получили экспоненциальную плотность распределения или экспоненциальный закон распределения.

Вид кривой экспоненциальной плотности распределения при различных значениях параметра λ представлен на рис. 13.

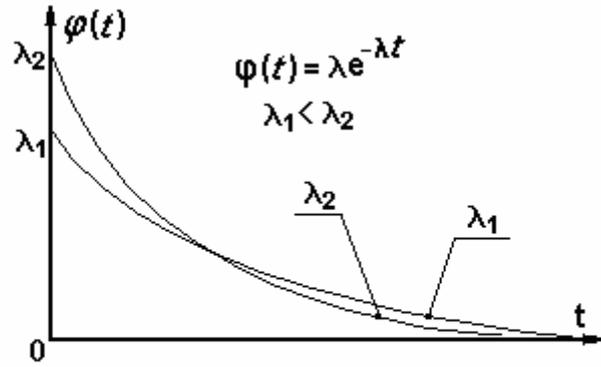


Рис. 13. Экспоненциальная плотность распределения

Фигуры, ограниченные кривой плотности распределения и осью абсцисс, равновелики, их площади равны 1.

Характеристическая функция экспоненциального распределения вычисляется, по определению (разд. 1.6.4), как интеграл:

$$g_{\tau}(v) = \int_0^{\infty} e^{jvt} \lambda e^{-\lambda t} dt = \lambda \int_0^{\infty} e^{(jv-\lambda)t} dt = \frac{\lambda}{\lambda - jv}.$$

Математическое ожидание и дисперсию случайного интервала времени между событиями (заявками, отказами, импульсами) находим, пользуясь свойствами характеристических функций:

$$j\mu_1(\tau) = \frac{dg_{\tau}(v)}{dv} = j \frac{\lambda}{(\lambda - jv)^2} \Big|_{v=0} = j \frac{1}{\lambda},$$

откуда $M[\tau] = \mu_1(\tau) = 1/\lambda$;

$$(j)^2 \mu_2(\tau) = \frac{d^2 g_{\tau}(v)}{dv^2} = (j)^2 \frac{2 \cdot \lambda}{(\lambda - jv)^3} \Big|_{v=0} = (j)^2 \frac{2}{\lambda^2},$$

откуда $\mu_2(\tau) = 2/\lambda^2$ и $D[\tau] = \mu_2(\tau) - \mu_1^2(\tau) = 1/\lambda^2$. Значит, среднеквадратическое значение $\sigma_{\tau} = 1/\lambda = M[\tau]$.

При статистической обработке экспериментальных данных, когда возникает необходимость идентификации вида плотности распределения исследуемой случайной величины, последнее соотношение может служить одним из признаков экспоненциального распределения.

Сформированная таким образом математическая модель случайной последовательности событий называется *простейшим потоком* или *потоком Пуассона* и применяется в *теории массового обслуживания* и в *теории надежности* для вероятностного описания, например:

потока вызовов на телефонную станцию;

потока обращений к серверу коллективного пользования в компьютерной сети;

потока автомобилей на пограничный пункт таможенного досмотра;

потока клиентов на обслуживающее предприятие;

потока отказов технического устройства и т.д.

Наряду с простейшим потоком в теории массового обслуживания используются и более сложные модели потоков заявок, например, *потоки Эрланга и Пальма*, однако наиболее полные аналитические описания всех процедур и характеристик обслуживания заявок, потоков необслуженных заявок, длины очередей на обслуживание и тому подобных характеристик существуют пока только для простейших потоков.

В теории надежности $F(t)$ называется *функцией распределения времени безотказной работы* устройства, а вероятность $P(t) = 1 - F(t) = e^{-\lambda t}$ – *функцией надежности*.

Математические модели случайных потоков событий применяются для анализа и синтеза систем, в которых подобные потоки возникают. Целью анализа и расчета таких систем может быть, например:

оценка вероятности того, что обслуживание заявки не состоится, и заявка пропадает;

оценка средней длины очереди заявок на обслуживание;

анализ и синтез стратегии обслуживания заявок;

расчет времени безотказной работы сложных систем и т.д.

1.6.6.2. Равномерная плотность распределения

Вид плотности распределения представлен на рис. 14. Аналитическая запись плотности распределения:

$$\varphi(x) = \begin{cases} \frac{1}{2a}, & |x - c| \leq a, \\ 0, & |x - c| > a. \end{cases}$$

Видно, что условие нормировки соблюдены, поскольку площадь под плотностью распределения равна 1.

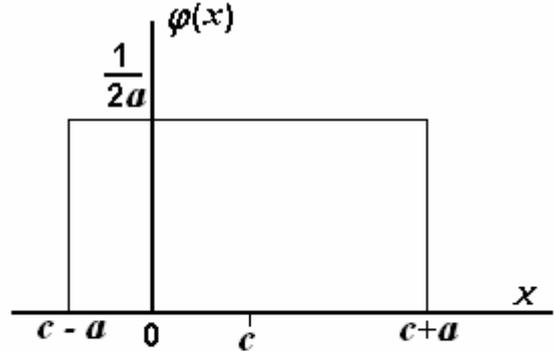


Рис. 14. Равномерная плотность распределения

Плотность распределения симметрична относительно математического ожидания,

$$M[\xi] = c, \quad D[\xi] = \overset{\circ}{\mu}_2 = \frac{1}{2a} \int_{-a}^a x^2 dx = \frac{a^2}{3}, \quad \overset{\circ}{\mu}_3 = 0, \quad \overset{\circ}{\mu}_4 = \frac{1}{2a} \int_{-a}^a x^4 dx = \frac{a^4}{5},$$

$$As = 0, \quad ex = \frac{\overset{\circ}{\mu}_4}{\overset{\circ}{\mu}_2^2} = \frac{9a^4}{5a^4} = 1,8.$$

$$\text{Дифференциальная энтропия } H_\varepsilon(\xi) = -\frac{1}{2a} \int_{-a}^a \log_2 \left(\frac{1}{2a} \right) dx = \log_2 2a.$$

Случайная величина с равномерной плотностью распределения экстремальна по энтропии в следующем смысле.

Пусть о случайной величине известно лишь то, что значения, которые она может принимать, ограничены по модулю некоторым числом A , то есть $P(|\xi| > A) = 0$. Тогда среди всех таких случайных величин максимальной относительной энтропией обладает случайная величина с равномерной плотностью распределения.

1.6.6.3. Плотность распределения Arcsin

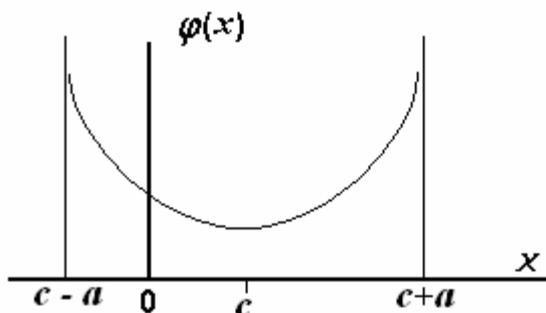


Рис. 15. Плотность распределения Arcsin

График плотности распределения приведен на рис. 15. Плотность распределения симметрична относительно математического ожидания. Кривая плотности распределения описывается функцией

$$\varphi(x) = \frac{1}{\pi\sqrt{a^2 - (x - c)^2}}.$$

Математическое ожидание и медиана совпадают:

$$M[\xi] = \mu_1[\xi] = x_{0,5} = c.$$

Дисперсия, четвертый центральный момент, асимметрия и эксцесс:

$$D[\xi] = \overset{\circ}{\mu}_2[\xi] = \sigma_\xi^2 = \frac{a^2}{2}, \quad \overset{\circ}{\mu}_4[\xi] = \frac{3a^4}{8}, \quad A_s = 0,$$

$$ex = \frac{\overset{\circ}{\mu}_4[\xi]}{\overset{\circ}{\mu}_2[\xi]^2} = \frac{3a^4 \cdot 4}{8a^4} = 1,5.$$

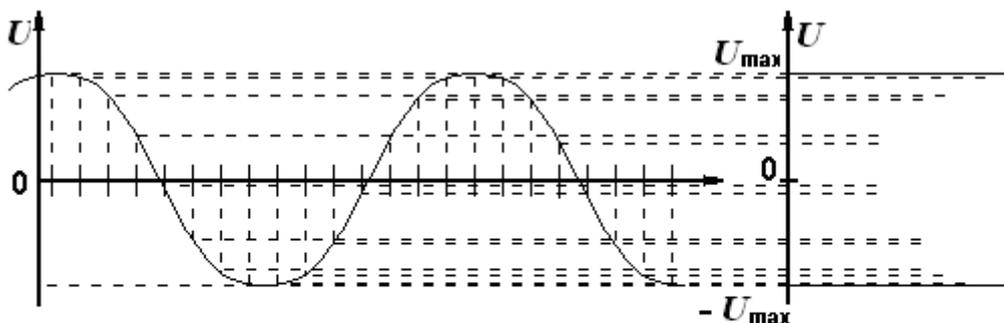


Рис. 16. К происхождению случайных погрешностей, распределенных по закону Arcsin

Случайная величина с такой плотностью распределения возникает при аналого-цифровом преобразовании, например, синусоидального напряжения, действующего на входе аналого-цифрового преобразователя (АЦП). При запуске АЦП в случайные моменты времени или в моменты времени,

не синхронизированные с частотой напряжения помехи (см. рис. 16), из-за плоских максимумов и минимумов синусоиды плотность значений выходного сигнала АЦП возрастает к краям интервала $[-U_{\max}, U_{\max}]$ и минимальна в середине этого интервала. Интегральная функция распределения описывается функцией, обратной функции $y = \sin x$, то есть функцией

$$F(x) = \begin{cases} 0, & x < c - a, \\ \frac{1}{2} + \frac{1}{\pi} \arcsin\left(\frac{x - c}{a}\right), & c - a \leq x \leq c + a, \\ 1, & x > c + a. \end{cases}$$

1.6.6.4. Нормальная плотность распределения (Гаусса)

Кривая нормальной плотности распределения описывается функцией

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - c)^2}{2\sigma^2}\right\},$$

которая называется *гауссианой* по имени германского математика К. Ф. Гаусса.

Позднее были открыты многие замечательные свойства случайных величин, распределенных по нормальному закону. Мы познакомимся с большинством этих свойств.

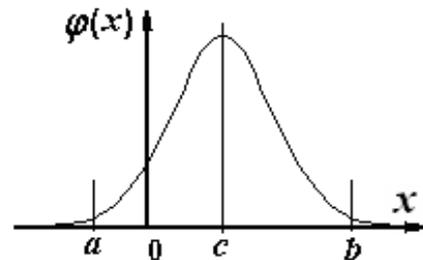


Рис. 17. Нормальная плотность распределения

Форма кривой нормальной плотности представлена на рис. 17. Это симметричная одномодальная плотность распределения, поэтому математическое ожидание, мода и медиана совпадают:

$$M[\xi] = \mu_1[\xi] = x_{0,5} = x_{mod} = c.$$

Дисперсия, четвертый центральный момент, асимметрия и эксцесс:

$$D[\xi] = \mu_2^0[\xi] = \sigma_\xi^2 = \sigma^2, \quad \mu_4^0[\xi] = 3\sigma^4, \quad As = 0, \quad ex = 3.$$

В некоторых источниках эксцесс плотности распределения определяют по отношению к эксцессу нормальной плотности распределения. Тогда

эксцесс любого распределения уменьшается на 3, а эксцесс нормального распределения равен 0:

$$\text{ex} = \frac{\mu_4[\xi] - 3\mu_2[\xi]^2}{\mu_2[\xi]^3}.$$

Дифференциальная энтропия нормально распределенной случайной величины

$$H_\varepsilon[\xi] = \log_2 \sqrt{2\pi e \sigma^2}.$$

Случайная величина с нормальной плотностью распределения экстремальна по энтропии в смысле следующей задачи. Пусть о случайной величине известно, что $D[\xi] \leq b$, где b – любое положительное конечное число. Найти плотность распределения, при которой дифференциальная энтропия этой случайной величины достигает максимального значения на множестве всех плотностей распределения.

Задача решается методами вариационного исчисления. Единственное решение задачи – нормальная плотность распределения. Иными словами, среди случайных величин, о которых известно лишь то, что дисперсия каждой из них ограничена одним числом, максимальной энтропией обладает случайная величина, распределенная нормально.

Принадлежность случайной величины к нормальной генеральной совокупности с математическим ожиданием c и дисперсией σ^2 будем обозначать следующим образом

$$\xi \in N(c, \sigma^2).$$

Характеристическую функцию нормальной случайной величины приведем здесь без вывода:

$$g_\xi(v) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{jvx} e^{-\frac{(x-c)^2}{2\sigma^2}} dx = e^{jcv} e^{-\frac{\sigma^2 v^2}{2}} = \exp\left\{jcv - \frac{\sigma^2 v^2}{2}\right\}.$$

Обратим внимание на то, что при отсутствии сдвига плотности распределения относительно начала координат, то есть при $c = 0$, множитель e^{jcv} пропадает.

Интегральная функция нормального распределения выражается интегралом от плотности распределения:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(y-c)^2}{2\sigma^2}} dy.$$

Этот интеграл не может быть записан в конечной форме, поэтому его значения табулируются. В связи с тем, что этот интеграл зависит от параметров c и σ^2 , которые могут принимать бесчисленное множество значений, в целях удобства табулирования эти параметры исключаются путем замены переменной интегрирования:

$$\left(\frac{y-c}{\sigma}\right) = t; \quad dy = \sigma dt.$$

Тогда, пользуясь тем, что нормальная плотность распределения симметрична и $F(c)=0,5$, получим для $x \geq c$:

$$\begin{aligned} F(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(y-c)^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-c}{\sigma}} e^{-\frac{t^2}{2}} dt = \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{\frac{x-c}{\sigma}} e^{-\frac{t^2}{2}} dt = \frac{1}{2} + \Phi\left(\frac{x-c}{\sigma}\right), \end{aligned}$$

где функция

$$\Phi\left(\frac{x-c}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{x-c}{\sigma}} e^{-\frac{t^2}{2}} dt$$

называется *функцией Лапласа*.

Таблицы этой функции приводятся во всех без исключения справочниках, учебниках и учебных пособиях по теории вероятностей и математической статистике. Если $x < c$, $F(x) = \frac{1}{2} - \Phi\left(\frac{x-c}{\sigma}\right)$. Оба эти выражения распространяются на всю ось путем объединения с использованием знаковой функции $\text{sign}[\bullet]$:

$$F(x) = \frac{1}{2} + \text{sign}(x-c)\Phi\left(\frac{|x-c|}{\sigma}\right).$$

Вероятностная мера полуоткрытого интервала $(a, b]$ вычисляется, как

$$P(a < \xi \leq b) = F(b) - F(a) = \text{sign}(b - c) \Phi\left(\frac{|b - c|}{\sigma}\right) - \text{sign}(a - c) \Phi\left(\frac{|a - c|}{\sigma}\right).$$

Если интервал симметричен относительно математического ожидания, то есть границы интервала $(c - a, c + a]$, то

$$P(c - a < \xi \leq c + a) = \text{sign}(a) \Phi\left(\frac{a}{\sigma}\right) - \text{sign}(-a) \Phi\left(\frac{a}{\sigma}\right) = 2\Phi\left(\frac{a}{\sigma}\right).$$

На практике часто используются интервалы, ширина которых исчисляется целыми значениями среднеквадратического отклонения σ . Наиболее популярными среди них являются: $a = \sigma$, $a = 2\sigma$, $a = 3\sigma$. Для этих значений $\Phi(1) \cong 0.34$, $\Phi(2) \cong 0.475$, $\Phi(3) \cong 0.4985$. Соответственно вероятностные меры интервалов

$$P(c - \sigma < \xi \leq c + \sigma) \cong 0,68, \quad P(c - 2\sigma < \xi \leq c + 2\sigma) \cong 0,95,$$

$$P(c - 3\sigma < \xi \leq c + 3\sigma) \cong 0,997.$$

Приведем два из многих исключительных свойств нормального распределения вероятностей: *интегральную теорему Муавра – Лапласа и центральную предельную теорему (ЦПТ)*.

Интегральная теорема Муавра-Лапласа основывается на локальной теореме Муавра-Лапласа, которая сообщена выше в разд. 1.3.7.

Напомним, что если в схеме Бернулли (см. разд. 1.2.4) количество испытаний возрастает, то есть $n \rightarrow \infty$, а вероятность p появления одного из двух противоположных событий не изменяется, то в соответствии с локальной теоремой Муавра – Лапласа

$$P_n(m) = C_n^m p^m (1 - p)^{n-m} \xrightarrow[\substack{n \rightarrow \infty \\ p = \text{const}}]{\quad} \frac{1}{\sqrt{2\pi} \sqrt{np(1-p)}} e^{-\frac{(m-np)^2}{2np(1-p)}}.$$

Интегральная теорема Муавра – Лапласа посвящена задаче упрощенной оценки вероятности $P(N < m \leq M) = P_n(N) + P_n(N + 1) + \dots + P_n(M)$ без необходимости трудоемких вычислений числа сочетаний.

На основании локальной теоремы

$$P(N \leq m \leq M) = P_n(N) + P_n(N+1) + \dots + P_n(M) = \sum_{m=N}^M \frac{1}{\sqrt{2\pi\sqrt{np(1-p)}}} e^{-\frac{(m-np)^2}{2np(1-p)}} \Delta m.$$

Последнее равенство является точным в условиях действия локальной теоремы Муавра – Лапласа в связи с тем, что из-за дискретности случайной величины $\Delta m = 1$. В то же время последняя сумма есть не что иное, как формула прямоугольников приближенного вычисления интеграла:

$$\begin{aligned} \int_N^M \frac{1}{\sqrt{2\pi\sqrt{np(1-p)}}} e^{-\frac{(t-np)^2}{2np(1-p)}} dt &\cong \int_{\frac{N-np}{\sqrt{np(1-p)}}}^{\frac{M-np}{\sqrt{np(1-p)}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \\ &= \text{sign}(M - np) \Phi\left(\frac{|M - np|}{\sqrt{np(1-p)}}\right) - \text{sign}(N - np) \Phi\left(\frac{|N - np|}{\sqrt{np(1-p)}}\right). \end{aligned}$$

В конечном итоге при $n \rightarrow \infty$ и $p = \text{const}$ интегральная теорема Муавра – Лапласа утверждает следующую асимптотику:

$$P(N < m \leq M) \cong \text{sign}(M - np) \Phi\left(\frac{|M - np|}{\sqrt{np(1-p)}}\right) - \text{sign}(N - np) \Phi\left(\frac{|N - np|}{\sqrt{np(1-p)}}\right).$$

Центральная предельная теорема (ЦПТ). Приведем упрощенную формулировку теоремы.

Плотность распределения суммы независимых произвольно распределенных случайных величин, дисперсии которых различаются не слишком сильно, при увеличении числа слагаемых стремится к нормальной плотности распределения.

Более точная формулировка приводится, например, в книге [6] на стр. 268.

1.6.6.5. Плотность распределения Лапласа

График плотности распределения приведен на рис. 18. Плотность распределения симметрична относительно математического ожидания. Кривая плотности распределения описывается функцией

$$\varphi(x) = \frac{\lambda}{2} e^{-\lambda|x-c|}.$$

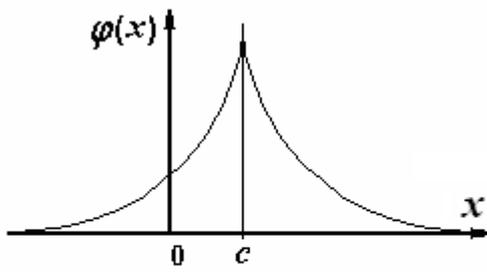


Рис. 18. Плотность распределения Лапласа

Математическое ожидание и медиана совпадают:

$$M[\xi] = \mu_1[\xi] = x_{0.5} = x_{\text{mod}} = c.$$

Для определения дисперсии воспользуемся характеристической функцией. Поскольку плотность распределения Лапласа симметрична и центральные моменты не зависят от смещения плотности распре-

деления, примем $c = 0$. Тогда характеристическая функция получается из выражения

$$\begin{aligned} g_{\xi}(v) &= \frac{\lambda}{2} \int_{-\infty}^{\infty} e^{jvx} e^{-\lambda|x|} dx = \frac{\lambda}{2} \left[\int_{-\infty}^0 e^{jvx} e^{\lambda x} dx + \int_0^{\infty} e^{jvx} e^{-\lambda x} dx \right] = \\ &= \frac{\lambda}{2} \left(\frac{1}{\lambda + jv} + \frac{1}{\lambda - jv} \right) = \frac{\lambda^2}{\lambda^2 + v^2}. \end{aligned}$$

Если $c \neq 0$, то в соответствии со свойством б) характеристических функций (см. разд. 1.6.4) появится экспоненциальный множитель

$$g_{\xi}(v) = \frac{\lambda^2}{\lambda^2 + v^2} e^{jvc}.$$

По свойству характеристических функций при $c = 0$

$$(j)^2 \overset{\circ}{\mu}_2(\xi) = \frac{\partial^2}{\partial v^2} g_{\xi}(v) \Big|_{v=0} = - \frac{2\lambda^2(\lambda^2 + v^2)^2 - 8\lambda^2 v^2(\lambda^2 + v^2)}{(\lambda^2 + v^2)^4} \Big|_{v=0} = - \frac{2}{\lambda^2},$$

откуда следует, что дисперсия $D[\xi] = \frac{2}{\lambda^2}$.

Четвертый центральный момент, асимметрия и эксцесс:

$$\overset{\circ}{\mu}_4[\xi] = \frac{4!}{\lambda^4}, \quad As = 0, \quad ex = \frac{\overset{\circ}{\mu}_4[\xi]}{\overset{\circ 2}{\mu}_2[\xi]} = \frac{4! \lambda^4}{4\lambda^4} = 3! = 6.$$

1.6.6.6. Плотность распределения Коши

График плотности распределения приведен на рис. 19. Плотность распределения симметрична относительно математического ожидания. Кривая плотности распределения описывается функцией

$$\varphi(x) = \frac{\Delta}{\pi(\Delta^2 + (x - c)^2)}.$$

Плотность распределения симметрична, одномодальна.

Из общего выражения для моментов

$$\mu_k[\xi] = \int_{-\infty}^{\infty} \frac{x^k \Delta}{\pi(\Delta^2 + (x - c)^2)} dx$$

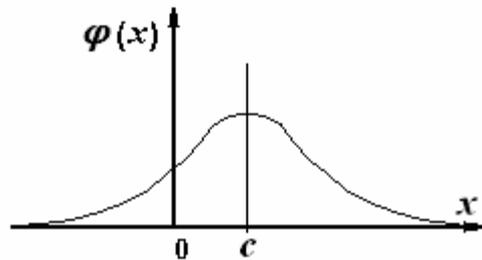


Рис. 19. Плотность распределения Коши

видно, что при всех целых значениях $k > 0$ этот несобственный интеграл не существует, поскольку порядок убывания подынтегральной функции на бесконечности при $k > 0$ не превышает 1. При $k = 0$ этот интеграл равен 1, что и должно следовать из условия нормировки плотности распределения.

В связи с этим математическое ожидание, дисперсия и моменты более высоких порядков случайной величины, распределенной по Коши, отсутствуют. Параметры плотности распределения Коши имеют названия: c – *параметр сдвига* (совпадает с модой и медианой), Δ – *параметр масштаба*.

Эксцесс случайной величины, распределенной по Коши, равен ∞ .

Характеристическая функция

$$g_{\xi}(v) = e^{jvc} e^{-\Delta|v|}.$$

Справедливость этого выражения обнаруживается при сопоставлении пары преобразований Фурье, связывающих плотность распределения Лапласа с ее характеристической функцией при $c = 0$. Здесь вид характеристической функции с точностью до множителя совпадает с видом плотности распределения Лапласа, а вид характеристической функции плотности

Лапласа с точностью до того же множителя совпадает с видом плотности распределения Коши.

Заметим, что характеристическая функция случайной величины, распределенной по Коши, недифференцируема в нуле, и это еще один признак отсутствия моментов данной случайной величины.

1.6.6.7. Экспоненциальное семейство распределений

Представленное здесь семейство плотностей распределений описывается функцией:

$$\varphi_p(x) = \frac{p}{2d\Gamma\left(\frac{1}{p}\right)} \exp\left\{-\left|\frac{x-c}{d}\right|^p\right\}, \text{ где } d = \sigma \sqrt{\frac{\Gamma\left(\frac{1}{p}\right)}{\Gamma\left(\frac{3}{p}\right)}},$$

p – параметр семейства, $\Gamma[\bullet]$ – гамма-функция (будет представлена в разделе 1.6.7). Все плотности распределения семейства симметричны, математическое ожидание совпадает с медианой и модой и равно c , эксцесс плотностей семейства равен

$$ex = \frac{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{5}{p}\right)}{\Gamma^2\left(\frac{3}{p}\right)}.$$

При $p = 1$ членом этого семейства является плотность распределения Лапласа, при $p = 2$ – нормальная плотность распределения.

В самом деле, при $p = 1$ $\Gamma(1) = 1$, $\Gamma(3) = 2$, $\Gamma(5) = 4! = 24$, $d = \frac{\sigma}{\sqrt{2}}$ и плотность распределения приобретает вид

$$\varphi_1(x) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{\sqrt{2}}{\sigma}|x-c|},$$

а это и есть плотность распределения Лапласа, параметр которой в этом случае $\lambda = \sqrt{2}/\sigma$, $ex = 6$.

При $p = 2$ $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(3/2) = \sqrt{\pi}/2$, $\Gamma(5/2) = 3\sqrt{\pi}/4$, $d = \sigma\sqrt{2}$ и плотность распределения принимает вид:

$$\varphi_2(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-c)^2}{2\sigma^2}\right\}, \quad \text{ex} = 3,$$

что и требовалось показать.

При дальнейшем увеличении параметра p плотность распределения приобретает форму, приближающуюся к форме равномерной плотности. При этом все плотности семейства симметричны с математическим ожиданием, равным c .

1.6.7. Функции от непрерывных случайных величин

Вначале введем и кратко опишем гамма-функцию, которая встретилась в разд. 1.6.6.7 и будет встречаться в дальнейшем.

Введем математическое определение гамма-функции и применим интегрирование по частям:

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt = uv \Big|_0^{\infty} - \int_0^{\infty} v du,$$

где $u = t^{x-1}$, $dv = e^{-t} dt$, $du = (x-1)t^{x-2} dt$, $v = -e^{-t}$.

Выполним подстановку:

$$\Gamma(x) = -t^{x-1} e^{-t} \Big|_0^{\infty} + \int_0^{\infty} (x-1)t^{x-2} e^{-t} dt = (x-1)\Gamma(x-1).$$

В результате мы получили рекуррентную формулу для вычисления значений гамма-функции.

Если x – целое положительное число, $x = n$, то в соответствии с этой рекуррентной формулой

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \dots = (n-1)(n-2)\dots 1 \cdot \Gamma(1).$$

Вычислим $\Gamma(1)$ отдельно:

$$\Gamma(1) = \int_0^{\infty} e^{-t} dt = -e^{-t} \Big|_0^{\infty} = 1.$$

Поэтому $\Gamma(n) = (n-1)!$ В частности, $\Gamma(1) = 0 \cdot \Gamma(0) = 0! = 1$.

В дальнейшем нам понадобятся следующие значения гамма-функции от дробных аргументов:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\sqrt{\pi}, \quad \Gamma\left(\frac{5}{2}\right) = \frac{3}{4}\sqrt{\pi}.$$

Последние два значения гамма-функции получены с использованием только что выведенной рекуррентной формулы.

Займемся теперь основной задачей. Пусть задана непрерывная дифференцируемая функция от случайной величины ξ : $\eta = f(\xi)$. Известна плотность распределения случайной величины ξ : $\varphi(x)$. Задача состоит в том, чтобы найти плотность распределения $\psi(y)$ случайной величины η .

Подобная задача возникает в технике, когда случайные процессы или измеряемые величины, возмущенные случайными помехами, претерпевают нелинейные преобразования, и возникает задача прогнозирования характеристик сигнала, который получается в результате этого преобразования.

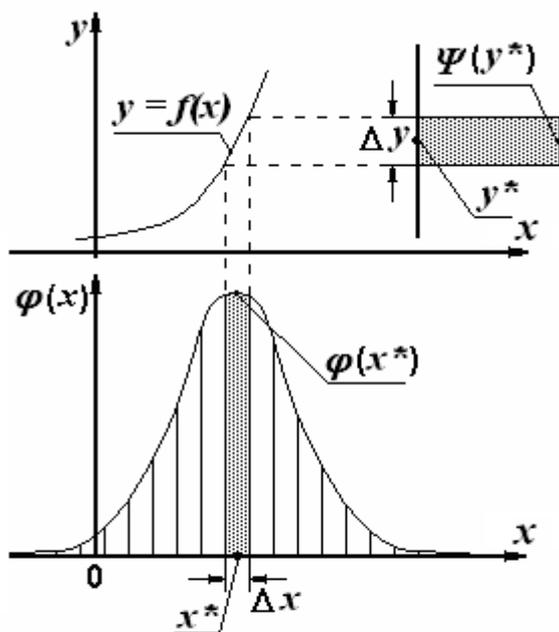


Рис. 20. Нелинейное преобразование случайной величины

Для вывода необходимого соотношения воспользуемся рис. 20, на котором представлены функция преобразования $y = f(x)$ и плотность распределения $\varphi(x)$. Функция преобразования предполагается монотонной, и это свойство функции преобразования практически всегда имеет место в технических устройствах: средствах измерения, измерительных преобразователях и регуляторах. В силу взаимной однозначности преобразования случайная величина η принимает значения из интервала Δy , в точности с той же вероятностью, с которой случайная величина ξ принимает значения из интервала Δx . Поскольку вероят-

ностная мера интервала есть площадь под кривой плотности распределения на этом интервале, это означает, что площади заштрихованных фигур на рис. 20 должны быть равны

$$\psi(y^*)\Delta y = \varphi(x^*)\Delta x,$$

где x^* , y^* – точки, находящиеся внутри выделенных интервалов Δx и Δy , $\psi(y^*)$ – значение искомой плотности распределения в точке y^* .

Из этого выражения следует:

$$\psi(y^*) = \varphi(x^*) \frac{\Delta x}{\Delta y}.$$

Заметим здесь, что ширина интервала Δy , в который преобразуется интервал Δx , не зависит от знака производной функции преобразования, и это обстоятельство мы учтем при выполнении предельного перехода

$$\psi(y) = \lim_{\Delta x \rightarrow 0} \psi(y^*) = \lim_{\Delta x \rightarrow 0} \left\{ \varphi(x^*) \left| \frac{\Delta x}{\Delta y} \right| \right\} = \varphi(x) \frac{1}{|f'_x(x)|}.$$

В силу инвариантности первого дифференциала производная f'_x выражается через производную от обратной функции:

$$f'_x = \frac{1}{(f^{-1}(y))'_y}.$$

Кроме того в выражении для $\psi(y)$ необходимо выразить аргумент x плотности распределения $\varphi(x)$ через y с помощью обратной функции: $x = f^{-1}(y)$. В итоге окончательно получим:

$$\psi(y) = \varphi[f^{-1}(y)] \left| [f^{-1}(y)]'_y \right|.$$

Примеры.

1. Случайная величина ξ распределена нормально: $\xi \in N(0,1)$. Функция преобразования $\eta = \xi^2$. Эта ситуация представлена на рис. 21. Видно, что в силу двузначности обратной функции случайная величина η принимает значения из интервала Δy , когда случайная величина ξ принимает значения в одном из двух выделенных интервалов Δx . Поэтому для данно-

го примера исходное выражение должно быть изменено следующим образом:

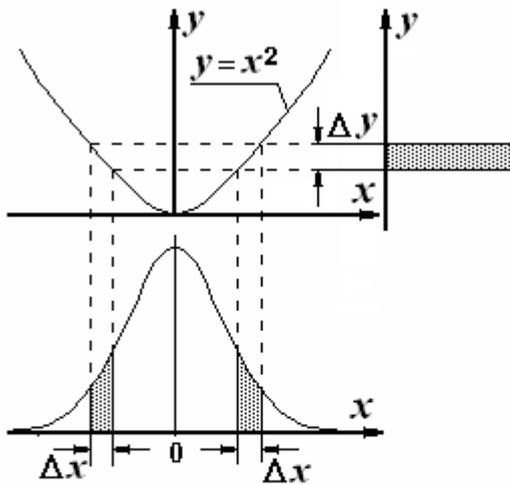


Рис. 21. Возведение нормальной случайной величины в квадрат

$$\psi(y^*) \Delta y = 2\varphi(x^*) \Delta x.$$

Из этого следует соответствующее изменение общей формулы:

$$\psi(y) = 2\varphi[f^{-1}(y)] \left| [f^{-1}(y)]_y' \right|$$

Для данного примера

$$y = x^2, \quad x = f^{-1}(y) = \pm\sqrt{y},$$

$$\left| [f^{-1}(y)]_y' \right| = (\sqrt{y})_y' = \frac{1}{2\sqrt{y}},$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right)$$

В конечном итоге после подстановки в общую формулу получим искомую плотность распределения

$$\psi(y) = \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right).$$

Найдем характеристическую функцию этого распределения:

$$g_{\eta}(v) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{\sqrt{y}} \exp\left(jvy - \frac{y}{2}\right) dy.$$

Сделаем замену переменной интегрирования:

$$\frac{y}{2} - jvy = t, \quad dt = \left(\frac{1}{2} - jv\right) dy, \quad dy = dt \left(\frac{1}{2} - jv\right)^{-1}, \quad \frac{1}{\sqrt{y}} = \frac{1}{\sqrt{t}} \sqrt{\frac{1}{2} - jv}.$$

В результате этой замены получим выражение с участием гамма-функции:

$$g_{\eta}(v) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{1}{2} - jv}} \int_0^{\infty} t^{-1/2} e^{-t} dt = \frac{1}{\sqrt{\pi} \sqrt{1 - 2jv}} \Gamma\left(\frac{1}{2}\right).$$

Поскольку $\Gamma(1/2) = \sqrt{\pi}$, окончательно получим:

$$g_{\eta}(v) = (1 - 2jv)^{-1/2}.$$

2. Пусть $F_{\xi}(x)$ – интегральная функция распределения случайной величины ξ . образуем случайную величину η , как функцию от случайной величины ξ : $\eta = F_{\xi}(\xi)$. Задача состоит в том, чтобы найти плотность распределения $\psi(y)$ случайной величины η .

Воспользуемся полученным ранее выражением

$$\psi(y) = \varphi(x) \frac{1}{|f'_x(x)|}.$$

В нашем случае в качестве функции $y = f(x)$ выступает функция $y = F_{\xi}(x)$, производная от которой по x есть плотность распределения $\varphi(x)$. Поэтому

$$\psi(y) = \varphi(x) \frac{1}{|(F'_{\xi}(x))|} = \frac{\varphi(x)}{\varphi(x)} = 1.$$

Таким образом, оказывается, что случайная величина η , полученная в результате функционального преобразования любой непрерывной случайной величины ξ путем ее подстановки в ее же интегральную функцию распределения, *распределена равномерно в интервале $[0,1]$ вне зависимости от вида функции распределения величины ξ .*

Полученный результат имеет два полезных применения.

Первое. Машинное моделирование случайных чисел с заданной интегральной функцией распределения. Технология моделирования следующая:

задается функция распределения $F(x)$;

по стандартным программам генерируются случайные числа $\{z_i\}$, распределенные равномерно в интервале $[0, 1]$;

случайные числа x_i , распределенные в соответствии с заданной функцией распределения $F(x)$, получают как решения уравнений

$$F(x_i) = z_i.$$

Второе. Статистическое оценивание параметров и характеристик случайных величин по результатам экспериментов, вне зависимости от вида

распределения исследуемой случайной величины. Это применение будет изложено в разделе 2. Математическая статистика.

3. Случайная величина η образуется в результате линейного преобразования случайной величины ξ : $\eta = a\xi + b$.

В данном случае реализуется функциональное преобразование $y = ax + b$, обратная функция $x = \frac{1}{a}(y - b)$, производная от нее по y равна

$[f^{-1}(y)]'_y = \frac{1}{a}$. В результате подстановки в общую формулу получим:

$$\psi(y) = \varphi\left(\frac{1}{a}(y - b)\right) \left| \frac{1}{a} \right| = \left| \frac{1}{a} \right| \varphi\left(\frac{y - b}{a}\right).$$

Это означает, что *любое линейное преобразование не изменяет вид плотности распределения* случайной величины. Изменяется лишь масштаб и смещение от начала координат.

1.6.8. Неравенство Чебышева

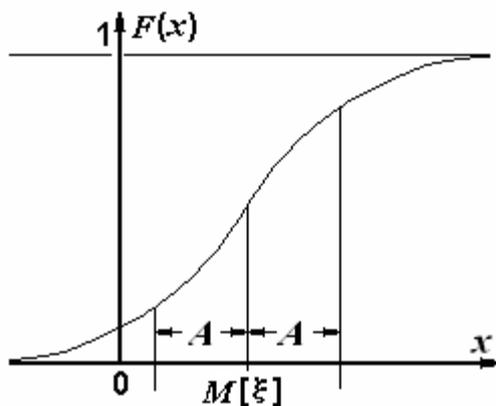


Рис. 22. К выводу неравенства Чебышева

Неравенство Чебышева позволяет сделать грубую, но быструю оценку вероятности отклонения значений случайной величины от ее математического ожидания, превышающих некоторое заданное положительное значение A :

$$P\{\xi - M[\xi] \geq A\}.$$

Несмотря на грубость получаемой оценки неравенство Чебышева имеет достаточно широкое и полезное применение благодаря тому, что оно справедливо для любых законов распределения, кроме закона распределения Коши. С одним из таких применений мы познакомимся в разделе 2 «Математическая статистика».

Запишем вероятность, подлежащую оценке, как интеграл от плотности распределения случайной величины ξ по отрезкам оси

$$(-\infty, M[\xi]-A], (M[\xi]+A, \infty),$$

то есть по множеству $|M[\xi] - \xi| \geq A$, как это видно из рис. 22:

$$P\{\xi - M[\xi] \geq A\} = \int_{|\xi - M[\xi]| \geq A} \varphi(x) \cdot dx \leq \int_{\frac{|\xi - M[\xi]|}{A} \geq 1} \frac{|x - M[\xi]|^2}{A^2} \varphi(x) dx \leq \frac{1}{A^2} \int_{-\infty}^{\infty} |x - M[\xi]|^2 \varphi(x) dx = \frac{\sigma^2}{A^2}.$$

Окончательно $P\{\xi - M[\xi] \geq A\} \leq \frac{\sigma^2}{A^2}$.

Существуют другие записи неравенства Чебышева:

если $A = k\sigma$, то $P\{\xi - M[\xi] \geq k\sigma\} \leq \frac{1}{k^2}$

или $P\{\xi - M[\xi] \leq k\sigma\} \geq 1 - \frac{1}{k^2}$.

Для того, чтобы оценить степень грубости оценки, вытекающей из неравенства Чебышева, сопоставим ее с точными значениями вероятностей, установленными в разд. 1.6.6.4 для нормально распределенной случайной величины.

Точные значения:

$$P\{\xi - M[\xi] \geq 2\sigma\} = 1 - 0,95 = 0,05,$$

$$P\{\xi - M[\xi] \geq 3\sigma\} = 1 - 0,997 = 0,003$$

Значения вероятностей, полученных из неравенства Чебышева для тех же отклонений значений случайной величины от математического ожидания:

$$P\{\xi - M[\xi] \geq 2\sigma\} \leq 0,25, \quad P\{\xi - M[\xi] \geq 3\sigma\} \leq 1/9.$$

1.7. ДВУМЕРНЫЕ НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ (ДВУМЕРНЫЕ СЛУЧАЙНЫЕ ВЕКТОРЫ)

1.7.1. Функции распределения и плотности распределения

Рассмотрим двумерный случайный вектор, то есть двумерный вектор, каждая составляющая которого есть непрерывная случайная величина:

$$\zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix}.$$

Как и ранее, случайный вектор и его случайные компоненты обозначим греческими буквами, а значения, которые может принимать вектор и его компоненты – соответствующими латинскими буквами, то есть будем считать, что случайный вектор ζ принимает значения

$$\mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}.$$

Функция распределения двумерного случайного вектора есть вероятность совместного осуществления событий:

$$F_{\zeta}(x, y) = P(\xi \leq x \cap \eta \leq y).$$

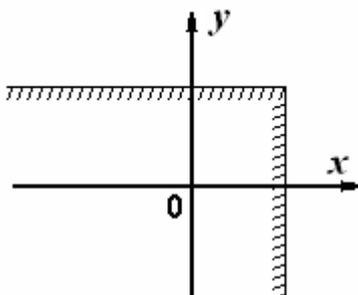


Рис. 23. Область определения функции распределения

Плотность распределения, как и ранее, есть производная от функции распределения по обоим аргументам:

$$\varphi(x, y) = \frac{d^2 F(x, y)}{dxdy},$$

поэтому

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x \varphi(u, v) dudv.$$

Область интегрирования показана на рис. 23.

