Б. А. СЕВАСТЬЯНОВ

КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

*Допущено Министерством высшего
и среднего специального образования СССР
в качестве учебника для студентов вузов,
обучающихся по специальностям
«Математика» и «Механика»*

МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1982

22.17
С 28
УДК 519.2

Севастьянов Б. А. **Курс теории вероятностей и математической статистики.**— М.: Наука. Главная редакция физико-математической литературы, 1982. — 256 с.

В основу книги положен годовой курс лекций, читавшихся автором в течение ряда лет на отделении математики механико-математического факультета МГУ. Основные понятия и факты теории вероятностей вводятся первоначально для конечной схемы. Математическое ожидание в общем случае определяется так же, как интеграл Лебега, однако у читателя не предполагается знание никаких предварительных сведений об интегрировании по Лебегу.

В книге содержатся следующие разделы: независимые испытания и цепи Маркова, предельные теоремы Муавра — Лапласа и Пуассона, случайные величины, характеристические и производящие функции, закон больших чисел, центральная предельная теорема, основные понятия математической статистики, проверка статистических гипотез, статистические оценки, доверительные интервалы.

Для студентов младших курсов университетов и втузов, изучающих теорию вероятностей.

С $\frac{1702060000 - 143}{053(02)-82}$ 12-82

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Первоначальный курс теории вероятностей и математической статистики должен удовлетворять двум условиям. С одной стороны, он должен помогать развитию теоретико-вероятностной интуиции, т. е. умения строить математические модели, правильно отражающие те или иные стороны реальных случайных явлений. При этом надо иметь в виду, что теория вероятностей и математическая статистика тесно связаны с различными приложениями, с некоторыми из которых выпускникам математических отделений университетов с большой вероятностью придется столкнуться в своей работе. С другой стороны, теория вероятностей должна развиваться как математическая наука, построенная на точных определениях и аксиомах. Однако многие существующие руководства по теории вероятностей придерживаются одной из двух крайностей. В одних курсах, нацеленных на приложения, нет четкого разделения реальных случайных явлений и их математических моделей. В частности, важное в теории вероятностей понятие независимости молчаливо смешивается с причинной независимостью реальных явлений. Другие курсы посвящены, главным образом, строгому изложению математических основ теории вероятностей, поэтому они либо очень велики по объему, либо в значительной степени опираются на такие понятия функционального анализа, как мера и интеграл Лебега, и поэтому не могут быть использованы при обучении студентов младших курсов.

Содержание данного учебника соответствует годовому курсу теории вероятностей и математической статистики, который автор читал в течение ряда лет на механико-математическом факультете Московского государственного университета студентам-математикам 4-го и 5-го семестров. Для преодоления указанных выше трудностей автор придерживается некоторого компромиссного направления. Первоначально многие теоретико-вероятностные понятия введены в простом случае конечного вероятностного пространства. Приведен ряд примеров, в которых указана связь вводимых математических понятий с теми или иными свойствами реальных явлений. Общий случай основан на способе изложения, который связан с введением интеграла Лебега без теории меры. На 4-м семестре, когда студенты еще не знакомы с соответствующими понятиями функционального анализа, аксиоматически вводится понятие вероятностной меры и на ее основе определяется математическое ожидание как интеграл Лебега. Теорема Каратеодори о продолжении меры формулируется без доказательства. Понятия условного распределения вероятностей и условного математического ожидания даны не в полном объеме, а лишь в простых случаях дискретных и абсолютно непрерывных распределений. В основном автор старался опираться лишь на знание студентами классического математического анализа.

Главы 1—5 связаны в основном с конечными вероятностными пространствами. В этих главах введены основные понятия вероятности, математического ожидания, независимости, случайной величины. Распространение этих понятий на общий случай дано в главах 6—12. Главы 13—16 посвящены некоторым задачам математической статистики. Каждая глава сопровождается небольшим количеством задач. Однако автор предполагает, что читатель использует какой-нибудь задачник (например, Севастьянов Б. А., Чистяков В. П., Зубков А. М. Сборник задач по теории вероятностей. — М.: Наука, 1980).

Москва,
май 1981 г.

Б. А. Севастьянов

Глава 1. ВЕРОЯТНОСТНОЕ ПРОСТРАНСТВО

§ 1. Предмет теории вероятностей

Сочетание слов «теория вероятностей» на неискушенного человека производит несколько странное впечатление. В самом деле, слово «теория» связывается с наукой, а наука изучает закономерные явления; слово «вероятность» в обычном языке связывается с чем-то неопределенным, случайным, незакономерным. Поэтому люди, знающие о существовании теории вероятностей только понаслышке, говорят о ней часто иронически. Однако теория вероятностей — это большой, интенсивно развивающийся раздел математики, изучающий случайные явления. Так в чем же тут дело? Как разрешить это противоречие между тем, что теория вероятностей — это наука, а ее предмет — случайность, которая, казалось бы, не поддается никакому научному предсказанию? Как мы увидим ниже, противоречие здесь только кажущееся, так как теория вероятностей изучает закономерности случайных явлений.

Математика, как и любая другая наука, изучает закономерные явления реального мира. Связь между математикой и объектом исследования можно изобразить схематически следующим образом (см. рис. 1). Классическим примером такой схемы является механика, созданная Ньютоном. На основе многовековых наблюдений движений небесных тел, а также практической деятельности людей, связанной со строительством и производством, Ньютон сформулировал несколько простых законов механики в виде аксиом и закон всемирного тяготения, из которых дедуктивными рассуждениями можно было объяснить все явления, которые наблюдались ранее, а также предсказать многие новые факты. Построение математических моделей реальных механических и физических процессов привело к созданию математического анализа.

Закономерное событие — это событие, которое всегда осуществляется, как только создаются определенные условия. Закономерное явление — это система закономерных событий. Роль математики, в частности теории дифференциальных уравнений, при изучении реальных закономерных явлений общеизвестна. Но наряду с закономерными мы все время сталкиваемся в практической деятельности с событиями незакономерными или, иначе, случайными. Это события, которые при одних и тех же условиях иногда происходят, а иногда — нет. Например, человек, заболевший гриппом в период эпидемии, может выздороветь, может получить те или иные тяжелые осложнения, или умереть. Таким образом, исход заболевания гриппом случаен.

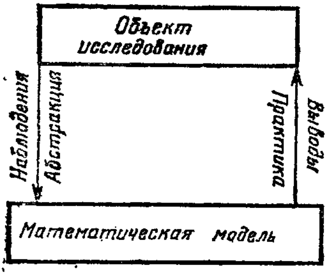

Рис. 1.

Казалось бы, что там, где мы имеем дело со случайными событиями, науке, в частности математике, делать нечего. Ведь наука открывает научные законы, которые помогают предсказывать течение того или иного процесса или явления, а случайное явление — это как раз такое явление, предсказать исход которого невозможно. Однако и случайные события подчиняются некоторым закономерностям, которые мы назовем *вероятностными* закономерностями. Прежде всего условимся, что мы будем иметь дело не со всякими случайными событиями, а с *массовыми* случайными событиями, т. е. мы будем предполагать, что в принципе возможно создать много раз одни и те же условия, при каждом из которых может произойти или нет некоторое случайное событие. Пусть при осуществлении некоторых условий N раз случайное событие A осуществляется $N(A)$ раз. Число $N(A)$ называется *частотой* события A, а отношение $N(A)/N$ — *относительной частотой* события A. Оказывается, при больших N относительная частота $N(A)/N$ для случайных массовых событий обладает так называемым свойством устойчивости, которое состоит в том, что в нескольких сериях из достаточно больших N_1, N_2, ..., N_s

наблюдений события A в одних и тех же условиях мы обычно имеем приближенные равенства

$$\frac{N_1(A)}{N_1} \approx \frac{N_2(A)}{N_2} \approx \ldots \approx \frac{N_s(A)}{N_s}.$$

Таким образом, относительная частота события A колеблется около одного и того же числа, которое характеризует данное случайное событие A. Это число $\mathsf{P}(A)$ в соответствующей математической модели мы будем называть вероятностью события A. Например, мы можем много раз подбрасывать одну и ту же монету. Пусть случайное событие A — это выпадение герба при одном бросании. В случае бросания «правильной» (симметричной, однородной) монеты $\mathsf{P}(A) = 1/2$. Статистика рождений показывает, что мальчиков рождается несколько больше, чем девочек, причем наблюдаемая доля рождений мальчиков равна 0,51—0,52 (в разные периоды, в разных странах могут быть колебания). Медицинская статистика свидетельствует о том, что смертность от гриппа имеет малую, но ненулевую вероятность (поэтому в условиях массовой эпидемии число смертных случаев от гриппа становится заметным).

Устойчивость частот — это объективное свойство массовых случайных явлений реального мира. Отсутствие устойчивости частот в сериях испытаний свидетельствует о том, что условия, при которых производятся испытания, претерпевают значительные изменения. Теория вероятностей — это математическая наука, которая изучает математические модели случайных явлений. Если говорить более подробно, то теория вероятностей устанавливает такие связи между вероятностями случайных событий в математических моделях, которые позволяют вычислять вероятности сложных событий по вероятностям более простых событий.

В теории вероятностей используются результаты и методы многих областей математики (комбинаторики, математического анализа, алгебры, логики и т. п.). Однако теория вероятностей обладает некоторым своеобразием, поскольку она очень тесно связана с различными приложениями, причем приложения эти не столь привычны, как, например, приложения дифференциальных уравнений. Поэтому овладеть теорией вероятностей может лишь тот, кто решает много задач (эти задачи часто имеют нематематическую постановку, и надо уметь построить соответствующую математическую модель) и приобретает, таким образом, теоретико-вероятностную интуицию.

§ 2. События

Одним из основных понятий теории вероятностей является случайное событие или, как мы будем чаще говорить, просто событие. В реальном мире случайное событие — это исход (какого-либо испытания, наблюдения, эксперимента), который может произойти (наступить, осуществиться) или не произойти (не наступить, не осуществиться).

Пример 1. При бросании игральной кости[1]) может выпасть число очков, равное какому-либо числу из множества чисел 1, 2, 3, 4, 5, 6. Событиями в этом случае будут, например,

$A=$ {выпадает четное число очков},

$B=$ {выпадает число очков, не большее трех}.

В математической модели можно принять понятие события как первоначальное, которому не дается определения и которое характеризуется лишь своими свойствами. Исходя из реального смысла понятия события, мы можем определить следующие частные случаи понятия события и следующие операции над событиями. В тех случаях, когда мы одновременно рассматриваем несколько событий, мы всегда будем предполагать, что эти события могут произойти или не произойти при одном и том же испытании (т. е. при осуществлении одних и тех же условий).

Достоверным событием будем называть событие, которое всегда происходит, и будем его обозначать Ω. *Невозможным* событием назовем событие, которое никогда не происходит. Обозначать невозможное событие будем $\varnothing$. Событие $\bar{A}$ назовем событием, *противоположным* A,

[1]) Игральной костью называется кубик, сделанный из однородного материала, грани которого занумерованы цифрами 1, 2, 3, 4, 5, 6. Число очков, выпавшее при бросании игральной кости, — это цифра на верхней грани кубика.

если оно происходит тогда и только тогда, когда не происходит A. *Суммой* или *объединением* событий A и B назовем событие, обозначаемое $A \cup B$ или $A + B$, которое происходит тогда и только тогда, когда происходят или A, или B (или оба вместе). *Произведением* или *пересечением* событий A и B назовем событие, обозначаемое $A \cap B$ или AB, которое происходит тогда и только тогда, когда происходят и A и B вместе. *Разностью* $A \setminus B$ событий A и B назовем событие, которое происходит тогда и только тогда, когда происходит A и не происходит B. События A и B назовем *несовместными*, если $AB = \varnothing$. Мы будем писать $A \subseteq B$ и говорить, что событие A *влечет за собой* событие B, если из наступления события A следует наступление события B. Если $A \subseteq B$ и $B \subseteq A$, то мы будем говорить, что события A и B равносильны, и писать $A = B$.