В силу монотонности вероятностной меры функция распределения – неубывающая функция по каждому аргументу, поэтому плотность распределения есть неотрицательная функция двух аргументов, которая описывает некоторую поверхность над координатной плоскостью. Эта поверхность

приближается к плоскости xOy при удалении значений аргументов от начала координат в любом направлении. Понятно, что

$$F(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) dx dy = 1.$$

Если по одному из аргументов имеются ограничения, то

$$F(x, \infty) = P(\xi \leq x \cap \eta \leq \infty) = P(\xi \leq x) = F_{\xi}(x) = \int_{-\infty}^x du \int_{-\infty}^{\infty} \varphi(u, v) dv.$$

$$F(\infty, y) = P(\xi \leq \infty \cap \eta \leq y) = P(\eta \leq y) = F_{\eta}(y) = \int_{-\infty}^y dv \int_{-\infty}^{\infty} \varphi(u, v) du.$$

Таким образом мы получили *маргинальные (частные)* функции распределения $F_{\xi}(x)$ и $F_{\eta}(y)$. Дифференцирование этих функций по их аргументам, то есть дифференцирование соответствующих интегралов по их верхним пределам, по определению, дает маргинальные (частные) плотности распределения:

$$\varphi_{\xi}(x) = \frac{dF_{\xi}(x)}{dx} = \int_{-\infty}^{\infty} \varphi(x, v) dv, \quad \varphi_{\eta}(y) = \frac{dF_{\eta}(y)}{dy} = \int_{-\infty}^{\infty} \varphi(u, y) du.$$

Определим *условную функцию распределения*, то есть функцию распределения одной из случайных величин при условии, что другая случайная величина принимает некоторое конкретное значение, например,

$$F(y / \xi = x) = P(-\infty < \eta \leq y / \xi = x).$$

Выделим на координатной плоскости область, показанную на рис. 24.

Вероятность того, что случайный вектор принимает значения из этой области, равна $P(-\infty < \eta \leq y / x < \xi \leq x + \Delta x)$. В соответствии с формулой для условной вероятности из разд. 1.2.3

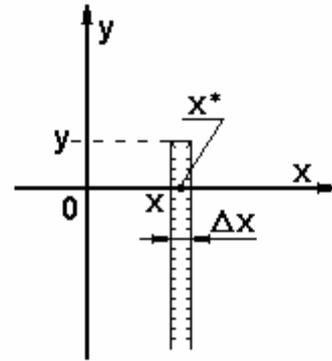


Рис. 24. К выводу условной плотности распределения

$$P(-\infty < \eta \leq y / x < \xi \leq x + \Delta x) = \frac{P(-\infty < \eta \leq y \cap x < \xi \leq x + \Delta x)}{P(x < \xi \leq x + \Delta x)}.$$

Условная функция распределения получается в результате предельного перехода:

$$\begin{aligned}
 F(y/\xi = x) &= \lim_{\Delta x \rightarrow 0} P(-\infty < \eta \leq y/x < \xi \leq x + \Delta x) = \\
 &= \lim_{\Delta x \rightarrow 0} \frac{P(-\infty < \eta \leq y \cap x < \xi \leq x + \Delta x)}{P(x < \xi \leq x + \Delta x)} = \lim_{\Delta x \rightarrow 0} \frac{\int_{-\infty}^y dv \int_x^{x+\Delta x} \varphi(u, v) du}{\int_{-\infty}^{\infty} dv \int_x^{x+\Delta x} \varphi(u, v) du}.
 \end{aligned}$$

По теореме о среднем, внутри интервала $(x, x + \Delta x]$ найдется точка x^* , такая, что

$$\int_x^{x+\Delta x} \varphi(u, v) du = \Delta x \varphi(x^*, v),$$

поэтому

$$F(y/\xi = x) = \lim_{\Delta x \rightarrow 0} \frac{\int_{-\infty}^y dv \Delta x \varphi(x^*, v)}{\int_{-\infty}^{\infty} dv \Delta x \varphi(x^*, v)} = \frac{\int_{-\infty}^y \varphi(x, v) dv}{\int_{-\infty}^{\infty} \varphi(x, v) dv} = \frac{\varphi_\xi(x)}{\varphi_\xi(x)}.$$

Условная плотность распределения есть производная от условной функции распределения:

$$\varphi(y/\xi = x) = \frac{dF(y/\xi = x)}{dy} = \frac{\varphi(x, y)}{\varphi_\xi(x)}.$$

Аналогично

$$\varphi(x/\eta = y) = \frac{\varphi(x, y)}{\varphi_\eta(y)}.$$

Обычно обозначают $\varphi(x/\eta = y) = \varphi(x/y)$ и $\varphi(y/\xi = x) = \varphi(y/x)$. В этих обозначениях из полученных формул следует

$$\varphi(x, y) = \varphi(x/y)\varphi_\eta(y), \quad \varphi(x, y) = \varphi(y/x)\varphi_\xi(x).$$

С учетом этих соотношений перепишем формулы для маргинальных распределений в виде:

$$\varphi_\xi(x) = \int_{-\infty}^{\infty} \varphi(x, y) dy = \int_{-\infty}^{\infty} \varphi(x/y)\varphi_\eta(y) dy, \quad \varphi_\eta(y) = \int_{-\infty}^{\infty} \varphi(x, y) dx = \int_{-\infty}^{\infty} \varphi(y/x)\varphi_\xi(x) dx.$$

Это формулы полной вероятности для непрерывных случайных величин.

Поскольку $\varphi(x/y)\varphi_\eta(y) = \varphi(y/x)\varphi_\xi(x)$, получаем формулу Байеса для непрерывных случайных величин:

$$\varphi(y/x) = \frac{\varphi(x/y)\varphi_\eta(y)}{\varphi_\xi(x)} = \frac{\varphi(x/y)\varphi_\eta(y)}{\int_{-\infty}^{\infty} \varphi(x/y)\varphi_\eta(y)dy}.$$

Если ξ и η независимы, то $\varphi(x/y) = \varphi_\xi(x)$, $\varphi(y/x) = \varphi_\eta(y)$ и поэтому $\varphi(x, y) = \varphi_\xi(x)\varphi_\eta(y)$.

Справедливо и обратное: если $\varphi(x, y) = \varphi_\xi(x)\varphi_\eta(y)$, то из этого с необходимостью следует независимость ξ и η .

Признак независимости случайных величин: две случайные величины независимы тогда и только тогда, когда их совместная плотность распределения может быть представлена как произведение маргинальных плотностей распределения этих величин (см. также разд. 1.2.3).

1.7.2. Числовые характеристики

Моменты случайных величин определяются, как и ранее, следующими формулами

начальные моменты k -го порядка

$$\begin{aligned} \mu_k[\xi] &= \int_{-\infty}^{\infty} x^k \varphi_\xi(x) dx = \int_{-\infty}^{\infty} x^k \int_{-\infty}^{\infty} \varphi(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k \varphi(x, y) dx dy, \\ \mu_k[\eta] &= \int_{-\infty}^{\infty} y^k \varphi_\eta(y) dy = \int_{-\infty}^{\infty} y^k \int_{-\infty}^{\infty} \varphi(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^k \varphi(x, y) dx dy; \end{aligned}$$

центральные моменты k -го порядка:

$$\overset{0}{\mu}_k[\xi] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_1[\xi])^k \varphi(x, y) dx dy, \quad \overset{0}{\mu}_k[\eta] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_1[\eta])^k \varphi(x, y) dx dy.$$

Среди этих моментов самыми употребительными являются математические ожидания

$$M[\xi] = \mu_1[\xi], \quad M[\eta] = \mu_1[\eta]$$

и дисперсии

$$D[\xi] = \sigma_{\xi}^2 = \overset{0}{\mu}_2[\xi], \quad D[\eta] = \sigma_{\eta}^2 = \overset{0}{\mu}_2[\eta].$$

Математическое ожидание случайного вектора есть вектор, компонентами которого являются математические ожидания соответствующих составляющих:

$$M[\vec{\xi}] = M \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} M[\xi] \\ M[\eta] \end{pmatrix}.$$

Из условных моментов выделим лишь первые начальные (условные математические ожидания) и вторые центральные (условные дисперсии):

$$M[\xi / \eta = y] = M[\xi / y] = \int_{-\infty}^{\infty} x \varphi(x / y) dx, \quad M[\eta / \xi = x] = M[\eta / x] = \int_{-\infty}^{\infty} y \varphi(y / x) dy,$$

$$D[\xi / \eta = y] = D[\xi / y] = \int_{-\infty}^{\infty} (x - M[\xi / y])^2 \varphi(x / y) dx,$$

$$D[\eta / \xi = x] = D[\eta / x] = \int_{-\infty}^{\infty} (y - M[\eta / x])^2 \varphi(y / x) dy.$$

Как и ранее, во всех случаях

$$D[\xi] = \sigma_{\xi}^2 = \mu_2[\xi] - M^2[\xi], \quad D[\eta] = \sigma_{\eta}^2 = \mu_2[\eta] - M^2[\eta],$$

$$D[\xi / y] = \mu_2[\xi / y] - M^2[\xi / y], \quad D[\eta / x] = \mu_2[\eta / x] - M^2[\eta / x].$$

Для двумерных случайных величин вводятся *смешанные моменты*: начальные порядка k, r

$$\mu_{k,r}[\xi, \eta] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^r \cdot \varphi(x, y) \cdot dx dy;$$

центральные порядка k, r

$$\overset{0}{\mu}_{k,r}[\xi, \eta] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M[\xi])^k (y - M[\eta])^r \cdot \varphi(x, y) \cdot dx dy.$$

Из них наиболее употребительным является центральный смешанный момент порядка (1, 1), который называется *ковариацией* и обозначается $\text{cov}(\xi, \eta)$:

$$\overset{0}{\mu}_{1,1}[\xi, \eta] = \text{cov}(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M[\xi])(y - M[\eta]) \varphi(x, y) dx dy.$$

Выясним связь между этим и начальным смешанным моментом того же порядка:

$$\begin{aligned}
\overset{0}{\mu}_{1,1}[\xi, \eta] &= \text{cov}(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M[\xi])(y - M[\eta]) \varphi(x, y) dx dy = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy - yM[\xi] - xM[\eta] + M[\xi]M[\eta]) \varphi(x, y) dx dy = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \varphi(x, y) dx dy - M[\xi] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \varphi(x, y) dx dy - M[\eta] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \varphi(x, y) dx dy + \\
&+ M[\xi]M[\eta] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) dx dy = \overset{0}{\mu}_{1,1}[\xi, \eta] - 2M[\xi]M[\eta] + M[\xi]M[\eta].
\end{aligned}$$

В итоге получаем, что $\overset{0}{\mu}_{1,1}[\xi, \eta] = \text{cov}(\xi, \eta) = \overset{0}{\mu}_{1,1}[\xi, \eta] - M[\xi]M[\eta]$.

Если случайные величины ξ и η независимы, в соответствии с признаком независимости, сформулированным ранее,

$$\begin{aligned}
\overset{0}{\mu}_{1,1}[\xi, \eta] &= \text{cov}(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M[\xi])(y - M[\eta]) \varphi_{\xi}(x) \varphi_{\eta}(y) dx dy = \\
&= \int_{-\infty}^{\infty} (x - M[\xi]) \varphi_{\xi}(x) dx \cdot \int_{-\infty}^{\infty} (y - M[\eta]) \varphi_{\eta}(y) dy,
\end{aligned}$$

то есть мы видим, что двукратный интеграл в этих условиях преобразуется в произведение однократных интегралов, каждый из которых равен 0. В самом деле,

$$\begin{aligned}
\int_{-\infty}^{\infty} (x - M[\xi]) \varphi_{\xi}(x) dx &= \int_{-\infty}^{\infty} x \varphi_{\xi}(x) dx - \int_{-\infty}^{\infty} M[\xi] \varphi_{\xi}(x) dx = \\
&= M[\xi] - M[\xi] \int_{-\infty}^{\infty} \varphi_{\xi}(x) dx = M[\xi] - M[\xi] = 0.
\end{aligned}$$

Поэтому при условии независимости случайных величин ξ и η их первый центральный смешанный момент, или ковариация, равна 0.

При взаимно однозначной зависимости между ξ и η , например, линейной $\eta = a\xi$, ковариация

$$\begin{aligned} \mu_{1,1}[\xi, \eta] = \text{cov}(\xi, \eta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M[\xi])(ax - aM[\xi]) \varphi(x, y) dx dy = \\ &= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - M[\xi])^2 \varphi(x, y) dx dy = aD[\xi] = a\sigma_{\xi}^2. \end{aligned}$$

Это означает, что ковариация есть характеристика степени зависимости между случайными величинами. Для того, чтобы избавиться от масштаба значений, принимаемых случайными величинами, в качестве показателя линейной зависимости используется частное от деления ковариации на произведение среднеквадратических значений случайных величин:

$$\rho(\xi, \eta) = \rho_{\xi\eta} = \frac{\text{cov}(\xi, \eta)}{\sigma_{\xi} \sigma_{\eta}}.$$

Эта величина называется *коэффициентом корреляции*. Случайные величины, у которых коэффициент корреляции равен нулю, называются *некоррелированными*. Независимые случайные величины с необходимостью некоррелированы. Обратное, вообще говоря, неверно! **Из некоррелированности случайных величин их независимость, вообще говоря, не следует.** Это понятно хотя бы потому, что из равенства нулю центрального смешанного момента порядка (1, 1) вовсе *не следует*, что все центральные смешанные моменты более высоких порядков также равны нулю.

Определим максимально возможное значение коэффициента корреляции. Естественно предположить, что своего максимального значения коэффициент корреляции достигает при взаимно однозначной связи между ξ и η , например, линейной $\eta = a\xi$. Для этого случая нам известна ковариация между ξ и η , а из разд. 1.6.5 следует, что $\sigma_{\eta}^2 = a^2 \sigma_{\xi}^2$, то есть $\sigma_{\eta} = |a| \sigma_{\xi}$.

В результате получаем

$$\rho(\xi, \eta) = \rho_{\xi\eta} = \frac{\text{cov}(\xi, \eta)}{\sigma_{\xi} \sigma_{\eta}} = \frac{a\sigma_{\xi}^2}{|a|\sigma_{\xi}^2} = 1 \cdot \text{sign}(a),$$

а это означает, что $|\rho(\xi, \eta)| \leq 1$.

1.7.3. Ковариационная матрица случайного вектора

Компонентами ковариационной матрицы являются центральные моменты компонент случайного вектора: вторые центральные моменты и ковариации. Математическое определение ковариационной матрицы:

$$\begin{aligned}\Sigma_{\zeta} &= M \left[(\zeta - M[\zeta])(\zeta - M[\zeta])^T \right] = M \left[\begin{pmatrix} \xi - M[\xi] \\ \eta - M[\eta] \end{pmatrix} (\xi - M[\xi] \quad \eta - M[\eta]) \right] = \\ &= M \begin{pmatrix} (\xi - M[\xi])^2 & (\xi - M[\xi])(\eta - M[\eta]) \\ (\xi - M[\xi])(\eta - M[\eta]) & (\eta - M[\eta])^2 \end{pmatrix}.\end{aligned}$$

Математическое ожидание случайной матрицы есть матрица, каждый элемент которой есть математическое ожидание соответствующего элемента исходной случайной матрицы. В итоге ковариационная матрица двумерного случайного вектора приобретает вид

$$\Sigma_{\zeta} = \begin{pmatrix} \sigma_{\xi}^2 & \rho(\xi, \eta)\sigma_{\xi}\sigma_{\eta} \\ \rho(\xi, \eta)\sigma_{\xi}\sigma_{\eta} & \sigma_{\eta}^2 \end{pmatrix}.$$

Как видно, это симметричная квадратная матрица, ее размер соответствует размерности исходного случайного вектора. Определитель этой матрицы вычисляется достаточно просто:

$$\text{Det } \Sigma_{\zeta} = |\Sigma_{\zeta}| = \sigma_{\xi}^2 \sigma_{\eta}^2 - \rho^2(\xi, \eta)\sigma_{\xi}^2 \sigma_{\eta}^2 = \sigma_{\xi}^2 \sigma_{\eta}^2 (1 - \rho^2(\xi, \eta)).$$

Если случайные компоненты вектора ζ независимы или хотя бы некоррелированы, ковариационная матрица становится диагональной, а ее определитель равен произведению диагональных элементов:

$$\Sigma_{\zeta} = \begin{pmatrix} \sigma_{\xi}^2 & 0 \\ 0 & \sigma_{\eta}^2 \end{pmatrix}, \quad \text{Det } \Sigma_{\zeta} = |\Sigma_{\zeta}| = \sigma_{\xi}^2 \sigma_{\eta}^2.$$

При взаимно однозначной связи между ξ и η , например линейной $\eta = a\xi$, коэффициент корреляции $|\rho(\xi, \eta)| = 1$, а из этого следует, что определитель ковариационной матрицы равен нулю, то есть в этом случае матрица оказывается особенной. Этого следовало ожидать, поскольку если между составляющими случайного вектора существует взаимно однозначная связь, то, по сути дела, существует всего одна случайная величина, случайный вектор вырождается в скалярную (одномерную) случайную

величину, и ковариационная матрица содержит только один элемент, то есть также вырождается в скаляр, а именно в значение дисперсии.

Для представления степени взаимной зависимости между компонентами случайного вектора применяется *корреляционная матрица*, элементами которой являются коэффициенты корреляции

$$C_{\zeta} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Если компоненты случайного вектора независимы или хотя бы некоррелированы, корреляционная матрица становится единичной матрицей.

1.7.4. Линейные функции случайных векторов

Задан общий вид линейной функции случайного вектора:

$$\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\zeta} + \mathbf{d},$$

где \mathbf{d} – неслучайный вектор, \mathbf{A} – квадратная матрица:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Найти математическое ожидание $M[\boldsymbol{\psi}]$ и ковариационную матрицу $\boldsymbol{\Sigma}_{\boldsymbol{\psi}}$ случайного вектора $\boldsymbol{\psi}$.

$$M[\boldsymbol{\psi}] = M \begin{pmatrix} a_{11}\xi + a_{12}\eta \\ a_{21}\xi + a_{22}\eta \end{pmatrix} + M[\mathbf{d}] = \begin{pmatrix} a_{11}M[\xi] + a_{12}M[\eta] \\ a_{21}M[\xi] + a_{22}M[\eta] \end{pmatrix} + \mathbf{d} = \mathbf{A}M[\boldsymbol{\zeta}] + \mathbf{d},$$

то есть, как и ранее, математическое ожидание линейной функции от случайного вектора есть функция от математического ожидания.

Для нахождения ковариационной матрицы вектора $\boldsymbol{\psi}$ воспользуемся математическим определением ковариационной матрицы, приведенным в разд. 1.7.3, и только что полученным выражением для математического ожидания линейной функции случайного вектора:

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{\psi}} &= M[(\boldsymbol{\psi} - M[\boldsymbol{\psi}])(\boldsymbol{\psi} - M[\boldsymbol{\psi}])^T] = M[(\mathbf{A}\boldsymbol{\zeta} + \mathbf{d} - \mathbf{A}M[\boldsymbol{\zeta}] - \mathbf{d})(\mathbf{A}\boldsymbol{\zeta} + \mathbf{d} - \mathbf{A}M[\boldsymbol{\zeta}] - \mathbf{d})^T] = \\ &= M[\mathbf{A}(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])^T \mathbf{A}^T] = \mathbf{A}M[(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])(\boldsymbol{\zeta} - M[\boldsymbol{\zeta}])^T] \mathbf{A}^T. \end{aligned}$$

В конечном итоге получим важную формулу:

$$\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{\zeta} + \mathbf{b}} = \mathbf{A}\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\mathbf{A}^T.$$

Как и ранее, мы видим, что ковариационная матрица не зависит от вектора \mathbf{d} , то есть от систематического (не случайного) сдвига случайного вектора.

Возможно так подобрать матрицу \mathbf{A} , чтобы ковариационная матрица Σ_{Ψ} оказалась диагональной, и тем самым компоненты вектора Ψ – некоррелированными.

Рассмотрим важный частный случай, когда матрица $\mathbf{A} = (a, b)$, а вектор \mathbf{d} состоит из одной компоненты. В этом случае вместо вектора Ψ имеем скаляр, и рассматриваемое преобразование выглядит следующим образом:

$$\psi = (ab) \begin{pmatrix} \xi \\ \eta \end{pmatrix} + d = a\xi + b\eta + d.$$

Пользуясь только что полученной формулой, получим

$$M[\psi] = aM[\xi] + bM[\eta] + d.$$

Ковариационная матрица (в нашем случае – дисперсия) получается после преобразований, которые мы осуществим по формуле для ковариационной матрицы:

$$\begin{aligned} \Sigma_{\Psi} = \sigma_{\Psi}^2 &= \mathbf{A}\Sigma_{\xi}\mathbf{A}^T = (ab) \begin{pmatrix} \sigma_{\xi}^2 & \rho(\xi, \eta)\sigma_{\xi}\sigma_{\eta} \\ \rho(\xi, \eta)\sigma_{\xi}\sigma_{\eta} & \sigma_{\eta}^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \\ &= \left(a\sigma_{\xi}^2 + b\rho(\xi, \eta)\sigma_{\xi}\sigma_{\eta} + a\rho(\xi, \eta)\sigma_{\xi}\sigma_{\eta} + b\sigma_{\eta}^2 \right) \begin{pmatrix} a \\ b \end{pmatrix}. \end{aligned}$$

Перемножив эти два вектора, окончательно получим

$$\sigma_{\Psi}^2 = a^2\sigma_{\xi}^2 + b^2\sigma_{\eta}^2 + 2ab\rho(\xi, \eta)\sigma_{\xi}\sigma_{\eta}.$$

В частном случае, когда ξ и η независимы, дисперсия случайной величины $\psi = a\xi + b\eta$

$$D[a\xi + b\eta] = \sigma_{\Psi}^2 = a^2\sigma_{\xi}^2 + b^2\sigma_{\eta}^2.$$

Если $a = b = 1$, эта формула приобретает совсем простой вид и означает, что дисперсия суммы двух независимых случайных величин равна сумме их дисперсий: $D[\xi + \eta] = \sigma_{\xi}^2 + \sigma_{\eta}^2$.

Полученные формулы справедливы вне зависимости от вида плотности распределения участвующих случайных величин.

1.7.5. Характеристическая функция двумерного случайного вектора

Двумерный случайный вектор ζ задан плотностью распределения $\varphi(x, y)$ и в результате “испытаний” принимает значения \mathbf{z} .

Обозначение и математическое определение характеристической функции вектора ζ :

$$g_{\zeta}(v_1, v_2) = M \left[e^{jv_1\xi} e^{jv_2\eta} \right] = M \left[e^{j\mathbf{v}^T \zeta} \right],$$

где \mathbf{v} – двумерный вектор, с компонентами v_1 и v_2 .

В соответствии с математическим определением математического ожидания

$$g_{\zeta}(v_1, v_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{jv_1x + jv_2y} \varphi(x, y) dx dy.$$

Свойства характеристических функций двумерного вектора аналогичны свойствам, которые были установлены в разд. 1.6.4 для характеристических функций случайной величины:

$$g_{\zeta}(0, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) dx dy = 1;$$

$$\left. \frac{\partial g_{\zeta}^k(v_1, v_2)}{\partial v_1^k} \right|_{\substack{v_1=0 \\ v_2=0}} = (j)^k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k \varphi(x, y) dx dy = (j)^k \mu_k(\xi);$$

$$\left. \frac{\partial g_{\zeta}^k(v_1, v_2)}{\partial v_2^k} \right|_{\substack{v_1=0 \\ v_2=0}} = (j)^k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^k \varphi(x, y) dx dy = (j)^k \mu_k(\eta);$$

$$\left. \frac{\partial g_{\zeta}^{k+r}(v_1, v_2)}{\partial v_1^k \partial v_2^r} \right|_{\substack{v_1=0 \\ v_2=0}} = (j)^{k+r} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^r \varphi(x, y) dx dy = (j)^{k+r} \mu_{k,r}(\xi, \eta).$$

К этим свойствам добавляется еще одно. Пусть ξ и η независимы и $\psi = \xi + \eta$. Тогда

$$\begin{aligned}
g_{\psi}(v) &= M\left[e^{jv(\xi+\eta)}\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{jv(x+y)} \varphi(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{jv(x+y)} \varphi_{\xi}(x) \varphi_{\eta}(y) dx dy = \\
&= \int_{-\infty}^{\infty} e^{jvx} \varphi_{\xi}(x) dx \int_{-\infty}^{\infty} e^{jvy} \varphi_{\eta}(y) dy = g_{\xi}(v) g_{\eta}(v),
\end{aligned}$$

то есть *характеристическая функция суммы независимых случайных величин равна произведению характеристических функций этих величин.*

Это свойство не могло быть обнаружено в разд. 1.6.4, поскольку не была определена совместная плотность распределения случайных величин.

1.7.6. Плотность распределения суммы двух независимых случайных величин

Поскольку ξ и η независимы, совместная плотность распределения $\varphi(x, y) = \varphi_{\xi}(x) \varphi_{\eta}(y)$. Сделаем замену $x = z - y$, тогда $\varphi(z - y, y) = \varphi_{\xi}(z - y) \varphi_{\eta}(y)$, после чего найдем маргинальную плотность распределения

$$\varphi_{\psi}(z) = \int_{-\infty}^{\infty} \varphi(z - y, y) dy = \int_{-\infty}^{\infty} \varphi_{\xi}(z - y) \varphi_{\eta}(y) dy.$$

Если вначале сделать замену $y = z - x$, то получим

$$\varphi_{\psi}(z) = \int_{-\infty}^{\infty} \varphi_{\eta}(z - x) \varphi_{\xi}(x) \cdot dx.$$

Рассмотрим эту задачу с других позиций.

Ранее мы обращали внимание на то, что интегральное преобразование, которое связывает плотность распределения и характеристическую функцию случайной величины, аналогично преобразованию Фурье и обладает его свойствами. В настоящем разделе мы используем следующее свойство преобразования Фурье: преобразование Фурье свертки двух функций есть произведение образов Фурье каждой из этих функций. В силу взаимной однозначности преобразования Фурье справедливо и обратное, то есть преобразование Фурье произведения функций есть свертка их Фурье-образов.

Если $\psi = \xi + \eta$ и величины ξ и η независимы, то характеристическая функция их суммы есть произведение характеристических функций слага-

емых $g_\psi(v) = g_\xi(v)g_\eta(v)$. Прямое преобразование Фурье этого произведения, в соответствии с указанным выше свойством, есть свертка плотностей распределения слагаемых:

$$\varphi_\psi(z) = \int_{-\infty}^{\infty} \varphi_\xi(z-y)\varphi_\eta(y)dy = \int_{-\infty}^{\infty} \varphi_\eta(z-x)\varphi_\xi(x)dx.$$

Применительно к плотностям распределения эту операцию иногда называют *композицией*.

Здесь мы еще раз вспомним о центральной предельной теореме, в соответствии с которой с увеличением числа независимых слагаемых случайных величин их сумма асимптотически нормальна, а это значит, что многократная свертка плотностей распределений должна приводить к плотности распределения, приближающейся к нормальной.

1.7.7. Двумерная нормальная плотность распределения

Вновь обозначим: $\zeta^T = (\xi, \eta)$ – нормальный случайный вектор, $\mathbf{z}^T = (x, y)$ – значения, которые он принимает. Двумерная нормальная плотность распределения записывается в виде

$$\varphi(\mathbf{z}) = \varphi(x, y) = \frac{1}{2\pi|\Sigma_\zeta|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{c})^T \Sigma_\zeta^{-1}(\mathbf{z} - \mathbf{c})\right\},$$

где $\mathbf{c} = M[\zeta]$ – двумерный вектор – математическое ожидание случайного вектора ζ , Σ_ζ – ковариационная матрица вектора ζ , $|\Sigma_\zeta|$ – ее определитель

$$\mathbf{c} = \begin{pmatrix} M[\xi] \\ M[\eta] \end{pmatrix}, \quad \Sigma_\zeta = \begin{pmatrix} \sigma_\xi^2 & \rho\sigma_\xi\sigma_\eta \\ \rho\sigma_\xi\sigma_\eta & \sigma_\eta^2 \end{pmatrix}, \quad |\Sigma_\zeta| = \sigma_\xi^2\sigma_\eta^2(1 - \rho^2).$$

Индексы у обозначения коэффициента корреляции ρ опущены, поскольку здесь это не вызовет никаких осложнений.

В дальнейшем принадлежность случайного вектора двумерному нормальному распределению будем записывать $\zeta \in N_2(\mathbf{c}, \Sigma_\zeta)$.

Форма и свойства двумерной нормальной плотности распределения определяются, в основном, показателем степени экспоненты и не зависят

от ее расположения, а потому примем $\mathbf{c} = \mathbf{0}$. В связи с этим рассмотрим показатель степени экспоненты отдельно:

$$\frac{1}{2} \mathbf{z}^T \boldsymbol{\Sigma}_{\zeta}^{-1} \mathbf{z} = \frac{1}{2} (x \quad y) \begin{pmatrix} \sigma_{\xi}^2 & \rho \sigma_{\xi} \sigma_{\eta} \\ \rho \sigma_{\xi} \sigma_{\eta} & \sigma_{\eta}^2 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Обращение матрицы выполним через алгебраические дополнения и определитель:

$$\begin{aligned} \begin{pmatrix} \sigma_{\xi}^2 & \rho \sigma_{\xi} \sigma_{\eta} \\ \rho \sigma_{\xi} \sigma_{\eta} & \sigma_{\eta}^2 \end{pmatrix}^{-1} &= \frac{1}{\sigma_{\xi}^2 \sigma_{\eta}^2 (1 - \rho^2)} \begin{pmatrix} \sigma_{\eta}^2 & -\rho \sigma_{\xi} \sigma_{\eta} \\ -\rho \sigma_{\xi} \sigma_{\eta} & \sigma_{\xi}^2 \end{pmatrix} = \\ &= \frac{1}{(1 - \rho^2)} \begin{pmatrix} \sigma_{\xi}^{-2} & -\rho \sigma_{\xi}^{-1} \sigma_{\eta}^{-1} \\ -\rho \sigma_{\xi}^{-1} \sigma_{\eta}^{-1} & \sigma_{\eta}^{-2} \end{pmatrix}. \end{aligned}$$

Тогда

$$\begin{aligned} \frac{1}{2} \mathbf{z}^T \boldsymbol{\Sigma}_{\zeta}^{-1} \mathbf{z} &= \frac{1}{2(1 - \rho^2)} (x \quad y) \begin{pmatrix} \sigma_{\xi}^{-2} & -\rho \sigma_{\xi}^{-1} \sigma_{\eta}^{-1} \\ -\rho \sigma_{\xi}^{-1} \sigma_{\eta}^{-1} & \sigma_{\eta}^{-2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \\ &= \frac{1}{2(1 - \rho^2)} \left(\frac{x^2}{\sigma_{\xi}^2} - 2\rho \frac{xy}{\sigma_{\xi} \sigma_{\eta}} + \frac{y^2}{\sigma_{\eta}^2} \right). \end{aligned}$$

Если нормальные случайные величины ξ и η некоррелированы, то есть $\rho = 0$, то показатель степени экспоненты упрощается очевидным образом, и двумерная нормальная плотность распределения распадается на два сомножителя:

$$\varphi(x, y) = \frac{1}{2\pi\sigma_{\xi}\sigma_{\eta}} \exp\left\{-\frac{1}{2}\left(\frac{x^2}{\sigma_{\xi}^2} + \frac{y^2}{\sigma_{\eta}^2}\right)\right\} = \frac{1}{\sqrt{2\pi}\sigma_{\xi}} \exp\left\{-\frac{x^2}{2\sigma_{\xi}^2}\right\} \frac{1}{\sqrt{2\pi}\sigma_{\eta}} \exp\left\{-\frac{y^2}{2\sigma_{\eta}^2}\right\},$$

каждый из которых есть не что иное, как маргинальная плотность распределения каждой из компонент вектора ζ . Из этого факта мы получаем полезный вывод:

Из некоррелированности двух нормально распределенных случайных величин с необходимостью следует их независимость; это исключительное свойство нормальных случайных величин, которое является их характеристическим признаком, а именно, если для двух случайных величин

установлено, что из некоррелированности следует их независимость, то эти случайные величины распределены нормально.

Проанализируем геометрическую форму поверхности, которая задана нормальной плотностью распределения.

Очевидно, что эта поверхность имеет максимум в точке $\mathbf{z} = \mathbf{c}$, поскольку при любом отклонении от этой точки абсолютная величина показателя степени экспоненты возрастает, а из-за того, что этот показатель степени меньше нуля, функция в целом убывает. Значение максимума в указанной точке равно множителю перед экспонентой.

Для исследования формы поверхности примем $\mathbf{c} = \mathbf{0}$ и рассечем ее плоскостью, параллельной плоскости xOy на уровне B , ниже максимального значения. Указанное рассечение выражается равенством

$$\varphi(x, y) = \frac{1}{2\pi\sigma_\xi\sigma_\eta\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_\xi^2} - 2\rho\frac{xy}{\sigma_\xi\sigma_\eta} + \frac{y^2}{\sigma_\eta^2}\right)\right\} = B.$$

В сечении получим фигуру, описываемую уравнением

$$\frac{1}{(1-\rho^2)}\left(\frac{x^2}{\sigma_\xi^2} - 2\rho\frac{xy}{\sigma_\xi\sigma_\eta} + \frac{y^2}{\sigma_\eta^2}\right) = -2 \cdot \ln\left(B \cdot 2\pi\sigma_\xi\sigma_\eta\sqrt{(1-\rho^2)}\right).$$

Обозначим $-2 \ln\left(B \cdot 2\pi\sigma_\xi\sigma_\eta\sqrt{(1-\rho^2)}\right) = r^2$, тогда уравнение фигуры, получившейся в сечении, принимает знакомый вид:

$$\frac{x^2}{\sigma_\xi^2} - 2\rho\frac{xy}{\sigma_\xi\sigma_\eta} + \frac{y^2}{\sigma_\eta^2} = r^2(1-\rho^2).$$

Это уравнение эллипса, размеры и расположение которого определяются его параметрами: радиусом $r\sqrt{1-\rho^2}$, который увеличивается с уменьшением B , то есть с приближением секущей плоскости к плоскости xOy , длинами полуосей $r\sigma_\xi$ и $r\sigma_\eta$, а также коэффициентом корреляции ρ . Пример одного из сечений приведен на рис. 25. Непосредственно из уравнения эллипса видно, что точки его пересечения с осями координат суть $(-\sigma_\xi r\sqrt{1-\rho^2}, 0)$, $(\sigma_\xi r\sqrt{1-\rho^2}, 0)$ – с осью x , и $(-\sigma_\eta r\sqrt{1-\rho^2}, \sigma_\eta r\sqrt{1-\rho^2})$, $(\sigma_\eta r\sqrt{1-\rho^2}, -\sigma_\eta r\sqrt{1-\rho^2})$ – с осью y . Угол наклона большой оси к оси абсцисс

$$\alpha = \frac{1}{2} \operatorname{arctg} \left(2\rho \frac{\sigma_\xi \sigma_\eta}{|\sigma_\xi^2 - \sigma_\eta^2|} \right).$$

При изменении коэффициента корреляции в диапазоне от -1 до 1 все эллипсы остаются вписанными в прямоугольник со сторонами $2r\sigma_\xi, 2r\sigma_\eta$. При $\rho = 0$ оси эллипса совпадают с осями координат, при $\rho = -1$ и при $\rho = +1$ точки его пересечения с осями перемещаются в начало координат, и эллипс вырождается в прямую линию. Это говорит о том, что в таких крайних случаях из-за жесткой связи между случайными величинами для их размещения хватает одной прямой, то есть по существу, мы имеем дело не с двумя, а с одной случайной величиной, как это уже указывалось в разд. 1.7.3.

На рис. 25 показана еще одна прямая, поименованная как *линия регрессии*. Эта линия есть геометрическое место точек,

которые представляют все условные математические ожидания $M[\eta/\xi = x]$.

Для того чтобы получить уравнение этой линии, найдем выражение для условной плотности распределения, для чего воспользуемся соответствующей формулой из разд. 1.7.1, помня о том, что маргинальные плотности двумерного нормального распределения также нормальны:

$$\varphi(y/x) = \frac{\varphi(x, y)}{\varphi_\xi(x)} = \frac{\sqrt{2\pi}\sigma_\xi}{2\pi\sigma_\xi\sigma_\eta\sqrt{1-\rho^2}} \exp \left\{ - \left[\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_\xi^2} - 2\rho \frac{xy}{\sigma_\xi\sigma_\eta} + \frac{y^2}{\sigma_\eta^2} \right) - \frac{1}{2} \frac{x^2}{\sigma_\xi^2} \right] \right\} =$$

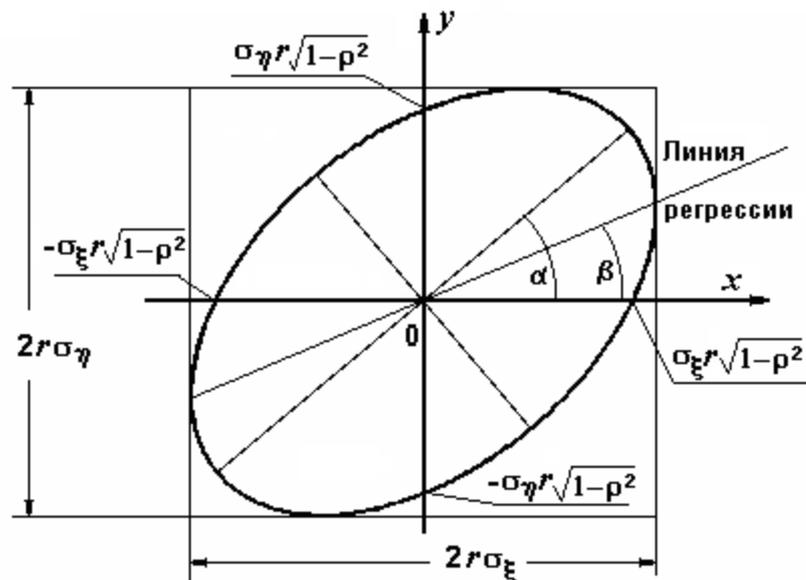


Рис. 25. Сечение двумерной нормальной плотности распределения

$$= \frac{1}{\sqrt{2\pi}\sigma_\eta\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\rho^2\frac{x^2}{\sigma_\xi^2} - 2\rho\frac{xy}{\sigma_\xi\sigma_\eta} + \frac{y^2}{\sigma_\eta^2}\right)\right\}.$$

В фигурных скобках имеет место полный квадрат разности, поэтому

$$\varphi(y/x) = \frac{1}{\sqrt{2\pi}\sigma_\eta\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_\eta^2(1-\rho^2)}\left(y - \rho\frac{\sigma_\eta}{\sigma_\xi}x\right)^2\right\}.$$

Полученное выражение есть каноническая запись одномерной нормальной плотности распределения с параметрами:

дисперсия равна $\sigma_\eta^2(1-\rho^2)$;

математическое ожидание равно $\rho\frac{\sigma_\eta}{\sigma_\xi}x$.

Таким образом, мы получили условную плотность распределения, которая оказывается также нормальной, и условное математическое ожидание $M[\eta/x] = \rho\frac{\sigma_\eta}{\sigma_\xi}x$, которое является линейной функцией x . Напомним,

что в начале настоящего пункта мы приняли равными нулю математические ожидания обеих компонент случайного вектора ζ . Настало время вспомнить об этом и записать условное математическое ожидание для общего случая:

$$M[\eta/x] = M[\eta] + \rho\frac{\sigma_\eta}{\sigma_\xi}(x - M[\xi]).$$

Полученное выражение называется *уравнением линейной регрессии случайной величины η на случайную величину ξ* . Угол наклона линии регрессии к оси абсцисс

$$\beta = \text{arctg}\left(\rho\frac{\sigma_\eta}{\sigma_\xi}\right).$$

Регрессия между двумя нормальными случайными величинами (если она существует, то есть когда $\rho \neq 0$) всегда линейна. Это естественно, поскольку только линейное преобразование не изменяет вид плотности распределения.

Прямая регрессии обязательно проходит через середины вертикальных хорд эллипса, потому что условные нормальные распределения, которые являются сечениями двумерной плотности вертикальными плоскостями, симметричны, и их математические ожидания совпадают с модами и медианами этих сечений. И поэтому прямая регрессии пересекается с эллипсом в точках касания вертикальных касательных к нему.

Заметим также, что частные (маргинальные) плотности двумерной нормальной величины также нормальны.

1.8. МНОГОМЕРНЫЙ СЛУЧАЙНЫЙ ВЕКТОР

Как и ранее, будем обозначать случайный вектор и его компоненты греческими буквами, а значения, которые они принимают, – латинскими:

$$\xi^T = (\xi_1 \xi_2 \dots \xi_n), \quad \mathbf{x}^T = (x_1 x_2 \dots x_n).$$

Функция распределения и плотность распределения многомерного случайного вектора определяются по аналогии с такими же характеристиками двумерного вектора (см. разд. 1.7.1):

$$F(x_1, x_2, \dots, x_n) = P(\xi_1 \leq x_1 \cap \xi_2 \leq x_2 \cap \dots \cap \xi_n \leq x_n),$$

$$\varphi(x_1, x_2, \dots, x_n) = \frac{d^n F(x_1, x_2, \dots, x_n)}{dx_1 dx_2 \dots dx_n}.$$

Моменты отдельных компонент случайного вектора вычисляются по формулам, аналогичным формулам разд. 1.7.2.