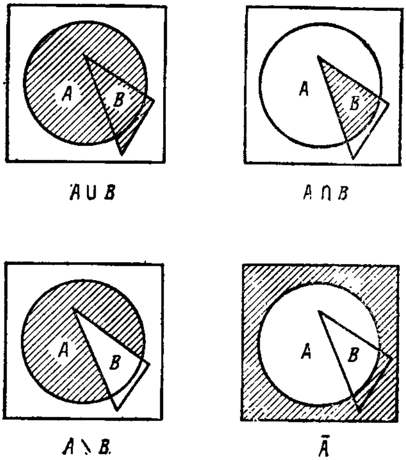

Рис. 2. Сумма, произведение, разность событий A и B; событие $\bar{A}$ противоположно A.

В примере 1 с бросанием игральной кости имеем следующие события:

$$A \cup B = \{\text{выпадает число очков, отличное от пяти}\},$$
$$A \cap B = \{\text{выпадает число очков, равное двум}\},$$
$$A \setminus B = \{\text{выпадает число очков, равное 4 или 6}\},$$
$$\bar{A} = \{\text{выпадает нечетное число очков}\}.$$

Пример 2. На квадрат случайно бросается частица (см. рис. 2);

событие $A = \{\text{частица попадает в круг } A\}$,

событие $B = \{\text{частица попадает в треугольник } B\}$.

События $A \cup B$, $A \cap B$, $A \setminus B$ и $\bar{A}$ в этом случае — это попадание частицы в области, получаемые объединением, пересечением, разностью областей A и B и дополнением $\bar{A}$ в квадрате (на рис. 2 соответствующие области заштрихованы).

В настоящее время в теории вероятностей наиболее распространенным является подход, в котором событие определяется через неопределяемое понятие элементарного события. Наиболее употребительная теоретико-вероятностная модель в простых случаях — это урновая модель. Пусть имеется урна с n одинаковыми шарами. Испытание состоит в том, что мы случайно выбираем из урны один шар. Обозначим $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ множество шаров в урне. Если из урны при испытании мы вынимаем шар $\omega_i \in A$, где A — некоторое подмножество множества шаров Ω, то мы будем говорить, что произошло событие A; если же $\omega_i \notin A$, то мы будем говорить, что событие A не произошло. В данном случае событие A отождествляется с подмножеством A множества всех возможных исходов или, как мы будем далее говорить, элементарных событий.

В общем случае мы будем в каждой теоретико-вероятностной модели рассматривать некоторое основное множество $\Omega = \{\omega\}$. Будем называть его элементы ω *элементарными событиями*, само множество Ω — *пространством элементарных событий*, а некоторые его подмножества $A \subseteq \Omega$ — *событиями*. Операции над событиями — это операции над подмножествами. При этом в теории вероятностей употребляется своя терминология, связь которой с теоретико-множественной терминологией отражена в таблице 1.

Операции суммы и произведения событий можно распространить на любое конечное или бесконечное множество событий $\bigcup_\alpha A_\alpha$, $\bigcap_\alpha A_\alpha$. Обычные свойства операций над множествами переносятся на операции над событиями, например,

$$\overline{\bigcup_\alpha A_\alpha} = \bigcap_\alpha \bar{A}_\alpha, \quad \overline{\bigcap_\alpha A_\alpha} = \bigcup_\alpha \bar{A}_\alpha, \quad \bar{\bar{A}} = A,$$

$$\bar{A} = \Omega \setminus A, \quad \bar{\varnothing} = \Omega, \quad \bar{\Omega} = \varnothing,$$

$$A \setminus B = A \setminus AB = A\overline{B}, \quad A \setminus (A \setminus B) = AB,$$

$$A \subseteq B \Rightarrow \overline{B} \subseteq \overline{A}$$

и т. д. Иногда придерживаются следующего соглашения: если $A_1, A_2, \ldots, A_n$ попарно несовместны, то вместо $\bigcup\limits_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$ пишут $\sum\limits_{i=1}^{n} A_i = A_1 + A_2 + \ldots + A_n$.

В общем случае бесконечного пространства Ω мы рассматриваем не все подмножества Ω, а лишь некоторые классы этих подмножеств, называемые алгебрами и σ-алгебрами множеств.

Таблица 1

Обозначение	Терминология в теории множеств	Терминология в теории вероятностей
Ω	пространство (основное множество)	пространство элементарных событий, достоверное событие
$\omega, \omega \in \Omega$	элемент пространства ω	элементарное событие ω
$A, A \subseteq \Omega$	множество A	событие A
$A \cup B$, $A + B$	сумма или объединение множеств A и B	сумма событий A и B
$A \cap B, AB$	пересечение множеств A и B	произведение событий A и B
$A \setminus B$	разность множеств A и B	разность событий A и B
$\varnothing$	пустое множество	невозможное событие
$\overline{A}$	дополнительное множество A	противоположное A событие
$AB = \varnothing$	A и B не пересекаются	A и B несовместны
$A \subseteq B$	A есть подмножество B	A влечет событие B
$A = B$	A и B равны	A и B равносильны

Определение 1. Назовем класс $\mathscr{A}$ подмножеств пространства Ω *алгеброй множеств*, если

1) $\varnothing \in \mathscr{A}$, $\Omega \in \mathscr{A}$;

2) из $A \in \mathscr{A}$ следует $\bar{A} \in \mathscr{A}$;

3) из $A, B \in \mathscr{A}$ следует $A \cup B \in \mathscr{A}$, $A \cap B \in \mathscr{A}$.

Определение 2. Алгебру множеств $\mathscr{A}$ назовем *σ-алгеброй*, если из $A_n \in \mathscr{A}$, $n = 1, 2, \ldots$, следует $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{A}$, $\bigcap\limits_{n=1}^{\infty} A_n \in \mathscr{A}$.

§ 3. Вероятностное пространство

Определение 3. Тройку $(\Omega, \mathscr{A}, \mathsf{P})$, где Ω — пространство элементарных событий, $\mathscr{A}$ — σ-алгебра подмножеств Ω, называемых *событиями*, P — числовая функция, определенная на событиях и называемая *вероятностью*, будем называть *вероятностным пространством*, если выполнены следующие аксиомы:

1°. $\mathsf{P}(A) \geqslant 0$ для всех $A \in \mathscr{A}$ (*неотрицательность* P);

2°. $\mathsf{P}(\Omega) = 1$ (*нормированность* P);

3°. $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B)$, если $AB = \varnothing$ (*аддитивность* P);

4°. Если $A_n \downarrow \varnothing$, т. е. $A_1 \supseteq A_2 \supseteq \ldots$ и $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$, то $\lim\limits_{n\to\infty} \mathsf{P}(A_n) = 0$ (*непрерывность* P).

Из этих аксиом вытекают следующие свойства вероятности.

1) *Если* $A \subseteq B$, *то* $\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A)$.

Так как $B = A + (B \setminus A)$ и $A \cap (B \setminus A) = \varnothing$, то по аксиоме 3°

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

2) *Если* $A \subseteq B$, *то* $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

Следует из (1).

3) *Для любого* $A \in \mathscr{A}$

$$0 \leqslant \mathsf{P}(A) \leqslant 1.$$

Следует из 2), так как $\varnothing \subseteq A \subseteq \Omega$.

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

Следует из аксиомы 3°, так как $A + \bar{A} = \Omega$, $A\bar{A} = \varnothing$.

5) $\mathsf{P}(\varnothing) = 0$.

Следует из 4) и аксиомы 2°.

6) *Имеет место* *конечная аддитивность: если $A_iA_j = \varnothing$ для любых $i \neq j$, то*

$$\mathsf{P}(A_1 + A_2 + \ldots + A_n) = \sum_{k=1}^{n} \mathsf{P}(A_k). \tag{2}$$

Следует из аксиомы 3°. Доказывается по индукции.

7) *Для любых событий $A_1, \ldots, A_n$*

$$\mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right) \leqslant \sum_{k=1}^{n} \mathsf{P}(A_k). \tag{3}$$

Представим $\bigcup\limits_{k=1}^{n} A_k$ в виде суммы попарно несовместных событий $B_k = A_k \setminus (A_1 \cup A_2 \cup \ldots \cup A_{k-1})$:

$$\bigcup_{k=1}^{n} A_k = \sum_{k=1}^{n} B_k.$$

По свойству аддитивности 6) имеем

$$\mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathsf{P}(B_k),$$

откуда следует (3), так как $\mathsf{P}(B_k) \leqslant \mathsf{P}(A_k)$.

8) *Для любых событий A и B*

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

Следует из $A \cup B = A + (B \setminus AB)$, аксиомы 3°: $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B \setminus AB)$ и свойства 1): $\mathsf{P}(B \setminus AB) = \mathsf{P}(B) - \mathsf{P}(AB)$.

Аксиомы 3° и 4° можно заменить одной аксиомой *счетной аддитивности* (или, как еще говорят, аксиомой *σ-аддитивности*).

3*. *Если события A_n в последовательности $A_1, A_2, \ldots$ попарно несовместны, то*

$$\mathsf{P}\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathsf{P}(A_n). \tag{4}$$

Теорема 1. *Система аксиом* 1°, 2°, 3°, 4° *равносильна системе аксиом* 1°, 2°, 3*.

Доказательство. Пусть справедливы аксиомы 1°, 2°, 3°, 4°, и пусть A_n — последовательность попарно несовместных событий. Обозначим $B_n = \sum_{k=n+1}^{\infty} A_k$, $A = \sum_{n=1}^{\infty} A_n$. Тогда A при любом n разлагается на конечную сумму попарно несовместных событий

$$A = A_1 + A_2 + \ldots + A_n + B_n,$$

поэтому

$$\mathsf{P}(A) = \sum_{k=1}^{n} \mathsf{P}(A_k) + \mathsf{P}(B_n).$$

Так как $B_n \downarrow \varnothing$, т. е. $B_1 \supseteq B_2 \supseteq \ldots$ и $\bigcap_{n=1}^{\infty} B_n = \varnothing$, то по аксиоме 4° имеем $\mathsf{P}(B_n) \downarrow 0$. Отсюда вытекает счетная аддитивность (4).

Пусть теперь выполнены аксиомы 1°, 2°, 3*, и пусть $B_n \downarrow \varnothing$. Обозначим $A_n = B_n \setminus B_{n+1}$, $n = 1, 2, \ldots$ События A_n попарно несовместны и

$$B_1 = \sum_{n=1}^{\infty} A_n, \quad B_n = \sum_{k=n}^{\infty} A_k,$$

поэтому по аксиоме 3* ряд $\mathsf{P}(B_1) = \sum_{n=1}^{\infty} \mathsf{P}(A_n)$ сходится, и сумма остатка этого ряда $\mathsf{P}(B_n) = \sum_{k=n}^{\infty} \mathsf{P}(A_k) \to 0$. Теорема доказана.

Система аксиом 1°, 2°, 3°, 4° или 1°, 2°, 3* определяет вероятностную меру на σ-алгебре $\mathscr{A}$ пространства Ω. Эта система аксиом предложена А. Н. Колмогоровым.

Происхождение аксиом 1°, 2°, 3° можно объяснить, исходя из свойства статистической устойчивости частот. Пусть A и B — несовместные события, $N(A)/N$ и $N(B)/N$ — их относительные частоты в какой-либо длинной серии наблюдений. Так как $N(A) \geqslant 0$, то $N(A)/N \geqslant 0$, следовательно, то число $\mathsf{P}(A)$, к которому близко отношение $N(A)/N$, должно быть неотрицательным, Для достоверного события $N(\Omega)=N$, поэтому надо потребовать $\mathsf{P}(\Omega)=1$. Для несовместных событий $N(A+B)=N(A)+N(B)$, откуда

$$\frac{N(A+B)}{N}=\frac{N(A)}{N}+\frac{N(B)}{N},$$

что и приводит к аксиоме 3°.

Аксиома 4° (или 3*) имеет несколько другое происхождение, связанное не с реальным свойством устойчивости частот, а с нуждами развиваемой на основе аксиоматики математической теории. Поясним сказанное на примере. Пусть на единичный квадрат бросается случайно частица, причем вероятность попадания в любой внутренний квадрат со сторонами, параллельными сторонам основного квадрата, равна площади меньшего квадрата. С помощью аксиомы 3° отсюда можно получить вероятность попадания в любую фигуру, составленную из суммы конечного числа квадратов. Но нам хотелось бы иметь также возможность находить вероятность попадания в более сложные фигуры, например, в круг. Это можно сделать с помощью аксиомы 3*, приближая круг фигурами, составленными из конечных сумм таких квадратов (см. рис. 3).

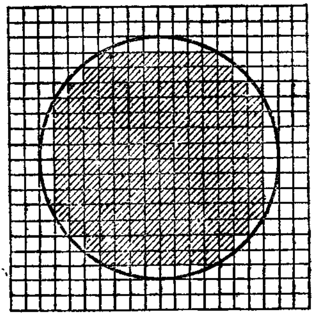

Рис. 3.

§ 4. Конечное вероятностное пространство. Классическое определение вероятности

Рассмотрим простой случай конечного вероятностного пространства. В этом случае $\Omega=\{\omega\}$ — конечное пространство, $\mathscr{A}$ — алгебра всех подмножеств множества Ω (ввиду конечности $\mathscr{A}$ эта алгебра автоматически представляет собой σ-алгебру). Вероятность $\mathsf{P}(A)$ для любого подмножества A из Ω в этом случае можно задать следующим образом. Пусть заданы неотрицательные числа p_ω такие, что $\sum\limits_{\omega\in\Omega} p_\omega=1$. Вероятность $\mathsf{P}(A)$

определим как сумму

$$\mathsf{P}(A) = \sum_{\omega \in A} p_\omega. \tag{5}$$

Легко видеть, что так определенная вероятность (вместе с $\mathsf{P}(\varnothing) = 0$) удовлетворяет всем аксиомам. Обозначим $|A|$ число элементов в множестве A. Частным случаем определения вероятности (5) будет так называемое *классическое определение вероятности*, когда все p_ω равны друг другу. Так как $1 = \sum_{\omega \in \Omega} p_\omega = p_\omega |\Omega|$, то в этом случае $p_\omega = \frac{1}{|\Omega|}$ и

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|}. \tag{6}$$

Модель вероятностного пространства, приводящая к классическому определению вероятности, используется в тех случаях, когда элементарные события обладают свойством «симметрии» в том смысле, что все элементарные события находятся в одинаковом отношении к тем условиям, которые определяют характер испытания. Например, бросание игральной кости или монеты обладает свойством «симметрии» по отношению к выпадению того или иного числа очков на кости или той или иной стороны монеты, если, конечно, при броске они были достаточно высоко от горизонтальной поверхности и им было придано в начале броска вращательное движение (но не вокруг оси симметрии). Таким же свойством симметрии обладают правильно организованная жеребьевка и тираж лотереи.

При нахождении вероятностей в схеме классического определения широко используется комбинаторика. Мы часто будем использовать комбинаторные понятия размещения, перестановки и сочетания. Будем исходить из конечного множества $X = \{x_1, x_2, \ldots, x_N\}$, состоящего из N элементов x_j. Пусть $1 \leqslant n \leqslant N$. *Размещением* из N элементов множества X по n элементам (коротко, размещением из N по n) назовем любой упорядоченный набор $\left(x_{j_1}, x_{j_2}, \ldots, x_{j_n}\right)$ элементов множества X. Два размещения $\left(x_{j_1}, x_{j_2}, \ldots, x_{j_n}\right)$ и $\left(x_{i_1}, x_{i_2}, \ldots, x_{i_n}\right)$

равны тогда и только тогда, когда все $x_{i_k} = x_{j_k}$, $k=1, \ldots$ $\ldots, n$. Число всех различных размещений из N элементов по n обозначается A_N^n и равно $N(N-1)\ldots(N-n+1)$. Последнее произведение мы иногда будем обозначать как обобщенную степень $N^{[n]}$. Таким образом, для числа всех размещений из N элементов по n мы имеем формулу

$$A_N^n = N^{[n]} = N(N-1)\ldots(N-n+1). \tag{7}$$

В дальнейшем будем полагать $A_N^0 = N^{[0]} = 1$ при любом целом $N \geqslant 1$. Формула (7) легко доказывается по индукции. Частный случай размещения при $N = n$ называется *перестановкой* из N элементов. Число всех перестановок из N элементов равно

$$A_N^N = N^{[N]} = N(N-1)\ldots 2 \cdot 1 = N! \tag{8}$$

Из (7) и (8) следует также формула

$$A_N^n = \frac{N!}{(N-n)!}. \tag{9}$$

Сочетанием из N элементов множества X по n называется любое подмножество $\{x_{j_1}, \ldots, x_{j_n}\}$ мощности n множества X. Общее число всех сочетаний из N по n обозначается C_N^n и равно

$$C_N^n = \frac{A_N^n}{n!} = \frac{N!}{n!\,(N-n)!}. \tag{10}$$

Из (10) имеем соотношение $C_N^n = C_N^{N-n}$. В дальнейшем будем полагать $0! = 1$, $C_N^0 = 1$ и $C_N^k = 0$, если k — целое и $k < 0$ или $k > N$.

Пример 3. *Выборка без возвращения.* Пусть имеется урна с N шарами, которые мы занумеруем числами $1, 2, \ldots, N$. Предположим, что шары с номерами $1, 2, \ldots, M$ белого цвета, остальные — черного. Выборка без возвращения состоит в том, что мы наугад вынимаем из урны последовательно n шаров, не возвращая их обратно. В этом случае за пространство элементарных событий $\Omega = \{\omega\}$ естественно принять множество всех упорядоченных наборов

$$\omega = (\alpha_1, \alpha_2, \ldots, \alpha_n) \tag{11}$$

чисел α_i, $1 \leqslant \alpha_i \leqslant N$, не равных друг другу. Мощность множества Ω равна в этом случае

$$|\Omega| = N(N-1)\ldots(N-n+1) = N^{[n]} \qquad (12)$$

— числу размещений N элементов по n.

Вычислим вероятность события A_m, состоящего в том, что среди выбранных n шаров имеется ровно m белых. Для этого подсчитаем $|A_m|$:

$$|A_m| = C_n^m M^{[m]}(N-M)^{[n-m]}. \qquad (13)$$

В самом деле, число элементарных событий (11), у которых ровно в m случаях $1 \leqslant \alpha_j \leqslant M$, определяется как произведение: C_n^m — числа способов выбора m координат из общего количества их n, на которые мы помещаем $1 \leqslant \alpha_j \leqslant M$; $M^{[m]}$ — числа различных наборов $1 \leqslant \alpha_j \leqslant M$, попадающих на отмеченные m мест; $(N-M)^{[n-m]}$ — числа различных наборов $M+1 \leqslant \alpha_j \leqslant \leqslant N$, попадающих на остальные места. Из (12) и (13) получаем

$$\mathsf{P}(A_m) = \frac{C_n^m M^{[m]}(N-M)^{[n-m]}}{N^{[n]}}.$$

Пользуясь (10), мы можем вероятность $\mathsf{P}(A_m)$ выразить в следующих эквивалентных видах:

$$\mathsf{P}(A_m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \frac{C_n^m C_{N-n}^{M-m}}{C_N^M}. \qquad (14)$$

Пример 4. *Выборка с возвращением.* Пусть имеется та же урна, но выборка n шаров из нее происходит последовательно по одному шару, и при этом каждый раз фиксируется номер шара, а сам шар возвращается обратно в урну. В этом случае пространство элементарных событий состоит из всевозможных векторов (11), у которых координаты не имеют никаких дополнительных ограничений, кроме $1 \leqslant \alpha_j \leqslant N$. В этом случае

$$|\Omega| = N^n,$$

а вероятность события A_m, вычисляемая аналогичным способом, равна

$$\mathsf{P}(A_m) = C_n^m \frac{M^m(N-M)^{n-m}}{N^n} = C_n^m \left(\frac{M}{N}\right)^m \left(1-\frac{M}{N}\right)^{n-m}. \qquad (15)$$

§ 5. Геометрические вероятности

Еще один важный класс моделей вероятностных пространств дают так называемые *геометрические вероятности*. Пусть $\Omega=\{\omega\}$ — область евклидова n-мерного пространства с конечным n-мерным объемом. Событиями назовем подмножества Ω, для которых можно определить n-мерный объем. За множество событий можно принять так называемую σ-алгебру $\mathscr{B}$ борелевских подмножеств Ω (подробнее об этом см. гл. 6, § 27). За вероятность события $A \in \mathscr{B}$ примем

$$\mathsf{P}(A)=\frac{|A|}{|\Omega|}, \qquad (16)$$

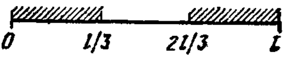

Рис. 4.

где $|V|$ означает n-мерный объем множества V. Понимая под n-мерным объемом соответствующую меру Лебега, мы получаем вероятностное пространство $(\Omega, \mathscr{B}, \mathsf{P})$, где вероятность P определена равенством (16). Это вероятностное пространство служит моделью задач, в которых частица случайно бросается в область Ω. Предполагается, что ее положение равномерно распределено в этой области, т. е. вероятность попасть частице в область A пропорциональна n-мерному объему этой области.

Пример 5. Стержень разламывается на две части в случайной точке, равномерно распределенной по длине стержня. Найти вероятность того, что меньший обломок имеет длину, не превосходящую одной трети длины стержня. Обозначим длину стержня l, а расстояние точки разлома от одного (фиксированного) конца стержня — x. Тогда описанное событие произойдет тогда и только тогда, когда либо $x \leqslant l/3$, либо $x \geqslant 2l/3$. Искомая вероятность равна отношению $(l/3+l/3):l=2/3$ (см. рис. 4).

Пример 6. *Задача Бюффона.* На плоскость, расчерченную параллельными прямыми, находящимися на расстоянии a друг от друга, случайно брошена игла длины $l<a$. Найти вероятность пересечения иглы с какой-нибудь из параллельных прямых. Обозначим y расстояние от середины иглы до ближайшей прямой, x — острый угол между иглой и перпендикуляром к парал-

лельным прямым (рис. 5). Координаты (x, y), определяющие положение иглы относительно параллельных прямых, удовлетворяют условиям $0 \leqslant x \leqslant \pi/2$, $0 \leqslant y \leqslant l/2$. На плоскости (x, y) они образуют прямоугольник Ω. Попадание точки (x, y) в заштрихованную область A (см. рис. 6) приводит к пересечению иглы с

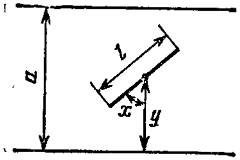

Рис. 5.

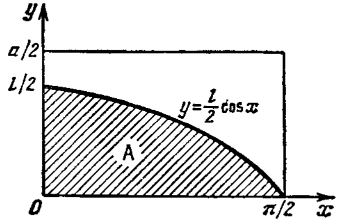

Рис. 6.

одной из параллельных прямых. По формуле (16) искомая вероятность равна

$$\frac{|A|}{|\Omega|} = \frac{\int\limits_0^{\pi/2} \frac{l}{2} \cos x \, dx}{a/2 \cdot \pi/2} = \frac{2l}{a\pi}.$$

Задачи

1. События A и B несовместны. Доказать, что $B = \overline{A}$ тогда и только тогда, когда $A + B = \Omega$.

2. Известно, что $A \cap B = \varnothing$ и $\overline{A} \cap \overline{B} = \varnothing$. Доказать, что в этом случае $B = \overline{A}$.

3. Доказать, что события $\overline{\overline{A}\overline{B} \cup A}$ и $B \setminus A$ равносильны.