Начальные моменты k -го порядка

$$\mu_k[\xi_i] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i^k \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

центральные моменты k -го порядка

$$\overset{0}{\mu}_k[\xi_i] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - \mu_1[\xi_i])^k \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n.$$

В частности математические ожидания и дисперсии i -х компонент:

$$M[\xi_i] = \mu_1[\xi_i] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

$$D[\xi_i] = \sigma_i^2 = \overset{0}{\mu}_2[\xi_i] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - \mu_1[\xi_i])^2 \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n.$$

Из бесчисленного количества комбинаций смешанных моментов запишем только ковариации, то есть центральные смешанные моменты порядка 1, 1:

$$\text{cov}(\xi_i, \xi_j) = \rho_{ij} \sigma_i \sigma_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - \mu_1[\xi_i])(x_j - \mu_1[\xi_j]) \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

где ρ_{ij} – коэффициент корреляции компонент ξ_i и ξ_j , σ_i, σ_j – средне-квадратические значения этих компонент.

Математическое ожидание, ковариационная и корреляционная матрицы многомерного случайного вектора:

$$M[\xi^T] = \mathbf{c}^T = (M[\xi_1] \quad M[\xi_2] \quad \dots \quad M[\xi_n]),$$

$$\text{cov}(\xi) = \Sigma_{\xi} = M[(\xi - \mathbf{c})(\xi - \mathbf{c})^T] = \begin{pmatrix} \sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 & \dots & \rho_{1n} \sigma_1 \sigma_n \\ \rho_{12} \sigma_1 \sigma_2 & \sigma_2^2 & \dots & \rho_{2n} \sigma_2 \sigma_n \\ \dots & \dots & \dots & \dots \\ \rho_{1n} \sigma_1 \sigma_n & \rho_{2n} \sigma_2 \sigma_n & \dots & \sigma_n^2 \end{pmatrix},$$

$$\mathbf{C}_{\xi} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{12} & 1 & \dots & \rho_{2n} \\ \dots & \dots & \dots & \dots \\ \rho_{1n} & \rho_{2n} & \dots & 1 \end{pmatrix}.$$

Обе матрицы симметричны и неотрицательно определены. Если компоненты вектора независимы или хотя бы некоррелированы, то матрицы диагональны, а корреляционная матрица есть единичная матрица.

Маргинальные (частные) и условные плотности распределения могут быть многомерными:

$$\varphi_{\xi_1 \xi_2 \dots \xi_m}(x_1, x_2, \dots, x_m) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(x_1, x_2, \dots, x_n) dx_{m+1} dx_{m+2} \dots dx_n,$$

$$\varphi(x_{m+1}, x_{m+2}, \dots, x_n / \xi_1 = x_1, \xi_2 = x_2, \dots, \xi_m = x_m) = \frac{\varphi(x_1, x_2, \dots, x_n)}{\varphi_{\xi_1 \xi_2 \dots \xi_m}(x_1, x_2, \dots, x_m)},$$

где $m = 1, 2, \dots, n - 1$.

Характеристическая функция многомерного случайного вектора записывается в векторном виде так же, как в разд. 1.7.5:

$$g_{\xi} = M \left[e^{j\mathbf{v}^T \xi} \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{j\mathbf{v}^T \xi} \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

где вектор \mathbf{v} есть $\mathbf{v}^T = (v_1 v_2 \dots v_n)$.

Ее свойства, перечисленные в разд. 1.7.5, полностью сохраняются.

Математическое ожидание и ковариационная матрица вектора $\boldsymbol{\psi}$, который получается в результате линейного преобразования вектора ξ : $\boldsymbol{\psi} = \mathbf{A}\xi + \mathbf{d}$, выражаются, формулами, которые были получены в разд. 1.7.4:

$$M[\boldsymbol{\psi}] = M[\mathbf{A}\xi + \mathbf{d}] = \mathbf{A}M[\xi] + \mathbf{d}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\psi}} = \boldsymbol{\Sigma}_{\mathbf{A}\xi + \mathbf{b}} = \mathbf{A}\boldsymbol{\Sigma}_{\xi}\mathbf{A}^T.$$

Многомерная нормальная плотность распределения имеет вид

$$\varphi(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{\xi}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{c})^T \boldsymbol{\Sigma}_{\xi}^{-1} (\mathbf{x} - \mathbf{c}) \right\}.$$

Сечения этой поверхности плоскостями $\varphi(\mathbf{x}) = B$ – многомерные эллипсоиды.

Принадлежность случайного вектора n -мерному нормальному распределению будем обозначать, как $\xi \in N_n(\mathbf{c}, \boldsymbol{\Sigma}_{\xi})$.

Дифференциальная (относительная) энтропия нормального случайного n -мерного вектора равна

$$H_{\varepsilon}(\xi) = \frac{n}{2} \log_2 2\pi e + \frac{1}{2} \log_2 |\boldsymbol{\Sigma}_{\xi}|.$$

2. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

2.1. ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Основными задачами математической статистики являются:

- 1) определение числовых характеристик, параметров и свойств случайных величин по результатам экспериментов (испытаний);
- 2) проверка статистических гипотез о числовых характеристиках, параметрах и свойствах случайных величин по результатам экспериментов (испытаний);

Все значения, которые может принимать случайная величина ξ , образуют генеральную совокупность X . При выполнении экспериментов (испытаний) из генеральной совокупности извлекаются значения

$$x_1, x_2, \dots, x_n,$$

которые называются *выборочными значениями*, а массив выборочных значений называется *выборкой*. Количество выборочных значений n – *объем выборки*. В ряде случаев нам будет удобно обозначать выборку с помощью вектора выборочных значений $\mathbf{x}_n^T = (x_1 x_2 \dots x_n)$, где индекс n обозначает объем выборки, то есть размерность этого вектора.

Пусть $F(x)$, $f(x)$ – функция распределения и плотность распределения случайной величины ξ . В дальнейшем будем называть их, а также их числовые характеристики и параметры *генеральной функцией распределения*, *генеральной плотностью распределения*, *генеральными числовыми характеристиками* и *генеральными параметрами* соответственно. Оценки функции распределения, плотности распределения, числовых характеристик и параметров, полученные по выборочным значениям, будем называть *выборочными*.

В математической статистике принято, что все эксперименты (испытания) выполняются таким образом, чтобы выборочные значения были независимыми в смысле определения независимости, сформулированного в

разд. 1.2.3, 1.7.1. Поэтому в дальнейшем будем предполагать, что эксперимент (испытания) организован таким образом, чтобы это условие обеспечивалось.

В дальнейшем нам придется говорить о случайности выборочных значений и соответственно – о случайности оценок характеристик и параметров, полученных путем обработки выборочных значений.

Противоречивость ситуации состоит в следующем. С одной стороны, выборочные значения, которые получены в результате эксперимента, – это реализации значений, принятых случайной величиной, и они не являются случайными. С другой стороны, при повторении n экспериментов выборочные значения будут иными. То есть выборочные значения и вектор выборочных значений являются функциями случайных событий. Поэтому в дальнейшем будем считать *выборочные значения случайными на множестве групп экспериментов (испытаний) объемом n каждая, выполняемых в одних и тех же неизменных заранее сформулированных условиях*. В терминах и аксиоматике, введенных в разд. 1, это означает, что в качестве элементарного события ω_i будем рассматривать исход одного эксперимента, результатом которого является получение выборочного вектора x . В таких условиях это случайный n -мерный вектор с независимыми одинаково распределенными компонентами, поскольку все выборочные значения в соответствии с соглашениями, принятыми в математической статистике, извлечены из одной генеральной совокупности. Вероятностная мера на этом множестве порождена вероятностной мерой, которая определена генеральной совокупностью исследуемой случайной величины ξ .

2.2. КОНДИЦИОНИРОВАНИЕ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТОВ

Пусть случайная величина ξ непрерывна. В этом случае все выборочные значения будут различны.

После получения выборочных значений случайной величины в хронологическом порядке первым шагом в их обработке является сортировка в порядке возрастания. Это действие осуществляется автоматически при расстановке выборочных значений на числовую ось. Порядковые номера

выборочных значений изменяются, и новые порядковые номера указываются в скобках:

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}.$$

Выборка, отсортированная таким образом, называется *вариационным рядом*, отдельные элементы – *членами вариационного ряда*. Первый и последний члены называются *крайними членами вариационного ряда*. Если количество членов вариационного ряда нечетное, то существует единственный *средний член вариационного ряда*, номер которого $(2n + 1)/2$. Средний член вариационного ряда называется *выборочной медианой*.

Следующий шаг – построение *выборочной функции распределения* $\tilde{F}(x)$, которая является оценкой генеральной функции распределения $F(x)$. Пример выборочной функции распределения приведен на рис. 26.

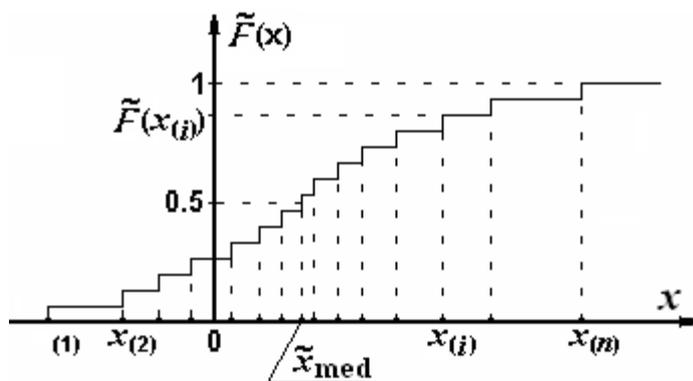


Рис. 26. Выборочная функция распределения

Выборочная функция распределения изображается ступенчатой линией. Абсциссами каждого скачка этой линии являются выборочные значения $x_{(i)}$. Высота всех ступеней одинакова и равна $1/n$. Такое построение выборочной функции распределения основано на результате, по-

лученном в примере 3 разд. 1.6.7.

В самом деле, при каждом значении $x_{(i)}$ ординатами каждой ступени являются значения выборочной функции распределения $\tilde{F}(x_{(i)})$. На множестве экспериментов, повторяющихся в неизменных условиях, выборочные значения являются случайными величинами, распределенными так же, как и генеральная совокупность, из которой они извлекаются, то есть $F(x)$ – их функция распределения. А это значит, что реализуются условия упомянутого примера, из которого следует, что значения $\tilde{F}(x_{(i)})$, которые являются случайными на множестве повторяющихся экспериментов, рас-

пределены равномерно в интервале $[0, 1]$. Поэтому высота всех ступенек принята одинаковой.

Оценкой плотности распределения является *гистограмма*. По сути дела, гистограмма является выборочной плотностью распределения, и в отличие от генеральной плотности распределения $\varphi(x)$ мы обозначим ее $\tilde{\varphi}(x)$. Пример гистограммы приведен на рис. 27.

Для построения гистограммы интервал между крайними членами вариационного ряда делится на M интервалов $(x_{m-1}, x_m]$ равной длины $\Delta = (x_m - x_{m-1})$. Подсчитывается количество n_m выборочных значений, попавших в каждый m -ой интервал и вычисляется частота n_m/n , которая является оценкой вероятностной меры каждого интервала.

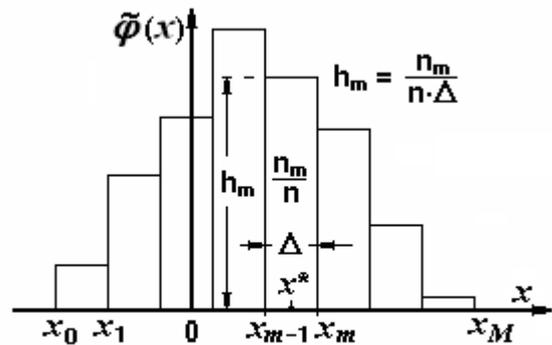


Рис. 27. Гистограмма

Количество интервалов или их ширина выбирается таким образом, чтобы самый “бедный” интервал содержал $3 \div 5$ выборочных значений. Для удобства построения гистограммы и последующих вычислений рекомендуется округлить значение ширины интервалов до ближайшего удобного числа. Далее на полученных интервалах, как на основаниях, строятся прямоугольники, высота каждого из которых должна быть

$$h_m = \frac{n_m}{n\Delta}.$$

Только при таком построении площадь гистограммы будет равна единице, точно так же, как и под генеральной плотностью распределения, оценкой которой является гистограмма. В самом деле,

$$\sum_{m=1}^M h_m \Delta = \sum_{m=1}^M \frac{n_m}{n} = \frac{1}{n} \sum_{m=1}^M n_m = 1.$$

Следует обратить внимание на то, что за счет группирования выборочных данных при построении гистограммы часть информации о случайной величине, содержащейся в выборочных данных, теряется. Напротив,

выборочная функция сохраняет в себе всю информацию, содержащуюся в выборочных данных.

2.3. ТОЧЕЧНОЕ ОЦЕНИВАНИЕ

Точечной статистической оценкой называется оценка числовой характеристики или параметра генеральной совокупности, выражаемая одним числом.

2.3.1. Оценивание квантилей

Оценивание квантилей выполняется по выборочной функции распределения, по сути дела, графическим способом. Это можно показать на примере точечного оценивания медианы, то есть 50%-ной квантили $x_{0,5}$. Для этого по заданной вероятности $q = 0,5$ на графике выборочной функции распределения (рис. 26) проводят горизонтальную прямую до ее пересечения с выборочной функцией распределения. Абсцисса точки пересечения есть искомая оценка $\tilde{x}_{0,5}$, которая является *выборочной медианой* \tilde{x}_{med} . Этой абсциссой при нечетном $n = 2k + 1$ обязательно окажется, по построению, выборочное значение – средний член вариационного ряда $x_{(k+1)}$.

Для оценки других квантилей, например, квартили $x_{0,25}$ или децили $x_{0,1}$ горизонтальные линии проводят на уровне вероятности $F(x_{0,25}) = 0,25$ или $F(x_{0,1}) = 0,1$ соответственно. Так же находятся точечные оценки высоковероятных квантилей $F(x_{0,75})$ или $F(x_{0,9})$. Из графика (см. рис. 26) видно, что абсциссой пересечения выборочной функции распределения с любой горизонтальной прямой обязательно окажется одно из выборочных значений. И здесь необходимо обратить внимание на следующее обстоятельство.

Пусть необходимо оценить квантиль x_α при $0 < \alpha \leq 1/n$ или квантиль x_β при $(n-1)/n \leq \beta < 1$. При задании вероятностей α и β из указанных промежутков оценками соответствующих квантилей неизменно окажется первый или последний член вариационного ряда независимо от заданных значений вероятностей, что будет свидетельствовать о крайне низкой достоверности этих оценок. Для того чтобы не сталкиваться с подобной си-

туацией, необходимо установить минимально допустимый объем выборки, обеспечивающий достаточную достоверность получаемых оценок квантилей в зависимости от задаваемой вероятности. Эту задачу решим, исходя из приведенных неравенств:

при задании значения вероятности, не превышающего 0,5

$$n \geq \frac{1}{\alpha};$$

при задании значения вероятности, превышающего 0.5

$$n \geq \frac{1}{1-\beta}.$$

Так, например, если задана вероятность $\alpha = 0,05$, то для точечного оценивания квантили $x_{0,05}$ объем выборки не должен быть меньше 20. Если задана вероятность $\beta = 0,99$, то для точечного оценивания квантили $x_{0,99}$ объем выборки не должен быть меньше 100. Рекомендуется увеличивать объем выборки сверх минимального, указанного этими неравенствами, чтобы застраховаться от возможных грубых промахов измерений, которые могут быть вызваны импульсными помехами или сбоями аппаратуры.

Пусть стоит задача получить точечные оценки границ интерквантильного промежутка J_P . Как следует из разд. 1.6.2, генеральными граничными квантилями этого промежутка являются $x_{\frac{1-P}{2}}$ и $x_{\frac{1+P}{2}}$. Поэтому точечное оценивание границ интерквантильного промежутка состоит в поочередном оценивании каждой из этих квантилей при соблюдении указанных ранее соотношений и рекомендаций относительно необходимого объема выборки, то есть

$$\text{для оценки нижней границы } n \geq \frac{2}{1-P};$$

$$\text{для оценки верхней границы также } n \geq \frac{2}{1-P}.$$

В частности, если задано $P = 0,95$, то минимальный объем выборки составит $n_{\min} = 40$.

2.3.2. Точечное оценивание моментов

Оценке подлежат начальные моменты $\mu_k[\xi]$ и центральные моменты $\overset{\circ}{\mu}_k[\xi]$ случайной величины ξ , плотность распределения которой $\varphi(x)$. Основное внимание будет обращено на оценивание первых четырех моментов.

Статистические оценки генеральных моментов будем обозначать латинскими буквами и при необходимости показывать их зависимость от вектора выборочных значений x_n . Соответствие между обозначениями генеральных моментов и их оценок представлено в таблице 2.

В разд. 2.1 было отмечено, что вектор выборочных значений является n – мерным случайным вектором на множестве групп экспериментов объемом n . Поэтому оценки моментов, вычисляемые по выборочным значениям, также случайны на том же множестве. Это означает, что оценки моментов, вычисленные по n выборочным значениям, будут различаться

Таблица 2

Обозначения генеральных моментов и их оценок

Генеральные моменты	$\mu_1[\xi] = c$	$\overset{\circ}{\mu}_2[\xi] = \sigma_\xi^2$	$\mu_k[\xi]$	$\overset{\circ}{\mu}_k[\xi]$
Оценки	$m_1(\mathbf{x}_n)$	$s^2(\mathbf{x}_n)$	$m_k(\mathbf{x}_n)$	$\overset{\circ}{m}_k(\mathbf{x}_n)$

между собой случайным образом при повторении тех же n экспериментов в неизменных условиях, и, значит, являются случайными. Для них, как и для всех случайных величин, могут быть определены такие характеристики, как математическое ожидание, дисперсия, квантили и т.д.

2.3.2.1. Оценивание моментов по выборочной функции распределения

Для получения оценок по выборочной функции распределения воспользуемся интегралом Стильеса (см., например, [4]).

Интеграл Стильеса определен как предел суммы Стильеса:

$$\int_a^b f(x)dF(x) = \lim_{\substack{n=(b-a)/\Delta x \\ \Delta x \rightarrow 0}} \sum_{i=0}^n f(x_i^*)\Delta F(x_i),$$

где $f(x)$ и $F(x)$ – две ограниченные функции, Δx – ширина участков, на которые разделен интервал $[a, b]$ (если эти участки разной ширины, то тогда Δx – максимальная ширина), x_i^* – точка внутри i -го участка, $\Delta F(x_i)$ – приращение функции $F(x)$ на i -ом участке $\Delta F(x_i) = F(x_i) - F(x_{i-1})$.

Когда функция $F(x)$ дифференцируема везде на $[a, b]$, и ее производная $F'(x) = \varphi(x)$, интеграл Стильеса обращается в интеграл Римана:

$$\int_a^b f(x)dF(x) = \int_a^b f(x)\varphi(x)dx.$$

Если функция $F(x)$ имеет ступенчатый характер, то есть в точках x_i она изменяется скачком, а в остальных точках постоянна, то интеграл Стильеса вычисляется как сумма:

$$\int_a^b f(x)dF(x) = \sum_{i=0}^n f(x_i)\Delta F(x_i),$$

где $\Delta F(x_i)$ – значение скачка функции $F(x)$ в точках x_i .

Применяя интеграл Стильеса для оценки начальных моментов по выборочной функции распределения, по определению моментов (разд. 1.6.2), получим

$$m_k = \int_{-\infty}^{\infty} x^k d\tilde{F}(x) = \sum_{i=1}^n x_{(i)}^k \Delta\tilde{F}(x_i).$$

Но, как мы выяснили в разд. 2.2, все скачки $\Delta\tilde{F}(x_i)$ выборочной функции распределения в точках x_i одинаковы, равны $1/n$ и их можно вынести за знак суммы. Кроме того, порядок перечисления слагаемых в сумме, стоящей справа, не имеет значения. Поэтому оценки начальных моментов порядка k вычисляются по формуле

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k .$$

В частности, оценкой математического ожидания служит среднее арифметическое:

$$m_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

Точно так же с помощью интеграла Стильеса получим оценки центральных моментов:

$$m_k = \int_{-\infty}^{\infty} (x - m_1)^k d\tilde{F}(x) = \sum_{i=1}^n (x_i - \bar{x})^k \Delta\tilde{F}(x_{(i)}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k .$$

В частности, оценка дисперсии вычисляется как

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

Эта же оценка может быть вычислена иначе с применением формулы из разд. 1.6.2:

$$s^2 = m_2 - m_1^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left[\frac{1}{n} \sum_{i=1}^n x_i \right]^2 .$$

Эта формула бывает полезной при вычислении оценок на компьютере в темпе получения данных путем накопления оценок начальных моментов при получении каждого i -го результата измерений. Однако здесь следует предостеречь от опасности, которая заключается в возможности получения отрицательного значения s^2 . Это может произойти из-за погрешности округления, когда выборочные значения очень велики, а дисперсия генеральной совокупности по сравнению с ними очень мала.

2.3.2.2. Оценивание моментов по выборочной плотности распределения (по гистограмме)

В отличие от разд. 1.6.2, где определены генеральные моменты, здесь для определения оценок моментов вместо плотности распределения генеральной совокупности будем использовать выборочную плотность, то есть гистограмму (см. рис. 27). В соответствии с математическими определени-

ями генеральных моментов их оценки по гистограмме приобретают иной вид:

оценки начальных моментов

$$m_k = \int_{-\infty}^{\infty} x^k \tilde{\varphi}(x) dx = \sum_{m=1}^M \frac{n_m}{n \cdot \Delta} \int_{x_{m-1}}^{x_m} x^k dx = \frac{1}{n(k+1)} \sum_{m=1}^M n_m \frac{x_m^{k+1} - x_{m-1}^{k+1}}{\Delta};$$

оценки центральных моментов

$$\begin{aligned} m_k^o &= \int_{-\infty}^{\infty} (x - m_1)^k \tilde{\varphi}(x) dx = \sum_{m=1}^M \frac{n_m}{n \Delta} \int_{x_{m-1}}^{x_m} (x - m_1)^k dx = \\ &= \frac{1}{n(k+1)} \sum_{m=1}^M n_m \frac{(x_m - m_1)^{k+1} - (x_{m-1} - m_1)^{k+1}}{\Delta}. \end{aligned}$$

Понятно, что потери информации, вызванные группированием выборочных значений при построении гистограммы, снижают качество оценок по сравнению с оценками по выборочной функции распределения.

Пользуясь этими общими формулами, найдем оценки математического ожидания и дисперсии.

$$m_1 = \frac{1}{2n} \sum_{m=1}^M n_m \frac{x_m^2 - x_{m-1}^2}{\Delta} = \frac{1}{n} \sum_{m=1}^M n_m \frac{(x_m + x_{m-1})(x_m - x_{m-1})}{2\Delta}.$$

Поскольку

$$(x_m - x_{m-1}) = \Delta, \quad (x_m + x_{m-1}) = 2 \cdot x_m^*,$$

где x_m^* – середина m -го отрезка, окончательно получим:

$$m_1 = \frac{1}{n} \sum_{m=1}^M n_m x_m^*.$$

Оценка дисперсии.

$$\begin{aligned} s^2 = m_2^o &= \frac{1}{3n} \sum_{m=1}^M \frac{n_m}{\Delta} [(x_m - m_1)^3 - (x_{m-1} - m_1)^3] = \\ &= \frac{1}{3n} \sum_{m=1}^M \frac{n_m}{\Delta} \Delta [(x_m - m_1)^2 + (x_{m-1} - m_1)^2 + (x_m - m_1)(x_{m-1} - m_1)]. \end{aligned}$$

Используя равенство $(x_m + x_{m-1}) = 2x_m^*$, сделаем замены:

$$y_m = x_m^* - m_1, \quad (x_m - m_1) = y_m + \Delta/2, \quad (x_{m-1} - m_1) = y_m - \Delta/2.$$

Тогда предыдущее равенство упрощается:

$$\begin{aligned}
s^2 &= \frac{1}{3n} \sum_{m=1}^M n_m \left[\left(y_m + \frac{\Delta}{2} \right)^2 + \left(y_m - \frac{\Delta}{2} \right)^2 + \left(y_m + \frac{\Delta}{2} \right) \left(y_m - \frac{\Delta}{2} \right) \right] = \\
&= \frac{1}{3n} \sum_{m=1}^M n_m \left[3y_m^2 + \left(\frac{\Delta}{2} \right)^2 \right] = \frac{1}{n} \sum_{m=1}^M n_m y_m^2 + \frac{\Delta^2}{12n} \sum_{m=1}^M n_m.
\end{aligned}$$

Окончательно получим

$$s^2 = \frac{1}{n} \sum_{m=1}^M n_m (x_m^* - m_1)^2 + \frac{\Delta^2}{12}.$$

Слагаемое $\Delta^2 / 12$ называется *поправкой Шеппарда*.

2.3.3. Оценивание параметров законов распределения

Рассмотрим задачу оценивания r параметров $\Theta^T = (\Theta_1, \Theta_2, \dots, \Theta_r)$ закона распределения (плотности распределения) $\varphi(\Theta, x)$ случайной величины ξ по выборочным данным.

Поскольку

$$\mu_k[\xi] = \int_{-\infty}^{\infty} x^k \varphi(\Theta, x) dx,$$

все начальные моменты являются функциями искомых параметров, и в общем случае можно записать

$$\mu_k[\xi] = f_k(\Theta).$$

Точно так же функциями параметров являются и центральные моменты. В качестве примеров таких функций можно привести экспоненциальный и равномерный законы распределения, для которых, как было установлено в разд. 1.6.6:

$$\mu_1 = 1/\lambda, \overset{\circ}{\mu}_2 = \sigma^2 = 1/\lambda^2 \text{ — для экспоненциального закона;}$$

$$\mu_1 = (a+b)/2, \overset{\circ}{\mu}_2 = \sigma^2 = (b-a)^2 / 12 \text{ — для равномерного закона.}$$

Указанные зависимости используются для оценивания параметров Θ в тех случаях, когда закон распределения генеральной совокупности известен. С этой целью по выборочным данным находят оценки моментов. Количество моментов, подлежащих определению, должно быть равно числу r искомых параметров. Таким образом будут получено r уравнений:

$$m_k = f_k(\Theta).$$

Искомые параметры находят путем решения этих уравнений.

2.3.4. Свойства точечных оценок

Поскольку выборочные значения (вектор выборочных значений) случайны на множестве исходов экспериментов, повторяющихся в неизменных условиях и в неизменном объеме, то оценки моментов и параметров законов распределения также случайны, а потому в качестве характеристик их свойств применяют вероятностные критерии.

Применим общее обозначение генеральных моментов и параметров случайных величин и их законов распределения: Θ . Для обозначения оценок моментов и параметров, вычисляемых по выборке объема n , будем использовать обозначение $\tilde{\Theta}_n$.

Важнейшими свойствами точечных оценок являются: *несмещенность*, *эффективность*, *состоятельность*. В последнее время возрастает роль четвертого свойства – *устойчивости (resistance)*.

а) Оценка момента (или параметра) является несмещенной, если ее математическое ожидание при фиксированном объеме выборки n равно оцениваемому генеральному моменту или параметру:

$$M[\tilde{\Theta}_n] = \Theta.$$

б) Оценка момента (или параметра) эффективна, если при фиксированном объеме выборки n она обладает минимальной дисперсией среди всех оценок данного момента (параметра):

$$D[\tilde{\Theta}_n] \rightarrow \min.$$

в) Оценка момента (или параметра) состоятельна, если с увеличением объема выборки n она стремится по вероятности к генеральному значению момента (или параметра), то есть если при любом сколько угодно малом положительном ε

$$\lim_{n \rightarrow \infty} P\{|\tilde{\Theta}_n - \Theta| \leq \varepsilon\} = 1.$$

г) *Устойчивые оценки* – оценки, нечувствительные или малочувствительные к возмущениям в виде грубых промахов измерений, погрешностям вычислений и т.п.

Оценки могут быть:
 несмещенные, эффективные, состоятельные;
 несмещенные, но неэффективные;
 эффективные, но смещенные;
 смещенные при каждом фиксированном n , но состоятельные, такие оценки являются асимптотически несмещенными.

2.3.4.1. Свойства оценок математического ожидания

Выборка $\mathbf{x}^T = (x_1, x_2, \dots, x_n)$ извлечена из генеральной совокупности \mathbf{X} , образованной случайной величиной ξ . Генеральные моменты: математическое ожидание $M[\xi] = \mu_1[\xi] = \mu_1$, дисперсия $D[\xi] = \sigma_\xi^2 = \sigma^2$.

Для оценки математического ожидания используют несколько видов оценок. Наиболее популярной оценкой является среднее арифметическое значение

$$m_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

Среди других оценок назовем выборочную медиану $\tilde{x}_{\text{med}} = \tilde{x}_{0.5}$ и *средину размаха* $x_{\text{cp}} = (x_{(1)} + x_{(n)})/2$, которые применяются в случаях, когда случайная величина ξ имеет симметричную плотность распределения.

Рассмотрим подробно свойства оценки \bar{x} .

а) П р о в е р и м н е с м е щ е н н о с т ь среднего арифметического, используя формулы для математического ожидания линейной функции от случайной величины, полученные в разд. 1.6.5:

$$M[\bar{x}] = M\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n M[x_i] = \frac{1}{n} \cdot n \cdot M[x_i] = M[x_i] = \mu_1,$$

а это означает, что *среднее арифметическое выборочных значений есть несмещенная оценка математического ожидания случайной величины независимо от вида закона распределения.*

б) Н а й д е м д и с п е р с и ю среднего арифметического, используя те же формулы разд. 1.6.5 с учетом следующих обстоятельств:

выборочные значения попарно независимы;

выборочные значения случайны на множестве групп экспериментов объемом n , и каждое из них распределено так же, как случайная величина ξ , каждое выборочное значение имеет числовые характеристики генеральной совокупности.

$$D[\bar{x}] = D\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n^2} \sum_{i=1}^n D[x_i].$$

Поскольку выборочные значения x_i изъяты из одной генеральной совокупности, их дисперсии одинаковы. Поэтому

$$D[\bar{x}] = \frac{1}{n^2} n D[\xi] = \frac{D[\xi]}{n} = \frac{\sigma^2}{n}.$$

Среднеквадратическое значение среднего арифметического

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

Из полученных соотношений видно, что обработка результатов многократных измерений путем вычисления среднего арифметического небесполезна, ибо показатель разброса значений средних арифметических убывает с увеличением объема выборки.

Из последующего материала (разд. 2.3.5) мы узнаем:

что среднее арифметическое – *эффективная оценка математического ожидания нормальной случайной величин;*

что выборочная медиана – *несмещенная оценка математического ожидания случайных величин с симметричной плотностью распределения, она является эффективной оценкой математического ожидания случайной величины, распределенной по Лапласу* (см. также разд. 2.3.5);

что *середины размаха – несмещенная оценка математического ожидания случайных величин с симметричной плотностью распределения и является эффективной оценкой математического ожидания случайной величины, распределенной равномерно или по закону Arcsin.*

в) Проверим состоятельность среднего арифметического, как оценки математического ожидания.

Применим неравенство Чебышева (разд. 1.6.8), приняв в качестве случайной величины ξ среднее арифметическое \bar{x} с дисперсией σ^2/n . Из неравенства

$$P\{|\bar{x} - M[\bar{x}]| > \varepsilon\} \leq \frac{\sigma_{\bar{x}}^2}{\varepsilon^2}$$

следует противоположное ему:

$$P\{|\bar{x} - M[\bar{x}]| < \varepsilon\} \geq 1 - \frac{\sigma_{\bar{x}}^2}{\varepsilon^2}.$$

Поскольку среднее арифметическое есть несмещенная оценка математического ожидания, $M[\bar{x}] = M[\xi]$. Поэтому

$$P\{|\bar{x} - M[\xi]| < \varepsilon\} \geq 1 - \frac{\sigma_{\xi}^2}{n\varepsilon^2}.$$

Каким бы малым не было ε , его значение фиксировано, поэтому при $n \rightarrow \infty$ вычитаемое в правой части неравенства стремится к нулю, а правая часть – к единице.

Таким образом, состоятельность среднего арифметического как оценки математического ожидания доказана независимо от вида закона распределения.

Мы получили пример полезного применения неравенства Чебышева, несмотря на грубость даваемой им оценки.

Выборочная медиана и середина размаха – состоятельные оценки математического ожидания случайных величин с симметричной плотностью распределения.

г) П л о т н о с т ь р а с п р е д е л е н и я среднего арифметического.

Примем вначале, что выборка извлечена из нормальной генеральной совокупности $N(\mu_1, \sigma^2)$. Тогда на множестве групп экспериментов каждое выборочное значение распределено также нормально, то есть $x_i \in N(\mu_1, \sigma^2)$. В силу безграничной делимости нормального распределения сумма выборочных значений распределена нормально, а умножение ее на множитель $1/n$ приводит лишь к изменению масштаба. Таким образом, на основании свойств а) и б) получаем, что

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \in N\left(\mu_1, \frac{\sigma^2}{n}\right).$$

Более того, по центральной предельной теореме (разд. 1.6.6.4), плотность распределения среднего арифметического ассимптотически нормальна вне зависимости от вида плотности распределения генеральных совокупностей, из которых изъяты выборочные значения.

2.3.4.2. Свойства оценок дисперсии

а) Рассмотрим вначале случай, когда математическое ожидание исследуемой случайной величины известно и равно μ_1 . Дисперсия генеральной совокупности \mathbf{X} , образованной случайной величиной ξ , равна σ^2 .

В этом случае оценка дисперсии вычисляется по формуле

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_1)^2.$$

Проверим несмещенность этой оценки.

$$M[s^2] = M\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu_1)^2\right] = \frac{1}{n} \sum_{i=1}^n M[(x_i - \mu_1)^2] = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2.$$

Когда математическое ожидание исследуемой случайной величины известно, оценка дисперсии не смещена.

б) Обычно на практике математическое ожидание исследуемой случайной величины **неизвестно**. В этом случае вместо него приходится использовать оценку, например среднее арифметическое:

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

Проверим несмещенность этой оценки.

$$\begin{aligned} M[s^2] &= \frac{1}{n} M\left[\sum_{i=1}^n (x_i - \bar{x})^2\right] = \frac{1}{n} M\left[\sum_{i=1}^n [(x_i - \mu_1) - (\bar{x} - \mu_1)]^2\right] = \\ &= \frac{1}{n} M\left[\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^n (\bar{x} - \mu_1)^2 - 2\sum_{i=1}^n (x_i - \mu_1)(\bar{x} - \mu_1)\right] = \end{aligned}$$

$$= \frac{1}{n} \left[M \left[\sum_{i=1}^n (x_i - \mu_1)^2 \right] + M \left[\sum_{i=1}^n (\bar{x} - \mu_1)^2 \right] - 2 \cdot M \left[\sum_{i=1}^n (x_i - \mu_1)(\bar{x} - \mu_1) \right] \right].$$

Первое слагаемое во внешних квадратных скобках есть сумма дисперсий генеральной совокупности, их в этой сумме n штук, то есть первое слагаемое равно $n\sigma^2$. Второе слагаемое есть сумма дисперсий средних арифметических, оно равно $n\sigma^2/n = \sigma^2$. Конструкция вычитаемого сложнее. Рассмотрим его отдельно. Запишем второй сомножитель под знаком суммы в виде:

$$(\bar{x} - \mu_1) = \frac{1}{n} \sum_{j=1}^n (x_j - \mu_1).$$

В справедливости такой записи можно убедиться, раскрыв скобки и подсчитав каждую сумму по отдельности.

Сделаем подстановку:

$$2M \left[\sum_{i=1}^n (x_i - \mu_1)(\bar{x} - \mu_1) \right] = 2 \frac{1}{n} M \left[\sum_{i=1}^n (x_i - \mu_1) \sum_{j=1}^n (x_j - \mu_1) \right].$$

Здесь в квадратных скобках суммируются произведения биномов, отличающихся только индексом. В результате внутри квадратных скобок образуется сумма n^2 слагаемых, среди которых встречаются слагаемые, сомножители которых имеют одинаковые индексы $i = j$ (таких слагаемых n штук), и слагаемые, сомножители которых имеют разные индексы (таких слагаемых $n^2 - n$ штук).

Математическое ожидание каждого слагаемого из первой группы (при $i = j$)

$$M[(x_i - \mu_1)^2] = \sigma^2.$$

Математическое ожидание второй группы слагаемых (при $i \neq j$) $M[(x_i - \mu_1)(x_j - \mu_1)]$ есть не что иное, как центральный смешанный момент порядка 1, 1, а это, как следует из разд. 1.7.2, ковариация, которая вследствие независимости выборочных значений равна нулю.

В результате получаем, что

$$2 \frac{1}{n} M \left[\sum_{i=1}^n (x_i - \mu_1)(\bar{x} - \mu_1) \right] = \frac{2 n \sigma^2}{n} = 2 \sigma^2.$$

Подводя окончательный итог, получаем

$$M[s^2] = \frac{1}{n} [n \sigma^2 + \sigma^2 - 2 \sigma^2] = \frac{n-1}{n} \sigma^2.$$

Таким образом мы выяснили, что когда математическое ожидание генеральной совокупности неизвестно, исследованная оценка смещена. При $n \rightarrow \infty$ смещение оценки убывает до нуля, поэтому такая оценка является *асимптотически несмещенной*. Обнаруженное смещение корректируется простым умножением на $n/(n-1)$, и мы получаем общеизвестную рабочую формулу для расчета несмещенной оценки дисперсии:

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

Обращает на себя внимание тот факт, что при количестве слагаемых, равном n , сумма делится на $n-1$. Это происходит потому, что каждое слагаемое содержит в себе одинаковые выборочные значения с коэффициентом $1/n$, которые были использованы при вычислении среднего арифметического значения \bar{x} , и значит, эти слагаемые не являются независимыми. Чтобы убедиться в этом, найдем ковариацию $\text{cov}(x_i - \bar{x})(x_j - \bar{x})$. В силу несмещенности среднего арифметического как оценки математического ожидания математические ожидания каждого из этих сомножителей равно 0. Поэтому

$$\begin{aligned} \text{cov}[(x_i - \bar{x})(x_j - \bar{x})] &= M[(x_i - \bar{x})(x_j - \bar{x})] = \\ &= M \{ [(x_i - \mu_1) - (\bar{x} - \mu_1)] [(x_j - \mu_1) - (\bar{x} - \mu_1)] \} = \\ &= M [(x_i - \mu_1)(x_j - \mu_1) + (\bar{x} - \mu_1)^2 - (x_i - \mu_1)(\bar{x} - \mu_1) - (x_j - \mu_1)(\bar{x} - \mu_1)] = \\ &= M [(x_i - \mu_1)(x_j - \mu_1)] + M [(\bar{x} - \mu_1)^2] - M [(x_i - \mu_1)(\bar{x} - \mu_1)] - M [(x_j - \mu_1)(\bar{x} - \mu_1)], \end{aligned}$$

где μ_1 – математическое ожидание генеральной совокупности.

Первое слагаемое в последней сумме равно нулю в силу независимости выборочных данных. Второе слагаемое есть дисперсия среднего арифметического, которая равна σ^2/n . Третье и четвертое слагаемые равны друг другу, поэтому можно записать:

$$\text{cov}[(x_i - \bar{x})(x_j - \bar{x})] = \frac{\sigma^2}{n} - 2M[(x_i - \mu_1)(\bar{x} - \mu_1)] = \frac{\sigma^2}{n} - 2M\left[(x_i - \mu_1)\frac{1}{n}\sum_{k=1}^n(x_k - \mu_1)\right].$$

Вычитаемое здесь представляет собой сумму произведений биномов. Математическое ожидание всех этих произведений в силу независимости выборочных значений равно нулю, кроме тех, у которых совпадают индексы. Математическое ожидание таких произведений равно дисперсии σ^2 , а их количество равно 1. Поэтому

$$\text{cov}[(x_i - \bar{x})(x_j - \bar{x})] = \frac{\sigma^2}{n} - \frac{2}{n}\sigma^2 = -\frac{1}{n}\sigma^2.$$

Коэффициент корреляции равен

$$\rho = \frac{\text{cov}[(x_i - \bar{x})(x_j - \bar{x})]}{\sigma_i \sigma_j} = \frac{\text{cov}[(x_i - \bar{x})(x_j - \bar{x})]}{\sigma^2} = -\frac{1}{n}.$$

Видно, что данные биномы коррелированы, и, значит, зависимы, что и требовалось доказать.

Переведем отмеченное обстоятельство на общепринятый язык математической статистики. Принято говорить, что оценка

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_1)^2$$

имеет n степеней свободы, а оценка

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

имеет $n - 1$ степеней свободы.

Этимологию этих понятий можно объяснить следующим образом. Если все выборочные значения независимы, то каждое из них обладает полной “свободой”. На n выборочных значений приходится n таких “свобод”. Столько же “свобод” приходится на n слагаемых первой из приведенных оценок дисперсии.

Как только на слагаемые накладываются какие-либо зависимости, количество “свобод” уменьшается. Во второй оценке степень зависимости между каждыми двумя биномами вида $x_i - \bar{x}$ мы оценили в $1/n$, поэтому в совокупности от всех n таких биномов отнята одна целая “свобода”, и общее количество “свобод” или степеней свободы осталось $n - 1$. Потому в

этой оценке сумма квадратов биномов делится на эквивалентное число оставшихся “свобод”.

Это подтверждается также следующим соображением: по выборке объема $n = 1$, то есть по одному значению (при этом в знаменателе формулы будет 0) невозможно судить о разбросе значений случайных величин.