4. Доказать, что $A \setminus (A \setminus B) = AB$.

5. Доказать, что: а) $AB = B$ тогда и только тогда, когда $B \subseteq A$; б) $A \cup B = B$ тогда и только тогда, когда $A \subseteq B$.

6. На карточке спортлото из 49 клеток отмечено шесть. Какова вероятность того, что ровно три из отмеченных клеток выпадут в очередном тираже? (В тираже производится случайная выборка шести элементов без возвращения из множества 49 клеток карточки спортлото.)

7. Трехзначное число случайно и равновероятно выбирается из всего множества трехзначных чисел. Найти вероятность того, что оно делится: а) на 3; б) на 5.

8. Деталь с вероятностью 0,01 имеет дефект A, с вероятностью 0,02 имеет дефект B и с вероятностью 0,005 имеет оба дефекта. Найти вероятность того, что деталь имеет хотя бы один дефект.

9. При жеребьевке N человек тянут билеты с номерами 1, 2, ..., N. Первые три человека вытянули номера x_1, x_2, x_3. Какова вероятность того, что

$$\min(x_1, x_2) < x_3 < \max(x_1, x_2)?$$

10. Из кармана, в котором находится 10 монет достоинством 20 коп. и 10 монет достоинством 3 коп., вынимается пригоршня из 10 случайно взятых монет. Какова вероятность того, что в кармане осталась сумма денег, не меньшая той, что вынута?

11. Из 10^4 чисел 0000, 0001, 0002, ..., 9999 случайно и равновероятно выбирается число. Какова вероятность того, что в выбранном числе: а) все цифры разные; б) имеются только 3 разные цифры; в) имеются только 2 разные цифры; г) все цифры одинаковые?

12. На бесконечную шахматную доску со стороной квадрата a бросается наудачу монета радиуса r, $2r < a$. Найти вероятность p_k того, что монета будет иметь общие точки с k квадратами, $k = 1,2,3,4$.

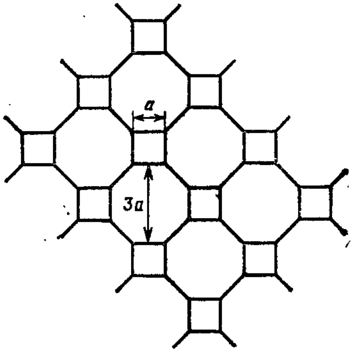

Рис. 7.

13. На паркет, изображенный на рис. 7, случайно падает монета радиуса r, $2r < a$. Найти вероятность того, что монета целиком окажется внутри маленького квадрата.

14. На квадрат случайно с равномерным распределением бросается частица. Найти вероятность того, что она удалена от вершин квадрата на расстояние, не меньшее половины длины стороны квадрата.

Глава 2. УСЛОВНЫЕ ВЕРОЯТНОСТИ. НЕЗАВИСИМОСТЬ

§ 6. Условные вероятности

Пусть при N испытаниях события A, B и AB произошли с частотами $N(A)$, $N(B)$ и $N(AB)$. Назовем отношение $N(AB)/N(B)$ *условной относительной частотой* события A при условии, что произошло событие B. Если имеет место устойчивость частот

$$\frac{N(A)}{N} \approx \mathsf{P}(A), \quad \frac{N(B)}{N} \approx \mathsf{P}(B), \quad \frac{N(AB)}{N} \approx \mathsf{P}(AB)$$

и $\mathsf{P}(B) > 0$, то относительная частота $N(AB)/N(B)$ тоже устойчива:

$$\frac{N(AB)}{N(B)} = \frac{N(AB)/N}{N(B)/N} \approx \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}. \tag{1}$$

Соотношение (1) приводит к следующему естественному определению.

Определение 1. Пусть $\mathsf{P}(B) > 0$. *Условной вероятностью* $\mathsf{P}(A \mid B)$ события A при условии, что произошло событие B (или просто: при условии B), назовем отношение

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}. \tag{2}$$

Для условной вероятности $\mathsf{P}(A \mid B)$ применяется также обозначение $\mathsf{P}_B(A)$.

Если B фиксировано, а $A \in \mathscr{A}$ из некоторого вероятностного пространства $(\Omega, \mathscr{A}, \mathsf{P})$, то условная вероятность $\mathsf{P}_B(A)$, рассматриваемая как функция P_B от события $A \in \mathscr{A}$, определяет новое вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P}_B)$. Для того чтобы это установить, надо проверить, что P_B удовлетворяет аксиомам $1^\circ - 4^\circ$. Это легко делается, так как в силу (2):

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} \geqslant 0; \quad \mathsf{P}_B(\Omega) = \frac{\mathsf{P}(\Omega B)}{\mathsf{P}(B)} = 1;$$

если $A_1A_2=\varnothing$, то $(A_1B)\cap(A_2B)=\varnothing$ и

$$\mathsf{P}_B(A_1+A_2)=\frac{\mathsf{P}(A_1B+A_2B)}{\mathsf{P}(B)}=\frac{\mathsf{P}(A_1B)}{\mathsf{P}(B)}+\frac{\mathsf{P}(A_2B)}{\mathsf{P}(B)}=$$
$$=\mathsf{P}_B(A_1)+\mathsf{P}_B(A_2);$$

и, наконец, из $A_n\downarrow\varnothing$ следует $BA_n\downarrow\varnothing$, поэтому

$$\mathsf{P}_B(A_n)=\frac{\mathsf{P}(BA_n)}{\mathsf{P}(B)}\downarrow 0.$$

Переписывая (2) в форме

$$\mathsf{P}(AB)=\mathsf{P}(B)\,\mathsf{P}_B(A), \tag{3}$$

мы получаем равенство, которое называют *теоремой умножения*. Если исходить из определения (2), то содержательность теоремы умножения (3) представляется весьма невысокой. Однако в применениях мы часто условную вероятность $\mathsf{P}_B(A)$ будем вычислять, исходя не из формулы (2), а из каких-либо других соображений. В этом случае формула (3) уже определяет $\mathsf{P}(AB)$ с помощью $\mathsf{P}(B)$ и $\mathsf{P}(B\,|\,A)$, а не наоборот.

Пример 1. В урне находится M белых и $N-M$ черных шаров. По схеме выборки без возвращения последовательно выбираются два шара. Найдем вероятность того, что оба шара будут белыми. Эту вероятность можно найти с помощью теоремы умножения (3). Обозначим события $A=\{$первый вынутый шар — белый$\}$, $B=\{$второй вынутый шар — белый$\}$. Тогда вычисление вероятностей $\mathsf{P}(A)=\frac{M}{N}$ и $\mathsf{P}_A(B)=\frac{M-1}{N-1}$ сводится к более простым задачам о вынимании белого шара из урны, содержащей M белых и $N-M$ черных шаров (соответственно во втором случае $M-1$ белых и $N-M$ черных шаров). Имеем окончательно $\mathsf{P}(AB)=$ $=\mathsf{P}(A)\,\mathsf{P}_A(B)=\frac{M(M-1)}{N(N-1)}$.

С помощью (3) по индукции легко доказывается более общая

Теорема 1. (Теорема умножения.) *Пусть события $A_1,\ \ldots,\ A_n$ таковы, что $\mathsf{P}(A_1\ldots A_{n-1})>0$. Тогда*

$$\mathsf{P}(A_1\ldots A_n)=$$
$$=\mathsf{P}(A_1)\,\mathsf{P}_{A_1}(A_2)\,\mathsf{P}_{A_1A_2}(A_3)\ldots\mathsf{P}_{A_1\ldots A_{n-1}}(A_n). \tag{4}$$

Доказательство. Из условия теоремы вытекает, что существуют все условные вероятности в (4). Для доказательства (4) по индукции обозначим $B = A_1 \ldots$ $\ldots A_{n-1}$, $A = A_n$ и применим (3) и индукционное предположение о справедливости (4), когда n заменяется на $n-1$. Справедливость (4) при $n=2$ также следует из (3).

Формулы типа (3) и (4) показывают, что на одном и том же пространстве элементарных событий Ω с σ-алгеброй $\mathscr{A}$ удобно рассматривать, наряду с вероятностью P, условные вероятности P_B.

§ 7. Формула полной вероятности

Определение 2. Систему событий $A_1, A_2, \ldots, A_n$ будем называть *конечным разбиением* (в дальнейшем — просто *разбиением*), если они попарно несовместны и

$$A_1 + A_2 + \ldots + A_n = \Omega. \tag{5}$$

Теорема 2. (Формула полной вероятности.) *Если* $A_1, \ldots, A_n$ — *разбиение и все* $\mathsf{P}(A_k) > 0$, *то для любого события B имеет место формула*

$$\mathsf{P}(B) = \sum_{k=1}^{n} \mathsf{P}(A_k)\,\mathsf{P}(B \mid A_k), \tag{6}$$

называемая формулой полной вероятности.

Доказательство. Из (5) следует разложение B на сумму

$$B = B\Omega = BA_1 + BA_2 + \ldots + BA_n$$

попарно несовместных событий, поэтому

$$\mathsf{P}(B) = \sum_{k=1}^{n} \mathsf{P}(BA_k).$$

Применяя к слагаемым $\mathsf{P}(BA_k)$ теорему умножения, получаем (6).

Пример 2. Вычислим в урновой схеме примера 1 вероятность события $B = \{$второй вынутый шар — белый$\}$. Из классического определения вероятности имеем

$$\mathsf{P}(A) = \frac{M}{N}, \quad \mathsf{P}(\bar{A}) = \frac{N-M}{N},$$

$$\mathsf{P}_A(B) = \frac{M-1}{N-1}, \quad \mathsf{P}_{\bar{A}}(B) = \frac{M}{N-1}.$$

По формуле полной вероятности

$$\mathsf{P}(B) = \mathsf{P}(A)\,\mathsf{P}_A(B) + \mathsf{P}(\bar{A})\,\mathsf{P}_{\bar{A}}(B) = \frac{M}{N}\,\frac{M-1}{N-1} + \frac{N-M}{N}\,\frac{M}{N-1} = \frac{M}{N},$$

т. е. $\mathsf{P}(A) = \mathsf{P}(B)$. Аналогично можно установить, что вынимая последовательно без возвращения шары, мы получаем одну и ту же вероятность вынуть белый шар на любом месте. Таким образом, при правильно организованной жеребьевке шансы всех участников одинаковы, независимо от того, в какой очередности они тянут жребий. Эту же задачу можно интерпретировать как вычисление вероятности вытащить белый шар из урны, из которой был случайно утерян один или несколько шаров.

§ 8. Формулы Байеса

Теорема 3. *Если выполнены условия теоремы* 2 *и* $\mathsf{P}(B) > 0$, *то имеют место формулы*

$$\mathsf{P}(A_k \mid B) = \frac{\mathsf{P}(A_k)\,\mathsf{P}(B \mid A_k)}{\sum\limits_{i=1}^{n} \mathsf{P}(A_i)\,\mathsf{P}(B \mid A_i)}, \tag{7}$$

называемые формулами Байеса.

Доказательство. По теореме умножения

$$\mathsf{P}(A_k B) = \mathsf{P}(A_k)\,\mathsf{P}(B \mid A_k) = \mathsf{P}(B)\,\mathsf{P}(A_k \mid B),$$

откуда имеем

$$\mathsf{P}(A_k \mid B) = \frac{\mathsf{P}(A_k)\,\mathsf{P}(B \mid A_k)}{\mathsf{P}(B)}.$$

Применяя к знаменателю $\mathsf{P}(B)$ формулу полной вероятности (6), получаем (7).

Формулы Байеса можно интерпретировать следующим образом. Назовем события A_k гипотезами. Пусть событие B — результат некоторого эксперимента. Вероятности $\mathsf{P}(A_k)$ — это *априорные* вероятности гипотез, вычисляемые до произведения опыта, а условные вероятности $\mathsf{P}(A_k \mid B)$ — это *апостериорные* вероятности гипотез, вычисляемые после того, как стал известен исход эксперимента B. Формулы Байеса позволяют по априорным вероятностям гипотез и по условным вероятностям события B при гипотезах A_k вычислять апостериорные вероятности $\mathsf{P}(A_k \mid B)$.

Пример 3. Пусть имеются две урны, в каждой из которых по N шаров, причем в первой урне M_1 белых шаров, а во второй урне M_2 белых шаров. Проводимый эксперимент состоит в том, что мы сначала с вероятностью 1/2 выбираем первую или вторую урну, а затем из выбранной урны случайно вынимаем (с возвращением) n шаров. Пусть событие B состоит в том, что все вынутые шары — белые. В этом случае имеем две гипотезы: A_1 — выбор первой урны и A_2 — выбор второй урны. По условиям задачи априорные вероятности равны друг другу: $\mathsf{P}(A_1) = \mathsf{P}(A_2) = 1/2$. Далее, легко вычисляются условные вероятности $\mathsf{P}(B \mid A_k) = \left(\frac{M_k}{N}\right)^n$. Формулы Байеса дают нам априорные вероятности:

$$\mathsf{P}(A_k \mid B) = \frac{\frac{1}{2}\left(\frac{M_k}{N}\right)^n}{\frac{1}{2}\left(\frac{M_1}{N}\right)^n + \frac{1}{2}\left(\frac{M_2}{N}\right)^n} = \frac{M_k^n}{M_1^n + M_2^n}, \qquad k = 1, 2.$$

Если $M_2 < M_1$, то при $n \to \infty$ $\mathsf{P}(A_1 \mid B) = \dfrac{1}{1 + \left(\frac{M_2}{M_1}\right)^n} \to 1$,

таким образом, знание исхода B эксперимента в этом случае дает нам возможность существенным образом изменить наши априорные сведения о гипотезах A_1 и A_2.

§ 9. Независимость событий

Понятие независимости относится к одному из основных в теории вероятностей. Если события A и B таковы, что $\mathsf{P}(B) > 0$, то существует условная вероятность $\mathsf{P}(A \mid B)$. В случае, когда $\mathsf{P}(A \mid B) = \mathsf{P}(A)$, мы говорим, что *событие A не зависит от события B*. Если и $\mathsf{P}(A) > 0$, то в этом случае

$$\mathsf{P}(B \mid A) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)} = \frac{\mathsf{P}(B)\,\mathsf{P}(A \mid B)}{\mathsf{P}(A)} = \mathsf{P}(B),$$

и из независимости A от B следует независимость B от A, т. е. понятие независимости A и B симметрично. Из теоремы умножения вероятностей (3) следует, что для независимых событий A и B имеет место равенство $\mathsf{P}(AB)=\mathsf{P}(A)\mathsf{P}(B)$. Это приводит нас к следующему определению независимости.

Определение 3. События A и B называются *независимыми*, если

$$\mathsf{P}(AB)=\mathsf{P}(A)\mathsf{P}(B). \tag{8}$$

Если равенство (8) не выполняется, то события будем называть *зависимыми*.

Это определение уже не содержит ограничений типа $\mathsf{P}(A)>0$. В частности, если $\mathsf{P}(A)=0$, то из $AB\subseteq A$ следует, что и $\mathsf{P}(AB)=0$, а тогда, в силу (8), A и B независимы. Из определения (8) следует $\mathsf{P}(A)=\mathsf{P}(A\,|\,B)$ и $\mathsf{P}(B)=\mathsf{P}(B\,|\,A)$, если эти условные вероятности существуют (т. е. $\mathsf{P}(B)>0$ и $\mathsf{P}(A)>0$ соответственно).

Обычно независимость A и B, которую иногда называют *теоретико-вероятностной*, или *статистической*, независимостью (в отличие от причинной независимости реальных явлений), не устанавливается с помощью равенства (8), а постулируется на основе каких-либо внешних соображений. С помощью же равенства (8) мы вычисляем вероятность $\mathsf{P}(AB)$, зная вероятности $\mathsf{P}(A)$ и $\mathsf{P}(B)$ двух независимых событий. При установлении независимости событий A и B часто используют следующий принцип: *события A и B, реальные прообразы которых $\tilde{A}$ и $\tilde{B}$ причинно независимы, независимы в теоретико-вероятностном смысле.* Реальный смысл этого принципа можно связать со свойством устойчивости частот. Пусть при N наблюдениях $N(\tilde{A})$, $N(\tilde{B})$, $N(\tilde{A}\tilde{B})$ — частоты событий $\tilde{A}$, $\tilde{B}$ и $\tilde{A}\tilde{B}$. Так как из устойчивости частот следует

$$\frac{N(\tilde{A})}{N}\approx\mathsf{P}(A),\quad \frac{N(\tilde{B})}{N}\approx\mathsf{P}(B),\quad \frac{N(\tilde{A}\tilde{B})}{N}\approx\mathsf{P}(AB),$$

$$\frac{N(\tilde{A}\tilde{B})}{N(\tilde{B})}\approx\mathsf{P}(A\,|\,B)=\frac{\mathsf{P}(AB)}{\mathsf{P}(B)},$$

то из независимости событий A и B, т. е. из $\mathsf{P}(A\,|\,B)=\mathsf{P}(A)$, вытекает

$$\frac{N(\tilde{A}\tilde{B})}{N(\tilde{B})}\approx\frac{N(\tilde{A})}{N}$$

или, что равносильно,

$$\frac{N(\tilde{A}\tilde{B})}{N} \approx \frac{N(\tilde{A})}{N}\frac{N(\tilde{B})}{N}. \tag{9}$$

Свойство (9) для причинно независимых реальных событий $\tilde{A}$ и $\tilde{B}$ установлено многовековой практикой человека. Это и позволяет нам сформулировать приведенный выше принцип.

Надо отметить, что этот принцип ни в коем случае не является теоремой. Так как он сформулирован не в терминах математической модели, то он и не может быть теоремой. И, конечно, из теоретико-вероятностной независимости событий A и B не следует причинная независимость их реальных прообразов $\tilde{A}$ и $\tilde{B}$. Следующий пример показывает, что независимость может исчезнуть, если незначительно изменить вероятностную модель.

Пример 4. Из колоды в 52 карты (состоящей из 13 карт каждой из четырех мастей) случайно вынимается карта. Рассмотрим события $A = \{$вынут туз$\}$ и $B = \{$вынута карта бубновой масти$\}$. Тогда событие $AB = \{$вынут туз бубновой масти$\}$. Поскольку в этом случае

$$\mathsf{P}(A) = 4/52 = 1/13, \quad \mathsf{P}(B) = 13/52 = 1/4,$$
$$\mathsf{P}(AB) = 1/52 = \mathsf{P}(A)\,\mathsf{P}(B),$$

то события A и B независимы. Если же колода карт содержит еще и джокер, то A и B станут зависимыми, так как $\mathsf{P}(A) = 4/53$, $\mathsf{P}(B) = 13/53$, $\mathsf{P}(AB) = 1/53$ и $\mathsf{P}(AB) \neq \mathsf{P}(A)\,\mathsf{P}(B)$.

Понятие независимости двух событий распространяется на случай нескольких событий.

Определение 4. События $A_1, A_2, \ldots, A_n$ называются *независимыми*, если для любых $1 \leqslant i_1 < i_2 < \ldots \ldots < i_m \leqslant n$, $2 \leqslant m \leqslant n$, выполняются равенства

$$\mathsf{P}(A_{i_1}A_{i_2}\ldots A_{i_m}) = \mathsf{P}(A_{i_1})\,\mathsf{P}(A_{i_2})\ldots\mathsf{P}(A_{i_m}); \tag{10}$$

в противном случае события называются *зависимыми*. Независимость нескольких событий называется иногда *независимостью* событий в *совокупности*.

Из определения 4 сразу следует, что события любого подмножества $A_{j_1}, A_{j_2}, \ldots, A_{j_r}$ независимых событий $A_1, A_2, \ldots, A_n$ также независимы.

Нижеследующий пример показывает, что независимость событий $A_1, A_2, \ldots, A_n$ в совокупности — более сильное свойство, чем попарная их независимость.

Пример 5. Пусть из чисел 2, 3, 5 и 30 выбирается одно число, причем каждое из чисел может быть выбрано с вероятностью 1/4. Обозначим событие $A_k =$ $= \{$выбранное число делится на $k\}$. Легко видеть, что события A_2, A_3, A_5 попарно независимы, но зависимы в совокупности, так как

$$\mathsf{P}(A_2) = \mathsf{P}(A_3) = \mathsf{P}(A_5) = 1/2, \quad \mathsf{P}(A_2A_3) = \mathsf{P}(A_2A_5) = \\ = \mathsf{P}(A_3A_5) = 1/4 \quad \text{и} \quad \mathsf{P}(A_2A_3A_5) = 1/4.$$

Из определения 2 вытекает следующее свойство условных вероятностей.

Теорема 4. *Если события* $A_1, A_2, \ldots, A_n$ *независимы, индексы* $i_1, i_2, \ldots, i_r, j_1, j_2, \ldots, j_s$ *все различны, вероятность* $\mathsf{P}(A_{i_1}A_{i_2}\ldots A_{i_r}) > 0$, *то*

$$\mathsf{P}(A_{j_1}\ldots A_{j_s} \mid A_{i_1}\ldots A_{i_r}) = \mathsf{P}(A_{j_1}\ldots A_{j_s}). \tag{11}$$

Доказательство. Из независимости событий $A_1, \ldots, A_n$ следует

$$\mathsf{P}(A_{i_1}\ldots A_{i_r}) = \mathsf{P}(A_{i_1})\ldots \mathsf{P}(A_{i_r}),$$
$$\mathsf{P}(A_{j_1}\ldots A_{j_s}) = \mathsf{P}(A_{j_1})\ldots \mathsf{P}(A_{j_s})$$

и

$$\mathsf{P}(A_{i_1}\ldots A_{i_r}A_{j_1}\ldots A_{j_s}) = \\ = \mathsf{P}(A_{i_1})\ldots \mathsf{P}(A_{i_r})\mathsf{P}(A_{j_1})\ldots \mathsf{P}(A_{j_s}),$$

поэтому $\mathsf{P}(A_{i_1}\ldots A_{i_r} \cap A_{j_1}\ldots A_{j_s}) = \mathsf{P}(A_{i_1}\ldots A_{i_r}) \times$ $\times \mathsf{P}(A_{j_1}\ldots A_{j_s})$, а отсюда вытекает (11).

§ 10. Независимость разбиений, алгебр и σ-алгебр

Определение 5. Пусть γ — некоторая система множеств. Наименьшая алгебра множеств $\mathscr{A}(\gamma)$, содержащая γ, называется *алгеброй, порожденной системой* γ.

Аналогично определяется *σ-алгебра, порожденная* γ, как наименьшая σ-алгебра, содержащая γ.

Если мы за систему множеств α возьмем разбиение $A_1, A_2, \ldots, A_n$, т. е. такие множества A_i, что $A_1 + A_2 + \ldots + A_n = \Omega$ и $A_i A_j = \varnothing$ при $i \neq j$, то нетрудно видеть, что алгебра $\mathscr{A}(\alpha)$, порожденная разбиением α, является конечной (т. е. в нее входит лишь конечное число множеств) и состоит только из пустого множества и множеств вида

$$A_{i_1} + A_{i_2} + \ldots + A_{i_m}.$$

Имеет место обратное свойство.