в) Плотность распределения оценки s^2 .

Будем считать, что выборка извлечена из нормальной генеральной совокупности с математическим ожиданием μ_1 и дисперсией σ^2 .

Рассмотрим вначале первую из оценок, несмещенную при априори известном математическом ожидании μ_1 . Преобразуем ее следующим образом:

$$\zeta = \frac{s^2 n}{\sigma^2} = \sum_{i=1}^n \frac{(x_i - \mu_1)^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - \mu_1}{\sigma} \right)^2.$$

Из разд. 1.6.5. следует, что

$$M \left[\frac{x_i - \mu_1}{\sigma} \right] = 0, \quad D \left[\frac{x_i - \mu_1}{\sigma} \right] = 1.$$

Это значит, что

$$\eta = \frac{x_i - \mu_1}{\sigma} \in N(0,1).$$

В разд. 1.6.7 мы нашли, что при возведении в квадрат нормальной случайной величины с параметрами $(0, 1)$ получается новая случайная величина, плотность распределения и характеристическая функция которой:

$$\varphi_{\eta^2}(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad g_{\eta^2}(v) = (1 - 2jv)^{-1/2}.$$

По свойству характеристических функций (см. разд. 1.7.5) характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых. В нашем случае характеристическая функция случайной величины

$$\zeta = \frac{s^2 n}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - \mu_1}{\sigma} \right)^2$$

равна $g_{\zeta}(v) = (1 - 2jv)^{-n/2}$. Плотность распределения этой случайной величины отыскивается обратным преобразованием характеристической функции и имеет следующий вид:

$$\varphi_{\zeta}(z) = \left(\frac{1}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)} z^{\frac{n}{2}-1} e^{-\frac{z}{2}}.$$

В справедливости этого выражения можно убедиться путем определения характеристической функции этой плотности подобно тому, как это было сделано в примере 1 разд. 1.6.7.

Полное название приведенной плотности: *плотность распределения “хи-квадрат” с n степенями свободы*. Эта плотность распределения имеет всего один параметр n – число степеней свободы.

Принадлежность случайной величины ζ генеральной совокупности с плотностью распределения “хи-квадрат” и числом степеней свободы n будем обозначать так:

$$\zeta \in \chi^2(n).$$

Найдем математическое ожидание и дисперсию случайной величины, распределенной по χ^2 с помощью характеристической функции.

Первая производная от характеристической функции по v

$$\frac{d}{dv}(1 - 2jv)^{-n/2} = nj(1 - 2jv)^{-\frac{n}{2}-1} \Big|_{v=0} = jn = jM\left[\frac{s^2 n}{\sigma^2}\right],$$

откуда

$$M[\zeta] = M\left[\frac{s^2 n}{\sigma^2}\right] = n.$$

Вторая производная

$$\frac{d^2}{dv^2}(1 - 2jv)^{-n/2} = -2n\left(\frac{n}{2} + 1\right)(1 - 2jv)^{-\frac{n}{2}-2} \Big|_{v=0} = -n(n+2) = -\mu_2\left[\frac{s^2 n}{\sigma^2}\right],$$

откуда

$$D[\zeta] = D\left[\frac{s^2 n}{\sigma^2}\right] = \mu_2\left[\frac{s^2 n}{\sigma^2}\right] - \mu_1^2\left[\frac{s^2 n}{\sigma^2}\right] = n(n+2) - n^2 = 2n.$$

В тех случаях, когда оценка дисперсии вычисляется при неизвестном математическом ожидании, количество степеней свободы уменьшается на единицу. Тогда случайная величина ζ распределена в соответствии с плотностью распределения χ^2 с $n - 1$ степенями свободы:

$$\zeta = \frac{s^2(n-1)}{\sigma^2} \in \chi^2(n-1).$$

Математическое ожидание и дисперсия этой случайной величины

$$M[\zeta] = M\left[\frac{s^2(n-1)}{\sigma^2}\right] = n-1, \quad D[\zeta] = D\left[\frac{s^2(n-1)}{\sigma^2}\right] = 2(n-1).$$

Плотность распределения имеет вид

$$\varphi_{\zeta}(z) = \left(\frac{1}{2}\right)^{\frac{n-1}{2}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} z^{\frac{n-3}{2}} e^{-\frac{z}{2}}.$$

Характерная особенность плотности распределения χ^2 : дисперсия равна удвоенному математическому ожиданию.

Плотность распределения χ^2 безгранично делима. Пусть $\zeta_1 \in \chi^2(n_1)$ с характеристической функцией $g_{\zeta_1}(v) = (1 - 2jv)^{-n_1/2}$ и $\zeta_2 \in \chi^2(n_2)$ с характеристической функцией $g_{\zeta_2}(v) = (1 - 2jv)^{-n_2/2}$. Тогда характеристическая функция случайной величины $\zeta = \zeta_1 + \zeta_2$ есть $g_{\zeta}(v) = (1 - 2jv)^{-n_1/2} (1 - 2jv)^{-n_2/2} = (1 - 2jv)^{-(n_1+n_2)/2}$, то есть это также характеристическая функция распределения χ^2 с $n_1 + n_2$ степенями свободы, а это значит, что $\zeta_1 + \zeta_2 \in \chi^2(n_1 + n_2)$.

Еще раз напомним, что распределение χ^2 получено нами при условии, что выборка изъята из нормальной генеральной совокупности.

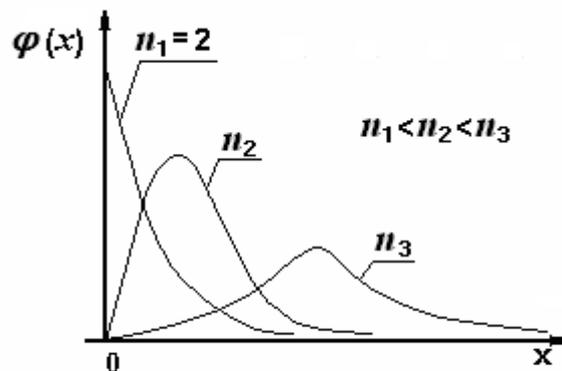


Рис. 28. Плотности распределения χ^2

В свою очередь, *распределение χ^2 асимптотически нормально.*

2.3.4.3. Свойства оценки математического ожидания случайного вектора

Из многомерной генеральной совокупности \mathbf{X} , образованной случайным вектором $\xi^T = (\xi_1 \ \xi_2 \ \dots \ \xi_k)$ с математическим ожиданием $\mathbf{c}^T = (c_1 \ c_2 \ \dots \ c_k)$ и ковариационной матрицей Σ_ξ извлечена выборка векторов $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, которые попарно независимы в соответствии с исходным требованием математической статистики об обеспечении независимости выборочных данных. Несмещенной состоятельной оценкой математического ожидания c_j каждой компоненты случайного вектора ξ_j , как следует из разд. 2.3.4.1, является среднее арифметическое значение, вычисленное по выборочным значениям соответствующих компонент выборочных векторов:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ji}, \quad j = 1, 2, \dots, k.$$

Значит, несмещенной состоятельной оценкой математического ожидания является вектор, составленный из таких компонент:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i.$$

Ковариационная матрица этого вектора

$$\Sigma_{\bar{\mathbf{x}}} = \frac{1}{n} \Sigma_\xi.$$

Это следует хотя бы из того, что каждый j -й диагональный элемент матрицы $\Sigma_{\bar{\mathbf{x}}}$, который является дисперсией $\sigma_{\bar{x}_j}^2$ каждой j -й компоненты вектора средних арифметических $\bar{\mathbf{x}}$, в n раз меньше дисперсии j -го компонента случайного вектора ξ , то есть j -го диагонального элемента ковариационной матрицы Σ_ξ (см. разд. 2.3.4.1):

$$\sigma_{\bar{x}_j}^2 = \frac{1}{n} \sigma_{\xi_j}^2.$$

Точно так же пропорционально изменяются и остальные элементы матрицы Σ_{ξ} , поэтому

$$\Sigma_{\bar{x}} = \frac{1}{n} \Sigma_{\xi}.$$

2.3.4.4. Оценка ковариационной матрицы случайного вектора

Из многомерной генеральной совокупности \mathbf{X} , образованной случайным вектором $\xi^T = (\xi_1 \ \xi_2 \ \dots \ \xi_k)$ с математическим ожиданием $\mathbf{c}^T = (c_1 \ c_2 \ \dots \ c_k)$ и ковариационной матрицей Σ_{ξ} , извлечена выборка векторов $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, которые попарно независимы в соответствии с исходным требованием математической статистики об обеспечении независимости выборочных данных.

Оценкой ковариационной матрицы Σ_{ξ} является матрица \mathbf{S}_{ξ} той же размерности, диагональные элементы которой – оценки s_j^2 диагональных элементов матрицы Σ_{ξ} , то есть дисперсий σ_j^2 j -х компонент случайного вектора ξ . Желательно получить несмещенные оценки этих дисперсий, как и всех остальных элементов матрицы Σ_{ξ} , то есть ковариаций.

Вычисление оценки ковариационной матрицы выполняется в соответствии с ее определением (см. разд. 1.8), где вместо символа математического ожидания, как и ранее, используются суммирование и усреднение. Поэтому несмещенная оценка ковариационной матрицы, все элементы которой должны быть несмещенными оценками элементов матрицы Σ_{ξ} , вычисляется по формуле

$$\mathbf{S}_{\xi} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

В знаменателе этой формулы n уменьшается на единицу, как и в разд. 2.3.4.2, поскольку каждый элемент матрицы \mathbf{S}_{ξ} “истратил” по одной степени свободы из-за предварительного вычисления среднего арифметического значения соответствующей компоненты. Это понятно и из такого соображения: по выборке объема $n = 1$ (при этом в знаменателе формулы будет 0) невозможно судить ни о разбросе значений случайных величин,

ни о степени их коррелированности. Кроме того, при оценивании ковариационной матрицы необходимо обеспечить, чтобы объем n выборки был не меньше размерности вектора ξ или матрицы S_ξ , то есть чтобы выполнялось неравенство $n \geq k + 1$, иначе определитель матрицы S_ξ будет равен 0, а сама матрица окажется особенной.

2.3.5. Метод максимального правдоподобия

Точечные оценки, полученные *методом максимального правдоподобия* (ММП-оценки), являются эффективными оценками, хотя и могут быть смещенными. Для получения ММП-оценок необходима априорная информация о плотности распределения.

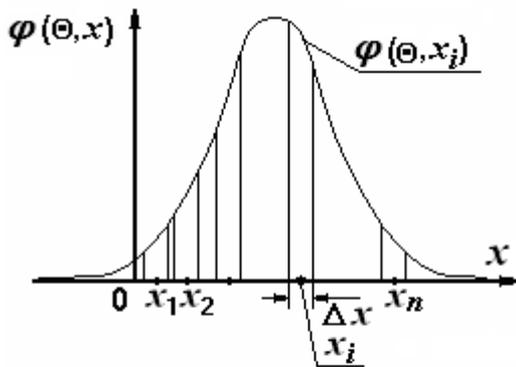


Рис. 29. К выводу функции правдоподобия

Итак, пусть плотность распределения $\varphi_\xi(\Theta, x)$ генеральной совокупности \mathbf{X} известна с точностью до параметров Θ , подлежащих оцениванию. Из этой генеральной совокупности извлечена выборка

$$\mathbf{x}^T = (x_1 \quad x_2 \quad \dots \quad x_n).$$

Образуем около каждого выборочного значения x_i *непересекающиеся* интервалы J_i равной ширины, такие, чтобы каждое выборочное значение находилось в середине своего интервала (рис. 29). Обозначим ширину этих интервалов Δx . Рассчитаем вероятность того, что каждое из выборочных значений попадет в свой интервал:

$$P(x_1 \in J_1 \cap x_2 \in J_2 \cap \dots \cap x_n \in J_n) = \prod_{i=1}^n \varphi(\Theta, x_i) \Delta x.$$

В соответствии с принципом максимального правдоподобия в качестве оценок параметров Θ предлагается найти такие значения параметров, при которых имеющаяся выборка оказывается наиболее правдоподобной или, что то же самое, наиболее вероятной. Для этого необходимо найти

максимум указанной выше вероятности. Предварительно сделаем два упрощения.

Первое – отбрасываем постоянные множители Δx , как не влияющие на положение максимума.

Второе – будем отыскивать максимум логарифма этой вероятности, ибо логарифмирование функции по основанию, большему единицы, не влияет на положение ее максимума, но в большинстве случаев приводит к значительному упрощению выкладок.

В результате этих действий получим *логарифмическую функцию правдоподобия*:

$$L(\Theta, \mathbf{x}) = \ln \left[\prod_{i=1}^n \varphi(\Theta, x_i) \right] = \sum_{i=1}^n \ln[\varphi(\Theta, x_i)].$$

ММП-оценки параметров Θ находят путем поиска максимума функции правдоподобия

$$\hat{\Theta} = \operatorname{argmax}_{\Theta} \sum_{i=1}^n \ln[\varphi(\Theta, x_i)].$$

Пример 1. Случайная величина ξ распределена нормально: $\xi \in N(a, \sigma)$.

Требуется найти ММП-оценку математического ожидания случайной величины ξ по выборочным значениям $\mathbf{x}^T = (x_1 \ x_2 \ \dots \ x_n)$.

Функция правдоподобия для поставленной задачи

$$L(a, \sigma, \mathbf{x}) = \sum_{i=1}^n \ln \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x_i - a)^2}{2\sigma^2} \right] \right\} = -n \ln[\sqrt{2\pi}\sigma] - \frac{1}{2} \sum_{i=1}^n \frac{(x_i - a)^2}{\sigma^2}.$$

Будем отыскивать максимум функции правдоподобия с помощью производной по a . Первое слагаемое не зависит от a , поэтому его учитывать не будем. Остается найти максимум отрицательного выражения, который достигается там, где достигается минимум его абсолютной величины. Поэтому ММП-оценка математического ожидания есть значение, при котором достигается максимум функции правдоподобия:

$$\hat{a} = \operatorname{argmax}_a L(a, \sigma, \bar{x}) = \operatorname{argmin}_a \sum_{i=1}^n (x_i - a)^2.$$

Продифференцируем минимизируемое выражение по a и приравняем производную нулю:

$$\frac{\partial}{\partial a} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - a)^2 = - \frac{2}{\sigma^2} \sum_{i=1}^n (x_i - \hat{a}) = - \frac{2}{\sigma^2} \left[\sum_{i=1}^n x_i - n \hat{a} \right] = 0,$$

откуда сразу получаем ММП-оценку математического ожидания

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Из этого факта с необходимостью следует эффективность среднего арифметического как оценки математического ожидания по выборке, извлеченной из нормальной генеральной совокупности. Об этом уже было упомянуто выше в разд. 2.3.4.1.

Пример 2. Измерения одной и той же величины, значение которой равно a , выполняются различными средствами измерений, обладающими различными случайными погрешностями, среднеквадратические значения которых $\sigma_1^2, \sigma_2^2, \dots, \sigma_i^2, \dots, \sigma_n^2$. Измерения, выполняемые каждым средством измерений, однократные, погрешности распределены нормально, систематические составляющие погрешностей отсутствуют.

Требуется найти ММП-оценку значения измеряемой величины.

Подобная ситуация в математической статистике классифицируется как случай *неравноточных измерений*, в отличие от примера 1, который классифицируется как случай *равноточных измерений*.

В этой ситуации каждое выборочное значение извлекается из своей генеральной совокупности, то есть $x_i \in N(a, \sigma_i^2)$. Функция правдоподобия теперь будет иметь иной вид:

$$L(a, \sigma, \mathbf{x}) = \sum_{i=1}^n \ln \left\{ \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[- \frac{(x_i - a)^2}{2\sigma_i^2} \right] \right\} = -n \sum_{i=1}^n \ln[\sqrt{2\pi}\sigma_i] - \frac{1}{2} \sum_{i=1}^n \frac{(x_i - a)^2}{\sigma_i^2}.$$

Дифференцируем функцию правдоподобия по a и приравниваем производную нулю:

$$\frac{\partial}{\partial a} \sum_{i=1}^n \frac{(x_i - a)^2}{2\sigma_i^2} = - \sum_{i=1}^n \frac{(x_i - \hat{a})}{\sigma_i^2} = \left[\sum_{i=1}^n \frac{x_i}{\sigma_i^2} - \sum_{i=1}^n \frac{\hat{a}}{\sigma_i^2} \right] = 0,$$

откуда получаем ММП-оценку математического ожидания при неравно- точных измерениях:

$$\hat{a} = \sum_{i=1}^n w_i x_i,$$

где веса w_i обратно пропорциональны дисперсиям случайных погрешно- стей измерений и в сумме составляют единицу:

$$w_i = \frac{1}{\sigma_i^2} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^{-1}, \quad \sum_{i=1}^n w_i = \sum_{i=1}^n \frac{1}{\sigma_i^2} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^{-1} = 1.$$

Пример 3. Выборка $\mathbf{x}^T = (x_1 \ x_2 \ \dots \ x_n)$ извлечена из генераль- ной совокупности, распределенной по Лапласу. Плотность распределения

$$\varphi(x) = \frac{\lambda}{2} e^{-\lambda|x-a|}.$$

Необходимо найти ММП-оценки параметров a (математическое ожи- дание) и λ .

Функция правдоподобия имеет вид

$$L(a, \lambda, \mathbf{x}) = \sum_{i=1}^n \ln(\lambda / 2) - \lambda \sum_{i=1}^n |x_i - a|.$$

Найдем вначале ММП- оценку математического ожида- ния a . Поскольку первое слагае- мое не содержит a , его значение не влияет на положение макси- мума, поэтому мы его исключим. Останется одно отрицательное слагаемое, минимум модуля ко- торого совпадает по положению с максимумом функции правдоподобия. Поэтому будем отыскивать ММП-

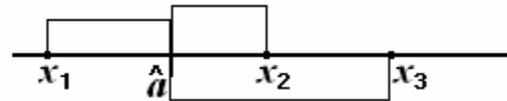


Рис. 30. Определение ММП-оценки математического ожидания случайной величины, распределенной по Лапласу

оценку \hat{a} путем поиска минимума суммы $\sum_{i=1}^n |x_i - a|$. Эта сумма недиффе- ренцируема, и нам придется находить искомую оценку с привлечением геометрических построений.

Пусть в результате эксперимента получено всего три выборочных значения x_1, x_2, x_3 , которые разместились на вещественной оси так, как показано на рис. 30.

Каждое слагаемое минимизируемой суммы есть расстояние от каждого выборочного значения до математического ожидания. В представленной ситуации (см. рис. 30), на рисунке, одно из этих расстояний входит в сумму дважды. Это расстояние $|x_2 - \hat{a}|$. Видно, что рассматриваемая сумма достигнет минимума только тогда, когда оценка \hat{a} совместится с x_2 , то есть с выборочной медианой. Добавляя к этой небольшой выборке четное количество элементов, мы приходим к тому же выводу, что ММП-оценкой математического ожидания случайной величины, распределенной по Лапласу, является выборочная медиана

$$\hat{a} = x_{\text{med}}.$$

А это означает, что эта оценка эффективна, о чем было упомянуто в разд. 2.3.4.1.

Теперь найдем ММП-оценку параметра λ .

$$\frac{\partial}{\partial \lambda} L(\hat{a}, \lambda, \mathbf{x}) = \frac{n}{\lambda} - \sum_{i=1}^n |x_i - \hat{a}| = 0 \quad \Rightarrow \quad \frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{a}|.$$

Отсюда следует, что

$$\hat{\lambda} = n \left(\sum_{i=1}^n |x_i - \hat{a}| \right)^{-1}.$$

2.3.6. Метод минимума χ^2

Как и в разд. 2.3.5, ставится задача оценки параметров Θ плотности распределения $\varphi(\Theta, x)$, вид которой известен. Исходными данными являются выборочные значения, по которым строится гистограмма (см. разд. 2.2, рис. 27). Идея метода заключается в подборе таких значений искомых параметров, при которых достигается минимальное отличие кривой плотности распределения от гистограммы. В качестве меры этого отличия чаще всего используется квадратичный функционал, наиболее удобный для реализации аналитических и численных методов поиска экстремума (минимума или максимума).

В нашей задаче в качестве такого функционала используется сумма, обозначенная здесь и далее χ^2 .

$$\chi^2 = \sum_{k=1}^K \frac{n}{P_k} \left(P_k - \frac{n_k}{n} \right)^2 = \frac{1}{n} \sum_{k=1}^K \frac{(nP_k - n_k)^2}{P_k}.$$

Причина такого обозначения станет ясна в дальнейшем из материала разд. 2.5.5.1.

В этой сумме K – общее количество интервалов, на которых построена гистограмма, n – объем выборки, n_k – количество выборочных значений, попавших в k -й интервал гистограммы, P_k – вероятностная мера k -го интервала гистограммы:

$$P_k = \int_{x_{k-1}}^{x_k} \varphi(\vec{\Theta}, x) dx.$$

(эти обозначения иллюстрируются рис. 38 п. 2.5.5.1.)

Таким образом, слагаемые этой суммы представляют собой квадраты разностей между вероятностными мерами k -х интервалов, порожденными генеральной плотностью распределения, и частотными оценками $\tilde{P}_k = n_k / n$ этих вероятностных мер. Знаменателем каждого слагаемого является вероятность P_k , благодаря чему в процессе поиска значений параметров Θ повышается вес интервалов с низкой вероятностью и тем самым обеспечивается повышенная точность подгонки в области этих интервалов. Как правило, эти интервалы находятся на удалении от центра распределения (на так называемых “хвостах” распределений).

Вероятности P_k – функции искомым параметров Θ , от тех же параметров зависит и величина χ^2 , и процедура оценивания параметров по методу минимума χ^2 формально записывается в виде

$$\tilde{\Theta} = \operatorname{argmin}_{\Theta} \chi^2 = \operatorname{argmin}_{\Theta} \left[\sum_{k=1}^K \frac{(nP_k(\Theta) - n_k)^2}{nP_k(\Theta)} \right].$$

В большинстве случаев этот минимум и оценки находятся численными методами. Доказано (см., например, [4] стр. 461, 549 и [5], стр. 303), что оценки, полученные методом минимума χ^2 , обладают свойствами, со-

поставимыми со свойствами ММП-оценок, а именно, эти оценки асимптотически эффективны.

2.3.7. Оценивание коэффициентов аппроксимирующих полиномов

2.3.7.1. Формулировка задачи

Будем рассматривать следующую ситуацию. Объективно существует функция $y = f(x)$, ограниченная и дифференцируемая не менее $q + 1$ раз на интервале $[x_{\min}, x_{\max}]$. Природа и происхождение этой функции могут быть различными:

этой функцией связаны между собой естественные параметры и(или) явления в природе, в обществе, в экономике и т.п.;

этой функцией описывается преобразование физических величин, происходящее в технических устройствах, таких, как регуляторы, датчики, измерительные преобразователи, устройства телекоммуникаций и т.п.;

этой функцией описываются взаимосвязи параметров технических объектов, в том числе технологических процессов, в различных режимах (штатный режим работы, испытания, нештатные режимы);

этой функцией, по мнению исследователя, описываются объекты, явления, процессы, которые он моделирует на компьютере.

По теореме Вейерштрасса, ограниченные и $q + 1$ раз дифференцируемые в интервале $[x_{\min}, x_{\max}]$ функции могут быть сколь угодно точно аппроксимированы в этом интервале степенным (и даже обобщенным) полиномом степени q .

Понятно, что в реальном исследовании сколь угодно высокая точность достигнута быть не может, хотя стремление к максимально достижимой точности у каждого исследователя имеется. Пусть это естественное стремление выражается следующим образом.

Желательно аппроксимировать реальную функцию $y = f(x)$ полиномом степени q так, чтобы максимальное расхождение между реальной функцией и этим полиномом не превосходило пренебрежимо малой, с точки зрения исследователя, величины $\delta > 0$:

$$\left| a_0 + a_1x + a_2x^2 + \dots + a_qx^q - f(x) \right| \leq \delta.$$

Будем называть этот полином степени q “точным”.

Поскольку, по теореме Вейерштрасса, такой полином существует при любом сколь угодно малом значении δ , будем считать коэффициенты a_0, a_1, \dots, a_q “истинными” и для их обозначения введем вектор этих коэффициентов $\mathbf{a}^T = (a_0 \ a_1 \ \dots \ a_q)$. Степень полинома q будем считать известной. В этом случае говорят: “*модель объекта известна с точностью до параметров*”. Объектом для нас является полином, аппроксимирующий функцию $y = f(x)$.

Теперь задачей исследователя является организация такого эксперимента, в результате которого он смог бы определить значения этих коэффициентов. Необходимыми условиями выполнения такого эксперимента являются

воспроизведение с необходимой точностью заданных значений аргумента x_1, x_2, \dots, x_k , $k \geq q + 1$, в диапазоне $[x_{\min}, x_{\max}]$, или, по крайней мере, фиксация фактически реализующихся этих значений с помощью измерений с заданной или хотя бы с известной точностью;

измерение с необходимой или хотя бы с известной точностью значений функции $y_1 = f(x_1), y_2 = f(x_2), \dots, y_k = f(x_k)$ при всех заданных (зафиксированных) значениях аргумента.

Пример графического представления результатов подобного эксперимента приведен на рис. 31. Непрерывной кривой изображен график исследуемой функции $y = f(x)$, точки на этой кривой – суть значения

$$y_1 = f(x_1),$$

$$y_2 = f(x_2), \dots, y_k = f(x_k).$$

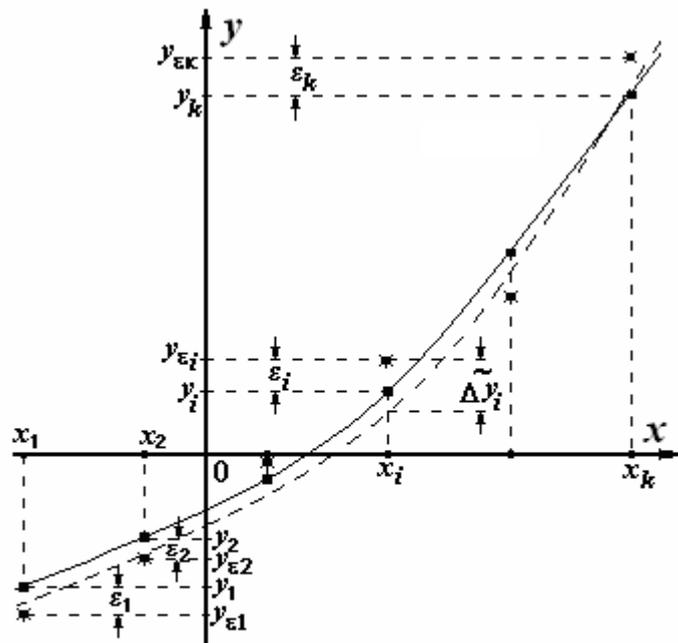


Рис. 31. Графики полиномов и экспериментальные точки

Для обозначения всей совокупности этих значений введем вектор \mathbf{y} :
 $\mathbf{y}^T = (y_1 \ y_2 \ \dots \ y_k)$. Точки вне этой кривой – результаты измерений, для обозначения которых введем вектор \mathbf{y}_ε :

$$\mathbf{y}_\varepsilon^T = (y_{\varepsilon 1} \ y_{\varepsilon 2} \ \dots \ y_{\varepsilon k}).$$

Будем считать, что воспроизведение (измерения) значений аргумента x_1, x_2, \dots, x_k , выполняются с настолько высокой точностью, что погрешностью результатов можно пренебречь.

Будем также считать, что погрешности $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k$ измерения значений функции суть компоненты случайного вектора $\boldsymbol{\varepsilon}$, не содержащие систематических составляющих (математическое ожидание всех компонент равно нулю), вектор $\boldsymbol{\varepsilon}$ распределен в соответствии с нормальным законом: $\boldsymbol{\varepsilon} \in N_k(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, где $\boldsymbol{\Sigma}_\varepsilon$ – его ковариационная матрица. Диагональными элементами ковариационной матрицы $\boldsymbol{\Sigma}_\varepsilon$ являются дисперсии погрешностей измерений $\sigma_{\varepsilon i}^2$. Результаты измерений образуют в совокупности случайный вектор $\mathbf{y}_\varepsilon = \mathbf{y} + \boldsymbol{\varepsilon}$, который распределен нормально $\mathbf{y}_\varepsilon \in N_k(\mathbf{y}, \boldsymbol{\Sigma}_\varepsilon)$. В случае независимости измерений ковариационная матрица $\boldsymbol{\Sigma}_\varepsilon$ диагональна.

Задача состоит в том, чтобы выполнить *полиномиальную аппроксимацию* исследуемой функции $y = f(x)$, то есть найти оценки коэффициентов $\mathbf{a}^T = (a_0 \ a_1 \ \dots \ a_q)$ полинома, аппроксимирующего эту функцию.

Измерения значений функции при каждом значении аргумента могут быть однократными или многократными.

Рассмотрим вначале случай однократных измерений, которым можно ограничиться только, если ковариационная матрица $\boldsymbol{\Sigma}_\varepsilon$ известна априори.

2.3.7.2. Измерения однократные

В соответствии с формулировкой задачи (разд. 2.3.7.1) в результате эксперимента при фиксированных значениях x_1, x_2, \dots, x_k мы получаем значения $y_{\varepsilon 1}, y_{\varepsilon 2}, \dots, y_{\varepsilon k}$, $k \geq q + 1$:

$$\begin{cases} y_{\varepsilon 1} = a_0 + a_1 x_1 + a_2 x_1^2 + \dots + a_q x_1^q + \varepsilon_1 \\ y_{\varepsilon 2} = a_0 + a_1 x_2 + a_2 x_2^2 + \dots + a_q x_2^q + \varepsilon_2 \\ \vdots \\ y_{\varepsilon k} = a_0 + a_1 x_k + a_2 x_k^2 + \dots + a_q x_k^q + \varepsilon_k \end{cases}$$

Эти значения представлены точками (см. рис. 31), лежащими вне кривых.

Приведенная система равенств есть система k уравнений, из которой нам необходимо получить оценки $q + 1$ коэффициентов a_0, a_1, \dots, a_q .

Для того чтобы эту систему записать в матричном виде, введем матрицу

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^q \\ 1 & x_2 & x_2^2 & \dots & x_2^q \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_k & x_k^2 & \dots & x_k^q \end{pmatrix}.$$

Тогда система уравнений записывается в виде

$$\mathbf{y}_{\varepsilon} = \mathbf{X}\mathbf{a} + \boldsymbol{\varepsilon},$$

где векторы \mathbf{a} , $\boldsymbol{\varepsilon}$, \mathbf{y}_{ε} определены в п. 2.3.7.1 и

$$\boldsymbol{\varepsilon} \in N_k(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}), \quad \mathbf{y}_{\varepsilon} \in N_k(\mathbf{X}\mathbf{a}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}).$$

Будем находить ММП-оценки неизвестных коэффициентов полинома. Для этого запишем k -мерную плотность распределения вектора \mathbf{y}_{ε} :

$$\varphi(\mathbf{y}_{\varepsilon}) = \frac{1}{(2\pi)^{k/2} |\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}|^{1/2}} \exp\left\{-\frac{1}{2} (\mathbf{X}\mathbf{a} - \mathbf{y}_{\varepsilon})^T \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} (\mathbf{X}\mathbf{a} - \mathbf{y}_{\varepsilon})\right\}.$$

Функция правдоподобия в этом случае

$$L(\mathbf{a}, \mathbf{y}_{\varepsilon}) = -\ln \sqrt{(2\pi)^k |\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}|} - \frac{1}{2} [(\mathbf{X}\mathbf{a} - \mathbf{y}_{\varepsilon})^T \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} (\mathbf{X}\mathbf{a} - \mathbf{y}_{\varepsilon})].$$

Максимум функции правдоподобия находится там же, где находится минимум квадратичной формы $(\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)^T \boldsymbol{\Sigma}_\varepsilon^{-1} (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)$:

$$\operatorname{argmax}_{\mathbf{a}} [L(\mathbf{a}, \mathbf{y}_\varepsilon)] = \operatorname{argmin}_{\mathbf{a}} \left[(\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)^T \boldsymbol{\Sigma}_\varepsilon^{-1} (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon) \right],$$

поэтому для нахождения ММП-оценок искоемых коэффициентов будем их отыскивать путем минимизации указанной квадратичной формы. С этой целью продифференцируем ее по вектору \mathbf{a} и приравняем производную нулю. Напомним предварительно правила дифференцирования по вектору (см., например, [5], стр. 73):

$$\frac{d}{d\mathbf{u}} (\mathbf{B}\mathbf{u}) = \mathbf{B}^T, \quad \frac{d}{d\mathbf{u}} (\mathbf{u}^T \mathbf{B}) = \mathbf{B}.$$

Пользуясь этими правилами после раскрытия скобок, получим

$$\begin{aligned} \frac{d}{d\mathbf{a}} (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)^T \boldsymbol{\Sigma}_\varepsilon^{-1} (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon) &= \\ &= \frac{d}{d\mathbf{a}} \left[\mathbf{a}^T \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X}\mathbf{a} + \mathbf{y}_\varepsilon^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{y}_\varepsilon - \mathbf{a}^T \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{y}_\varepsilon - \mathbf{y}_\varepsilon^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X}\mathbf{a} \right] = \\ &= 2 \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X}\mathbf{a} - 2 \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{y}_\varepsilon = 0, \end{aligned}$$

откуда находим вектор ММП-оценок коэффициентов аппроксимирующего полинома:

$$\hat{\mathbf{a}} = (\mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{y}_\varepsilon = \mathbf{D} \mathbf{y}_\varepsilon,$$

где $\mathbf{D} = (\mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1}$.

Очевидно, что эта оценка линейно зависит от результатов измерений, а размер матрицы \mathbf{D} равен $(q + 1)k$.

Являясь ММП-оценкой, полученный вектор есть эффективная оценка вектора коэффициентов полинома. Проверим ее несмещенность:

$$\begin{aligned} M[\hat{\mathbf{a}}] &= M \left[(\mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{y}_\varepsilon \right] = (\mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} M[\mathbf{y}_\varepsilon] = \\ &= (\mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X}\mathbf{a} = \mathbf{a}, \end{aligned}$$

поскольку произведение взаимно обратных матриц есть единичная матрица. Несмещенность полученной оценки доказана.

Как показано в разд.3.7.1, ковариационная матрица вектора \mathbf{y}_ε есть Σ_ε . Тогда, поскольку $\hat{\mathbf{a}} = \mathbf{D}\mathbf{y}_\varepsilon$, то в соответствии с разд. 1.7.4 ковариационная матрица вектора оценок коэффициентов

$$\Sigma_{\hat{\mathbf{a}}} = \mathbf{D}\Sigma_\varepsilon\mathbf{D}^\top = (\mathbf{X}^\top \Sigma_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Sigma_\varepsilon^{-1} \Sigma_\varepsilon \Sigma_\varepsilon^{-1} \mathbf{X} (\mathbf{X}^\top \Sigma_\varepsilon^{-1} \mathbf{X})^{-1}.$$

Производя перемножение ряда взаимно обратных матриц, находящихся в середине правой части, окончательно получим:

$$\Sigma_{\hat{\mathbf{a}}} = (\mathbf{X}^\top \Sigma_\varepsilon^{-1} \mathbf{X})^{-1}.$$

Напомним также, что при линейном преобразовании случайных величин вид плотности распределения не изменяется (разд. 1.6.7). Поэтому в связи с обнаруженной нами несмещенностью оценки

$$\hat{\mathbf{a}} \in N_{q+1} \left(\mathbf{a}, (\mathbf{X}^\top \Sigma_\varepsilon^{-1} \mathbf{X})^{-1} \right).$$

Если измерения равноточные, то есть при всех $i = 1, 2, \dots, k$ $\sigma_{\varepsilon i}^2 = \sigma_\varepsilon^2$, а компоненты вектора ε независимы, то ковариационная матрица $\Sigma_\varepsilon = \sigma_\varepsilon^2 \mathbf{E}$, где \mathbf{E} – единичная матрица, $\Sigma_\varepsilon^{-1} = \sigma_\varepsilon^{-2} \mathbf{E}$. В этих условиях квадратичная форма, подлежащая минимизации, принимает вид

$$\sigma_\varepsilon^{-2} (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)^\top (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon).$$

В результате минимизации получаем решение и его характеристики

$$\hat{\mathbf{a}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}_\varepsilon, \quad \Sigma_{\hat{\mathbf{a}}} = \sigma_\varepsilon^2 (\mathbf{X}^\top \mathbf{X})^{-1}, \quad \hat{\mathbf{a}} \in N_{q+1} \left(\mathbf{a}, \sigma_\varepsilon^2 (\mathbf{X}^\top \mathbf{X})^{-1} \right).$$

Компонентами вектора $(\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)$ являются разности

$$\Delta y_i = a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_q x_i^q - y_{\varepsilon i}, \quad i = 1, 2, \dots, k,$$

и минимизируемая квадратичная форма представляет собой сумму квадратов этих разностей:

$$(\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)^\top (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon) = \sum_{i=1}^k (\Delta y_i)^2 = \sum_{i=1}^k \left(a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_q x_i^q - y_{\varepsilon i} \right)^2.$$

По этой причине приведенный метод определения коэффициентов аппроксимирующих полиномов называется *методом наименьших квадратов* (МНК), а получаемые этим методом оценки коэффициентов полиномов называются *МНК-оценками*.

Общий метод оценивания коэффициентов аппроксимирующих полиномов путем минимизации квадратичной формы $(\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)^T \Sigma_\varepsilon^{-1} (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)$ называется *обобщенным методом наименьших квадратов* (ОМНК), а оценки, вычисляемые по формуле $\hat{\mathbf{a}} = (\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{y}_\varepsilon$, – *ОМНК-оценками*.

Конечным итогом и целью оценивания является полином, коэффициентами которого являются найденные оценки:

$$\hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2 + \dots + \hat{a}_q x^q.$$

При значениях аргумента x_1, x_2, \dots, x_k этот полином принимает значения, которые суть компоненты вектора $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{a}}$. График этого полинома представлен на рис. 31 пунктирной линией. В точке $x = x_i$ показана разность $\Delta\hat{y}_i = (\hat{a}_0 + \hat{a}_1 x_i + \hat{a}_2 x_i^2 + \dots + \hat{a}_q x_i^q) - y_{\varepsilon i}$ между значением в этой точке построенного полинома и результатом измерений. Все эти разности в совокупности для $i = 1, 2, \dots, k$ составляют вектор

$$\Delta\hat{\mathbf{y}}^T = (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon)^T = (\Delta\hat{y}_1 \quad \Delta\hat{y}_2 \quad \dots \quad \Delta\hat{y}_k).$$

Итак, в результате выполненных операций мы определили, что квадратичная форма $(\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)^T \Sigma_\varepsilon^{-1} (\mathbf{X}\mathbf{a} - \mathbf{y}_\varepsilon)$ принимает минимальное значение при $\hat{\mathbf{a}} = (\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{y}_\varepsilon$. Обозначим это минимальное значение R^2 . При ОМНК

$$R^2 = (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon)^T \Sigma_\varepsilon^{-1} (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon).$$

Применительно к МНК, когда $\Sigma_\varepsilon = \sigma_\varepsilon^2 \mathbf{E}$,

$$R^2 = \sigma_\varepsilon^{-2} (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon)^T (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon).$$

В обоих вариантах величина R^2 случайна, поскольку зависит от выборочных данных. Естественно выяснить плотность распределения величины R^2 . Этот вопрос мы рассмотрим отдельно в следующем разделе.

2.3.7.3. Плотность распределения величины R^2

Начнем с рассмотрения величины

$$R^2 = \sigma_\varepsilon^{-2} (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon)^T (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon),$$

которая вычисляется в рамках применения МНК, когда для получения оценок коэффициентов \mathbf{a} достаточно знать лишь о факте равноточности измерений, а значение дисперсии σ_ε^2 погрешностей измерений может быть неизвестным.

В выражении для R^2 \mathbf{y}_ε – вектор выборочных значений, $M[\mathbf{y}_\varepsilon] = \mathbf{X}\mathbf{a}$.

Вектор $\mathbf{X}\hat{\mathbf{a}}$ – несмещенная ММП – оценка своего математического ожидания $M[\mathbf{X}\hat{\mathbf{a}}] = \mathbf{X}M[\hat{\mathbf{a}}] = \mathbf{X}\mathbf{a}$. Поэтому $M[(\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon)] = 0$. В соответствии с МНК и постановкой задачи в разд. 2.3.7.1 компоненты ε_i вектора случайных погрешностей $\boldsymbol{\varepsilon}$ распределены нормально с одинаковыми параметрами $M[\varepsilon_i] = 0$, $D[\varepsilon_i] = \sigma_\varepsilon^2$. Поэтому компоненты вектора $\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon$, то есть разности

$$\Delta \hat{y}_i = (\hat{a}_0 + \hat{a}_1 x_i + \hat{a}_2 x_i^2 + \dots + \hat{a}_q x_i^q) - y_{\varepsilon i}$$

При равноточных измерениях выборочные значения погрешностей разумно считать изъятыми из одной нормальной генеральной совокупности $\Delta \hat{y}_{\varepsilon i} \in N(0, \sigma_\varepsilon^2)$. Тогда реализуется МНК, и значение R^2 представляет собой сумму

$$R^2 = \sigma_\varepsilon^{-2} (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon)^T (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_\varepsilon) = \sum_{i=1}^k \left(\frac{\Delta \hat{y}_i}{\sigma_\varepsilon} \right)^2,$$

где дробь $\frac{\Delta \hat{y}_i}{\sigma_\varepsilon} \in N(0, 1)$, а сумма квадратов этих дробей есть не что иное,

как сумма квадратов нормальных случайных величин с нулевым математическим ожиданием и единичной дисперсией.