Теорема 5. *Каждая конечная алгебра множеств порождается некоторым разбиением.*

Доказательство. Пусть $\mathscr{B}$ — конечная алгебра событий. Обозначим $\mathscr{B}_\omega$ совокупность всех $B \in \mathscr{B}$, для которых $\omega \in B$. Для каждого $\omega \in \Omega$ введем $B_\omega = \bigcap\limits_{B \in \mathscr{B}_\omega} B$. Покажем, что для двух $\omega \neq \omega'$ либо $B_\omega = B_{\omega'}$, либо $B_\omega \cap B_{\omega'} = \varnothing$. Для любых $\omega \in \Omega$ и $B \in \mathscr{B}$ имеет место следующее свойство: если $\omega \in B$, то $B_\omega \subseteq B$. Пусть теперь $\omega \in B_{\omega'}$; тогда $B_\omega \subseteq B_{\omega'}$. Далее, если $\omega' \in B_\omega$, то $B_{\omega'} \subseteq B_\omega$ и, следовательно, $B_{\omega'} = B_\omega$. Случай $\omega' \in \overline{B}_\omega$ невозможен, так как приводит к противоречию $B_{\omega'} \subseteq \overline{B}_\omega$ (а мы уже доказали, что $B_\omega \subseteq B_{\omega'}$). Выберем среди B_ω разные множества $B_1, B_2, \ldots, B_r$. Они образуют разбиение, так как $B_1 + \ldots + B_r = \Omega$ и $B_i B_j = \varnothing$ при $i \neq j$. Поскольку любое $B \in \mathscr{B}$ представимо в виде $B = \bigcup\limits_{\omega \in B} B_\omega$, то это разбиение порождает алгебру $\mathscr{B}$, что и требовалось доказать.

Пример 6. Разбиение $A + \overline{A} = \Omega$ порождает алгебру $\mathscr{B} = \{\varnothing, \Omega, A, \overline{A}\}$.

Пример 7. Разбиение $A_1 + A_2 + A_3 = \Omega$ порождает алгебру $\mathscr{B} = \{\varnothing, \Omega, A_1, A_2, A_3, A_1 + A_2, A_1 + A_3, A_2 + A_3\}$.

Определение 6. Разбиения

$$\alpha_k\colon A_{k1} + A_{k2} + \ldots + A_{kr_k} = \Omega, \qquad k = 1, \ldots, n,$$

называются *независимыми*, если для любых i_k, $1 \leqslant i_k \leqslant r_k$, $k = 1, \ldots, n$,

$$\mathsf{P}(A_{1i_1} A_{2i_2} \ldots A_{ni_n}) = \mathsf{P}(A_{1i_1}) \mathsf{P}(A_{2i_2}) \ldots \mathsf{P}(A_{ni_n}).$$

Определение 7. Алгебры (или σ-алгебры) событий $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n$ называются *независимыми*, если для любых $A_i \in \mathscr{A}_i$

$$\mathsf{P}(A_1 A_2 \ldots A_n) = \mathsf{P}(A_1)\mathsf{P}(A_2) \ldots \mathsf{P}(A_n).$$

Теорема 6. *Конечные алгебры $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n$ независимы тогда и только тогда, когда независимы порождающие их разбиения $\alpha_1, \alpha_2, \ldots, \alpha_n$.*

Доказательство. Так как порождающее $\mathscr{A}_i$ разбиение α_i есть подсистема $\mathscr{A}_i$, т. е. $\alpha_i \subseteq \mathscr{A}_i$, то из независимости $\mathscr{A}_1, \ldots, \mathscr{A}_n$ следует независимость $\alpha_1, \ldots, \alpha_n$. Каждое $A \in \mathscr{A}_i$ есть сумма попарно несовместных событий из α_i, поэтому обратное заключение получаем из следующей леммы.

Лемма 1. 1°. *Если события A и B независимы, то события $\bar{A}$ и B также независимы.* 2°. *Если A_1 и B независимы и A_2 и B независимы, а $A_1 A_2 = \varnothing$, то $A_1 + A_2$ и B независимы.*

Доказательство. 1°. Из независимости A и B следует

$$\mathsf{P}(B\bar{A}) = \mathsf{P}(B \setminus AB) = \mathsf{P}(B) - \mathsf{P}(AB) =$$
$$= \mathsf{P}(B) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(B)(1 - \mathsf{P}(A)) = \mathsf{P}(B)\mathsf{P}(\bar{A}),$$

т. е. B и $\bar{A}$ также независимы. 2°. Из независимости A_i и B имеем $\mathsf{P}(A_i B) = \mathsf{P}(A_i)\mathsf{P}(B)$, откуда вытекает $\mathsf{P}((A_1 + A_2)B) = \mathsf{P}(A_1 B) + \mathsf{P}(A_2 B) = \mathsf{P}(A_1)\mathsf{P}(B) + \mathsf{P}(A_2) \times \mathsf{P}(B) = (\mathsf{P}(A_1) + \mathsf{P}(A_2))\mathsf{P}(B) = \mathsf{P}(A_1 + A_2)\mathsf{P}(B)$, т. е. $A_1 + A_2$ и B независимы.

Следствие. Каждое событие A порождает разбиение $A + \bar{A} = \Omega$, которое в свою очередь порождает алгебру $\mathscr{A}(A)$. Из леммы 1 вытекает, что *независимость событий $A_1, \ldots, A_n$ и независимость порожденных ими алгебр $\mathscr{A}(A_1), \ldots, \mathscr{A}(A_n)$ эквивалентны.*

§ 11. Независимые испытания

Под испытанием мы будем понимать некоторый эксперимент, исходами которого служат те или иные случайные события. В принятой нами аксиоматике испытание — это некоторое вероятностное пространство. Пусть даны n испытаний, т. е. даны вероятностные пространства

$$(\Omega_1, \mathscr{A}_1, \mathsf{P}_1), \ldots, (\Omega_n, \mathscr{A}_n, \mathsf{P}_n). \tag{12}$$

Если эти вероятностные пространства есть модели некоторых причинно независимых испытаний, то σ-алгебры $\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n$ должны быть независимыми. Но для того чтобы иметь возможность говорить о теоретико-вероятностной независимости, мы должны рассматривать $\mathscr{A}_i$ как σ-подалгебры σ-алгебры $\mathscr{A}$ одного общего вероятностного пространства $(\Omega, \mathscr{A}, \mathsf{P})$. Такое вероятностное пространство всегда можно построить. Мы проделаем это построение в частном случае, когда вероятностные пространства (12) конечны.

Итак, пусть $(\Omega_i, \mathscr{A}_i, \mathsf{P}_i)$ — конечное вероятностное пространство, $\Omega_i = \{\omega_i\}$, $\mathscr{A}_i$ состоит из всех подмножеств Ω_i, а вероятность $\mathsf{P}_i(A) = \sum\limits_{\omega_i \in A} p_i(\omega_i)$ задается с помощью вероятностей элементарных событий $p_i(\omega_i)$, $\omega_i \in \Omega_i$. Построим прямое произведение вероятностных пространств (12) $(\Omega, \mathscr{A}, \mathsf{P})$, полагая $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$, точки которого $\omega \in \Omega$ есть векторы $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ с компонентами $\omega_i \in \Omega_i$, $i = 1, \ldots, n$, $\mathscr{A}$ — алгебра всех подмножеств Ω,

$$p(\omega) = p_1(\omega_1) \ldots p_n(\omega_n), \quad \mathsf{P}(A) = \sum_{\omega \in A} p(\omega). \tag{13}$$

Построенная так вероятность P называется *прямым произведением вероятностей* P_i и обозначается $\mathsf{P} = \mathsf{P}_1 \times \ldots \times \mathsf{P}_n$. Аналогично в этом случае $\mathscr{A} = \mathscr{A}_1 \times \ldots \times \mathscr{A}_n$ есть *прямое произведение алгебр*. В построенном вероятностном пространстве выделим класс событий A, называемых *прямоугольниками*, определяемый следующим образом. Пусть $A_i \in \mathscr{A}_i$, $i = 1, \ldots, n$. Прямоугольник

$$A = A_1 \times A_2 \times \ldots \times A_n \tag{14}$$

состоит из тех и только тех $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, для которых $\omega_i \in A_i$, $i = 1, \ldots, n$. Из определения вероятности (13) следует, что вероятность прямоугольника (14) равна

$$\mathsf{P}(A)=\sum_{\omega\in A}p(\omega)=\sum_{\omega_1\in A_1}p_1(\omega_1)\ldots\sum_{\omega_n\in A_n}p_n(\omega_n)=$$

$$=\prod_{k=1}^{n}\mathsf{P}_k(A_k). \qquad (15)$$

Обозначим $\mathscr{A}'_k$ подалгебру алгебры $\mathscr{A}$, состоящую из всех тех прямоугольников (14), у которых $A_i=\Omega_i$ для $i\neq k$. Нетрудно видеть, что между событиями

$$A'_i=\Omega_1\times\ldots\times\Omega_{i-1}\times A_i\times\Omega_{i+1}\times\ldots\times\Omega_n\in\mathscr{A}'_i$$

и $A_i\in\mathscr{A}_i$ устанавливается естественный изоморфизм $A'_i\sim A_i$, поэтому вместо событий A_i из вероятностного пространства $(\Omega_i,\ \mathscr{A}_i,\ \mathsf{P}_i)$ можно рассматривать изоморфные события A'_i из подалгебры $\mathscr{A}'_i$ вероятностного пространства $(\Omega,\ \mathscr{A},\ \mathsf{P})$, Из определения вероятности (15) следует $\mathsf{P}(A'_i)=\mathsf{P}_i(A_i)$. Так как $A=A_1\times\ldots\times A_n=$ $=\bigcap\limits_{k=1}^{n}A'_k$, то из (15) получаем для любых $A'_k\in\mathscr{A}'_k$

$$\mathsf{P}\left(\bigcap_{k=1}^{n}A'_k\right)=\prod_{k=1}^{n}\mathsf{P}(A'_k),$$

т. е. алгебры $\mathscr{A}'_1,\ \ldots,\ \mathscr{A}'_n$ независимы.

Схема Бернулли. Частный случай независимых испытаний, с двумя исходами в каждом из испытаний, строится следующим образом. Пусть вероятностные пространства в (12) таковы, что $\Omega_i=\{0,1\}$, $\mathscr{A}_i=\{\varnothing,\{0\},\{1\},\Omega_i\}$, $p(0)=p$, $p(1)=q$, $p+q=1$. Тогда в прямом произведении $(\Omega,\ \mathscr{A},\ \mathsf{P})$ имеем $\Omega=\{\omega\}$, $\omega=(\omega_1,\omega_2,\ldots,\omega_n)$, $\omega_i=0,\ 1$,

$$p(\omega)=\prod_{i=1}^{n}p^{\omega_i}q^{1-\omega_i}. \qquad (16)$$

Построенная схема независимых испытаний называется *схемой Бернулли*. Обычно она трактуется следующим образом. Пусть некоторый исход A, который мы будем называть *успехом*, может произойти при каждом испытании с одной и той же вероятностью p; противоположный исход $\bar{A}$ (*неуспех*) может произойти при каждом испытании с дополнительной вероятностью $q = 1 - p$. В элементарном событии $\omega = (\omega_1, \ldots, \omega_n)$ имеем $\omega_i = 1$, если при i-м испытании произошел успех, и $\omega_i = 0$ в противоположном случае. Обозначим $B_k = \{\omega: \omega_1 + \ldots + \omega_n = k\}$ событие, состоящее в том, что при n независимых испытаниях в схеме Бернулли произошло ровно k успехов. Поскольку из (16) следует, что при $\omega \in B_k$ $p(\omega) = p^k q^{n-k}$, то $\mathsf{P}(B_k) = p^k q^{n-k} \times$ (число элементарных событий $\omega \in B_k$). Итак имеем,

$$\mathsf{P}(B_k) \doteq C_n^k p^k q^{n-k}, \quad k = 0, 1, \ldots, n. \tag{17}$$

Вероятности (17) называются *биномиальным распределением*. Примерами, в которых появляется биномиальное распределение, служат: выборка с возвращением (§ 4, формула (15)), выпадение шестерки m раз при n бросаниях игральной кости (вероятность этого события $C_n^m \left(\frac{1}{6}\right)^m \left(\frac{5}{6}\right)^{n-m}$), рождение m мальчиков при регистрации n рождений (если вероятность рождения мальчика $p = 0{,}51$, то вероятность рождения m мальчиков при регистрации n рождений равна $C_n^m (0,51)^m \cdot (0,49)^{n-m}$; обширный статистический материал, собранный в разное время и в разных странах, свидетельствует о том, что вероятность $p > 1/2$ и примерно равна 0,51—0,52).

Полиномиальная схема. Более сложная схема n независимых испытаний получается, когда при каждом испытании возможно появление одного из r попарно несовместных исходов. Пусть i-е испытание связано с вероятностным пространством $(\Omega_i, \mathscr{A}_i, \mathsf{P}_i)$, где $\Omega_i = \{1, 2, \ldots, r\}$ состоит из номеров $1, 2, \ldots, r$ исходов. Пусть $p_1, \ldots, p_r$ — вероятности этих исходов, $p_1 + \ldots + p_r = 1$, а $\mathscr{A}_i$ состоит из всех подмножеств Ω_i. В прямом произведении вероятностных пространств $(\Omega, \mathscr{A}, \mathsf{P})$ элементарное событие $\omega \in \Omega$ равно $\omega = (\omega_1, \ldots, \omega_n)$, где ω_i — номер исхода при i-м испытании. Полагая $p(\omega) = p_{\omega_1} p_{\omega_2} \ldots p_{\omega_n}$, вычислим вероятность события

$$B_{n_1 \ldots n_r} = \{\text{в } n \text{ независимых испытаниях}$$
$$\text{произошло ровно по } n_k \ k\text{-х исходов}\},$$
$$n_1 + n_2 + \ldots + n_r = n.$$

Так как для любого $\omega \in B_k$

$$p(\omega) = \prod_{k=1}^{r} p_k^{n_k},$$

а количество точек в $B_{n_1 \dots n_r}$ равно полиномиальному коэффицненту $\frac{n!}{n_1! \dots n_r!}$, то

$$\mathsf{P}\left(B_{n_1 \dots n_r}\right) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}, \tag{18}$$

$$n_1 + \dots + n_r = n.$$

Распределение (18) называется *полиномиальным*; описанная схема независимых испытаний с r исходами также называется *полиномиальной*. При $r = 2$ эта схема превращается в биномиальную схему Бернулли.

Задачи

1. Из множества чисел 000, 001, ..., 999 равновероятно выбирается одно число. Какова вероятность того, что это число не содержит цифру 1, если все его цифры различны?

2. Из урны, содержащей M белых и $N - M$ черных шаров, случайно последовательно по схеме выборки без возвращения извлекаются три шара. С помощью теоремы умножения найти вероятность того, что появится последовательность шаров: белый, черный, белый.

3. Показать, что любая конечная алгебра событий состоит из 2^k событий, где k — натуральное число.

4. Плоскость расчерчена параллельными прямыми, расстояния между соседними прямыми, чередуясь, равны a и b. На эту плоскость случайно бросается игла длины $l < \min\{a, b\}$. Пользуясь решением задачи Бюффона и формулой полной вероятности, найти вероятность того, что игла пересечет одну из этих прямых.

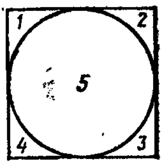

Рис. 8.

5. На бесконечную шахматную доску с длиной стороны квадрата a случайно бросается монета радиуса $r < a/2$. Найти вероятность того, что монета пересечет сторону какого-либо квадрата.

6. В последовательности n независимых испытаний с вероятностью p успеха в каждом из испытаний произошел ровно один успех. Какова вероятность того, что успех произошел при втором испытании?

7. В схеме испытаний задачи 6 произошло ровно два успеха. Найти вероятность того, что успехи произошли в соседних испытаниях.

8. На паркет, составленный из прямоугольников со сторонами a и b, $a < b$, случайно бросается монета радиуса r, $2r < \min\{a, b\}$. Найти вероятность того, что монета заденет меньшую сторону

какого-нибудь прямоугольника, если известно, что она какую-то сторону задела.

9. Для перехода улицы пешеходу нужно три секунды. Каждую секунду с вероятностью p по улице проезжает автомобиль и с вероятностью $q = 1 - p$ улица свободна. Будем считать время дискретным (по секундам), а наличие или отсутствие автомобиля на улице в разные моменты времени независимыми испытаниями. Пешеход начинает переходить улицу лишь в том случае, если в течение трех секунд она будет свободна от автомобилей на переходе. Найти вероятность того, что пешеходу придется ждать перехода а) больше двух секунд; б) больше трех секунд.

10. Бросаются две игральные кости. Какова вероятность того, что на первой кости выпала 1, если известно, что на второй кости выпало число очков больше, чем на первой? (Применить формулу Байеса.)

11. В единичный квадрат со вписанным в него кругом независимо с равномерным распределением случайно бросается 6 частиц. Найти вероятность того, что ни одна из пяти частей квадрата не будет свободна от частиц (см. рис. 8).

Глава 3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ (КОНЕЧНАЯ СХЕМА)

§ 12. Случайные величины. Индикаторы

Рассмотрим конечное вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P})$. Числовую функцию от элементарного события $\xi = \xi(\omega)$, $\omega \in \Omega$, назовем *случайной величиной.* Мы будем обычно обозначать случайные величины греческими буквами ξ, η, ζ, μ, ν, ... и т. п. (в англо-американской литературе и иногда у нас случайные величины обозначаются прописными латинскими буквами X, Y, Z и т. п.).

Пример 1. В схеме независимых испытаний Бернулли в § 11 множество Ω состоит из элементарных событий $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, где $\omega_i = 1$, если при i-м испытании произошел успех, и $\omega_i = 0$ в случае неуспеха. Случайная величина $\mu \leftrightharpoons \mu(\omega) = \omega_1 + \omega_2 + \ldots + \omega_n$ равна числу успехов при n испытаниях в схеме Бернулли.

Пример 2. Рассмотрим следующую урновую схему. Пусть в урне имеется N шаров, из них M белых, остальные — черные. По схеме выборки без возвращения из урны извлекаются n шаров (см. § 4, пример 3). Перенумеруем все N шаров числами $1, 2, \ldots, N$ так, чтобы белые шары получили номера $1, 2, \ldots, M$. Тогда множество Ω можно составить из элементарных событий, состоящих из подмножеств $\omega = \{i_1, i_2, \ldots, i_n\}$, $i_1 < i_2 < \ldots < i_n$, мощности n множества целых чисел $\{1, 2, \ldots, N\}$. Элементарное событие $\omega = \{i_1, i_2, \ldots, i_n\}$ соответствует выборке, в которую вошли шары с номерами $i_1, i_2, \ldots, i_n$. Случайная величина ξ, равная числу белых шаров в выборке, определяется как функция от ω следующим образом: $\xi = \xi(\omega) = m$, если в $\omega = \{i_1, \ldots, i_n\}$, $i_m \leqslant M < i_{m+1}$ при $1 \leqslant m < n$; $\xi(\omega) = 0$, если $M < i_1$; $\xi(\omega) = n$, если $i_n \leqslant M$.

Пусть $g(x_1, \ldots, x_r)$ — числовая функция от числовых аргументов $x_1, \ldots, x_r$, а $\xi_1, \ldots, \xi_r$ — случайные величины. Тогда сложная функция $\eta = \eta(\omega) = g(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_r(\omega))$ также будет случайной величиной. В частности, так определяются случайные величины, равные сумме $\sum_{k=1}^{r} \xi_k$ и произведению $\prod_{k=1}^{r} \xi_k$ случайных величин.

С каждым событием $A \in \mathscr{A}$ можно связать случайную величину

$$I_A = I_A(\omega) = \begin{cases} 1, & \text{если} \quad \omega \in A, \\ 0, & \text{если} \quad \omega \notin A, \end{cases}$$

называемую *индикатором* события A. Индикаторы удовлетворяют следующим легко проверяемым свойствам:

$$I_\varnothing \equiv 0, \quad I_\Omega \equiv 1, \quad I_{AB} = I_A I_B, \quad I_{\bar{A}} = 1 - I_A. \tag{1}$$

Если события $A_1, \ldots, A_n$ попарно несовместны, то нетрудно установить, что

$$I_{\sum\limits_{k=1}^{n} A_k} = \sum_{k=1}^{n} I_{A_k}.$$

Выведем формулу для индикатора объединения $\bigcup\limits_{k=1}^{n} A_k$ любых событий. Так как $\overline{\bigcup\limits_k A_k} = \bigcap\limits_k \bar{A}_k$, то учитывая свойства (1), мы имеем

$$I_{\bigcup\limits_{k=1}^{n} A_k} = 1 - I_{\overline{\bigcup\limits_{k=1}^{n} A_k}} = 1 - I_{\bigcap\limits_{k=1}^{n} \bar{A}_k} =$$

$$= 1 - \prod_{k=1}^{n} I_{\bar{A}_k} = 1 - \prod_{k=1}^{n} (1 - I_{A_k}),$$

откуда следует

$$I_{\bigcup\limits_{k=1}^{n} A_k} = \sum_{k=1}^{n} I_{A_k} - \sum_{1 \leqslant k \leqslant l \leqslant n} I_{A_k A_l} +$$

$$+ \sum_{1 \leqslant k < l < m \leqslant n} I_{A_k A_l A_m} - \ldots + (-1)^{n-1} I_{A_1 A_2 \ldots A_n}. \tag{2}$$

Обозначим $x_1 < x_2 < \ldots < x_k$ **всевозможные** значения, которые принимает случайная величина ξ. С каждой случайной величиной ξ можно связать разбиение α_ξ, состоящее из событий $A_i = \{\omega: \xi(\omega) = x_i\}$. В самом деле, так как $x_i \neq x_j$, то $A_i A_j = \varnothing$ для $i \neq j$; сумма $A_1 + A_2 + \ldots + A_k$ есть достоверное событие Ω, так как $x_1, x_2, \ldots, x_k$ — *все* значения случайной величины ξ. Разбиение α_ξ порождает алгебру событий $\mathscr{A}_\xi$, которая состоит из событий, представимых в виде

$$\{\xi \in B\} = \{\omega: \xi(\omega) \in B\},$$

где B — любое числовое множество. Разбиение α_ξ и алгебру $\mathscr{A}_\xi$ мы будем называть *порожденными случайной величиной* ξ. Любое событие $\{\xi \in B\}$ представимо в виде суммы $\sum\limits_i A_i$, где суммирование ведется по тем i, для которых $x_i \in B$.

Случайную величину ξ можно выразить с помощью индикаторов разбиения $A_1 + \ldots + A_k = \Omega$ через сумму

$$\xi = \xi(\omega) = \sum_{i=1}^{k} x_i I_{A_i}(\omega), \tag{3}$$

так как левая и правая части (3) принимают одно и то же значение x_i при $\omega \in A_i$.

Законом распределения случайной величины ξ мы будем называть вероятность $\mathsf{P}\{\xi \in B\}$, рассматриваемую как функцию числового множества B. Закон распределения ξ определяется значениями $x_1, x_2, \ldots, x_k$, которые принимает ξ, и вероятностями $\mathsf{P}\{\xi = x_i\}$ этих значений. Обозначим $\mathsf{P}\{\xi = x_i\} = p_i$. Тогда закон распределения $\mathsf{P}\{\xi \in B\}$ можно определить с помощью табл. 2, верхний ряд которой состоит из различных

Таблица 2

Закон распределения случайной величины

x_1	x_2	x_3	...	x_k
p_1	p_2	p_3	...	p_k

чисел x_i, а числа нижнего ряда удовлетворяют условиям

$$p_i \geqslant 0, \quad \sum_{i=1}^{k} p_i = 1. \tag{4}$$

С помощью табл. 2 можно определить вероятность

$$\mathsf{P}\{\xi \in B\} = \sum_{x_i \in B} p_i \tag{5}$$

для любого числового множества B.