Такая сумма распределена по закону χ^2 с числом степеней свободы, равным количеству слагаемых k , уменьшенному на количество связей между выборочными значениями. В данном случае количество таких связей равно количеству уравнений, из которых получены оценки коэффици-

ентов \mathbf{a} , то есть $q + 1$. Поэтому число степеней свободы $k - q - 1$. На этом основании заключаем, что

$$R^2 = \sigma_{\varepsilon}^{-2} (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_{\varepsilon})^T (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_{\varepsilon}) \in \chi^2(k - q - 1).$$

Ранее, в одномерном случае в разд. 2.3.4.2. в) было обнаружено, что распределению χ^2 с $n - 1$ степенью свободы подчиняется величина $(n - 1)s^2 / \sigma^2$. Это также сумма

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

Поэтому по аналогии с величиной R^2 мы можем заключить, что при однократных равноточных измерениях несмещенной оценкой дисперсии погрешностей этих измерений может служить величина

$$s^2 = \frac{1}{k - q - 1} \sum_{i=1}^k (\Delta \hat{y}_i)^2 = \frac{1}{k - q - 1} (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_{\varepsilon})^T (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_{\varepsilon})$$

При $k = q + 1$ это выражение теряет смысл. Подобная ситуация возникает в одномерном случае, когда математическое ожидание исследуемой случайной величины неизвестно и выполняется только одно измерение, то есть $n = 1$. Тогда оценить характеристику разброса, каковой является дисперсия, принципиально невозможно (см. также разд. 2.3.4.2. б). Поэтому при организации эксперимента необходимо обеспечивать $k > q + 1$. При применения ОМНК также

$$R^2 = (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_{\varepsilon})^T \Sigma_{\varepsilon}^{-1} (\mathbf{X}\hat{\mathbf{a}} - \mathbf{y}_{\varepsilon}) \in \chi^2(k - q - 1).$$

По материалам настоящего раздела см. также [5, 7].

2.3.7.4. Практически важные замечания

З а м е ч а н и е 1. При однократных измерениях применение ОМНК затруднено тем, что ковариационная матрица в большинстве случаев неизвестна. Реально известными могут быть только дисперсии $\sigma_{\varepsilon_i}^2$ случайных погрешностей ε_i из нормативной документации на средство измерений, применяемое для измерения значений функции $y = f(x)$. В этой ситуации ковариационная матрица вынужденно оказывается неполной: в ней оста-

ются только диагональные элементы $\sigma_{\varepsilon_i}^2$, а внедиагональные элементы, которые характеризуют степень корреляции между результатами измерений, отсутствуют. Возникает естественный вопрос о том, как влияет недостаток этой информации на качество получаемых оценок?

Исследования показывают, что *неучет корреляции* в пределах $|\rho| \leq 0.8$ приводит лишь к *незначительной потере эффективности оценок*. *Оценки остаются несмещенными*.

В случаях, когда при неравноточных измерениях применяется МНК, то есть *когда неравноточность не учитывается, в оценках коэффициентов появляется нежелательное смещение*.

З а м е ч а н и е 2. Отличие фактической плотности распределения погрешностей ε_i от нормальной приводит к потере эффективности оценок среди всех возможных. Однако *на множестве всех линейных оценок*, то есть оценок, которые линейно зависят от экспериментальных данных, *ОМНК и МНК-оценки эффективны* для любых плотностей распределения погрешностей ε_i . Это означает, что при каждом конкретном виде плотности распределения может быть получена более эффективная оценка, чем оценка ОМНК или МНК, но она обязательно будет нелинейно зависеть от экспериментальных данных.

З а м е ч а н и е 3. Нестатистический вариант полиномиальной аппроксимации сложных функций.

Этот вариант может возникать при желании упростить представление и(или) вычисление сложных функций. В этом варианте в качестве исходных данных совместно используются:

значения аппроксимируемой функции, вычисленные на компьютере с высокой точностью, или табличные значения, взятые из математических или иных справочников;

заданная исследователем точность аппроксимации в виде неравенств типа $|\varepsilon_i| \leq \Delta_i$.

Если заданные значения Δ_i одинаковы, то можно применить МНК с подбором подходящей степени полинома по признаку удовлетворения требований к точности аппроксимации.

При различии значений Δ_i можно применить ОМНК, а в качестве ковариационной матрицы Σ_ε использовать диагональную матрицу с элементами в диагонали, равными Δ_i или Δ_i^2 – по усмотрению исследователя. Степень полинома, необходимая для достижения требуемой точности аппроксимации, подбирается путем проб.

В этой ситуации статистические и вероятностные термины и характеристики не применяются.

З а м е ч а н и е 4. Все установленные выше свойства ОМНК и МНК-оценок коэффициентов аппроксимирующих полиномов справедливы только тогда, когда модель, то есть вид полинома (степенной полином) и его степень известны. Если *фактическая* степень полинома меньше, чем q , то получаемые оценки коэффициентов будут смещены, и оценки не будут обладать теми свойствами, которые были обнаружены выше в разд. 2.3.7.2, 2.3.7.3. Оценки коэффициентов будут смещены, а плотность распределения величины R^2 не будет соответствовать распределению “хи-квадрат”.

2.3.7.5. Измерения многократные, характеристики погрешностей известны

Полагаем, что известны характеристики погрешностей измерения ε_i значений y_i аппроксимируемой функции:

при равноточных измерениях – дисперсия σ_ε^2 ,

при неравноточных измерениях – ковариационная матрица Σ_ε .

В частном случае ковариационная матрица может быть диагональной, i -ми элементами диагонали являются дисперсии $\sigma_{\varepsilon i}^2$ погрешностей измерения значений y_i аппроксимируемой функции.

При каждом значении аргумента x_i , $i = 1, 2, \dots, k$, выполняется n измерений функции. Обозначим результат каждого j – го измерения y_{ij} , где j – номер эксперимента, $j = 1, 2, \dots, n$. Вычисляются средние арифметические значения

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij},$$

из которых составляется вектор $\bar{\mathbf{y}}^T = (\bar{y}_1 \quad \bar{y}_2 \quad \dots \quad \bar{y}_k)$, после чего в зависимости от обстоятельств вычисляются МНК или ОМНК-оценки коэффициентов аппроксимирующего полинома по формулам разд. 2.3.7.2, где вместо вектора \mathbf{y}_ε следует использовать вектор $\bar{\mathbf{y}}$.

В силу центральной предельной теоремы плотность распределения среднего арифметического стремится к нормальной довольно быстро при любых плотностях распределения исходных погрешностей, которые не слишком сильно различаются по дисперсии (см. разд. 1.6.6.4). Поэтому при многократных измерениях требование к нормальности распределения погрешностей измерений значительно смягчается.

Как известно из разд. 2.3.4.1, дисперсии средних арифметических $\sigma_{\bar{y}_i}^2 = \sigma_{\varepsilon i}^2 / n$. Точно так же из разд. 2.3.4.4 следует, что ковариационная матрица вектора средних арифметических $\Sigma_{\bar{\mathbf{y}}} = (1/n)\Sigma_\varepsilon$. В связи с этими обстоятельствами формулы разд. 2.3.7.2, 2.3.7.3 несколько изменятся.

а) Применение МНК.

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{y}}, \quad \Sigma_{\hat{\mathbf{a}}} = \frac{\sigma_\varepsilon^2}{n} (\mathbf{X}^T \mathbf{X})^{-1}, \quad \hat{\mathbf{a}} \in N_{q+1} \left(\mathbf{a}, \frac{\sigma_\varepsilon^2}{n} (\mathbf{X}^T \mathbf{X})^{-1} \right),$$

$$R^2 = n \sigma_\varepsilon^{-2} (\mathbf{X}\hat{\mathbf{a}} - \bar{\mathbf{y}})^T (\mathbf{X}\hat{\mathbf{a}} - \bar{\mathbf{y}}),$$

но, как и прежде, $R^2 \in \chi^2(k - q - 1)$.

б) Применение ОМНК.

$$\hat{\mathbf{a}} = (\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_\varepsilon^{-1} \bar{\mathbf{y}}, \quad \Sigma_{\hat{\mathbf{a}}} = \frac{1}{n} (\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X})^{-1}, \quad \hat{\mathbf{a}} \in N_{q+1} \left(\mathbf{a}, \frac{1}{n} (\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X})^{-1} \right),$$

$$R^2 = n (\mathbf{X}\hat{\mathbf{a}} - \bar{\mathbf{y}})^T \Sigma_\varepsilon^{-1} (\mathbf{X}\hat{\mathbf{a}} - \bar{\mathbf{y}}),$$

но, как и прежде, $R^2 \in \chi^2(k - q - 1)$.

2.3.7.6. Измерения многократные, характеристики погрешностей неизвестны

В разделе 2.3.7.5 предполагалось, что ковариационная матрица Σ_{ε} или, по крайней мере, дисперсии $\sigma_{\varepsilon i}^2$ погрешностей результатов измерений известны, что на практике бывает достаточно редко, особенно в отношении ковариационной матрицы.

Однако при многократных измерениях предоставляется возможность оценить дисперсии $\sigma_{\varepsilon i}^2$ при каждом $i, i = 1, 2, \dots, k, k > p + 1$:

$$s_{\varepsilon i}^2 = \frac{1}{n-1} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$$

или ковариационную матрицу в целом.

Корректная оценка всех элементов ковариационной матрицы Σ_{ε} , не только диагональных, но и внедиагональных возможна лишь при выполнении специально организованного эксперимента.

Выполняется один цикл измерений в такой последовательности:

воспроизводится значение физической или иной величины, соответствующее первому значению аргумента x_1 и выполняется измерение (определение) значения функции, полученный результат – y_{11} ;

воспроизводится значение физической или иной величины, соответствующее второму значению аргумента x_2 и выполняется измерение (определение) значения функции, полученный результат – y_{21} ;

описанная процедура продолжается до достижения последнего, k -го значения аргумента x , таким образом будет получен первый вектор результатов измерений $\mathbf{y}_1^T = (y_{11} \ y_{21} \ \dots \ y_{i1} \ \dots \ y_{k1})$;

устанавливается значение физической величины x , превышающее значение x_k , затем вновь устанавливается значение x_k , и процесс повторяется, но в обратном порядке, при уменьшении значений x ; таким образом будет получен второй вектор результатов $\mathbf{y}_2^T = (y_{12} \ y_{22} \ \dots \ y_{i2} \ \dots \ y_{k2})$.

При повторении подобных циклов измерений в конечном итоге будет получено четное количество n векторов вида

$$\mathbf{y}_j^T = (y_{1j} \quad y_{2j} \quad \dots \quad y_{ij} \quad \dots \quad y_{kj}), j = 1, 2, \dots, n.$$

По этому массиву экспериментальных данных вычисляются оценки (см. разд. 2.3.4.3, 2.3.4.4) :

$$\bar{\mathbf{y}} = \frac{1}{n} \sum_{j=1}^n \mathbf{y}_j, \quad \mathbf{S}_\varepsilon = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}})^T.$$

Оценка ковариационной матрицы (см. также разд. 2.3.4.4) построена в соответствии с ее математическим определением, приведенным в разд. 1.7.3. Поскольку при реализации ОМНК эту матрицу придется обращать, она не должна быть особенной. Для этого **необходимо, чтобы $n > k$** . Но если по техническим, экономическим или иным причинам это условие выполнить невозможно, то придется ограничиться вычислением только оценок дисперсий $s_{\varepsilon i}^2$ при каждом значении x_i . По этим значениям строится диагональная матрица \mathbf{S}_ε , в диагонали которой на i -м месте стоит оценка дисперсии $s_{\varepsilon i}^2$:

$$\mathbf{S}_\varepsilon = \begin{pmatrix} s_{\varepsilon 1}^2 & 0 & \dots & 0 \\ 0 & s_{\varepsilon 2}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & s_{\varepsilon k}^2 \end{pmatrix}, \quad \bar{\mathbf{y}} = \frac{1}{n} \sum_{j=1}^n \mathbf{y}_j, \quad s_{\varepsilon i}^2 = \frac{1}{n-1} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2.$$

Матрица \mathbf{S}_ε может оказаться диагональной даже при $n > k$, если исследователь по каким-либо причинам решил пренебречь ковариациями.

В этих случаях не учитывается ковариация между измерениями в точках x_i , что приводит к незначительной потере в эффективности оценок коэффициентов, но они остаются несмещенными (см. разд. 2.3.7.4, замечание 1).

После этого для вычисления оценок коэффициентов аппроксимирующего полинома применяется ОМНК с заменой во всех формулах разд. 2.3.7.5 матрицы Σ_ε на \mathbf{S}_ε :

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \bar{\mathbf{y}}, \quad \Sigma_{\hat{\mathbf{a}}} = \frac{1}{n} (\mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1}, \quad \hat{\mathbf{a}} \in N_{q+1} \left(\mathbf{a}, \frac{1}{n} (\mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1} \right),$$

$$R^2 = n (\mathbf{X} \hat{\mathbf{a}} - \bar{\mathbf{y}})^T \mathbf{S}_\varepsilon^{-1} (\mathbf{X} \hat{\mathbf{a}} - \bar{\mathbf{y}}),$$

Вследствие случайности исходных данных величина R^2 также случайна. Из-за того, что в формуле для R^2 вместо генеральной ковариационной матрицы участвует ее оценка, распределение “хи-квадрат” неприменимо. Вместо него здесь применяется плотность F -распределения Фишера (иногда она именуется, как плотность распределения Фишера-Снедекора), и обозначается, как $F(k_1, k_2)$, где k_1 и k_2 – количества степеней свободы. Плотность распределения Фишера имеет случайная величина [5]

$$F = \frac{(n - k + q + 1)}{(k - q - 1)(n - 1)} R^2,$$

что записывается в виде

$$F = \frac{(n - k + q + 1)}{(k - q - 1)(n - 1)} R^2 \in F(k - q - 1, n - k + q + 1),$$

Где $k - q - 1$ и $n - k + q + 1$ – количества степеней свободы, $k > q + 1$.

Эта плотность распределения широко используется для сопоставления дисперсий генеральных совокупностей при дисперсионном анализе посредством исследования отношения оценок этих дисперсий. Нетрудно увидеть, что величина R^2 также в некотором смысле есть отношение дисперсий. Функция распределения Фишера табулирована, таблицы приводятся в таблицах математической статистики (например, [1,13, 14]).

Напомним, что для предотвращения особенности матрицы S_{ε} необходимо обеспечить, чтобы количество повторных измерений при каждом значении x_i превышало количество значений аргумента, то есть должно выполняться неравенство $n > k$. Если это неравенство выполняется, да к тому же для полиномиальной аппроксимации всегда должно быть, чтобы $k > q + 1$, тогда количество степеней свободы $n - k + q + 1$ будет неотрицательным, что и требуется. Если же по техническим, экономическим или иным причинам необходимое количество n повторных измерений выполнить невозможно, то придется формировать неособенную диагональную матрицу с диагональными элементами $s_{\varepsilon i}^2$, как это было только что показано.

В этой ситуации F -распределению Фишера подчиняется случайная величина

$$F = \frac{R^2}{(k - q - 1)} \in F(k - q - 1, n - 1) .$$

Число степеней свободы $k - q - 1$ и $n - 1$.

В частном случае, когда в этих же условиях гипотеза о равенстве дисперсий $\sigma_{\varepsilon i}^2$ по критерию Кочрена (п. 2.5.6.1) не будет отклонена, тогда принимается решение о равноточности измерений и об аппроксимации в соответствии с МНК, для чего вычисляется средняя оценка дисперсии

$$s_{\varepsilon}^2 = \frac{1}{k} \sum_{i=1}^k s_{\varepsilon i}^2 ,$$

которая подставляется вместо σ_{ε}^2 во всех соответствующих формулах разд. 2.3.7.5

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{y}}, \quad \Sigma_{\hat{\mathbf{a}}} = \frac{s_{\varepsilon}^2}{n} (\mathbf{X}^T \mathbf{X})^{-1}, \quad \hat{\mathbf{a}} \in N_{q+1} \left(\mathbf{a}, \frac{s_{\varepsilon}^2}{n} (\mathbf{X}^T \mathbf{X})^{-1} \right),$$

$$R^2 = n s_{\varepsilon}^{-2} (\mathbf{X} \hat{\mathbf{a}} - \bar{\mathbf{y}})^T (\mathbf{X} \hat{\mathbf{a}} - \bar{\mathbf{y}}),$$

И в этом случае случайная величина

$$F = \frac{R^2}{(k - q - 1)} \in F(k - q - 1, n - 1) .$$

распределена в соответствии с F -распределением Фишера с числом степеней свободы $k - q - 1$ и $n - 1$.

Все замечания, сделанные выше в разд. 2.3.7.4, распространяются на случаи многократных измерений в полном объеме.

2.3.7.7. Особенности вычислений при реализации МНК и ОМНК

Как следует из разд. 2.3.7.2, 2.3.7.3, в процессе оценивания коэффициентов аппроксимирующих полиномов приходится обращаться матрицы $(\mathbf{X}^T \mathbf{X})$ и $(\mathbf{X}^T \Sigma_{\varepsilon}^{-1} \mathbf{X})$. Из линейной алгебры и вычислительной математики известно, что *устойчивость* результатов подобных действий в сильной степени зависит от *обусловленности* обращаемых матриц. Что касается матриц Σ_{ε} и \mathbf{S}_{ε} , то здесь следует опасаться того, что они могут оказаться

особенными. В частности, одна из причин появления особенности у матрицы \mathbf{S}_ε указана в разд. 2.3.7.6.

Обусловленность матриц характеризуется *числом обусловленности*, которое есть не что иное, как коэффициент “усиления” погрешностей экспериментальных данных и погрешностей округления к погрешностям результатов вычислений.

Для квадратных симметричных матриц, каковыми являются матрицы, перечисленные выше, число обусловленности определено равенствами

$$\text{cond}(\mathbf{X}^T \mathbf{X}) = \sqrt{\frac{\lambda_{\max}(\mathbf{X}^T \mathbf{X})}{\lambda_{\min}(\mathbf{X}^T \mathbf{X})}}, \quad \text{cond}(\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X}) = \sqrt{\frac{\lambda_{\max}(\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X})}{\lambda_{\min}(\mathbf{X}^T \Sigma_\varepsilon^{-1} \mathbf{X})}}$$

где $\lambda_{\max}(\bullet)$, $\lambda_{\min}(\bullet)$ – наибольшее и наименьшее собственные числа соответствующей матрицы.

Число обусловленности матриц, используемых в МНК или в ОМНК, может достигать значений от 10^6 до 10^{12} и выше.

Известно, что число обусловленности указанных матриц монотонно возрастает с увеличением количества столбцов матрицы \mathbf{X} , то есть с увеличением порядка q или, что то же самое, с увеличением числа коэффициентов полинома (см. конструкцию матрицы \mathbf{X} в разд. 2.3.7.2). Максимального значения число обусловленности достигает при $q + 1 = k$, когда распределение Фишера уже неприменимо. Особенно опасной оказывается ситуация, когда количество оцениваемых коэффициентов превышает их фактическое количество, то есть при завышении степени полинома.

Можно рекомендовать три способа повышения устойчивости оценок коэффициентов МНК и ОМНК.

1. Не стремиться к излишне высокому порядку аппроксимирующего полинома, использовать априорную информацию о гладкости аппроксимируемой функции.

2. При необходимости аппроксимации функции $y = f(x)$ полиномом высокого порядка вплоть до $q = k - 1$ использовать *метод регуляризации Тихонова*.

Применительно к МНК и ОМНК этот метод заключается в следующем [7].

Исходное уравнение преднамеренно искажается таким образом, чтобы это искажение заведомо улучшало обусловленность. Таким регуляризирующим искажением является, по Тихонову, увеличение диагональных элементов матрицы системы уравнений. Регуляризованное таким образом решение имеет вид $\hat{\mathbf{a}}_\alpha = (\mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{X} + \alpha \mathbf{E})^{-1} \mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{y}_\varepsilon$, $\alpha > 0$.

Число α называется *параметром регуляризации*. Оценка $\hat{\mathbf{a}}_\alpha$ в англоязычной литературе называется *ридж-оценкой*. В отечественной литературе встречается калькоподобный перевод этого термина: “гребневая оценка”.

Эта оценка, конечно, смещена. Ее смещение примерно

$$\approx \alpha (\mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{y}_\varepsilon.$$

Проблема состоит в выборе такого значения параметра регуляризации, при котором результирующая погрешность, вызванная смещением оценки и плохой обусловленностью матрицы системы, была минимальной.

На рис. 32 показана принципиальная возможность такого выбора. Однако, универсального практического рецепта выбора оптимального значения параметра регуляризации пока не существует.

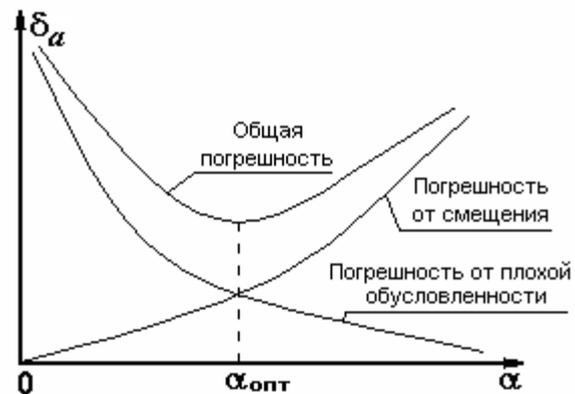


Рис. 32. Поведение погрешностей оценки при изменении параметра регуляризации

3. Третий способ заключается в таком размещении значений аргумента \mathbf{x} , при котором число обусловленности матрицы $(\mathbf{X}^T \mathbf{X})$ или $(\mathbf{X}^T \mathbf{S}_\varepsilon^{-1} \mathbf{X})$ будет минимальным. Этот способ рассматривается в следующем пункте.

2.3.7.8. Основные принципы планирования эксперимента, выполняемого с целью полиномиальной аппроксимации

Снова обратимся к конструкции матрицы \mathbf{X} , приведенной в разд. 2.3.7.2, и к одной из формул для ковариационной матрицы оценок коэффициентов полинома, например, $\Sigma_{\hat{\mathbf{a}}} = \frac{\sigma^2}{n} (\mathbf{X}^T \mathbf{X})^{-1}$. Из конструкции матрицы \mathbf{X} видно, что ее элементы изменяют свои значения в зависимости от значений аргумента полинома: x_1, x_2, \dots, x_k . Стало быть, можно надеяться на то, что существует такой план расстановки этих значений, при котором погрешности в каком-либо смысле будут минимальными.

Такие планы, действительно, существуют. Назовем некоторые из них (см., например, [15]).

A-оптимальный план эксперимента – план, при котором достигается минимум следа матрицы $(\mathbf{X}^T \mathbf{X})^{-1}$, то есть минимум суммы ее диагональных элементов.

D-оптимальный план эксперимента – план, при котором достигается минимальное значение определителя матрицы $(\mathbf{X}^T \mathbf{X})^{-1}$.

C-оптимальный план эксперимента – план, при котором достигается минимальное значение числа обусловленности матрицы $\mathbf{X}^T \mathbf{X}$.

C-оптимальный план эквивалентен **D**-оптимальному плану.

2.3.7.9. Расширение класса аппроксимирующих полиномов

Материал, изложенный выше в разд. 2.3.7.1 – 2.3.7.8, в равной степени относится к оценке коэффициентов обобщенных аппроксимирующих полиномов с заменой понятия “степень полинома” на “порядок полинома”.

Обобщенным полиномом называется полином вида

$$y(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + a_2 \varphi_2(x) + \dots + a_q \varphi_q(x),$$

$\{\varphi_i(x)\}, i = 0, 1, 2, \dots, q$ – система базисных функций.

Если эти функции ортогональны и соответствуют характеру аппроксимируемой зависимости лучше, чем степени x , то для достижения необходимой точности аппроксимации может понадобиться меньше членов, чем в случае аппроксимации степенным полиномом. А это обстоятельство способствует улучшению обусловленности задачи и является *четвертым средством повышения устойчивости оценок* МНК и ОМНК.

Матрица \mathbf{X} в этом случае будет иметь следующую конструкцию:

$$\mathbf{X} = \begin{pmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_q(x_1) \\ \varphi_0(x_2) & \varphi_1(x_2) & \cdots & \varphi_q(x_2) \\ \vdots & \vdots & \cdots & \vdots \\ \varphi_0(x_k) & \varphi_1(x_k) & \cdots & \varphi_q(x_k) \end{pmatrix}.$$

Это единственное отличие от изложенного ранее в данном разделе. Все остальные формулы, замечания и рекомендации остаются в силе без каких-либо изменений.

2.4. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ

2.4.1. Постановка задачи

Как и ранее, будем обозначать генеральные значения моментов, или параметров плотности распределения, или иные характеристики генеральной совокупности случайной величины Θ .

Точечные оценки генеральных характеристик (параметров) Θ недостаточно информативны, поскольку в них не содержится сведений о том, насколько далеки они от истинных значений. Напротив, интервальные оценки содержат такие сведения.

Целью интервального оценивания является вычисление по выборочным данным \mathbf{x} объема n такого интервала с границами: нижней $\underline{\Theta}(\mathbf{x})$ и верхней $\overline{\Theta}(\mathbf{x})$, чтобы

$$P(\underline{\Theta}(\mathbf{x}) < \Theta < \overline{\Theta}(\mathbf{x})) \geq Q,$$

где Q – вероятность, близкая к единице, например $Q = 0,8 - 0,95$.

Такой интервал называется *доверительным интервалом* (confidence interval), вероятность Q – *доверительной вероятностью*.

Поскольку границы доверительного интервала являются функциями выборочных значений, они случайны. Поэтому говорят, что доверительный интервал покрывает значение искомой генеральной характеристики (параметра) с вероятностью, не меньшей, чем Q .

2.4.2. Доверительный интервал для математического ожидания. Дисперсия генеральной совокупности известна

Пусть выборка $\mathbf{x}^T = (x_1 \ x_2 \ \dots \ x_n)$ извлечена из нормальной генеральной совокупности, образованной случайной величиной $\xi \in N(a, \sigma^2)$, σ^2 известна. Несмещенная, состоятельная и эффективная точечная оценка математического ожидания – среднее арифметическое

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

Это линейная функция выборочных значений, извлеченных из нормальной генеральной совокупности, нормальное распределение безгранично делимо, поэтому \bar{x} распределено нормально:

$$\bar{x} \in N\left(a, \frac{\sigma^2}{n}\right).$$

Вычтем из \bar{x} его математическое ожидание и разделим результат на его среднеквадратическое значение. Получим новую случайную величину, которая также нормально распределена:

$$\frac{\bar{x} - a}{\sigma / \sqrt{n}} = \sqrt{n} \frac{\bar{x} - a}{\sigma} \in N(0,1).$$

Пусть N_α – $\alpha \cdot 100$ -процентная квантиль, $N_{1-\alpha}$ – $(1 - \alpha) \cdot 100$ -процентная квантиль нормального распределения. Вероятностная мера интервального промежутка $[N_\alpha, N_{1-\alpha}]$ есть

$$P\left(N_\alpha < \frac{\bar{x} - a}{\sigma} \sqrt{n} \leq N_{1-\alpha}\right) = 1 - 2\alpha = Q.$$

Пусть Q – доверительная вероятность, $\alpha = (1 - Q)/2$, $(1 - \alpha) = (1 + Q)/2$.

В силу симметрии нормального распределения $N_\alpha = -N_{1-\alpha}$.

Решим неравенство, стоящее в скобках, относительно математического ожидания:

$$P\left(\bar{x} - \frac{\sigma}{\sqrt{n}} N_{\frac{1+Q}{2}} < a \leq \bar{x} + \frac{\sigma}{\sqrt{n}} N_{\frac{1+Q}{2}}\right) = Q.$$

Доверительный интервал для математического ожидания построен для случаев, когда выборка извлечена из нормальной генеральной совокупности. Но поскольку \bar{x} вычисляется путем суммирования выборочных значений, распределенных одинаково, условия центральной предельной теоремы соблюдаются, а это значит, что уже начиная с объемов выборки $n = 15 - 20$, плотность распределения среднего арифметического достаточно близка к нормальной вне зависимости от вида плотности распределения генеральной совокупности, из которой извлечена выборка. Начиная с таких объемов выборки, эта интервальная оценка может быть применена без ограничений на вид плотности распределения исследуемой случайной величины.

Следует обратить пристальное внимание на две тенденции поведения границ доверительного интервала:

1. С увеличением объема выборки при фиксированном значении доверительной вероятности ширина доверительного интервала уменьшается и в пределе стремится к нулю, что вполне естественно.

2. При фиксированном значении объема выборки с увеличением доверительной вероятности ширина доверительного интервала увеличивается, и в пределе при $Q = 1$ доверительным интервалом становится вся ось, что также вполне естественно, ибо для покрытия неизвестного значения с большой вероятностью требуется широкий интервал.

2.4.3. Доверительный интервал для математического ожидания.

Дисперсия генеральной совокупности неизвестна

Пусть как и ранее, $\xi \in N(a, \sigma^2)$, $\bar{x} \in N(a, \sigma^2 / n)$. Поскольку дисперсия неизвестна, будем использовать ее несмещенную оценку

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

По аналогии с разд. 2.4.3 сформируем случайную величину:

$$\tau = \frac{\bar{x} - a}{s / \sqrt{n}} = \frac{\bar{x} - a}{s} \cdot \sqrt{n} .$$

Плотность распределения этой случайной величины есть плотность распределения Стьюдента с $n - 1$ степенью свободы, которая имеет следующий вид:

$$\text{St}(t, n-1) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right) \sqrt{\pi(n-1)}} \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} .$$

Это одномодальная симметричная плотность распределения. При значительных объемах выборки n математическое ожидание, дисперсия и эксцесс случайной величины τ , распределенной по Стьюденту:

$$M[\tau] = 0, D[\tau] = \frac{n}{n-1}, ex = 3 \frac{n-2}{n-4} .$$

Единственный параметр плотности распределения Стьюдента – число степеней свободы.

Частный вид плотности распределения Стьюдента при $n = 2$ – плотность распределения Коши. При $n \rightarrow \infty$ плотность распределения Стьюдента стремится к нормальному распределению. Удовлетворительная близость к нормальному распределению начинается уже с $n = 20$.

Квантили распределения Стьюдента с числом степеней свободы $n - 1$: $t_\alpha(n-1)$ и $t_{1-\alpha}(n-1)$ – границы интерквантильного промежутка, такого, что

$$P\left(t_\alpha(n-1) < \frac{\bar{x} - a}{s} \cdot \sqrt{n} \leq t_{1-\alpha}(n-1)\right) = 1 - 2\alpha = Q .$$

В силу симметрии плотности распределения Стьюдента

$$t_\alpha(n-1) = - t_{1-\alpha}(n-1) .$$

С учетом этого факта и выражая вероятность α через Q , решим неравенство, стоящее в скобках, относительно a :

$$P\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\frac{1+Q}{2}}(n-1) < a \leq \bar{x} + \frac{s}{\sqrt{n}} \cdot t_{\frac{1+Q}{2}}(n-1)\right) = Q.$$

Получен доверительный интервал для математического ожидания в условиях, когда вместо дисперсии применяется ее несмещенная оценка.

В этом случае также проявляется полезное свойство центральной предельной теоремы, позволяющее при значительных объемах выборки (начиная с $n = 20$) пользоваться полученным доверительным интервалом для оценки математического ожидания широкого класса наиболее употребительных случайных величин, плотность распределения которых отличается от нормальной.

Как и в предыдущем разделе, обращаем внимание на две тенденции поведения границ доверительного интервала.

1. С увеличением объема выборки при фиксированном значении доверительной вероятности ширина доверительного интервала уменьшается и в пределе стремится к нулю, что вполне естественно.

2. При фиксированном значении объема выборки с увеличением доверительной вероятности ширина доверительного интервала увеличивается, и в пределе при $Q = 1$ доверительным интервалом становится вся ось, что также вполне естественно, ибо для покрытия неизвестного значения с большой вероятностью требуется широкий интервал.

2.4.4. Доверительный интервал для дисперсии

Пусть, как и ранее, $\xi \in N(a, \sigma^2)$. Математическое ожидание неизвестно, для его оценки применено среднее арифметическое \bar{x} . Точечная несмещенная оценка дисперсии

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

В разд. 2.3.4.2 с) мы выяснили, что плотность распределения случайной

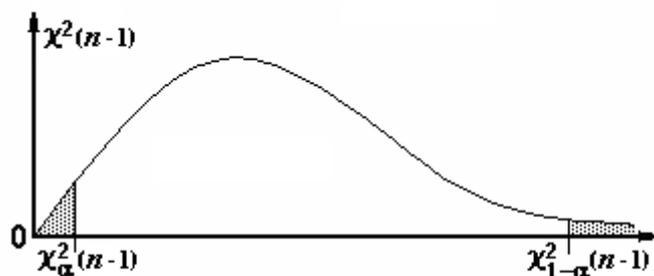


Рис. 33. Расположение квантилей распределения хи-квадрат

величины $\frac{s^2(n-1)}{\sigma^2}$ есть плотность распределения $\chi^2(n-1)$ с $n-1$ степенями свободы.

Квантили этой плотности распределения $\chi_\alpha^2(n-1)$ и $\chi_{1-\alpha}^2(n-1)$ – границы интерквантильного промежутка, который определяется равенством

$$P\left(\chi_\alpha^2(n-1) < \frac{s^2(n-1)}{\sigma^2} \leq \chi_{1-\alpha}^2(n-1)\right) = 1 - 2\alpha = Q.$$

Как видно из рис. 33 (см. также рис. 28), плотность распределения хи-квадрат несимметрична, поэтому после решения неравенства, стоящего в скобках, получим доверительные интервалы с несимметричными границами:

$$P\left(\frac{s^2(n-1)}{\chi_{1-\alpha}^2(n-1)} < \sigma^2 \leq \frac{s^2(n-1)}{\chi_\alpha^2(n-1)}\right) = 1 - 2\alpha = Q,$$

$$P\left(s \sqrt{\frac{n-1}{\chi_{1+Q}^2(n-1)}} < \sigma \leq s \sqrt{\frac{n-1}{\chi_{1-Q}^2(n-1)}}\right) = Q.$$

Расположение квантилей, использованных в этих формулах, показано на рис. 33.

Материалы настоящего раздела справедливы только для случаев, когда выборка извлечена из нормальной генеральной совокупности.

2.4.5. Доверительный интервал для вероятности

Выполняются n независимых испытаний, по результатам которых необходимо оценить вероятность p события **A**. Эти испытания соответствуют схеме Бернулли, исходами которых может быть одно из двух событий: **A** и $\bar{\mathbf{A}}$. Пусть m_0 – количество появления события **A** в n испытаниях. Точечной оценкой вероятности в соответствии с частотным определением (см. разд. 1.2.2) является отношение

$$\tilde{p} = \frac{m_0}{n}.$$

В соответствии со схемой Бернулли (см. разд. 1.2.4) вероятность того, что событие **A** появится ровно m_0 раз, составляет

$$P_n(m_0) = C_n^{m_0} p^{m_0} (1-p)^{n-m_0}.$$

Пусть M – произвольное целое число, $0 < M < n$. Рассмотрим событие B_m , которое заключается в том, что в результате n независимых испытаний событие **A** появилось ровно m раз. События $B_0, B_1, B_2, \dots, B_M$ не пересекаются, поэтому вероятность того, что при n испытаниях событие **A** произойдет не более M раз, выражается суммой вероятностей

$$P(B_0 \cup B_1 \cup \dots \cup B_M) = \sum_{m=0}^M C_n^m p^m (1-p)^{n-m} = S_M(p).$$

Эта сумма, как функция от p , монотонно убывает при любом значении $M < n$. В самом деле, при $p = 0$ и $m = 0$ $S_M(0) = 1$, поскольку при нулевой вероятности событие **A** невозможно, и уже первое слагаемое суммы, то есть $P(B_0) = 1$. В то же время вероятности событий $P(B_1) = P(B_2) = \dots = P(B_M) = 0$. Напротив, при $p = 1$ событие **A** – достоверное событие, поэтому только при n испытаниях событие **A** может произойти ровно n раз, и при фиксированном значении $M < n$ функция $S_M(1) = P(B_0 \cup B_1 \cup \dots \cup B_M) = 0$. И вообще, с увеличением вероятности события **A** вероятность появления малого количества этого события при n испытаниях может только уменьшаться. Для подтверждения этого возьмем производную от функции $S_M(p)$ по аргументу p . Первое слагаемое суммы есть $C_n^0 (1-p)^n = (1-p)^n$ потому, что $C_n^0 = 1$. Производная от него равна $-n(1-p)^{n-1}$. Поскольку каждое следующее слагаемое содержит произведение $p^m (1-p)^{n-m}$, производная от каждого такого слагаемого имеет вид

$$C_n^m m p^{m-1} (1-p)^{n-m} - C_n^m p^m (n-m) (1-p)^{n-m-1}.$$

Раскрывая выражения для числа сочетаний, получим, что производная от m -го слагаемого есть

$$\frac{n!}{m!(n-m)!} m p^{m-1} (1-p)^{n-m} - \frac{n!}{m!(n-m)!} (n-m) p^m (1-p)^{n-m-1} =$$

$$= \frac{n!}{(m-1)!(n-m)!} p^{m-1} (1-p)^{n-m} - \frac{n!}{m!(n-m-1)!} p^m (1-p)^{n-m-1} .$$

Теперь сложим производные от первого и второго слагаемых, когда $m = 0$ и $m = 1$:

$$\begin{aligned} & -n(1-p)^{n-1} + \frac{n!}{(n-1)!} (1-p)^{n-1} - \frac{n!}{(n-2)!} p^2 (1-p)^{n-2} = \\ & = -n(1-p)^{n-1} + n(1-p)^{n-1} - n(n-1)p^2 (1-p)^{n-2} = -n(n-1)p^2 (1-p)^{n-2} . \end{aligned}$$

Мы видим, что производная от второго слагаемого исходной суммы скомпенсировала производную от первого слагаемого.

Аналогичный результат получится, когда мы сложим производные от двух соседних слагаемых ($m-1$ и m) исходной суммы, то есть функции $S_M(p)$:

$$\begin{aligned} & \frac{n!}{(m-1)!(n-m+1)!} (m-1)p^{m-2}(1-p)^{n-m+1} - \frac{n!}{(m-1)!(n-m+1)!} (n-m+1)p^{m-1}(1-p)^{n-m} + \\ & + \frac{n!}{m!(n-m)!} mp^{m-1}(1-p)^{n-m} - \frac{n!}{m!(n-m)!} (n-m)p^m(1-p)^{n-m-1} = \\ & = \frac{n!}{(m-2)!(n-m+1)!} p^{m-2}(1-p)^{n-m+1} - \frac{n!}{(m-1)!(n-m)!} p^{m-1}(1-p)^{n-m} + \\ & + \frac{n!}{(m-1)!(n-m)!} p^{m-1}(1-p)^{n-m} - \frac{n!}{m!(n-m-1)!} p^m(1-p)^{n-m-1} = \\ & = \frac{n!}{(m-2)!(n-m+1)!} p^{m-2}(1-p)^{n-m+1} - \frac{n!}{m!(n-m-1)!} p^m(1-p)^{n-m-1} . \end{aligned}$$

И здесь мы видим, что следующие друг за другом внутренние слагаемые производной от $S_M(p)$ по p взаимно уничтожаются. В конечном итоге останется только последнее слагаемое при $m = M$, поэтому

$$\frac{dS_M(p)}{dp} = -\frac{n!}{M!(n-M-1)!} p^M (1-p)^{n-M-1} ,$$

а это значит, что функция $S_M(p)$ монотонно убывает от 1 до 0 при любом M .

В нашем случае при выполнении испытаний по схеме Бернулли с целью оцениванию вероятности это означает следующее.

Пусть выполняются n независимых испытаний, в результате которых событие **A** произошло m_0 раз. Тогда точечная оценка искомой вероятности есть $\tilde{p} = m_0 / n$. Примем, что $m_0 = M$. Тогда по заданной малой вероятности α может быть однозначно определена верхняя граница доверительного интервала \bar{p} для искомой вероятности p , как решение уравнения

$$S_{m_0}(\bar{p}) = \sum_{m=0}^{m_0} C_n^m (\bar{p})^m (1 - \bar{p})^{n-m} = \alpha.$$

В самом деле, если $p > \bar{p}$, то $S_{m_0}(p) < S_{m_0}(\bar{p})$, а это значит, что вероятность ошибочно отбросить, то есть не принимать во внимание любое значение вероятности события **A**, превышающее \bar{p} , меньше, чем α .

Рассуждая точно так же, как и в предыдущем случае, сконструируем несколько иную функцию от p :

$$S'_{m_0}(p) = P(B_{m_0} \cup B_{m_0+1} \cup \dots \cup B_n) = \sum_{m=m_0}^n C_n^m p^m (1-p)^{n-m}.$$

Эта функция монотонно возрастает с увеличением p от 0 до 1. В самом деле, при $p = 0$, то есть при невозможности события **A** функция $S'_{m_0}(p) = 0$. Напротив, если событие **A** достоверно, $p = 1$, и это событие произойдет при каждом испытании также с вероятностью 1, поскольку ненулевым в сумме окажется только одно последнее слагаемое при $m_0 = n$. В этой ситуации ненулевым в сумме окажется только последнее слагаемое, и значит функция $S'_{m_0}(1) = 1$. Анализ производной от этой функции по p , аналогичный только что выполненному, подтверждает монотонное возрастание функции $S'_M(p)$ от 0 до 1. И здесь простое рассуждение легко приводит к тому, что увеличением вероятности p события **A** вероятность появления увеличивающегося числа появления этого события при n испытаниях может только возрастать.

В связи с установленной монотонностью функции $S'_M(p)$ при числе независимых испытаний n , заданной малой вероятности β и количестве m_0 появления события **A** нижняя граница \underline{p} для искомой вероятности однозначно определяется из уравнения:

$$S'_{m_0}(\underline{p}) = \sum_{m=m_0}^n C_n^m (\underline{p})^m (1 - \underline{p})^{n-m} = \beta.$$

Вероятность отбросить любое значение вероятности p , меньшее, чем \underline{p} , не превышает β . В результате решения двух уравнений получены нижняя и верхняя границы доверительного интервала

$$P(\underline{p} \leq p \leq \bar{p}) \geq Q.$$

Доверительная вероятность равна $Q = 1 - (\alpha + \beta)$.

Обычно принимают $\alpha = \beta$, и $Q = 1 - 2\alpha$.

2.4.6. Доверительные интервалы для интерквантильного промежутка

2.4.6.1. Параметрические толерантные пределы

Выборка $x^T = (x_1 \ x_2 \ \dots \ x_n)$ извлечена из нормальной генеральной совокупности \mathbf{X} , образованной случайной величиной $\xi \in N(a, \sigma^2)$. Для этой генеральной совокупности определен (задан) симметричный относительно математического ожидания интерквантильный промежуток J_P в соответствии с разд. 1.6.2 (см. рис. 10 а), такой, что

$$P(\xi \in J_P) = P(M[\xi] - g < \xi \leq M[\xi] + g) = P.$$

Поскольку плотность распределения генеральной совокупности нормальна,

$$g = \sigma N_{\frac{1+P}{2}},$$

и границами этого промежутка являются

$$M[\xi] - \sigma N_{\frac{1+P}{2}} \quad \text{и} \quad M[\xi] + \sigma N_{\frac{1+P}{2}},$$

где $N_{\frac{1+P}{2}}$ это $\frac{1+P}{2}$ 100-процентная квантиль нормальной случайной величины, дисперсия которой $\sigma^2 = 1$. Например, при $P = 0,8$ квантиль $N_{0,9} = 1,6$, при $P = 0,95$ квантиль $N_{0,975} = 1,96$.

Наша задача состоит в том, чтобы по имеющейся выборке, используя точечные оценки математического ожидания и среднеквадратического

значения случайной величины, определить границы доверительного интервала, который накрывает искомый интерквантильный промежуток J_P с заданной доверительной вероятностью Q . Границы такого доверительного интервала называются *толерантными пределами*. Обозначим эти доверительные пределы, нижний и верхний соответственно через \underline{tol} и \overline{tol} . Поставленная задача будет решена, если с вероятностью, не меньшей Q , будут совместно выполняться неравенства

$$\underline{tol} \leq M[\xi] - \sigma N_{\frac{1+P}{2}} \quad \text{и} \quad M[\xi] + \sigma N_{\frac{1+P}{2}} \leq \overline{tol}.$$

Понятно, что расположение доверительных пределов на оси должно определяться тремя параметрами: вероятностью P , для которой определен искомый интерквантильный промежуток J_P , объемом выборки n , по которой мы будем находить точечные оценки $M[\xi]$ и σ^2 , а также значением доверительной вероятности Q .

Толерантные пределы определяются с помощью *толерантных множителей*, которые зависят от перечисленных параметров и обозначаются, как $\kappa(n, P, Q)$. Доверительный интервал, накрывающий интерквантильный промежуток J_P с доверительной вероятностью Q , определяется доверительными пределами (см. [6], стр. 353, а также [13,14]):

$$\bar{x} - \kappa(n, P, Q)s, \quad \bar{x} + \kappa(n, P, Q)s,$$

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

Эти пределы называются *параметрическими доверительными пределами*, поскольку определяются через точечные оценки параметров плотности распределения. Они могут использоваться *только для интервальной оценки интерквантильного промежутка нормальной генеральной совокупности*.

Значения доверительного множителя табулированы в таблицах математической статистики (см. [13] для $n \leq 50$ или [14]).

Сопоставим математическое определение генерального интерквантильного промежутка, ширина которого пропорциональна σ с коэффициентом пропорциональности $2 N_{\frac{1+P}{2}}$, и доверительный интервал для него,

Таблица 3

**Значения толерантного множителя и квантилей
нормальной плотности распределения для $P = 0.95$**

$Q \backslash n$	5	8	12	20	100	$N_{0.975}$
0.9	4.152	3.264	2.863	2.564	2.170	1.96
0.95	5.079	3.732	3.162	2.752	2.231	1.96

ширина которого пропорциональна s с коэффициентом пропорциональности $2 k(n, P, Q)$. С этой целью в табл. 3 приведено сравнение квантилей нормальной плотности распределения и толерантного множителя для одного значения вероятности $P = 0.95$.

Заметим, что, в отличие от доверительных интервалов для математического ожидания и дисперсии, границы доверительного интервала для интерквантильного промежутка, то есть толерантные пределы, с увеличением объема выборки приближаются к границам искомого интерквантильного промежутка. В пределе при $n \rightarrow \infty$ ширина этого доверительного интервала равна ширине искомого интерквантильного промежутка.

2.4.6.2. Непараметрические толерантные пределы

Непараметрические толерантные пределы являются границами доверительного интервала для интерквантильного промежутка. Для того, чтобы их определить, не требуется априорная информация о виде плотности распределения генеральной совокупности, а поэтому точечные оценки параметров не используются. В англоязычной литературе для обозначения подобного рода статистического оценивания характеристик случайных величин применяется термин “distribution-free estimation”.

В качестве непараметрических толерантных пределов служат непосредственно выборочные значения - члены вариационного ряда (см. [6], стр. 338).

Принцип нахождения непараметрических толерантных пределов, иными словами, границ доверительных интервалов для интерквантильного промежутка, не зависящих от вида плотности распределения, основывается на результате, полученном ранее в примере 2 разд. 1.6.7. Этот же результат использован при построении выборочной функции распределения в разд. 2.2 (см. рис. 26). Как следует из указанных материалов, вероятностные меры полуоткрытых интервалов, заключенных между двумя соседними членами вариационного ряда в среднем, по множеству групп однородных экспериментов объемом n одинаковы и равны $1/n$.

Это происходит по следующей причине. Выборка x_1, x_2, \dots, x_n извлекается из генеральной совокупности, образованной всеми значениями случайной величины ξ , интегральная функция распределения которой есть $F_\xi(x)$. Поскольку при любом распределении ξ случайная величина $\eta = F_\xi(\xi)$ распределена равномерно в интервале $(0, 1)$, значения функции $F_\xi(x_1), F_\xi(x_2), \dots, F_\xi(x_n)$ от выборочных значений x_1, x_2, \dots, x_n также распределены равномерно в том же интервале. При нанесении выборочных значений на числовую ось эти выборочные значения выстраиваются в вариационный ряд $x_{(1)}, x_{(2)}, \dots, x_{(n)}$. Вероятностные меры полуоткрытых интервалов $(x_{(1)}, x_{(2)}], (x_{(2)}, x_{(3)}], \dots, (x_{(n-1)}, x_{(n)}]$:

$$F_\xi(x_{(2)}) - F_\xi(x_{(1)}), F_\xi(x_{(3)}) - F_\xi(x_{(2)}), \dots, F_\xi(x_{(n)}) - F_\xi(x_{(n-1)}).$$

Поскольку случайная величина $\eta = F_\xi(\xi)$ распределена равномерно, эти вероятностные меры одинаковы на множестве всех возможных групп выборочных значений, извлеченных из генеральной совокупности, образованной случайной величиной ξ .

В связи с этим свойством полуоткрытые интервалы между соседними членами вариационного ряда $x_{(i)}$ и $x_{(i+1)}$ называются *статистически эквивалентными блоками*. Первый статистически эквивалентный блок $(-\infty, x_{(1)}]$. Последний статистически эквивалентный блок $(x_{(n-1)}, x_{(n)}]$.

Вероятностная мера полуоткрытого интервала $(x_{(1)}, x_{(n)}]$ равна, по частотному определению вероятности, $(n-1)/n$, поскольку один статистически эквивалентный блок, а именно, $(-\infty, x_{(1)}]$ не входит в интервал

$(x_{(1)}, x_{(n)}]$. Такой же результат мы получим при графическом определении этой же вероятности по выборочной функции распределения (см. разд. 2.2, рис. 26).

Вероятностная мера полуоткрытого интервала $(x_{(2)}, x_{(n-1)}]$ по той же причине равна $(n - 3)/n$, поскольку в этот интервал не входят статистически эквивалентные блоки $(-\infty, x_{(1)}]$, $(x_{(1)}, x_{(2)}]$, $(x_{(n-1)}, x_{(n)}]$.

Теперь уместно вспомнить, что нашей целью является определение границ такого доверительного интервала, который накрывает интерквартильный промежуток $J_P = (M[\xi] - g, M[\xi] + g]$ с вероятностью, не меньшей Q . Например, для нормального распределения $g = \sigma N_{0,5(1+P)}$. Понятно, что, как и в предыдущем разделе,

$$\underline{\text{tol}} \leq M[\xi] - g \quad \text{и} \quad M[\xi] + g \leq \overline{\text{tol}},$$

то есть нижний и верхний толерантные пределы должны охватывать истинный интерквартильный промежуток с вероятностью Q . Это обстоятельство эквивалентно тому, что в силу монотонности вероятностной меры (см. разд. 1.2.2) $P(M[\xi] - g < \xi \leq M[\xi] + g) \leq P(\underline{\text{tol}} < \xi \leq \overline{\text{tol}})$, то есть исходная задача трансформируется в следующую.

Необходимо найти такие значения $\underline{\text{tol}}$ и $\overline{\text{tol}}$, что вероятностная мера интервала между ними не меньше, чем вероятностная мера P искомого интерквартильного промежутка.

Из предыдущего материала настоящего раздела мы уже можем заключить, что вероятностная мера полуоткрытых интервалов, заключенных между элементами вариационного ряда, может быть просто определена путем подсчета относительного количества статистически эквивалентных блоков, находящихся (или попавших) в эти полуоткрытые интервалы.

По условию задачи требуется определить и обеспечить не точечную оценку этой вероятностной меры, а такое гарантированное значение этой меры, о котором с вероятностью, не меньшей Q , можно говорить, что истинное значение вероятности не меньше, чем заданное при определении искомого интерквартильного интервала, то есть P .

Материал разд. 2.4.5 предоставляет нам возможность решения этой задачи, которая является обратной по отношению к построению довери-

тельного интервала для вероятности. Основное отличие заключается в том, что в разд. 2.4.5 мы имели уже полученные выборочные значения и заданную доверительную вероятность. В данном случае нам задана нижняя граница вероятности в виде вероятности P , для которой определен искомый интерквантильный промежуток, и задача заключается в определении условий, при которых интервал между элементами вариационного ряда будет с вероятностью Q иметь вероятностную меру, не меньше, чем P .

Таким образом, неравенство для определения нижней границы доверительного интервала для вероятности, приведенное в разд. 2.4.5, а именно

$$\sum_{m=m_0}^n C_n^m (\underline{p})^m (1 - \underline{p})^{n-m} = \beta$$

имеет теперь для нас несколько иной, скорее обратный, смысл.

Поскольку нижняя граница нам задана и равна P , верхняя граница вероятности равна 1 и здесь нас не интересует, поэтому примем $Q = 1 - \alpha$. Теперь, руководствуясь настоящим неравенством, нам остается подобрать такие значения n и m_0 , чтобы обеспечить заданные значения P и Q .

Примеры точечных оценок вероятностной меры интервалов, заключенных между элементами вариационного ряда, свидетельствуют о том, что, по сути дела, это не что иное, как оценка вероятности по частоте, то есть по относительной частоте попадания (нахождения) статистически эквивалентных блоков внутри этих интервалов. На основании этого выясненного факта и, учитывая близость вероятностей P и Q к единице, перепишем последнее равенство в виде

$$\sum_{m=n-k}^n C_n^m P^m (1 - P)^{n-m} = 1 - Q,$$

где $n - k$ – число статистически эквивалентных блоков, находящихся внутри интервала между такими элементами вариационного ряда, которые желательно объявить толерантными пределами или, иными словами, границами доверительного интервала, который покрывает искомый интерквантильный промежуток, определенный при вероятности P .

В теории непараметрического интервального оценивания число k именуется, как количество отброшенных статистически эквивалентных

блоков. Понятно, что отбрасываемыми статистически эквивалентными блоками должны быть крайние блоки.

Это неравенство решают в двух вариантах постановки основной задачи:

1. Зафиксировано количество k статистически эквивалентных блоков, не попавших между теми элементами вариационного ряда, которые желательно принять в качестве толерантных пределов, и отыскивается минимальный объем выборки, необходимый для обеспечения заданных параметров P и Q .

2. Зафиксирован объем выборки, отыскивается количество k статистически эквивалентных блоков, которые необходимо отбросить, чтобы оставшиеся крайние члены вариационного ряда принять в качестве толерантных пределов, обеспечивающих заданные значения параметров P и Q .

В большинстве случаев задача ставится и решается в первом варианте, и мы вскоре увидим, почему.

Пусть при подготовке испытаний (измерений) в качестве непараметрических толерантных пределов планируется использовать крайние члены вариационного ряда. Это решение может быть принято при практическом отсутствии факторов, способных привести к резким выбросам отдельных результатов. На нашем языке это означает, что из всех статистически эквивалентных блоков отбрасывается один первый блок $(-\infty, x_{(1)}]$, и доверительным интервалом оказывается полуоткрытый интервал $(x_1, x_n]$. Необходимый объем выборки n находят как наименьшее решение неравенства

$$\sum_{m=n-1}^n C_n^m P^m (1-P)^{n-m} = P^{n-1} [n(1-P) + P] \leq 1 - Q.$$

При необходимости защиты от возможных импульсных помех или иных факторов, вызывающих резкие и значительные выбросы результатов измерений, пытаются принять в качестве толерантных пределов (доверительного интервала для интерквантильного промежутка) интервал $(x_{(2)}, x_{(n-1)}]$, что означает отбрасывание трех статистически эквивалентных блоков. Необходимый объем выборки n находят как наименьшее решение неравенства

$$\sum_{m=n-3}^n C_n^m P^m (1-P)^{n-m} \leq 1-Q.$$

Понятно, что в этом случае при фиксированных P и Q объем выборки должен возрасти.

В таблице 4 приведены результаты расчетов объема выборки, минимально необходимого для определения непараметрических толерантных пределов – границ доверительного интервала для интерквантильного промежутка J_P при $P = 0.95$ и $k = 1, 2, 3$.

Таблица 4

**Минимально необходимый объем выборки
для нахождения непараметрических толерантных пределов при $P = 0.95$**

Q	0,8	0,9	0,95	0,99	k
n	32	45	59	90	0
n	59	77	93	130	1
n	84	105	124	165	2
n	109	132	153	197	3

За независимость от вида плотности распределения исследуемой генеральной совокупности приходится “платить” существенным увеличением объема выборки (см. табл. 4). Кроме того, становится ясно, почему в качестве толерантных пределов выбираются члены вариационного ряда, расположенные поближе к его краям.

Ранее подробно был рассмотрен случай нахождения непараметрических толерантных пределов для интервального оценивания интерквантильного промежутка J_P , симметричного относительно математического ожидания и показанного на рис. 10, а в разд. 1.6.2.

В ряде случаев, когда математическое ожидание случайной величины невелико по сравнению с ее среднеквадратическим значением, объем выборки может быть уменьшен без ущерба для достоверности оценок. Как правило, такая ситуация встречается при экспериментальном определении характеристик погрешности средств измерений или результатов измерений, или характеристик точности регуляторов, позиционером, систем автоматического управления и регулирования, когда в составе погрешности

этих устройств имеются и случайные, и систематические составляющие, но в документации на эти средства установлены нормы на сумму этих составляющих. Пример такого интерквантильного промежутка приведен в разд. 1.6.2 (см. на рис. 10, б). Он обозначен там \hat{J}_P и имеет границы $[-g, g]$. В подобных случаях рекомендуется поступать следующим образом.

Исходная выборка значений погрешностей $\Delta_1, \Delta_2, \dots, \Delta_n$, среди которых имеются положительные и отрицательные значения, преобразуется в выборку, состоящую из абсолютных величин этих исходных значений. В результате весь вариационный ряд оказывается на правой полуоси, что снимает заботу об отбрасывании первого статистически эквивалентного блока, а именно, $(0, |\Delta|_{(1)}]$, поскольку обе границы доверительного интервала для интерквантильного промежутка, симметричного относительно начала координат, определяются конечными членами вариационного ряда, составленного из модулей выборочных значений $|\Delta|_{(i)}$.

В этом случае толерантные пределы $[-|\Delta|_{(n)}, |\Delta|_{(n)}]$, составленные на основе одного лишь максимального члена этого вариационного ряда, являются границами доверительного интервала для интерквантильного промежутка \hat{J}_P (см. разд. 1.6.2, рис.10, б), если, конечно, объем выборки достаточен для этого. Поскольку в этом конкретном случае не отбрасывается ни один статистически эквивалентный блок (ибо первый, как уже было сказано, нас не интересует, а последним блоком является $(|\Delta|_{(n-1)}, |\Delta|_{(n)}]$, минимально необходимый объем выборки находят при $k = 0$ из неравенства

$$\sum_{m=n}^n C_n^m P^m (1-P)^{n-m} = P^n \leq 1-Q.$$

Этому варианту соответствует первая строчка в табл. 4.

Понятно, что при желании застраховаться от вредного влияния грубых промахов или помех количество отбрасываемых членов может быть увеличено при том же объеме выборки. При этом доверительная вероятность не изменится, а объем выборки, необходимый при отбрасывании од-

ного, двух и более крайних членов, находится из последующих строк данной таблицы. Видно, что достигается существенная экономия объема выборки. “Платой” за это является расширение доверительного интервала.

2.4.7. Бутстреп – оценивание

*Бутстреп*¹-оценивание – метод нахождения доверительных интервалов для числовых характеристик случайных величин и любых параметров плотности распределения, не зависящий от вида плотности распределения генеральной совокупности из которой извлечена выборка. Этот метод предложен Б. Эфроном и описан в его книге [10] а также в серии статей отечественных авторов, опубликованных в журнале [11].

Метод заключается в следующем.

Пусть имеется выборка $x^T = (x_1 \ x_2 \ \dots \ x_n)$, извлеченная из генеральной совокупности X , плотность распределения которой $\varphi(\Theta, x)$, где Θ – числовая характеристика или параметр, для которого по этой выборке – необходимо определить границы доверительного интервала.

Согласно бутстреп-методу, из имеющейся выборки конструируются бутстреп-выборки путем комбинации исходных выборочных значений при условии, что бутстреп-выборки должны отличаться хотя бы одним элементом. Порядок расположения элементов бутстреп-выборок не имеет значения. Объем бутстреп-выборок может быть различным, но, по построению, он будет обязательно меньше n . По каждой из бутстреп-выборок вычисляется оценка числовой характеристики или параметра Θ .

Обозначим бутстреп-выборки b_1, b_2, \dots, b_r , объемы этих выборок – n_1, n_2, \dots, n_r , оценки параметра – $\tilde{\Theta}_1, \tilde{\Theta}_2, \dots, \tilde{\Theta}_r$. Объединим все эти обозначения в табл. 5.

Из полученных оценок строится вариационный ряд

$$\tilde{\Theta}_{(1)}, \tilde{\Theta}_{(2)}, \dots, \tilde{\Theta}_{(r)},$$

и задача определения границ доверительного интервала для параметра Θ при доверительной вероятности Q ставится, как задача оценки интеркван-

¹ Bootstrep (англ) – петля на верхней части голенища сапога, с помощью которой этот сапог надевается

тельного промежутка J_Q с помощью непараметрических толерантных пределов (см. разд. 2.4.6.2.), если, конечно, количество членов вариационного ряда достаточно для этого. Поскольку в этой процедуре не могут возникать грубые промахи и резкие выбросы значений оценок $\tilde{\Theta}_j$, здесь не имеет смысла страховаться от них, поэтому достаточно использовать

Таблица 5

Бутстреп-выборки и оценки

Обозначение выборки	Бутстреп-выборка	Объем выборки	Оценка числовой характеристики (параметра)
b_1	\mathbf{x}_1	n_1	$\tilde{\Theta}_1$
b_2	\mathbf{x}_2	n_2	$\tilde{\Theta}_2$
\vdots	\vdots	\vdots	\vdots
b_r	\mathbf{x}_r	n_r	$\tilde{\Theta}_r$

только крайние члены вариационного ряда, то есть только первую строчку табл. 4 (при оценивании интервала с симметричными границами) или вторую строчку (при оценивании правой и левой границы по отдельности). В качестве доверительной вероятности в данной ситуации принимается вероятность π из интервала значений (0.8 – 0.9), которая в дальнейшем не фигурирует, как характеристика качества оценки параметра Θ .

Таким образом, для реализации бутстреп-метода оценивания используются непараметрические толерантные пределы из п. 2.4.6.2. применительно к вариационному ряду $\tilde{\Theta}_{(1)}, \tilde{\Theta}_{(2)}, \dots, \tilde{\Theta}_{(r)}$ со следующей заменой: вероятность P (см. п. 2.4.6.2). заменяется на вероятность Q настоящего пункта, в качестве вероятности Q (см. п. 2.4.6.2) назначается вероятность π . Доверительной вероятностью для найденного доверительного интервала параметра считается вероятность Q .

На первый взгляд, метод кажется некорректным.

По теореме Шеннона, никакое преобразование информации не приводит к увеличению ее количества (эта теорема логически согласуется с по-

ложением физики о невозможности создания вечного двигателя). В данном случае, когда для нахождения доверительного интервала требуется значительное количество информации (как мы убедились хотя бы в предыдущем пункте), здесь она как будто добывается путем простого комбинирования элементов исходной выборки, информации в которой до того было недостаточно для решения той же задачи. И здесь не в пользу бутстреп-метода, но в согласии с английским бытовым значением этого слова возникают ассоциации с «опытом» барона Карла Иеронима Фридриха фон Мюнхгаузена самовынимания себя вместе с лошастью из болота за собственный парик.

Серьезные сомнения в корректности этого метода широко распространены в среде математиков – специалистов в области теории вероятностей и математической статистики.

Тем не менее публикации практических результатов применения этого метода и результаты статистического моделирования свидетельствуют о работоспособности метода в определенных условиях.

2.5. Статистические методы проверки гипотез

2.5.1. Постановка задачи и общие принципы

Задача проверки гипотез заключается в проверке согласования теоретических априорных предположений об объекте исследования (испытаний, управления, регулирования) с опытными данными в условиях действия случайных факторов. Первичными являются опытные данные, как фактически полученные, при условии, что эти данные получены с помощью исправных средств корректными методами.

В условиях действия случайных факторов методы проверки гипотез являются статистическими, основанными на применении теоретического аппарата теории вероятностей и математической статистики.

Обозначения гипотез:

H_0 – нулевая гипотеза, как правило, априорное предположение исследователя,

H_1 – гипотеза, альтернативная гипотезе H_0 .

Говорят о проверке гипотезы H_0 против H_1 .

Теория проверки гипотез в математической статистике и все выполняемые операции базируются на априорном предположении о справедливости нулевой гипотезы.

Основной принцип, лежащий в основе статистических методов проверки гипотез, идеологически близок принципу, по которому строились доверительные интервалы в разд. 2.4, например, доверительный интервал для вероятности в разд. 2.4.2.

Этот принцип заключается в следующем. Если при справедливости гипотезы H_0 вероятность появления полученных экспериментальных данных не слишком мала, то говорят об *отсутствии достаточных оснований для отклонения этой гипотезы*.

Напротив, гипотезу H_0 считают недостаточно обоснованной, если при предположении о ее справедливости появление полученных данных маловероятно. Это означает, что экспериментальные данные не подтверждают справедливость нулевой гипотезы. В таких случаях говорят об *отсутствии достаточных оснований для признания справедливости гипотезы H_0* .

Вообще *категорические выводы о бесспорной справедливости или несправедливости нулевой или альтернативной ей гипотезы при условии действия случайных факторов в математической статистике делать нельзя*.

2.5.2. Критическая область и критическое значение

Пусть $\mathbf{x}^T = (x_1 \ x_2 \ \dots \ x_n)$ – выборка, изъятая из генеральной совокупности \mathbf{X} , плотность распределения которой $\varphi(\Theta, \mathbf{x})$. Для проверки гипотезы H_0 против H_1 вычисляется некоторая функция от выборочных значений $z(x_1, x_2, \dots, x_n) = z(\mathbf{x})$, вид которой определяется характером проверяемой гипотезы и параметром (характеристикой), относительно которого выдвигается гипотеза.

Область определения этой функции – все значения, которые образуют генеральную совокупность \mathbf{X} .

Область значений функции $z(\mathbf{x})$ $Z : \{ z(\mathbf{x}); x_1, x_2, \dots, x_n \in \mathbf{X} \}$ делится на два непересекающихся множества Z_0 и $Z_{\text{крит}}$ так, что $Z_0 \cup Z_{\text{крит}} = Z$, $Z_0 \cap Z_{\text{крит}} = \emptyset$.

Обычно областью определения функции $z(\mathbf{x})$ является вся вещественная ось, или непрерывный ее отрезок, а каждая из областей Z_0 и $Z_{\text{крит}}$ – непрерывные части этого отрезка, разделенные точкой или третьим непрерывным отрезком. Точка, разделяющая эти области отображает на оси *критическое значение*, которое обозначается, как $z_{\text{крит}}$. Совокупность функции $z(\mathbf{x})$ и множеств $Z_0, Z_{\text{крит}}$ называется *критерием проверки гипотезы H_0 против H_1* , функция $z(\mathbf{x})$ – *статистика критерия*, множество $Z_{\text{крит}}$ – *критическая область* или *критическое множество*. Понятно, что будучи функцией от выборочных данных, статистика критерия является случайной функцией.

$Z_{\text{крит}}$ выбирается так, чтобы при предположении о справедливости гипотезы H_0 условная вероятность попадания статистики критерия в критическую область $P\{z(\mathbf{x}) \in Z_{\text{крит}} / H_0\}$ была небольшой.

Если оказалось, что при выдвинутом предположении о справедливости гипотезы H_0 и небольшой заранее назначенной вероятности статистика критерия попала в критическую область, что соответствует вероятности $P\{z(\mathbf{x}) \in Z_{\text{крит}} / H_0\}$, то делается вывод о том, что, повидимому, экспериментальные данные не подтверждают справедливость нулевой гипотезы. В этой ситуации говорят, что *нет достаточных оснований для признания справедливости гипотезы H_0* . В противном случае, когда статистика критерия попала в область Z_0 , говорят, что *нет достаточных оснований для отклонения гипотезы H_0* . Однако, из-за действия случайных факторов, особенно, если объем выборки недостаточен, возможны ошибки в таких суждениях, и вероятности этих ошибок необходимо, по крайней мере, знать.

В теории и практике статистических методов проверки гипотез используются следующие определения и понятия:

$P\{z(\mathbf{x}) \in Z_{\text{крит}} / H_0\} = P\{H_1 / H_0\} = \alpha$ – вероятность ошибки первого рода, или риск поставщика, или риск продавца, или уровень значимости, (может применяться любой из этих терминов применительно к конкретной ситуации),

$P\{z(\mathbf{x}) \in Z_0 / H_1\} = P\{H_0 / H_1\} = \beta$ – вероятность ошибки второго рода, или риск заказчика, или риск покупателя.

Естественно, что

$$P\{z(\mathbf{x}) \in Z_0 / H_0\} = P\{H_0 / H_0\} = 1 - \alpha,$$

$P\{z(\mathbf{x}) \in Z_{\text{крит}} / H_1\} = P\{H_1 / H_1\} = 1 - \beta$ – называется мощностью критерия проверки гипотез.

Вероятность $\alpha + \beta$ есть вероятность ошибки.

Вероятности α и β обычно выбираются в каждом конкретном случае из экономических или технических соображений, а также из соображений безопасности.

Приведем пример. При контроле выпускаемых изделий и действии случайных факторов вероятность α есть не что иное, как относительное количество годных изделий, ошибочно забракованных и направленных на утилизацию, или на дополнительную регулировку, или в ремонт. Ущерб производителя складывается из напрасно затраченных средств на изготовление, ремонт, регулировку изделия и нереализованной прибыли от его продажи.

Вероятность β – это не что иное, как относительное количество дефектных изделий, ошибочно выпущенных на реализацию, как годные. Потребитель, если он не предъявил рекламацию, несет прямой ущерб. Если же потребитель предъявил рекламацию, то ущерб несет производитель в размере затрат на транспортировку, ремонт, регулировку изделия, и кроме того производитель несет ущерб от утраты положительной репутации на рынке подобных изделий.

2.5.3. Простые гипотезы

Простой называется гипотеза H_0 о значении числовой характеристики или параметра, когда область значений, предполагаемых этой гипоте-

зой, и область альтернативных значений - каждая из них является одно-точечной.

В других случаях гипотеза называется сложной.

Пример простой гипотезы.

Плотность распределения генеральной совокупности – нормальна, дисперсия известна и равна σ^2 . Из генеральной совокупности X извлечена выборка x_1, x_2, \dots, x_n . Проверяется простая гипотеза

$$H_0: \mu_1[\xi] = a \text{ против альтернативы } H_1: \mu_1[\xi] = b.$$

Примеры необходимости проверки подобного рода гипотез в технике: при цифровой передаче данных в условиях действия сильных помех с целью различения передаваемых двоичных символов 0 и 1;

в системах управления технологическими процессами с целью выявления состояния двухпозиционных устройств, контроля исполнения дискретных управляющих воздействий, обнаружения разладки технологического процесса и т.д.

Для проверки этой гипотезы подходящей статистикой является среднее арифметическое значение, плотность распределения которого, как мы уже выяснили, также нормальна с дисперсией σ^2/n , а математическое ожидание равно a или b – в зависимости от того, какая из двух гипотез реально имеет место.

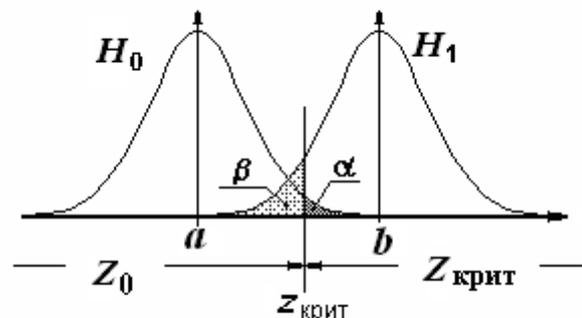


Рис. 34. К проверке простой гипотезы

В случае справедливости гипотезы H_0 $\bar{x} \in N(a, \sigma^2/n)$, в случае справедливости гипотезы H_1 $\bar{x} \in N(b, \sigma^2/n)$. Эти две плотности распределения статистики $z(\mathbf{x}) = \bar{x}$ приведены на рис. 34. Они пересекаются, и стоит задача определения границы областей Z_0 и $Z_{\text{крит}}$, то есть критического значения $z_{\text{крит}}$, чтобы обеспечить требуемые значения вероятностей ошибочных решений α и β . Эти вероятности показаны на рис.34, как площади

под кривыми плотностей распределения статистики, соответствующих гипотезам H_0 и H_1 .

В самом деле, если гипотеза H_0 справедлива, среднее арифметическое значение из-за действия случайных факторов может попасть в критическую область с вероятностью $\alpha = P\{\bar{x} \in Z_{\text{крит}} / \mu_1[\xi] = a\}$. Тогда, если так случится, будет сделан ошибочный вывод, который приведет к отклонению этой, на самом деле справедливой гипотезы.

В противном случае, когда справедлива гипотеза H_1 , среднее арифметическое значение из-за действия случайных факторов может попасть в область Z_0 с вероятностью $\beta = P\{\bar{x} \in Z_0 / \mu_1[\xi] = b\}$, что также приведет к ошибочному выводу.

Видно (см. рис. 34), что перемещением границы между областями Z_0 и $Z_{\text{крит}}$ можно устанавливать желательное соотношение между вероятностями α и β . Но может оказаться, что их значения слишком велики, и перемещение критического значения не приводит к их снижению. В таком случае необходимо увеличивать объем выборки. В результате дисперсия среднего арифметического уменьшится, его плотность распределения сузится, общая площадь под обеими кривыми уменьшится и, следовательно, уменьшатся вероятности ошибочных решений.

Этот пример показывает, что при проверке простой гипотезы контролируются значения обеих вероятностей ошибочных решений: α и β .

2.5.4. Сложные гипотезы

Пример сложной гипотезы. Плотность распределения генеральной совокупности нормальная, дисперсия известна и равна σ^2 . Из генеральной совокупности \mathbf{X} извлечена выборка x_1, x_2, \dots, x_n . Проверяется сложная гипотеза

$$H_0: \mu_1[\xi] < a \text{ против альтернативы } H_1: \mu_1[\xi] > a.$$

Примеры необходимости проверки подобного рода гипотез:

выявление разладки технологического процесса по параметру a ;

распознавание перехода значения регулируемого параметра через уставку;

выявление превышения некоторой характеристикой установленной нормы (при медицинской диагностике, экологическом мониторинге, контроле качества пищевых продуктов, при сертификационных и иных испытаниях изделий и т.д.).

Поскольку проверяются гипотезы о математическом ожидании, как и в предыдущем разделе, подходящей статистикой для проверки этой гипотезы является среднее арифметическое, плотность распределения которого нормальна с дисперсией σ^2 / n .

На рис. 35 приведены примеры двух вариантов расположения плотностей распределения средних арифметических значений, соответствующих гипотезам H_0 и H_1 : при значительном удалении друг от друга истинных значений математических ожиданий (см. рис. 35, а) и при близком их расположении.

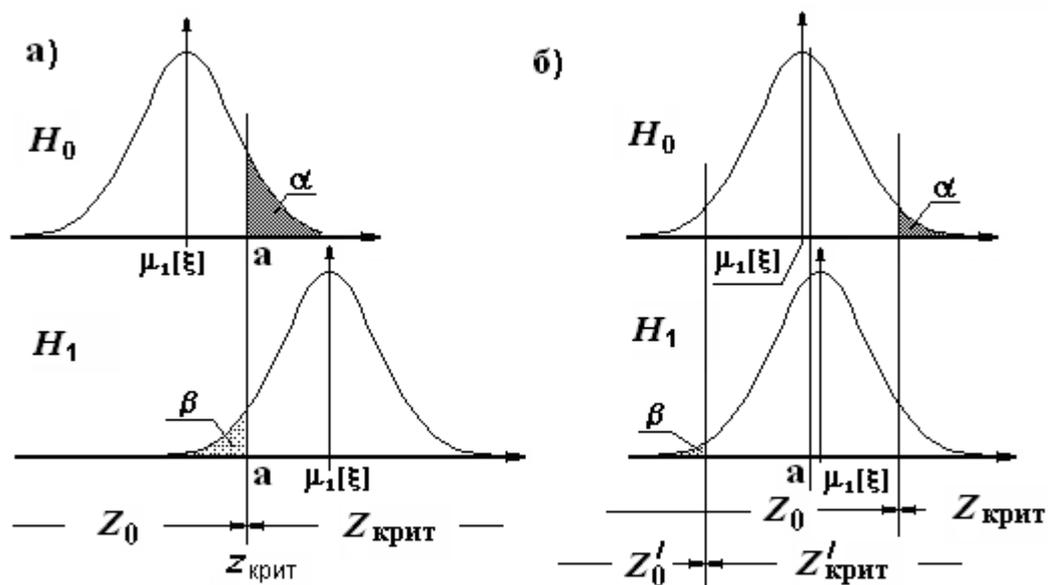


Рис. 35. К проверке сложной гипотезы

Если в первом варианте в качестве критического значения выбрать значение a , и сравнивать с ним получающееся среднее арифметическое значение, то возможно достижение достаточно малых значений вероятностей ошибочных решений α и β . Однако, формулировка сложной гипотезы допускает вариант, показанный на рис. 35, б. В этой ситуации, если крити-

ческим значением является a , вероятности α и β сколь угодно близки к 0,5, что для практики недопустимо. Можно уменьшить вероятность ошибки первого рода α путем перемещения границы между областями Z_0 и $Z_{\text{крит}}$ вправо. Но тогда вероятность ошибки второго рода $\beta = P\{\bar{x} \in Z_0 / \mu_1[\xi] > a\}$ будет недопустимо близка к единице, поскольку в этом случае она равна площади под кривой плотности распределения статистики \bar{x} , соответствующей гипотезе H_1 , по всей области Z_0 , то есть области, при попадании \bar{x} в которую принимается решение о непротиворечивости экспериментальным данным гипотезы H_0 .

Попытка уменьшить значение вероятности β путем перемещения влево критического значения, то есть границы между областями Z_0 и $Z_{\text{крит}}$ (на рис. 35 эти области обозначены, как Z_0' и $Z_{\text{крит}}'$), приводит к недопустимо близкому к единице значению вероятности ошибки первого рода $\alpha = P\{\bar{x} \in Z_{\text{крит}}' / \mu_1[\xi] < a\}$.

Из этого следует, что при проверке сложных гипотез может контролироваться только одна из вероятностей ошибочного решения: либо α , либо β . Значение другой из них не гарантируется и может быть сколь угодно близким к единице.

Это обстоятельство является существенным недостатком описанного метода проверки сложных гипотез.

2.5.5. Проверка гипотезы о виде плотности распределения

2.5.5.1. Критерий “хи-квадрат”

Из генеральной совокупности X , образованной случайной величиной ξ , извлечена выборка x_1, x_2, \dots, x_n . Выдвигается предположение о том, что плотность распределения случайной величины есть $\varphi(\Theta, x)$, где Θ – вектор параметров. Для проверки этого предположения по выборочным данным вычисляются оценки параметров $\tilde{\Theta}$ и проверяется сложная гипотеза:

H_0 : плотность распределения случайной величины ξ есть $\varphi(\tilde{\Theta}, x)$
против альтернативы

H_1 : плотность распределения случайной величины ξ не $\varphi(\tilde{\Theta}, x)$.

Поскольку эта гипотеза сложная, задается только вероятность ошибки первого рода α , которая в подобных случаях именуется *уровнем значимости*.

Для проверки этой гипотезы строится гистограмма на K интервалах одинаковой ширины, $|x_k - x_{k-1}| = \Delta x = \text{const}$, $k=1, 2, \dots, K$.

Степень различия между гистограммой и предполагаемой плотностью распределения выражается суммой квадратов разностей

$$\sum_{k=1}^K (P_k - \tilde{P}_k)^2 = \sum_{k=1}^K \left(\frac{nP_k - n_k}{n} \right)^2,$$

где

$$P_k = \int_{x_{k-1}}^{x_k} \varphi(\tilde{\Theta}, x) dx,$$

то есть вероятность попадания значения случайной величины в интервал $(x_{k-1}, x_k]$ при условии справедливости нулевой гипотезы, $\tilde{P}_k = n_k / n$ - оценки этих вероятностей, где n_k - количество выборочных значений, попавших в интервал $(x_{k-1}, x_k]$, n - объем выборки, K - общее количество интервалов, на которых построена гистограмма.

Каждое слагаемое этой суммы является случайной величиной, поскольку случайным является число n_k . Если выборочные значения независимы, A_k - событие, которое заключается в том, что выборочное значение попадает в интервал $(x_{k-1}, x_k]$, \bar{A}_k - противоположное событие. Поэтому в соответствии со схемой Бернулли вероятность того, что при n экспериментах событие A_k произойдет ровно n_k раз, равна (см. разд. 1.3.2)

$$P_n(n_k) = C_n^{n_k} P_k^{n_k} (1 - P_k)^{n - n_k}.$$

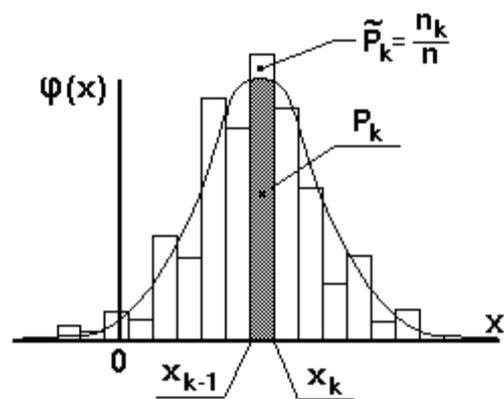


Рис. 36. Гистограмма и предполагаемая плотность распределения

Из результатов, полученных в примере разд. 1.3.5, следует, что

$$M[n_k] = nP_k, \quad D[n_k] = nP_k(1 - P_k).$$

Пользуясь формулами для моментов линейных функций от случайных величин, приведенными в разд. 1.3.4, можем записать, что

$$M\left[\frac{nP_k - n_k}{n}\right] = 0, \quad D\left[\frac{nP_k - n_k}{n}\right] = \frac{1}{n^2} D[n_k] = \frac{1}{n} P_k(1 - P_k).$$

Преобразуем исходную сумму путем деления каждого из слагаемых на его дисперсию. Получим сумму

$$\sum_{k=1}^K \left(\frac{nP_k - n_k}{n\sqrt{P_k(1 - P_k)}} \sqrt{n} \right)^2 = \sum_{k=1}^K \left(\frac{nP_k - n_k}{\sqrt{nP_k(1 - P_k)}} \right)^2.$$

Видно, что после этого деления

$$M\left[\frac{nP_k - n_k}{\sqrt{nP_k(1 - P_k)}}\right] = 0, \quad D\left[\frac{nP_k - n_k}{\sqrt{nP_k(1 - P_k)}}\right] = 1.$$

Строго говоря, случайная величина

$$\frac{nP_k - n_k}{\sqrt{nP_k(1 - P_k)}}$$

является дискретной из-за того, что порождена дискретной случайной величиной n_k , распределенной по биномиальному закону. При дискретности значений величины n_k , равной 1, дискретность значений вновь сформированной случайной величины равна $\frac{1}{\sqrt{nP_k(1 - P_k)}}$, и с ростом n убывает до нуля. Поэтому можно говорить, что эта случайная величина в асимптотике при $n \rightarrow \infty$ становится непрерывной.

С другой стороны, по теореме Муавра-Лапласа (см. разд. 1.3.7), распределение вероятностей случайной величины n_k при $n \rightarrow \infty$ аппроксимируется значениями

$$\frac{1}{\sqrt{2\pi}\sqrt{nP_k(1 - P_k)}} \exp\left\{-\frac{(nP_k - n_k)^2}{2nP_k(1 - P_k)}\right\}.$$

После выполненных преобразований и с учетом того, что при $n \rightarrow \infty$ дискретность значений случайной величины

$$\frac{nP_k - n_k}{\sqrt{nP_k(1 - P_k)}}$$

уменьшается до нуля, мы имеем право говорить, что эта случайная величина распределена асимптотически нормально с параметрами $(0, 1)$, то есть

$$\frac{nP_k - n_k}{\sqrt{nP_k(1 - P_k)}} \underset{n \rightarrow \infty}{\in} N(0, 1).$$

Как известно из разд. 2.3.4.2, в), плотность распределения суммы квадратов таких случайных величин есть плотность распределения хи-квадрат. Таким образом, окончательно можем записать формулу для вычисления статистики критерия “хи-квадрат”, плотность распределения которой при условии справедливости нулевой гипотезы есть плотность распределения хи-квадрат с числом степеней свободы $K - r$, где K – количество слагаемых в сумме (то есть число интервалов, на которых построена гистограмма), r – число параметров предполагаемой плотности распределения, которые были определены по выборочным данным (то есть число связей, наложенных на выборочные данные):

$$\sum_{k=1}^K \frac{(nP_k - n_k)^2}{nP_k} \in \chi^2(K - r).$$

Поскольку, как правило, $P_k \ll 1$ сомножитель $(1 - P_k)$ в знаменателях слагаемых опущен.

Подобный функционал был использован нами ранее в разд. 2.3.6 для нахождения оценок параметров плотности распределения методом минимума χ^2 .

При заданной вероятности ошибки первого рода α , здесь – уровня значимости, критическое значение $z_{\text{крит}}$ (нижняя граница критической области $Z_{\text{крит}}$) назначается из следующих соображений.

При справедливости нулевой гипотезы маловероятно, чтобы статистика критерия оказалась слишком большой. Ограничимся таким критическим значением, вероятность превышения которого будет не более заданного значения α . Поскольку нам известно, что при условии справедливости нулевой гипотезы статистика критерия распределена приблизительно

по закону χ^2 , мы можем принять в качестве критического значения $(1 - \alpha)100$ – процентную квантиль $\chi_{1-\alpha}^2(K - r)$.

Таким образом, сформирован критерий “хи-квадрат” проверки гипотезы о виде плотности распределения (или закона распределения) генеральной совокупности по экспериментальным данным.

Процедура проверки гипотезы о виде плотности распределения по критерию “хи-квадрат”.

1. Задают уровень значимости α
2. По выборочным данным строят гистограмму в соответствии с указаниями разд. 2.2.

3. Вычисляются точечные оценки моментов: m_1, s^2, m_3, m_4 .

4. Из теоретических соображений, по виду гистограммы, по соотношениям между моментами, по значениям асимметрии и эксцесса выдвигается гипотеза о виде плотности распределения $\varphi(\Theta, x)$.

5. Вычисляются оценки r параметров $\tilde{\Theta}$ предполагаемой плотности распределения, в результате будет получена плотность распределения $\varphi(\tilde{\Theta}, x)$.

6. С использованием $\varphi(\tilde{\Theta}, x)$ вычисляются вероятности

$$P_k = \int_{x_{k-1}}^{x_k} \varphi(\tilde{\Theta}, x) dx.$$

7. Вычисляется статистика критерия

$$\chi^2 = \sum_{k=1}^K \frac{(nP_k - n_k)^2}{nP_k}.$$

8. Полученное значение сравнивается с критическим значением

$$\chi_{1-\alpha}^2(K - r),$$

где r – количество оцениваемых параметров предполагаемой плотности распределения $\varphi(\tilde{\Theta}, x)$.

9. Если $\chi^2 > \chi_{1-\alpha}^2(K - r)$, делается вывод о том, что экспериментальные данные не подтверждают справедливость выдвинутой гипотезы или о том, что отсутствуют достаточные основания для того, чтобы считать ну-

левую гипотезу справедливой. Гипотеза пересматривается, выдвигается новая нулевая гипотеза, переход на п. 4 настоящей процедуры.

10. Если $\chi^2 < \chi_{1-\alpha}^2(K-r)$, делается вывод о том, что экспериментальные данные подтверждают справедливость выдвинутой гипотезы или о том, что имеются достаточные основания для того, чтобы считать нулевую гипотезу справедливой.

Сделаем замечание о том, что с уменьшением вероятности α возрастает критическое значение $\chi_{1-\alpha}^2(K-r)$, а это значит, что объективно растет вероятность ошибочного подтверждения нулевой гипотезы, когда она неверна. Крайний случай иллюстрирует это положение: если задать $\alpha = 0$, то критическое значение $\chi_1^2(K-r) = \infty$, а это означает, что нулевая гипотеза, какой бы она ни была, не будет подвергаться сомнению ни при каком значении статистики критерия.

2.5.5.2. Критерий Колмогорова – Смирнова

Из генеральной совокупности \mathbf{X} , образованной случайной величиной ξ , извлечена выборка x_1, x_2, \dots, x_n . По этим данным строится выборочная функция распределения $\tilde{F}(x)$, как это описано в разд. 2.2. По виду выборочной функции распределения выдвигается предположение о том, что функция распределения есть $F(\Theta, x)$, где Θ – вектор параметров. По выборочным данным вычисляются оценки параметров $\tilde{\Theta}$, по соотношениям между ними уточняется вид функции распределения, и, если это нужно, ранее выдвинутое предположение уточняется. Проверяется сложная гипотеза

H_0 : функция распределения случайной величины ξ есть $F(\Theta, x)$

против альтернативы

H_1 : функция распределения случайной величины ξ не $F(\Theta, x)$.

Поскольку эта гипотеза сложная, задается только вероятность ошибки первого рода α , которая в подобных случаях именуется *уровнем значимости*.

В соответствии с формулировкой гипотезы сравниваются две функции распределения: выборочная (см. разд. 2.2) и предполагаемая, пред-

ставленные на рис. 37. Различие между ними определено, как

$$D = \sup_i \left| \tilde{F}(x_{(i)}) - F(\tilde{\Theta}, x_{(i)}) \right|,$$

где $F(x_{(i)})$ – значения выборочной функции распределения при $x = x_{(i)}$.

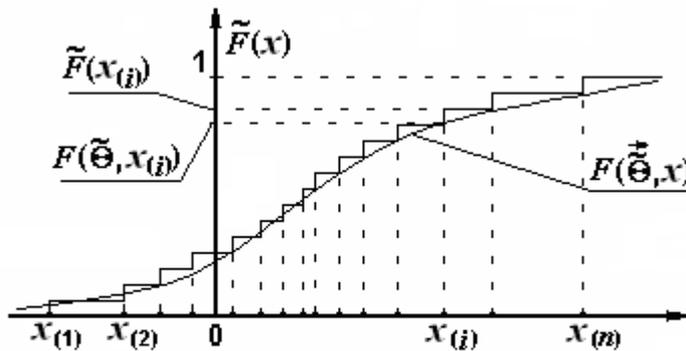


Рис. 37. Выборочная и предполагаемая функции распределения

Статистикой критерия является величина D . Критические значения табулированы. Таблицы критических значений D_α как функций от вероятности α ,

приводятся практически во всех учебниках и справочниках по математической статистике. В таблице 6 приводятся некоторые часто употребляемые критические значения.

Таблица 6

Критические значения критерия Колмогорова – Смирнова

α	$n = 25$	$n = 50$	$n = 80$	$n = 100$
0,2	0,209	0,148	0,118	0,106
0,1	0,237	0,169	0,135	0,121
0,05	0,264	0,188	0,150	0,134

Если $n > 10$ и вероятность α выражена в относительных единицах, для расчета критических значений можно пользоваться приближенной формулой

$$D_\alpha \approx \sqrt{\frac{-\ln 0,5\alpha}{2n}} - \frac{1}{6n}.$$

Процедура проверки гипотезы о виде функции распределения по критерию Колмогорова-Смирнова.

1. Задается уровень значимости α .
2. По выборочным данным строят выборочную функцию распределения в соответствии с указаниями разд. 2.2.

3. Вычисляют точечные оценки моментов : m_1, s^2, m_3, m_4 .

4. Из теоретических соображений, по виду выборочной функции распределения, по соотношениям между моментами, по значениям асимметрии и эксцесса, по результатам анализа других данных выдвигается гипотеза о виде функции распределения $F(\Theta, x)$ и тем самым – о виде плотности распределения $\varphi(\Theta, x)$.

5. Вычисляют r параметров предполагаемой функции распределения и ее значения $F(\tilde{\Theta}, x)$ при $x = x_{(i)}$.

6. Вычисляют статистику критерия $D = \sup_i |\tilde{F}(x_{(i)}) - F(\tilde{\Theta}, x_{(i)})|$.

7. Полученное значение сравнивают с критическим значением D_α .

8. Если $D > D_\alpha$ делают вывод о том, что экспериментальные данные не подтверждают справедливость выдвинутой гипотезы или о том, что отсутствуют достаточные основания для того, чтобы считать нулевую гипотезу справедливой. Гипотезу пересматривают, выдвигают новую нулевую гипотезу и переходят на п. 4 данной процедуры.

9. Если $D \leq D_\alpha$ делают вывод о том, что экспериментальные данные не противоречат выдвинутой гипотезе или о том, что имеются достаточные основания для того, чтобы считать нулевую гипотезу справедливой.

В некоторых источниках приводится такая рекомендация. Для корректного применения критерия Колмогорова – Смирнова: исходная выборка x_1, x_2, \dots, x_n делится на две части. По одной из них определяют параметры $\tilde{\Theta}$, по другой – строят выборочную и предполагаемую функции распределения и вычисляют статистику критерия. Это позволяет избавиться от необходимости учета зависимости между выборочными значениями, которая появляется в результате вычисления параметров предполагаемой плотности распределения и соответствующего уменьшения числа степеней свободы, как это было при использовании критерия χ^2 . В качестве n в этом случае принимается объем второй выборки.

2.5.5.3. Критерий ω^2 Мизеса

Из генеральной совокупности \mathbf{X} , образованной случайной величиной ξ , извлечена выборка x_1, x_2, \dots, x_n . Выдвигается предположение о том, что функция распределения случайной величины есть $F(\Theta, x)$, где Θ – вектор параметров. По выборочным данным вычисляют оценки параметров $\tilde{\Theta}$ и проверяют сложную гипотезу

H_0 : функция распределения случайной величины ξ есть $F(\Theta, x)$

против альтернативы

H_1 : функция распределения случайной величины ξ не $F(\Theta, x)$.

Поскольку эта гипотеза сложная, задают только вероятность ошибки первого рода α , которая в подобных случаях именуется *уровнем значимости*.

В соответствии с формулировкой гипотезы сравнивают две функции распределения: выборочную (разд. 2.2) и предполагаемую. Различие между ними определено, как

$$\omega^2 = \int_{-\infty}^{\infty} |\tilde{F}(x) - F(\tilde{\Theta}, x)|^2 \varphi(\tilde{\Theta}, x) dx,$$

где $\varphi(\tilde{\Theta}, x)$ – предполагаемая плотность распределения.

Этот интеграл вычисляется как сумма интегралов по интервалам между соседними членами вариационного ряда. Если на этих интервалах предполагаемая функция распределения интерполируется прямой линией, то этот интеграл выражается суммой

$$\omega^2 = \sum_{i=1}^n \int_{x(i)}^{x(i+1)} |\tilde{F}(x) - F(\tilde{\Theta}, x)|^2 \varphi(\tilde{\Theta}, x) dx = \frac{1}{12n^2} + \frac{1}{n} \cdot \sum_{i=1}^n \left[F(\tilde{\Theta}, x(i)) - \frac{2i-1}{2n} \right]^2.$$

. В качестве статистики критерия используется

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^n \left[F(\tilde{\Theta}, x(i)) - \frac{2i-1}{2n} \right]^2.$$

Критические значения $(n\omega^2)_\alpha$ приведены в таблицах математической статистики (см., например, [13,14]). В табл. 7 представлены некоторые часто употребляемые критические значения.

Критические значения критерия ω^2 Мизеса

α	0,03	0,05	0,1	0,2
$(n\omega^2)_\alpha$	0,55	0,4614	0,3473	0,2415

Процедура проверки гипотезы о виде функции распределения по критерию ω^2 Мизеса.

1. Задают уровень значимости α
 2. По выборочным данным строят выборочную функцию распределения $\tilde{F}(x)$ в соответствии с указаниями разд. 2.2

3. Вычисляют точечные оценки моментов : m_1, s^2, m_3, m_4 .

4. Из теоретических соображений, по виду выборочной функции распределения, по соотношениям между моментами, по значениям асимметрии и эксцесса, по результатам анализа других данных выдвигается гипотеза о виде функции распределения $F(\Theta, x)$ и тем самым – о виде плотности распределения $\varphi(\Theta, x)$.

5. Вычисляют r параметров предполагаемой функции распределения и ее значения $F(\tilde{\Theta}, x_i)$ при $x = x_i$.

6. Вычисляют статистику критерия

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^n \left[F(\tilde{\Theta}, x_{(i)}) - \frac{2i-1}{2n} \right]^2 .$$

7. Полученное значение сравнивают с критическим значением $(n\omega^2)_\alpha$

8. Если $n\omega^2 > (n\omega^2)_\alpha$ делают вывод о том, что экспериментальные данные не подтверждают справедливость выдвинутой гипотезы, или о том, что отсутствуют достаточные основания для того, чтобы считать нулевую гипотезу справедливой. Гипотеза пересматривается, выдвигается новая нулевая гипотеза и выполняется переход на п. 4 данной процедуры.

9. Если $n\omega^2 \leq (n\omega^2)_\alpha$, делают вывод о том, что экспериментальные данные подтверждают справедливость выдвинутой гипотезы, или о том,

что имеются достаточные основания для того, чтобы считать нулевую гипотезу справедливой.

Критерий ω^2 Мизеса – равномерно наиболее мощный критерий проверки гипотезы о виде функции распределения.

2.5.6. Проверка гипотез при полиномиальной аппроксимации

2.5.6.1. Критерий Кочрена проверки гипотезы о равенстве дисперсий

Проверка гипотезы о равенстве дисперсий погрешностей измерения значений y при различных значениях x в процессе полиномиальной аппроксимации. Это, по сути дела, проверка гипотезы о равнозначности измерений значений аппроксимируемой функции. От исхода проверки этой гипотезы зависит выбор метода аппроксимации: МНК или ОМНК (см. разд. 2.3.7.2). Применяется в тех случаях, когда характеристики погрешностей измерений аппроксимируемой функции неизвестны и трудоемкость выполнения ОМНК значительно выше трудоемкости выполнения МНК.

Формулировка гипотезы.

Имеется k выборок $y_{i1}, y_{i2}, \dots, y_{in}$, $i = 1, 2, \dots, k$, каждая из которых получена при значении аргумента x_i . Размерность каждой выборки n . Каждая выборка изъята из нормальной генеральной совокупности с дисперсиями $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$. По выборочным значениям при каждом i , $i = 1, 2, \dots, k$ вычисляются оценки математических ожиданий и дисперсий

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}, \quad s_i^2 = \frac{1}{n-1} \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2.$$

Выдвигается гипотеза

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ против альтернативы $H_1: \bar{H}_0$.

Статистика критерия Кочрена

$$g = \frac{s_{\max}^2}{\sum_{i=1}^k s_i^2}.$$

Критическое значение $g_\alpha(n-1, k)$ выбирается из таблиц критических значений критерия Кочрена. Эти таблицы приведены в учебниках и специальных таблицах математической статистики.

Рассчитанное значение статистики критерия сравнивается с критическим значением. Если $g \leq g_\alpha(n-1, k)$, делают вывод о том, что нулевая гипотеза не противоречит экспериментальным данным, и для аппроксимации применяется МНК (см. разд. 2.3.7.2, 2.3.7.6).

В противном случае, если $g > g_\alpha(n-1, k)$, делается противоположный вывод, и для аппроксимации применяется ОМНК (см. разд. 2.3.7.2, 2.3.7.6).

2.5.6.2. Проверка гипотезы о степени аппроксимирующего полинома, характеристики погрешностей измерений известны

а) *Измерения однократные.* Постановка задачи полиномиальной аппроксимации некоторой исследуемой функции $y = f(x)$ и метод ее решения изложены в разд. 2.3.7.1, 2.3.7.2.

Пусть в результате некоторого физического или численного эксперимента в точках $x_i, i = 1, 2, \dots, k$ получены однократные значения функции $y_i, i = 1, 2, \dots, k$. Будем считать, что существует полином неизвестной высокой степени q , который проходит через все точки с координатами (x_i, y_i) . Однако, результаты эксперимента возмущены погрешностями $\varepsilon_i, i = 1, 2, \dots, k$. Дисперсии этих погрешностей $\sigma_{\varepsilon_i}^2, i = 1, 2, \dots, k$ должны быть известны или при численном эксперименте – заданы. Если все дисперсии одинаковы и равны σ_ε^2 , а ковариации неизвестны или ими желательно пренебречь, то ковариационная матрица Σ_ε диагональна и равна $\sigma_\varepsilon^2 \mathbf{E}$, где \mathbf{E} – единичная матрица. Если в этой же ситуации дисперсии различаются, то матрица Σ_ε также диагональна, диагональными элементами являются дисперсии $\sigma_{\varepsilon_i}^2$. Когда известны и ковариации, то матрица Σ_ε будет полной. В конечном итоге исходными данными для аппроксимации будут дискретные значения аргумента x_i , дискретные значения функции $y_{\varepsilon_i} = y_i + \varepsilon_i$ и ковариационная матрица Σ_ε в любом из перечисленных

трех видов. Объединим значения функции в вектор \mathbf{y}_ε . Дискретные значения x_1, x_2, \dots, x_k аргумента используем для формирования матрицы \mathbf{X} , как это показано в разд. 2.3.7.2..

Если предполагается выполнить аппроксимацию полиномом степени p , то формируется матрица \mathbf{X} , состоящая из первых $p + 1$ столбцов. Отметим такую матрицу индексом $p + 1$: \mathbf{X}_{p+1} .

В соответствии с разд. 2.3.7.2 оценки $\hat{\mathbf{a}}_p$ коэффициентов полинома получаются путем минимизации квадратичных форм:

$$\min_{\mathbf{a}} (\mathbf{X}_{p+1} \mathbf{a} - \mathbf{y}_\varepsilon)^T (\mathbf{X}_{p+1} \mathbf{a} - \mathbf{y}_\varepsilon) \quad \text{— для равноточных измерений, когда}$$

все дисперсии σ_i^2 одинаковы и ковариации неизвестны, то есть $\Sigma_\varepsilon = \sigma_\varepsilon^2 \mathbf{E}$, при этом применяется МНК ;

$$\min_{\mathbf{a}} (\mathbf{X}_{p+1} \mathbf{a} - \mathbf{y}_\varepsilon)^T \Sigma_\varepsilon^{-1} (\mathbf{X}_{p+1} \mathbf{a} - \mathbf{y}_\varepsilon) \quad \text{— для неравноточных измерений,}$$

когда матрица Σ_ε диагональная (ковариации неизвестны) или когда она полная (ковариации известны), при этом применяется ОМНК.

Из-за случайности погрешностей измерений оценки $\hat{\mathbf{a}}_p$ также случайны, а значит, случайными являются векторы $(\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \mathbf{y}_\varepsilon)$ и вслед за ними – квадратичные формы, которые определены в разд. 2.3.7.2:

$$R^2 = \sigma^{-2} (\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \mathbf{y}_\varepsilon)^T (\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \mathbf{y}_\varepsilon) \quad \text{— при применении МНК;}$$

$$R^2 = (\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \mathbf{y}_\varepsilon)^T \Sigma_\varepsilon^{-1} (\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \mathbf{y}_\varepsilon) \quad \text{— при применении ОМНК.}$$

Когда степень полинома $p = q$, найденные оценки коэффициентов несмещены. В этом случае

$$M[\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \mathbf{y}_\varepsilon] = 0.$$

В разд. 2.3.7.3 приведены плотности распределения этих квадратичных форм в условиях, когда плотности распределения погрешностей нормальны и модель известна. Плотности распределения обеих квадратичных форм одинаковы: это плотность распределения хи-квадрат с $k - p - 1$ степенями свободы, то есть

$$R_p^2 \in \chi^2(k - p - 1).$$

На практике модель практически никогда не бывает известной. Тогда при ошибочном назначении степени p аппроксимирующего полинома, меньшей, чем истинная степень q , оценки коэффициентов оказываются смещенными (см. также п. 2.3.7.4, замечание 4), поэтому $M[\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \mathbf{y}_\varepsilon] \neq 0$, значения R^2 существенно возрастают, и плотность распределения изменяется. При увеличении степени аппроксимирующего полинома и в особенности при $p > q$ обусловленность матриц $\mathbf{X}_{p+1}^T \mathbf{X}_{p+1}$ и $\mathbf{X}_{p+1}^T \Sigma_\varepsilon^{-1} \mathbf{X}_{p+1}$ существенно ухудшается, и теряется вычислительная устойчивость МНК и ОМНК (см. разд. 2.3.7.7). Поэтому случаи, когда $p > q$, не рассматриваются.

Итак, пусть в изложенных условиях при истинной степени аппроксимирующего полинома q была назначена степень $p < q$, и с помощью МНК или ОМНК получены оценки $p + 1$ коэффициента. Обозначим вектор найденных таким образом оценок $\hat{\mathbf{a}}_p$, а квадратичные формы, вычисленные при этих значениях оценок – R_p^2 .

Сформулируем гипотезу.

H_0 : степень аппроксимирующего полинома $p = q$;

H_1 : степень аппроксимирующего полинома $p < q$.

Статистикой критерия проверки этой гипотезы является величина R_p^2 , которая при $p = q$, то есть при справедливости нулевой гипотезы, равна R^2 и распределена как $\chi^2(k - p - 1)$ (см. разд. 2.3.7.3). Поэтому в качестве критического значения при заданной вероятности α выбирается $(1 - \alpha)100$ -процентная квантиль из таблицы квантилей распределения хи-квадрат с $(k - p - 1)$ степенями свободы.

Процедура проверки гипотезы в условиях, когда заданы вид полинома (степенной, обобщенный по разд. 2.3.7.9) и вероятность α . При заданной степени полинома, равной p , матрица \mathbf{X} имеет $p + 1$ столбец, и впредь будет обозначаться, как \mathbf{X}_{p+1}

1. Оценивают коэффициенты полинома

$$\tilde{\mathbf{a}}_p = (\mathbf{X}_{p+1}^T \mathbf{X}_{p+1})^{-1} \mathbf{X}_{p+1}^T \mathbf{y}_\varepsilon \quad \text{или} \quad \tilde{\mathbf{a}}_p = (\mathbf{X}_{p+1}^T \Sigma_\varepsilon^{-1} \mathbf{X}_{p+1})^{-1} \mathbf{X}_{p+1}^T \Sigma_\varepsilon^{-1} \mathbf{y}_\varepsilon$$

2. Вычисляют статистику критерия

$$\text{при МНК } R_p^2 = \sigma_\varepsilon^{-2} (\mathbf{X}_{p+1} \tilde{\mathbf{a}}_p - \mathbf{y}_\varepsilon)^T (\mathbf{X}_{p+1} \tilde{\mathbf{a}}_p - \mathbf{y}_\varepsilon)$$

$$\text{при ОМНК } R_p^2 = (\mathbf{X}_{p+1} \tilde{\mathbf{a}}_p - \mathbf{y}_\varepsilon)^T \Sigma_\varepsilon^{-1} (\mathbf{X}_{p+1} \tilde{\mathbf{a}}_p - \mathbf{y}_\varepsilon).$$

3. Значение статистики R_p^2 сравнивают с критическим значением:

если $R_p^2 \leq \chi_{1-\alpha}^2(k-p-1)$, делают вывод о том, что нет достаточных оснований для того, чтобы отвергнуть нулевую гипотезу;

в противном случае, если $R_p^2 > \chi_{1-\alpha}^2(k-p-1)$, делают вывод о том, что нет достаточных оснований для того, чтобы считать нулевую гипотезу справедливой.

В заключение заметим, что случай однократных измерений – экзотический. Эта ситуация может возникнуть, когда эксперимент не может быть повторен многократно, например при запуске космического корабля, когда объектом исследований является наземный стартовый комплекс. Второй вариант может встретиться, например, при необходимости моделирования сложного объекта или функции, когда исходные данные (x_i, y_i) получаются в результате численного эксперимента или просто вычислений. Подробнее эти ситуации описаны в разд. 2.3.7.4.

b) *Измерения многократные.*

Этот случай отличается от предыдущего тем, что в качестве исходных данных для полиномиальной аппроксимации используется не вектор результатов однократных измерений \mathbf{y}_ε , а вектор средних арифметических значений $\bar{\mathbf{y}}$ результатов многократных измерений. В условиях, перечисленных ранее для однократных измерений, когда $p = q$, вектор $\bar{\mathbf{y}}$ средних арифметических значений обладает свойствами (см. также разд. 2.3.4.3, 2.3.7.5):

$$M[\bar{\mathbf{y}}] = \mathbf{X}\mathbf{a}, \quad M[\mathbf{X}\mathbf{a} - \bar{\mathbf{y}}] = 0, \quad \bar{\mathbf{y}} \in N_k \left(\mathbf{X}\mathbf{a}, \frac{\Sigma_\varepsilon}{n} \right), \quad \Sigma_{\bar{\mathbf{y}}}^{-1} = n\Sigma_\varepsilon^{-1}.$$

При равноточных измерениях, когда ковариации неизвестны или ими пренебрегли, тогда применяется МНК,

$$\Sigma_{\bar{\mathbf{y}}} = \frac{1}{n} \sigma_\varepsilon^2 \mathbf{E} \quad \Sigma_{\bar{\mathbf{y}}}^{-1} = n\sigma_\varepsilon^{-2} \mathbf{E},$$

где \mathbf{E} – единичная матрица (см. также разд. 2.3.7.2).

Если измерения неравноточные, ковариации известны и матрица Σ_ε полная, для аппроксимации назначена степень полинома p и $\hat{\mathbf{a}}_p$ – вектор оценок $(p + 1)$ коэффициента этого полинома, то матрица \mathbf{X}_{p+1} имеет $p + 1$ столбец, и статистикой критерия проверки гипотезы о степени полинома в общем случае будет

$$R_p^2 = n(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T \Sigma_\varepsilon^{-1}(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}}).$$

Когда измерения равноточные, ковариации неизвестны или исследователь ими пренебрег, тогда применяется МНК и статистикой критерия будет

$$R_p^2 = n\sigma_\varepsilon^{-2}(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}}).$$

Как и ранее, при условии справедливости нулевой гипотезы, то есть при $p = q$, $R_p^2 = R_q^2$ и $R_p^2 \in \chi^2(k - p - 1)$ Критическое значение при заданном значении вероятности α есть $R_p^2 \in \chi_{1-\alpha}^2(k - p - 1)$.

Гипотеза формулируется по аналогии с формулировкой, приведенной для однократных измерений.

H_0 : степень аппроксимирующего полинома $p = q$,

H_1 : степень аппроксимирующего полинома $p < q$.

Процедура проверки гипотезы в условиях, когда заданы вид полинома (степенной, обобщенный по разд. 2.3.7.9), его степень p и уровень значимости критерия α .

1. Оценивают коэффициенты полинома

$$\hat{\mathbf{a}}_p = (\mathbf{X}_{p+1}^T \mathbf{X}_{p+1})^{-1} \mathbf{X}_{p+1}^T \bar{\mathbf{y}} \quad \text{или} \quad \hat{\mathbf{a}}_p = (\mathbf{X}_{p+1}^T \Sigma_\varepsilon^{-1} \mathbf{X}_{p+1})^{-1} \mathbf{X}_{p+1}^T \Sigma_\varepsilon^{-1} \bar{\mathbf{y}}.$$

2. Вычисляют статистику критерия

$$\text{при МНК } R_p^2 = n\sigma_\varepsilon^{-2}(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}}) \quad \text{или}$$

$$\text{в других случаях } R_p^2 = n(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T \Sigma_\varepsilon^{-1}(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}}).$$

3. Значение статистики R_p^2 сравнивают с критическим значением:

если $R_p^2 \leq \chi_{1-\alpha}^2(k - p - 1)$, делают вывод о том, что нет достаточных оснований для того, чтобы отвергнуть нулевую гипотезу;

в противном случае, если $R_p^2 > \chi_{1-\alpha}^2(k-p-1)$, делают вывод о том, что нет достаточных оснований для того, чтобы считать нулевую гипотезу справедливой.

Заметим, что в соответствии с центральной предельной теоремой плотность распределения средних арифметических асимптотически нормальна. Поэтому, начиная с $n = 15-20$, требования к нормальности погрешностей измерений могут быть значительно смягчены. Кроме того, в разд. 2.3.7.4 (замечание 2) указано, что оценки МНК и ОМНК эффективны в классе всех линейных оценок вне зависимости от закона распределения погрешностей измерений.

2.5.6.3. Проверка гипотезы о степени аппроксимирующего полинома, характеристики погрешностей измерений неизвестны

Если характеристики погрешностей измерения значений аппроксимируемой функции неизвестны, то неизбежно приходится выполнять многократные измерения и по результатам этих измерений оценивать характеристики погрешностей: дисперсии $\sigma_{\varepsilon_i}^2$ или ковариационную матрицу \mathbf{S}_ε . Выполнение эксперимента в этой ситуации и вычисление оценок производится в соответствии с указаниями разд. 2.3.4.3, 2.3.4.4, 2.3.7.6. Выяснить равноточность или неравноточность выполненных измерений по критерию Кочрена (см. разд. 2.5.6.1) имеет смысл в том случае, если матрица \mathbf{S}_ε окажется диагональной. Это может произойти, когда $n < k + 1$ и матрица \mathbf{S}_ε окажется особенной (см. также разд. 2.3.7.6). Поскольку особенная матрица не может быть обращена, придется вычислять только элементы диагонали матрицы \mathbf{S}_ε по формуле разд. 2.3.7.6. Эта матрица не будет особенной. В этом случае, если гипотеза о равноточности подтвердилась, применяется МНК. Во всех остальных случаях следует применить ОМНК.

Матрица \mathbf{S}_ε может оказаться диагональной, если исследователь по каким-то причинам пренебрегает внедиагональными элементами, то есть ковариацией, даже, если $n > k$. Но и в этом случае следует применить для аппроксимации ОМНК. Это обстоятельство несколько снижает эффективность оценок коэффициентов аппроксимирующего полинома (см. разд.

2.3.7.4), но оценки коэффициентов остаются несмещенными, если, конечно, назначенная степень полинома p равна истинной степени q .

При многократных измерениях, как и ранее в разд. 2.5.6.2, “б”, в качестве исходных данных для полиномиальной аппроксимации используется не вектор результатов однократных измерений \mathbf{y}_ε , а вектор средних арифметических значений $\bar{\mathbf{y}}$ результатов многократных измерений. С учетом замены матрицы Σ_ε ее оценкой \mathbf{S}_ε (см. разд. 2.5.6.2, б)

$$M[\bar{\mathbf{y}}] = \mathbf{X}\mathbf{a}, \quad M[\mathbf{X}\mathbf{a} - \bar{\mathbf{y}}] = 0, \quad \bar{\mathbf{y}} \in N_k \left(\mathbf{X}\mathbf{a}, \frac{\mathbf{S}_\varepsilon}{n} \right), \quad \mathbf{S}_{\bar{\mathbf{y}}} = n^{-1}\mathbf{S}_\varepsilon.$$

Если $n < k + 1$ и гипотеза о равенстве дисперсий не отклонена, то есть измерения равноточные, принимается решение о применении МНК. Тогда

$$\mathbf{S}_{\bar{\mathbf{y}}} = n^{-1}s_\varepsilon^2\mathbf{E}, \quad \mathbf{S}_{\bar{\mathbf{y}}}^{-1} = ns_\varepsilon^{-2}\mathbf{E},$$

где \mathbf{E} – единичная матрица, s_ε^2 – оценка дисперсии равноточных измерений, которая вычисляется, как среднее арифметическое значение оценок дисперсий $s_{\varepsilon i}^2$, найденных при каждом значении x_i :

$$s_\varepsilon^2 = \frac{1}{k} \sum_{i=1}^k s_{\varepsilon i}^2.$$

В этом случае, если для аппроксимации назначена степень полинома p , и $\hat{\mathbf{a}}_p$ – вектор оценок $p + 1$ коэффициента этого полинома, статистикой критерия проверки гипотезы о степени полинома будет

$$F_p = \frac{1}{k - p - 1} R_p^2,$$

где $R_p^2 = n s_\varepsilon^{-2} (\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T (\mathbf{X}_{p+1} \hat{\mathbf{a}}_p - \bar{\mathbf{y}})$.

В тех случаях, когда

$n < k + 1$, гипотеза о равенстве дисперсий отклонена, и диагональные элементы матрицы \mathbf{S}_ε вычисляются по формуле разд. 2.3.7.6,

$n > k$ и получены оценки либо только диагональных элементов, либо всех элементов матрицы \mathbf{S}_ε , включая ковариации (то есть матрица \mathbf{S}_ε полная),

тогда для аппроксимации применяется обобщенный МНК (ОМНК) и статистика критерия проверки гипотезы о степени полинома есть

$$F_p = \frac{n - k + p + 1}{(k - p - 1)(n - 1)} R_p^2,$$

где $R_p^2 = n(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T \mathbf{S}_\varepsilon^{-1}(\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})$.

При условии справедливости нулевой гипотезы, то есть при $p = q$, как указано выше в разд. 2.3.7.6, статистика F_p во всех рассмотренных случаях распределена по закону распределения Фишера (см. [5], стр. 485):

если $n < k + 1$ и матрица \mathbf{S}_ε диагональная, то $F_p \in F(k - p - 1, n - 1)$ с количеством степеней свободы $(k - p - 1)$ и $(n - 1)$;

при применении ОМНК $F_p \in F(k - p - 1, n - k + p + 1)$ с количеством степеней свободы $(k - p - 1)$ и $(n - k + p + 1)$.

Гипотеза о степени полинома формулируется по аналогии с формулировкой, приведенной в разд. 2.5.6.2 :

H_0 : степень аппроксимирующего полинома $p = q$,

H_1 : степень аппроксимирующего полинома $p < q$.

Процедура проверки гипотезы в условиях, когда заданы вид полинома (степенной, обобщенный по разд. 2.3.7.9) и вероятность α , а также вычислены оценки $\bar{\mathbf{y}}$ и s_ε^2 или $\bar{\mathbf{y}}$ и \mathbf{S}_ε .

1. Оценивают коэффициенты полинома

$$\hat{\mathbf{a}}_p = (\mathbf{X}_{p+1}^T \mathbf{X}_{p+1})^{-1} \mathbf{X}_{p+1}^T \bar{\mathbf{y}} \quad \text{или} \quad \hat{\mathbf{a}}_p = (\mathbf{X}_{p+1}^T \mathbf{S}_\varepsilon^{-1} \mathbf{X}_{p+1})^{-1} \mathbf{X}_{p+1}^T \mathbf{S}_\varepsilon^{-1} \bar{\mathbf{y}}.$$

2. Вычисляют статистику критерия

$$\text{в случае МНК} \quad F_p = \frac{n}{k - p - 1} s_\varepsilon^{-2} (\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T (\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}});$$

в случае ОМНК при $n < k + 1$ и диагональной матрице \mathbf{S}_ε

$$F_p = \frac{n}{k - p - 1} (\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T \mathbf{S}_\varepsilon^{-1} (\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}});$$

в случае ОМНК при $n > k$

$$F_p = \frac{n(n - k + p + 1)}{(k - p - 1)(n - 1)} (\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}})^T \mathbf{S}_\varepsilon^{-1} (\mathbf{X}_{p+1}\hat{\mathbf{a}}_p - \bar{\mathbf{y}}).$$

3. Значение статистики F_p сравнивают с критическим значением распределения Фишера:

в случае МНК или при $n < k + 1$ и диагональной матрице S_ε , если

$$F_p \leq F_{1-\alpha}(k - p - 1, n - 1);$$

при ОМНК, если $F_p \leq F_{1-\alpha}(k - p - 1, n - k + p + 1)$;

делают вывод о том, что нет достаточных оснований для того, чтобы отвергнуть нулевую гипотезу о степени полинома;

в противном случае делают вывод о том, что нет достаточных оснований для того, чтобы считать нулевую гипотезу $H_0: p = q$ справедливой.

Критические значения плотности распределения Фишера приведены в таблицах математической статистики (см. [13, 14] и др.)

2.6. ПОСЛЕДОВАТЕЛЬНАЯ ПОЛИНОМИАЛЬНАЯ АППРОКСИМАЦИЯ С ПРОВЕРКОЙ ГИПОТЕЗЫ О СТЕПЕНИ ПОЛИНОМА

На практике при экспериментальном исследовании (идентификации) функциональных зависимостей, которые имеют место в природном или рукотворном объекте или в компьютерной модели, порядок степенного или обобщенного полинома, как правило, неизвестен, за исключением случаев, когда этот порядок задан заранее. Поэтому практически важной задачей является выбор такого порядка (степени) полинома, при котором обеспечиваются:

заданная точность аппроксимации исследуемой функции;

вычислительная устойчивость операций, выполняемых для нахождения оценок коэффициентов аппроксимирующего полинома.

Эти два требования противоположны.

С одной стороны, точность аппроксимации гладких функций степенным или обобщенным полиномом повышается с увеличением степени полинома.

С другой стороны, с увеличением степени полинома, как указано в разд. 2.3.7.7, увеличивается количество оцениваемых коэффициентов и в связи с этим ухудшается обусловленность матриц, участвующих в выполняемых операциях, что приводит к утрате вычислительной устойчивости этих операций. Неустойчивость вычислений доходит до экстремально

высокого уровня, когда заданная при аппроксимации степень полинома настолько высока, что приходится находить оценки несуществующих или пренебрежимо малых коэффициентов.

Ранее назывались средства повышения вычислительной устойчивости полиномиальной аппроксимации, такие, как C -оптимальное планирование эксперимента (см. разд. 2.3.7.8) и регуляризация по Тихонову (ридж-оценивание, см. разд. 2.3.7.7), но эти средства не являются радикальными по следующим причинам.

Далеко не всегда имеется возможность выполнять эксперимент при значениях x_i , расстановка которых соответствует C -оптимальному плану. Это может быть при исследовании природных зависимостей, когда воспроизведение заданных значений x_i неподвластно человеку, или когда воспроизведение этих значений с заданной точностью технически невозможно.

Применение метода тихоновской регуляризации (ридж-оценивания) затруднено неопределенностью в выборе параметра регуляризации, универсальные практические рецепты выбора его оптимального значения пока отсутствуют.

Практически реализуемой является стратегия полиномиальной аппроксимации (как и всякой другой), основанная на следующих принципах:

не стремиться к точности аппроксимации, превышающей точность исходных данных, которая определяется характеристиками погрешностей измерений ϵ и количеством независимых повторных измерений n ;

подбирать подходящую степень (порядок) полинома путем направленного перебора от наименьшей степени, например, $p = 1$ в сторону возрастания до того значения, при котором характеристики погрешности аппроксимации окажутся согласованными с погрешностями исходных данных.

При реализации такой стратегии погрешность аппроксимации монотонно уменьшается, а погрешность, порождаемая вычислительной неустойчивостью, монотонно повышается. М. А. Красносельский доказал, что подобная стратегия является регуляризирующей стратегией решения плохообусловленных и некорректных задач.

Может оказаться, что этот процесс наращивания степени полинома остановится, но, по мнению исследователя, требуемая точность аппроксимации недостаточна. Это значит, что при имеющейся погрешности исходных данных такая точность недостижима. Что делать в этом случае? Есть два пути: повысить точность измерений или увеличить количество независимых многократных измерений.

А л г о р и т м п о с л е д о в а т е л ь н о й п о л и н о м и а л ь н о й а п п р о к с и м а ц и и ф у н к ц и й с п р о в е р к о й г и п о т е з ы о с т е п е н и п о л и н о м а:

1. Выполняют эксперименты при значениях x_i , $i = 1, 2, \dots, k$, где k должно быть больше максимально возможного предполагаемого количества коэффициентов искомого полинома.

2. Задают вероятность $\alpha \in [0,05; \dots; 0,2]$.

3. В зависимости от ситуации вычисляются оценки $\bar{y}, s_{\varepsilon i}^2, s_{\varepsilon}^2, \mathbf{S}_{\varepsilon}$, при необходимости в зависимости от соотношения между n и k , в зависимости от результатов проверки гипотезы о равенстве дисперсий $\sigma_{\varepsilon i}^2$ и от решения об учете ковариаций принимают решение о применении МНК или ОМНК.

4. Задают начальную степень полинома p , например, $p = 1$,

5. Формируют матрицу с $(p + 1)$ столбцами \mathbf{X}_{p+1} .

6. По формулам разд. 2.5 в зависимости от ситуации вычисляются оценки коэффициентов аппроксимирующего полинома порядка (степени) p .

7. Вычисляется статистика R_p^2 или F_p (в зависимости от ситуации) критерия проверки гипотезы о степени полинома, значение этой статистики сравнивается с соответствующим критическим значением (см. разд. 2.5.6.2, 2.5.6.3).

8. Делается вывод о справедливости нулевой гипотезы.

Если принято решение об отклонении нулевой гипотезы, $p = p + 1$, возврат на 5.

В противном случае переход на 9.

9. По одной из подходящих формул разд. 2.5.6.2, 2.5.6.3 вычисляется ковариационная матрица $\Sigma_{\hat{a}}$ погрешностей определения коэффициентов аппроксимирующего полинома.

Замечание. Шаг 9 приведенного алгоритма обязателен, ибо всякий результат эксперимента или оценивания должен сопровождаться сообщением о характеристиках его погрешности.

2.7. ПРОВЕРКА СЛОЖНЫХ ГИПОТЕЗ О ЧИСЛОВЫХ ХАРАКТЕРИСТИКАХ СЛУЧАЙНЫХ ВЕЛИЧИН С КОНТРОЛЕМ ВЕРОЯТНОСТЕЙ ОШИБОК ПЕРВОГО И ВТОРОГО РОДА

2.7.1. Основные положения

Ранее в разд. 2.5.4 шла речь об основном недостатке методов проверки сложных гипотез, который заключается в невозможности одновременно контролировать обе вероятности ошибочных решений – первого и второго рода. Объем выборки был фиксирован.

Возможности контроля обеих вероятностей ошибок открываются при использовании последовательных методов проверки сложных гипотез, когда объем выборки не фиксирован, а наращивается при необходимости в зависимости исхода на последнем шаге процедуры.

Первым методом последовательной проверки статистической гипотезы был метод последовательного анализа, разработанный А. Вальдом [8]. Метод разработан для проверки сложной гипотезы о вероятности. Если на n -ом шаге отсутствуют основания, достаточные для принятия положительного или отрицательного решения, количество испытаний увеличивается на один шаг. За счет этого контролируются вероятности ошибок первого и второго рода. Этот метод излагается в конце раздела.

В основной части настоящего раздела с целью реализации последовательной проверки гипотез с контролем вероятностей ошибок первого и второго рода описаны приемы применения *фидуциальных вероятностей* (fiducial probabilities). Фидуциальные вероятности введены одним из выдающихся специалистов в области математической статистики Р. Фишером в 1935 году (см.[6], стр. 378 – 380). Эти вероятности определены на доверительных интервалах. Аксиоматику фидуциальных вероятностных

мер на полуоткрытых интервалах разработал Г. П. Климов в 1973 году [12]. Несмотря на это, идеология системы фидуциальных вероятностных мер не является общепризнанной.

2.7.2. Проверка сложной гипотезы о математическом ожидании

Плотность распределения генеральной совокупности X - нормальна, дисперсия неизвестна. Из генеральной совокупности извлечена выборка x_1, x_2, \dots, x_{n_1} объема n_1 . Как и в разд. 2.5.4, проверяется сложная гипотеза

а) $H_0: \mu_1[\xi] < a$ против альтернативы $H_1: \mu_1[\xi] > a$.

Заданы вероятности ошибок первого и второго рода α и β .

Естественной статистикой для проверки гипотезы о математическом ожидании, как и ранее, является среднее арифметическое значение выборочных данных, а для построения доверительного интервала для математического ожидания требуется вычислить оценку дисперсии:

$$\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i \quad s^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 .$$

Доверительный интервал для математического ожидания строится с помощью квантилей (коэффициентов) Стьюдента в соответствии с разд. 2.4.4:

$$P \left(\bar{x} - \frac{s}{\sqrt{n_1}} t_{1+Q} \frac{(n_1 - 1)}{2} < \mu_1[\xi] \leq \bar{x} + \frac{s}{\sqrt{n_1}} t_{1+Q} \frac{(n_1 - 1)}{2} \right) = Q .$$

Эту вероятность Фишер предложил трактовать, как вероятностную меру полуоткрытого интервала

$$\left[\bar{x} - \frac{s}{\sqrt{n_1}} t_{1+Q} \frac{(n_1 - 1)}{2}, \bar{x} + \frac{s}{\sqrt{n_1}} t_{1+Q} \frac{(n_1 - 1)}{2} \right],$$

из чего сразу же следует трактовка генеральной характеристики (математического ожидания $\mu_1[\xi]$), как случайной величины, и это заключение как раз является основной причиной неполного признания такой вероятностной меры, которую Фишер назвал *фидуциальной*.

Посмотрим, как, несмотря ни на что, применить эту меру с пользой для дела, а именно для проверки сформулированной гипотезы с контролем вероятностей α и β .

На рис. 38 показаны три варианта возможного расположения этого доверительного интервала относительно границы a , которая разделяет вещественную ось на две зоны, соответствующие гипотезам H_0 и H_1 .

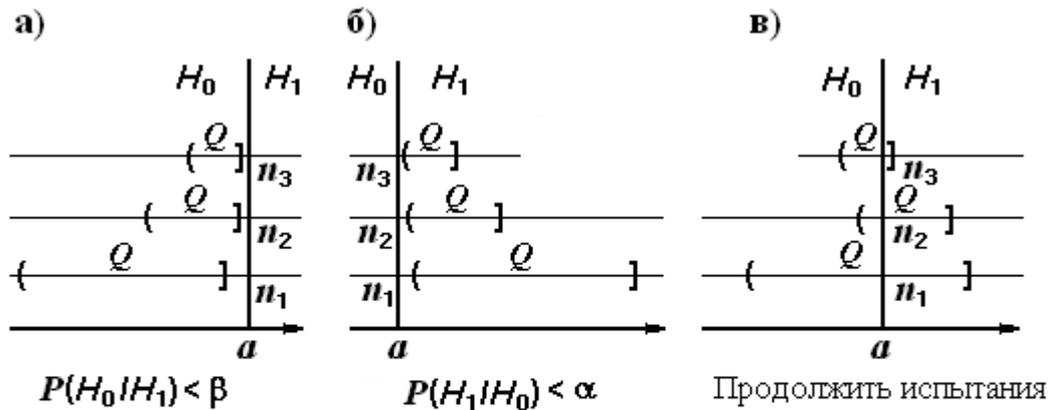


Рис. 38. Варианты расположения доверительных интервалов при проверке гипотезы а) о математическом ожидании

В соответствии с формулировкой гипотезы положение нижней границы доверительного интервала в зоне, соответствующей гипотезе H_0 , безразлично, и поэтому на рис. 38, а) нижняя граница обозначена условно. По указанной причине для этого варианта строится доверительный интервал, нижняя граница которого равна $-\infty$, а верхняя граница определяется заданной вероятностью β ошибки второго рода:

$$P\left(-\infty < \mu_1[\xi] \leq \bar{x} + \frac{s}{\sqrt{n_1}} t_{1-\beta}(n_1 - 1)\right) \geq Q = 1 - \beta,$$

где $t_{1-\beta}(n_1 - 1)$ – $(1 - \beta)100$ процентная квантиль плотности распределения Стьюдента.

Это равенство означает, что с позиций фидуциальной вероятностной меры, истинное значение $\mu_1[\xi]$ находится в зоне, соответствующей гипотезе H_0 с вероятностью, не меньшей, чем Q . В этой ситуации, когда

$$\bar{x} + \frac{s}{\sqrt{n_1}} t_{1-\beta}(n_1 - 1) < a,$$

принимается решение о справедливости гипотезы H_0 . Вероятность ошибки в этом решении не будет превышать значения вероятности, с которой

истинное значение $\mu_1[\xi]$ расположено справа от верхней границы доверительного интервала. Объясняется это тем, что здесь

$$P\left(\mu_1[\xi] > \bar{x} + \frac{s}{\sqrt{n_1}} t_{1-\beta}(n_1 - 1)\right) \leq 1 - Q = \beta,$$

и в силу монотонности вероятностной меры

$$P(\mu_1[\xi] > a) < P\left(\mu_1[\xi] > \bar{x} + \frac{s}{\sqrt{n_1}} t_{1-\beta}(n_1 - 1)\right),$$

а это означает, что $P(\mu_1[\xi] > a) < \beta$.

Поэтому в показанной ситуации вероятность ошибочного решения о признании гипотезы H_0 справедливой не превысит заданного значения β :

$$P(H_0 / H_1) \leq \beta.$$

Точно так же в соответствии с формулировкой гипотезы положение верхней границы доверительного интервала в зоне, соответствующей гипотезе H_1 , безразлично. По указанной причине для этого варианта строится доверительный интервал, верхняя граница которого равна ∞ , а нижняя граница определяется заданной вероятностью α ошибки первого рода:

$$P\left(\bar{x} - \frac{s}{\sqrt{n_1}} t_{1-\alpha}(n_1 - 1) < \mu_1[\xi] \leq \infty\right) \geq Q = 1 - \alpha,$$

где $t_{1-\alpha}(n_1 - 1)$ – $(1 - \alpha)$ 100-процентная квантиль плотности распределения Стьюдента с $n_1 - 1$ степенями свободы.

Это равенство означает, что с позиций фидуциальной вероятностной меры истинное значение $\mu_1[\xi]$ находится в зоне, соответствующей гипотезе H_1 с вероятностью, не меньшей, чем $Q = 1 - \alpha$. В этой ситуации, когда

$$a < \bar{x} - \frac{s}{\sqrt{n_1}} t_{1-\alpha}(n_1 - 1),$$

принимается решение о справедливости гипотезы H_1 . Вероятность ошибки в этом решении не будет превышать значения вероятности, с которой истинное значение $\mu_1[\xi]$ расположено слева от нижней границы доверительного интервала. Объясняется это тем, что здесь

$$P\left(\mu_1[\xi] < \bar{x} - \frac{s}{\sqrt{n_1}} t_{1-\alpha}(n_1 - 1)\right) \leq 1 - Q = \alpha,$$

и в силу монотонности вероятностной меры

$$P(\mu_1[\xi] < a) < P\left(\mu_1[\xi] < \bar{x} - \frac{s}{\sqrt{n_1}} t_{1-\alpha}(n_1 - 1)\right),$$

а это означает, что $P(\mu_1[\xi] > a) < \alpha$.

Поэтому в показанной ситуации вероятность ошибочного решения о признании гипотезы H_1 справедливой не превысит заданного значения α :

$$P(H_1 / H_0) \leq \alpha.$$

В третьем варианте, (см. рис. 38, в), у нас нет достаточных оснований для принятия никакого иного решения, кроме решения продолжить испытания с целью увеличения объема выборки. Надежда здесь возлагается на то, что с увеличением объема выборки ширина доверительного интервала уменьшается.

После увеличения объема выборки до n_2 снова строится доверительный интервал при тех же заданных значениях вероятностей α и β , и на втором этапе вновь возможны три варианта расположения нового, более узкого доверительного интервала относительно значения a . Все описанные рассуждения и действия повторяются.

В конечном итоге описанная последовательная процедура проверки сложной гипотезы о математическом ожидании должна завершиться выходом доверительного интервала целиком в одну из зон и принятием соответствующей гипотезы. Когда по экономическим, техническим и иным причинам дальнейшее продолжение испытаний (экспериментов) оказывается невозможным, придется принимать произвольное (так называемое волевое) решение об отклонении той или иной гипотезы в зависимости от того, риск какого рода (первого или второго) более оправдан.

Второй разновидностью сложной гипотезы о математическом ожидании, которая имеет важные практические приложения, является гипотеза

$$\text{б) } H_0: |\mu_1[\xi]| < a \text{ против альтернативы } H_1: |\mu_1[\xi]| > a.$$

К проверке такой гипотезы приходится прибегать, например, когда по результатам испытания продукции необходимо установить, что контроли-

руемый параметр или параметры находятся в пределах установленного и рекламируемого допуска.

Заданы вероятности ошибок первого и второго рода α и β , которые в этом случае приходится принимать равными, то есть $\alpha = \beta$.

Как и ранее, в соответствии с разд. 2.4.4, по выборке объема n_1 строится доверительный интервал

$$P\left(\bar{x} - \frac{s}{\sqrt{n_1}} t_{1-\frac{\alpha}{2}}(n_1 - 1) < \mu_1[\xi] \leq \bar{x} + \frac{s}{\sqrt{n_1}} t_{1-\frac{\alpha}{2}}(n_1 - 1)\right) = Q,$$

где на этот раз $\alpha = \beta$, $Q = 1 - \alpha$. Этот интервал располагается по отношению к интервалу $[-a, +a]$, соответствующему гипотезе H_0 так, как показано на рис. 39.

Также возможны три варианта расположения доверительного интервала.

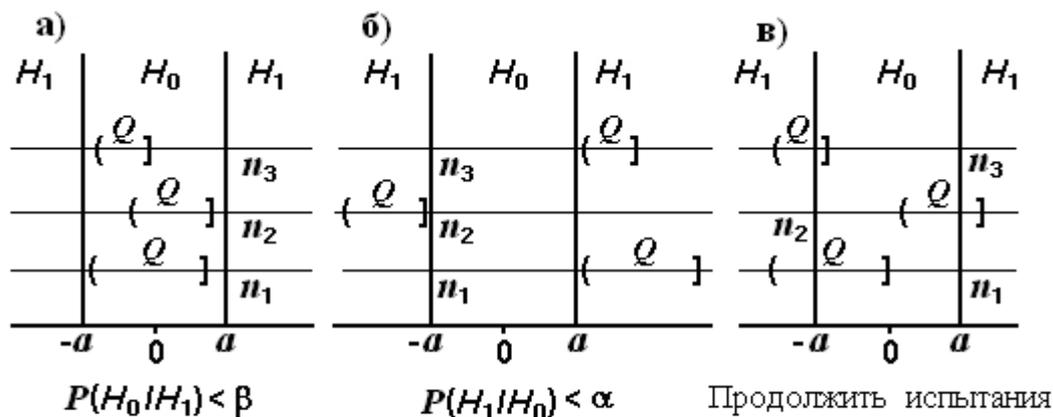


Рис. 39. Варианты расположения доверительных интервалов при проверке гипотезы б) о математическом ожидании

В первом варианте (см. рис. 39 а), когда доверительный интервал, который покрывает истинное значение $\mu_1[\xi]$ с вероятностью Q , целиком находится в зоне справедливости нулевой гипотезы, это означает, что истинное значение $\mu_1[\xi]$ находится в зоне, соответствующей гипотезе H_0 с вероятностью, превышающей Q . В этой ситуации принимается решение о принятии гипотезы H_0 . Вероятность ошибочного решения не будет превышать β , то есть значения вероятности, с которой истинное значение

$\mu_1[\xi]$ расположено вне доверительного интервала. Объясняется это тем, что в силу монотонности вероятностной меры в ситуации, показанной на рис. 39, а)

$$P(|\mu_1[\xi]| > a) < 1 - P\left(\bar{x} - \frac{s}{\sqrt{n_1}} \cdot t_{1-\frac{\beta}{2}}(n_1 - 1) < \mu_1[\xi] \leq \bar{x} + \frac{s}{\sqrt{n_1}} \cdot t_{1-\frac{\beta}{2}}(n_1 - 1)\right) = \beta,$$

а это означает, что $P(|\mu_1[\xi]| > a) < \beta = 1 - Q$.

Поэтому вероятность ошибочного решения о признании гипотезы H_0 справедливой, то есть вероятность ошибки второго рода не превысит заданного значения β :

$$P(H_0 / H_1) \leq \beta.$$

Когда во втором варианте (см. рис. 39, б) доверительный интервал целиком находится вне зоны, соответствующей гипотезе H_0 , справа или слева от нее, принимается решение о принятии гипотезы H_1 . Вероятность ошибочного решения не будет превышать α , то есть значения вероятности, с которой истинное значение $\mu_1[\xi]$ расположено вне доверительного интервала, поскольку в этой ситуации,

если доверительный интервал находится слева от зоны гипотезы H_0 ,

$$P(|\mu_1[\xi]| < a) < P\left(\mu_1[\xi] > \bar{x} + \frac{s}{\sqrt{n_1}} t_{1-\alpha}(n_1 - 1)\right) < \alpha;$$

если доверительный интервал находится справа от зоны гипотезы H_0 , то также

$$P(|\mu_1[\xi]| < a) < P\left(\mu_1[\xi] < \bar{x} - \frac{s}{\sqrt{n_1}} \cdot t_{1-\alpha}(n_1 - 1)\right) < \alpha.$$

Это означает, что $P(|\mu_1[\xi]| > a) < \alpha$.

Поэтому вероятность ошибочного решения о признании гипотезы H_1 справедливой, то есть вероятность ошибки первого рода не превысит заданного значения α :

$$P(H_1 / H_0) \leq \alpha.$$

В третьем варианте (см. рис. 39, в), как и прежде, принимается решение об увеличении объема выборки, если этому нет экономических, техни-

ческих или иных препятствий. Получив выборку увеличенного объема n_2 , процедуру повторяют до ее естественного завершения или до прекращения по соображениям экономического, технического или иного обоснованного характера.

2.7.3. Проверка сложной гипотезы о дисперсии

Плотность распределения генеральной совокупности \mathbf{X} нормальна, дисперсия σ^2 неизвестна. Из генеральной совокупности извлечена выборка x_1, x_2, \dots, x_{n_1} объема n_1 . Проверяется сложная гипотеза:

$$H_0: \sigma^2 \leq d \text{ против альтернативы } H_1: \sigma^2 > d.$$

Заданы вероятности ошибок первого и второго рода α и β .

Естественной статистикой для проверки гипотезы о дисперсии является ее несмещенная оценка:

$$s^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 .$$

Доверительный интервал для дисперсии определен в разд. 2.4.5 следующим образом:

$$P \left(\frac{s^2 (n-1)}{\chi_{\frac{1+Q}{2}}^2 (n-1)} < \sigma^2 \leq \frac{s^2 (n-1)}{\chi_{\frac{1-Q}{2}}^2 (n-1)} \right) = Q .$$

На рис. 40 показаны варианты расположения доверительного интервала для дисперсии относительно зон, соответствующих гипотезам H_0 и H_1 . В соответствии с формулировкой гипотезы и в связи с тем, что $\sigma^2 \geq 0$, положение нижней границы доверительного интервала в зоне, соответствующей гипотезе H_0 , безразлично. По указанным причинам для этого варианта строится доверительный интервал, нижняя граница которого равна 0, а верхняя граница определяется заданной вероятностью β ошибки второго рода:

$$P \left(0 < \sigma^2 \leq \frac{s^2 (n_1 - 1)}{\chi_{\beta}^2 (n_1 - 1)} \right) = Q = 1 - \beta$$

где $\chi_{\beta}^2(n_1 - 1)$ – β 100-процентная квантиль плотности распределения хи-квадрат.

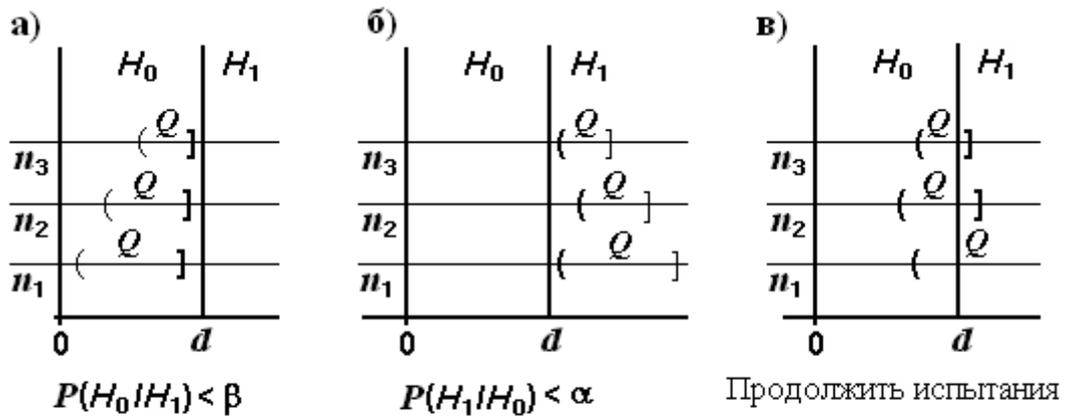


Рис. 40. Варианты расположения доверительных интервалов при проверке гипотезы о дисперсии

Таким образом, решение о принятии гипотезы H_0 принимается, когда верхняя граница доверительного интервала для дисперсии оказывается меньше заданного значения d , то есть когда

$$\frac{s^2(n_1 - 1)}{\chi_{\beta}^2(n_1 - 1)} < d.$$

В данной ситуации вероятность того, что истинное значение дисперсии находится в зоне, соответствующей гипотезе H_1 , не превышает β , поскольку в данной ситуации

$$P\left(\sigma^2 > \frac{s^2(n_1 - 1)}{\chi_{\beta}^2(n_1 - 1)}\right) \leq 1 - Q = \beta,$$

значит, в силу монотонности вероятностной меры

$$P(\sigma^2 > d) < P\left(\sigma^2 > \frac{s^2(n_1 - 1)}{\chi_{\beta}^2(n_1 - 1)}\right).$$

Из этого следует, что вероятность ошибки второго рода $P(H_0 / H_1) \leq \beta$.

В соответствии с формулировкой гипотезы положение верхней границы доверительного интервала в зоне, соответствующей гипотезе H_1 , безразлично. В связи с этим здесь строится доверительный интервал, верхняя граница которого равна ∞ , а нижняя граница определяется заданной вероятностью α ошибки первого рода:

$$P\left(\frac{s^2(n_1 - 1)}{\chi_{1-\alpha}^2(n_1 - 1)} < \sigma^2 \leq \infty\right) \geq Q = 1 - \alpha$$

где $\chi_{1-\alpha}^2(n_1 - 1)$ – $(1 - \alpha)100$ -процентная квантиль плотности распределения хи-квадрат.

Решение об отклонении нулевой гипотезы принимается, когда нижняя граница доверительного интервала для дисперсии превышает заданное значение d , то есть когда

$$d < \frac{s^2(n_1 - 1)}{\chi_{1-\alpha}^2(n_1 - 1)}.$$

Вероятность ошибки в этом решении не превышает заданного значения α , поскольку в данной ситуации

$$P\left(\sigma^2 \leq \frac{s^2(n_1 - 1)}{\chi_{1-\alpha}^2(n_1 - 1)}\right) \leq 1 - Q = \alpha,$$

значит, в силу монотонности вероятностной меры

$$P(\sigma^2 < d) < P\left(\sigma^2 < \frac{s^2(n_1 - 1)}{\chi_{1-\alpha}^2(n_1 - 1)}\right).$$

Из этого следует, что вероятность ошибки первого рода $P(H_1 / H_0) \leq \alpha$.

В третьем варианте, показанном на рис. 40, в), не имеется достаточных оснований для принятия никакого иного решения, кроме решения продолжить испытания с целью увеличения объема выборки. Надежда здесь возлагается на то, что с увеличением объема выборки ширина доверительного интервала уменьшается.

После увеличения объема выборки до n_2 снова строится доверительный интервал при тех же заданных значениях вероятностей α и β , и на втором этапе вновь возможны три варианта расположения нового, более

узкого доверительного интервала относительно значения d . Все описанные рассуждения и действия повторяются.

В конечном итоге описанная последовательная процедура проверки сложной гипотезы о дисперсии должна завершиться выходом доверительного интервала целиком в одну из зон и принятием соответствующей гипотезы. Когда по экономическим, техническим и иным причинам дальнейшее продолжение испытаний (экспериментов) оказывается невозможным, придется принимать решение об отклонении той или иной гипотезы в зависимости от того, какого рода риск (первого или второго) более оправдан.

2.7.4. Проверка сложной гипотезы об интерквартильном промежутке

Непрерывная случайная величина ξ образует генеральную совокупность \mathbf{X} , на которой определены вероятностная мера и плотность распределения $\varphi(x)$. Вид плотности распределения произвольный. Из генеральной совокупности извлечена выборка x_1, x_2, \dots, x_{n_1} объема n_1 .

Заданы допускаемые пределы $[\underline{x}_{\text{доп}}, \bar{x}_{\text{доп}}]$, внутри которых должен находиться генеральный интерквартильный промежуток J_P , содержащий не менее $P_{\text{доп}} 100\%$ значений этой генеральной совокупности. Это означает, что в силу монотонности вероятностная мера интервала $[\underline{x}_{\text{доп}}, \bar{x}_{\text{доп}}]$, которая определена на генеральной совокупности \mathbf{X} , не должна быть меньше, чем $P_{\text{доп}}$, поскольку предполагаемый интерквартильный промежуток вложен в этот интервал. Поэтому гипотеза об интерквартильном промежутке

$$H_0: J_P \subset (\underline{x}_{\text{доп}}, \bar{x}_{\text{доп}}] \text{ против альтернативы } H_1: J_P \not\subset (\underline{x}_{\text{доп}}, \bar{x}_{\text{доп}}]$$

эквивалентна гипотезе

$$H_0: p = P(\underline{x}_{\text{доп}} < \xi \leq \bar{x}_{\text{доп}}) \geq P_{\text{доп}} \text{ против альтернативы}$$

$$H_1: p = P(\underline{x}_{\text{доп}} < \xi \leq \bar{x}_{\text{доп}}) < P_{\text{доп}},$$

то есть гипотезе о вероятности

$$H_0: p \geq P_{\text{доп}} \text{ против альтернативы } H_1: p < P_{\text{доп}}.$$

Заданы вероятности ошибок первого и второго рода α и β .

Гипотезы такого рода проверяются при испытаниях продукции, на характеристики которых установлены пределы допускаемых отклонений от номинальных значений, объявленных на этикетках, в документах или в рекламе.

Естественной статистикой для проверки этой гипотезы является оценка вероятности в соответствии с ее частотным определением:

$$p = \frac{m_0}{n},$$

где m_0 – количество выборочных значений, попавших в интервал $[\underline{x}_{\text{доп}}, \bar{x}_{\text{доп}}]$, n – объем выборки.

Излагаемый метод проверки сформулированной гипотезы относится к distribution-free методам, то есть к методам, которые в отличие от предыдущих, пригодных только при нормальном распределении, не зависят от вида плотности распределения.

Будем использовать подход, который применялся в предыдущих пунктах. Для этого необходимо определить границы доверительного интервала для вероятности. В соответствии с разд. 2.4.2 и заданными вероятностями ошибочных решений α и β границы доверительного интервала для вероятности определяются из уравнений:

нижняя граница

$$\sum_{m=m_0}^n C_n^m p_\beta^m (1 - p_\beta)^{n-m} = \beta;$$

верхняя граница

$$\sum_{m=0}^{m_0} C_n^m p_\alpha^m (1 - p_\alpha)^{n-m} = \alpha.$$

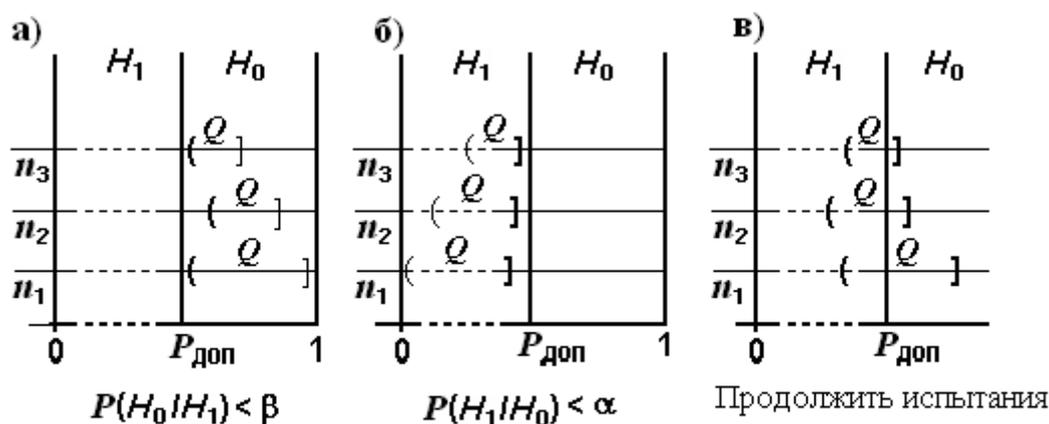


Рис. 41. Варианты расположения доверительных интервалов при проверке гипотезы о вероятности

На рис. 41 показаны варианты расположения доверительного интервала для вероятности относительно зон, соответствующих гипотезам H_0 и H_1 . В соответствии с формулировкой гипотезы и в связи с тем, что $P \leq 1$, положение верхней границы доверительного интервала в зоне, соответствующей гипотезе H_0 , безразлично. По указанным причинам для этого варианта строится доверительный интервал, верхняя граница которого равна 1, а нижняя граница определяется заданной вероятностью β ошибки второго рода:

$$P(p_\beta \leq p \leq 1) = Q = 1 - \beta.$$

Таким образом, решение о справедливости гипотезы H_0 принимается, когда нижняя граница доверительного интервала для вероятности оказывается больше заданного значения $P_{\text{доп}}$, то есть когда

$$P_{\text{доп}} \leq p_\beta.$$

В данной ситуации вероятность того, что истинное значение вероятности находится в зоне, соответствующей гипотезе H_1 , не превышает β , поскольку в данной ситуации $P(p < p_\beta) = \beta$, значит, в силу монотонности вероятностной меры $P(p < P_{\text{доп}}) \leq P(p < p_\beta)$.

Из этого следует, что вероятность ошибки второго рода $P(H_0 / H_1) \leq \beta$.

В соответствии с формулировкой гипотезы и поскольку $P \geq 0$ положение нижней границы доверительного интервала в зоне, соответствующей гипотезе H_1 , безразлично. В связи с этим здесь строится доверительный интервал, нижняя граница которого равна 0, а верхняя граница определяется заданной вероятностью α ошибки первого рода:

$$P(0 \leq p \leq p_\alpha) = Q = 1 - \alpha.$$

Решение об отклонении нулевой гипотезы принимается, когда верхняя граница доверительного интервала для вероятности меньше заданного значения $P_{\text{доп}}$, то есть когда

$$p_\alpha < P_{\text{доп}}.$$

Вероятность ошибки в этом решении не превышает заданного значения α , поскольку в данной ситуации $P(p > p_\alpha) = 1 - Q = \alpha$; значит, в силу монотонности вероятностной меры $P(p > P_{\text{доп}}) \leq P(p > p_\alpha)$.

Из этого следует, что вероятность ошибки первого рода $P(H_1 / H_0) \leq \alpha$.

В третьем варианте, показанном на рис. 41, в), не имеется достаточных оснований для принятия никакого иного решения, кроме решения продолжить испытания с целью увеличения объема выборки. Надежда здесь возлагается на то, что с увеличением объема выборки ширина доверительного интервала уменьшается.

После увеличения объема выборки до n_2 снова строятся доверительные интервалы при тех же заданных значениях вероятностей α и β , и на втором этапе вновь возможны три варианта расположения нового, более узкого доверительного интервала относительно значения $P_{\text{доп}}$. Все описанные рассуждения и действия повторяются.

В конечном итоге описанная последовательная процедура проверки сложной гипотезы о вероятности должна завершиться выходом доверительного интервала целиком в одну из зон и принятием соответствующей гипотезы. Когда по экономическим, техническим и иным причинам дальнейшее продолжение испытаний (экспериментов) оказывается невозможным, придется принимать произвольное (так называемое волевое) решение

об отклонении той или иной гипотезы в зависимости от того, какого рода риск (первого или второго) более оправдан.

Удобным для практического применения может быть иной метод проверки сложной гипотезы о вероятности, а именно, метод статистического последовательного анализа А. Вальда [8,9], которому посвящен отдельный раздел 2.7.6.

2.7.5. Проверка сложной гипотезы о числовых характеристиках случайных величин с применением доверительных интервалов, построенных методом бутстреп

Как следует из разд. 2.4.7, построение доверительных интервалов для любых числовых характеристик случайных величин и параметров плотности распределений при заданной доверительной вероятности Q методом бутстреп не зависит от вида плотности распределения. Однако этот метод не позволяет отдельно назначать нижнюю и верхнюю границы доверительного интервала по заданным значениям вероятностей ошибок первого и второго рода α и β , как в разд. 2.7.3, 2.7.4 и в 2.7.2, “а”). Поэтому здесь могут быть гарантированы только равные вероятности этих ошибок $\alpha = \beta$, как в разд. 2.7.2, “б”).

Пусть на вещественной оси задан некоторый интервал J , может быть полубесконечный. Применим для обозначения числовой характеристики или параметра плотности распределения генеральной совокупности \mathbf{X} общий символ Θ .

Пусть из генеральной совокупности \mathbf{X} извлечена выборка x_1, x_2, \dots, x_{n_1} объема n_1 . Задача состоит в проверке гипотезы

$$H_0: \Theta \in J \text{ против альтернативы } H_1: \Theta \in \bar{J}.$$

Сформулированная гипотеза проверяется по той же схеме, которая представлена в разд. 2.7.2. – 2.7.4.

Задаются равные вероятности ошибок первого и второго рода $\alpha = \beta$. По имеющейся выборке методом бутстреп (разд. 2.4.7) для доверительной вероятности $Q = 1 - \alpha$ строится доверительный интервал J_Q .

Если этот доверительный интервал целиком попадает в зону, соответствующую нулевой гипотезе, то есть если $J_Q \subset \underline{J}$, принимается решение о справедливости гипотезы H_0 . Вероятность ошибочности этого решения, как и в разд. 2.7.2, “б”), не превышает α .

Напротив, если доверительный интервал J_Q целиком попадает в зону, соответствующую альтернативной гипотезе, то есть если $J_Q \subset \bar{J}$, то принимается решение об отклонении гипотезы H_0 . Вероятность ошибочности этого решения, как и в разд. 2.7.2, б), также не превышает α .

Если не реализуется ни та, ни другая ситуация, принимается решение о наращивании объема выборки путем выполнения дополнительных экспериментов, если этому нет препятствий экономического, технического или иного характера.

После того как объем выборки увеличился, описанная процедура повторяется. Процедура останавливается в результате принятия решения о справедливости той или иной гипотезы или при невозможности или нецелесообразности ее продолжения (см. также разд. 2.7.2 – 2.7.4).

2.7.6. Проверка гипотезы о вероятности методом статистического последовательного анализа

Описывается некоторое событие \mathbf{A} , которое может произойти или не произойти в результате повторных независимых испытаний (экспериментов). Примером такого события может служить попадание выборочного значения в интервал с заданными границами $[\underline{x}_{\text{доп}}, \bar{x}_{\text{доп}}]$, как это было в разд. 2.7.4. Проверяется гипотеза о вероятности $P_{\mathbf{A}}$ события \mathbf{A} :

$$H_0: P(\mathbf{A}) = P_0 \text{ против альтернативы } H_1: P(\mathbf{A}) = P_1.$$

В процессе испытаний после каждого из них ведется подсчет количества k случаев, когда, например, событие \mathbf{A} не произошло, и это количество сравнивается со значениями, которые лежат на двух прямых линиях:

$$y(n) = \frac{1}{\ln \frac{P_0}{P_1} - \ln \frac{1-P_0}{1-P_1}} \left(n \ln \frac{1-P_1}{1-P_0} + \ln \frac{1-\beta}{\alpha} \right),$$

$$y(n) = \frac{1}{\ln \frac{P_0}{P_1} - \ln \frac{1-P_0}{1-P_1}} \left(n \ln \frac{1-P_1}{1-P_0} + \ln \frac{\beta}{1-\alpha} \right),$$

где α и β – вероятности ошибочных решений первого и второго рода соответственно, которые обеспечиваются, если каждая из них не превышает 0,15.

Пример таких прямых приведен на рис. 42. Они параллельны. Часть плоскости, лежащая над верхней прямой, – зона отклонения нулевой гипотезы. Часть плоскости, лежащая под нижней прямой, – зона принятия нулевой гипотезы. Зона между ними называется *зоной безразличия*. Ее ширина определяется вероятностями α и β , а также разностью между вероятностями P_0, P_1 . Для того, чтобы применить последовательный анализ к проверке сложной гипотезы об интерквантильном промежутке, описанной в разд. 2.7.5, ее нужно преобразовать к проверке простой гипотезы. В качестве P_0 следует принять $P_{\text{доп}}$ и взять $P_1 = (0,8 \div 0,9)P_0$.

Процесс последовательной проверки сформулированной гипотезы отображается на координатной плоскости (см. рис. 42) ступенчатой траекторией точки с координатами (n, k) , где n – номер очередного испытания (эксперимента), k – суммарное количество случаев неосуществления события **A**. Испытания продолжаются до тех пор, пока эта траектория не выйдет в зону той или иной гипотезы.

А. Вальдом [8] доказано, что при неограниченном количестве испытаний процесс заканчивается с вероятностью 1. Однако на практике количество испытаний ограничено по экономическим, техническим и иным причинам. Поэтому при выходе на ограничение приходится принимать то или иное решение в зависимости от того, какого рода риск (первого или второго) более оправдан.

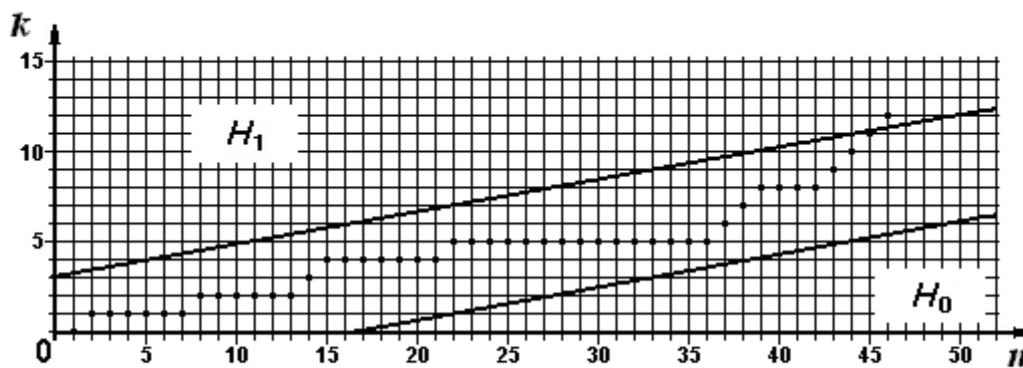


Рис. 42. Пример траектории процесса проверки гипотезы о вероятности методом статистического последовательного анализа

Преимущество статистического последовательного анализа перед подходом, изложенным в разд. 2.7.4, состоит в том, что здесь гипотеза проверяется при каждом новом результате непосредственно в процессе испытаний, и за счет этого общая трудоемкость в среднем по множеству подобных процессов уменьшается.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Основной

1. Солопченко Г.Н. Теория вероятностей и математическая статистика: учебное пособие/Г.Н.Солопченко. 2014, 214 с. Электронный ресурс <http://dl.unilib.neva.ru/dl/2/3614.pdf>.

2. Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман.– М. : Высш. шк., 2010.

3. Ивановский Р. И. Теория вероятностей и математическая статистика / Р. И. Ивановский.– СПб. : БХВ-Петербург, 2008.

Дополнительный

4. Крамер Г. Математические методы статистики / Г. Крамер.– М. : Мир, 1975.

5. Рао С. Р. Линейные статистические методы и их применение / С. Р. Рао.– М. : Наука. 1968.

6. Уилкс С. Математическая статистика / С. Уилкс.– М. : Наука, 1967.

7. Дрейпер Н. Прикладной регрессионный анализ. Т. 1,2 / Н. Дрейпер, Г. Смит.– М. : Финансы и статистика, 1987.

8. Вальд А. Последовательный анализ / А. Вальд.– М. : Физматгиз, 1960.

9. Башаринов А. Е. Методы последовательного статистического анализа и их приложения / А. Е. Башаринов, Б. С. Флейшман.– М. : Сов. Радио, 1962.

10. Эфрон Б. Нетрадиционные методы многомерного анализа / Б. Эфрон. – М. : Финансы и статистика. 1960.

11. Подборка статей по методу бутстреп // Заводская лаборатория. – 1987. № 10.

12. Климов Г. П. О фидуциальном подходе в статистике // Доклады АН СССР.– 1970.– т. 191, № 4.– с. 763-765.

13. Большев Л. Н. Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов; ВЦ АН СССР. М., 1968.

14. Оуэн Д. Б. Сборник статистических таблиц / Д. Б. Оуэн; ВЦ АН СССР. М., 1966.

15. Максимов Ю.Д. Математика. Вып. 6. Теория вероятностей. Контрольные задания с образцами тестов / Ю.Д. Максимов, А. Б. Куклин, Ю. А. Хватов. – СПб.: Изд-во СПбГПУ, 2008.