В теории вероятностей часто говорят о случайной величине ξ с законом распределения (5), не указывая ни вероятностного пространства $(\Omega, \mathscr{A}, \mathsf{P})$, ни функции $\xi(\omega)$, которая задает случайную величину. В этом случае предполагается, что существует какое-то вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P})$, на котором можно определить функцию $\xi = \xi(\omega)$ так, что табл. 2 будет задавать ее закон распределения. Выбор вероятностного пространства каждый раз определяется существом задачи или простотой получающейся схемы. Простейшим вероятностным пространством, связанным с законом распределения (5), будет множество элементарных событий $\Omega = \{x_1, x_2, \ldots, x_k\}$ с элементарными вероятностями $p(x_i) = p_i$. Случайная величина ξ определяется тогда функцией $\xi(x_i) = x_i$.

Закон распределения индикатора I_A события A определяется табл. 3. Каждой случайной величине соответствует закон распределения. Один и тот же закон

Таблица 3

Закон распределения индикатора I_A

0	1
$1-\mathsf{P}(A)$	$\mathsf{P}(A)$

распределения могут иметь разные случайные величины. Например, если события A и B разные, но $\mathsf{P}(A) = \mathsf{P}(B)$, то разные случайные величины I_A и I_B имеют один и тот же закон распределения.

Закон распределения ξ иногда называют кратко просто *законом* или *распределением*. Законом распределения случайиой величииы иногда изывают задающую его таблицу 2.

Примеры закоиов распределеиия.

1. *Биномиальный закон* для числа успехов μ при n независимых испытаииях в схеме Бериулли:

$$\mathsf{P}\{\mu=m\}=C_n^m p^m (1-p)^{n-m}, \qquad m=0, 1, \ldots, n$$

(см. §§ 11 и 12, пример 1).

2. *Гипергеометрическое распределение* — распределение числа белых шаров ξ в выборке без возвращения объема n из урны, содержащей M белых и $N-M$ черных шаров (см. § 4, пример 3 и § 12, пример 2):

$$\mathsf{P}\{\xi=m\}=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \qquad m=0, 1, \ldots, \min(n, M).$$

3. *Равномерное распределение на* $\{1, 2, \ldots, N\}$:

$$\mathsf{P}\{\xi=m\}=\frac{1}{N}, \qquad m=1, 2, \ldots, N.$$

§ 13. Математическое ожидаиие

Пусть вероятиость P иа коиечиом вероятиостиом простраистве $(\Omega, \mathscr{A}, \mathsf{P})$ определяется с помощью элементариых вероятностей $p(\omega)$. *Математическое ожидание* случайиой величииы $\xi=\xi(\omega)$ обозиачается $\mathsf{M}\xi$ и определяется как сумма

$$\mathsf{M}\xi=\sum_{\omega\in\Omega}\xi(\omega)\,p(\omega). \tag{6}$$

Математическое ожидаиие ξ изывают иногда *средним значением* ξ или просто *средним* ξ. Из этого определения вытекают следующие свойства математического ожидаиия:

1) $$\mathsf{M}I_A=\mathsf{P}(A).$$

В самом деле,

$$\mathsf{M}I_A=\sum_{\omega\in\Omega}I_A(\omega)\,p(\omega)=\sum_{\omega\in A}p(\omega)=\mathsf{P}(A). \tag{7}$$

2) *Аддитивность*: $\mathsf{M}(\xi+\eta)=\mathsf{M}\xi+\mathsf{M}\eta$.

Из определения $\mathsf{M}(\xi+\eta)$ получаем

$$\mathsf{M}(\xi+\eta)=\sum_{\omega\in\Omega}(\xi(\omega)+\eta(\omega))\,p(\omega)=$$
$$=\sum_{\omega\in\Omega}\xi(\omega)\,p(\omega)+\sum_{\omega\in\Omega}\eta(\omega)\,p(\omega)=\mathsf{M}\xi+\mathsf{M}\eta.$$

Из свойства 2) нетрудно по индукции вывести свойство конечной аддитивности математического ожидания:

$$\mathsf{M}(\xi_1+\ldots+\xi_n)=\mathsf{M}\xi_1+\ldots+\mathsf{M}\xi_n. \qquad (8)$$

3) *Для любой константы* c

$$\mathsf{M}(c\xi)=c\mathsf{M}\xi,\quad \mathsf{M}c=c.$$

Это свойство легко вытекает из определения $\mathsf{M}\xi$.

4) *Если* $\xi\geqslant\eta$, *то* $\mathsf{M}\xi\geqslant\mathsf{M}\eta$. *Если* $\xi\geqslant 0$ *и* $\mathsf{M}\xi=0$, *то* $\mathsf{P}\{\xi=0\}=1$.

Доказательство. В сумме $\mathsf{M}(\xi-\eta)=\sum_{\omega}(\xi(\omega)-\eta(\omega))p(\omega)$ при $\xi\geqslant\eta$ все слагаемые неотрицательны, поэтому $\mathsf{M}(\xi-\eta)\geqslant 0$, откуда по свойствам 2) и 3) вытекает $\mathsf{M}\xi\geqslant\mathsf{M}\eta$. Если $\xi\geqslant 0$ и $\mathsf{M}\xi=0$, то при любом $\omega\in\Omega$ $\xi(\omega)\,p(\omega)=0$, откуда из $p(\omega)>0$ следует $\xi(\omega)=0$.

5) *Математическое ожидание* ξ *выражается через закон распределения случайной величины* ξ *формулой*

$$\mathsf{M}\xi=\sum_{i=1}^{k}x_i\mathsf{P}\{\xi=x_i\}. \qquad (9)$$

Доказать (9) можно с помощью представления ξ в виде суммы (3)

$$\xi=\sum_{i=1}^{k}x_iI_{\{\xi=x_i\}},$$

свойства аддитивности (8) и свойств 1) и 3):

$$\mathsf{M}\xi=\sum_{i=1}^{k}x_i\mathsf{M}I_{\{\xi=x_i\}}=\sum_{i=1}^{k}x_i\mathsf{P}\{\xi=x_i\}.$$

Пусть $g(x)$ — некоторая числовая функция. Подставляя вместо x случайную величину ξ, мы получаем новую случайную величину $\eta=g(\xi)$. Вычислить $\mathsf{M}\eta$ можно или исходя из определения, или с помощью закона распределения η или с помощью формулы

$$\mathsf{M}\eta = \mathsf{M}g(\xi) = \sum_{i=1}^{k} g(x_i)\,\mathsf{P}\{\xi = x_i\}, \tag{10}$$

которая доказывается так же, как и (9). При этом надо воспользоваться равенством

$$g(\xi) = \sum_{i=1}^{k} g(x_i)\, I_{\{\xi = x_i\}}.$$

Полагая $g(\xi) = \xi^n$, мы получаем из (10):

$$\mathsf{M}\xi^n = \sum_{i=1}^{k} x_i^n \mathsf{P}\{\xi = x_i\}.$$

Математическое ожидание $\mathsf{M}\xi^n$ называется *n-м моментом* (или *моментом n-го порядка*) случайной величины ξ (или ее закона распределения). *Абсолютным n-м моментом* называется $\mathsf{M}|\xi|^n$. Обозначим $\mathsf{M}\xi = a$. *Центральным моментом n-го порядка* называется $\mathsf{M}(\xi - a)^n$, а *абсолютным центральным моментом n-го порядка* — $\mathsf{M}|\xi - a|^n$.

Центральный момент второго порядка называется *дисперсией* случайной величины и обозначается $\mathsf{D}\xi = \mathsf{M}(\xi - a)^2$. Корень квадратный $\sqrt{\mathsf{D}\xi}$ из дисперсии называется *средним квадратическим отклонением* (или иногда *стандартным отклонением*).

Дисперсия обладает следующими свойствами:

1) $$\mathsf{D}\xi = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2.$$

Доказательство. Имеем $\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2 = \mathsf{M}(\xi^2 - 2\mathsf{M}(\xi \cdot \mathsf{M}\xi) + (\mathsf{M}\xi)^2) = \mathsf{M}\xi^2 - 2 \cdot \mathsf{M}\xi \cdot \mathsf{M}\xi + (\mathsf{M}\xi)^2 = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2$.

2) *$\mathsf{D}\xi \geqslant 0$ и $\mathsf{D}\xi = 0$ тогда и только тогда, когда существует такая константа c, что $\mathsf{P}\{\xi = c\} = 1$.*

Следует из свойства 4) математического ожидания, так как $\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2$ и $(\xi - \mathsf{M}\xi)^2 \geqslant 0$.

3) *Для любой константы c*

$$\mathsf{D}(c\xi) = c^2\mathsf{D}\xi, \quad \mathsf{D}(\xi + c) = \mathsf{D}\xi.$$

Следует из определения и свойства 3) математического ожидания.

Многие известные в анализе неравенства для сумм и интегралов широко применяются в теории вероятностей, причем в этих неравенствах используется понятие математического ожидания. Приведем здесь некоторые из этих неравенств.

Неравенство Иенсена. *Если числовая функция $g(x)$ выпукла, то для любой случайной величины ξ*

$$\mathsf{M}g(\xi) \geqslant g(\mathsf{M}\xi). \tag{11}$$

Доказательство. Если $g(x)$ имеет производные g', g'', то из выпуклости g следует, что в любой точке x $g''(x) \geqslant 0$. Поэтому при любом a

$$g(\xi) \geqslant g(a) + g'(a)(\xi - a). \tag{12}$$

Полагая в (12) $a = \mathsf{M}\xi$ и беря математическое ожидание от обеих частей, получаем (11). В общем случае вместо (12) надо воспользоваться тем, что для любой выпуклой функции $g(x)$ и любой точки a найдется такая константа C, что для всех x

$$g(x) \geqslant g(a) + C(x - a). \tag{13}$$

Функция $g(x)$, определенная на интервале (c, d), где $-\infty \leqslant c < d \leqslant \infty$, называется *выпуклой* (или *выпуклой вниз*), если для любых $x_1, x_2 \in (c, d)$ и любого $0 \leqslant \theta \leqslant 1$ выполняется неравенство

$$g(\theta x_1 + (1-\theta)x_2) \leqslant \theta g(x_1) + (1-\theta)g(x_2). \tag{14}$$

Пусть g — выпуклая функция и $a \in (c, d)$. Возьмем любые x_1, x_2, удовлетворяющие неравенствам $c < x_1 < a < x_2 < d$. Покажем, что для них

$$\frac{g(x_1) - g(a)}{x_1 - a} \leqslant \frac{g(x_2) - g(a)}{x_2 - a}. \tag{15}$$

Нетрудно проверить, что неравенство (15) равносильно (14), если положить в нем $\theta = \dfrac{x_2 - a}{x_2 - x_1}$, $1 - \theta = \dfrac{a - x_1}{x_2 - x_1}$. Из (15) вытекает существование такой константы C, что

$$\sup_{x_1 < a} \frac{g(x_1) - g(a)}{x_1 - a} \leqslant C \leqslant \inf_{x_2 > a} \frac{g(x_2) - g(a)}{x_2 - a},$$

а это равносильно утверждению (13).

Неравенство Ляпунова. *Для любых положительных $\alpha \leq \beta$*

$$(\mathsf{M}|\xi|^\alpha)^{1/\alpha} \leqslant (\mathsf{M}|\xi|^\beta)^{1/\beta}. \tag{16}$$

Для доказательства надо применить к выпуклой функции $g(x) = x^{\beta/\alpha}$ и случайной величине $|\xi|^\alpha$ неравенство Иенсена (11).

Неравенство Коши — Буняковского. *Для любых двух случайных величин* ξ, η

$$|\mathsf{M}\xi\eta| \leqslant \sqrt{\mathsf{M}\xi^2 \cdot \mathsf{M}\eta^2}. \tag{17}$$

Доказательство. Для любых чисел x, y по свойству 4) математического ожидания $\mathsf{M}(x\xi + y\eta)^2 \geqslant 0$. Отсюда следует, что квадратичная формула $x^2\mathsf{M}\xi^2 + 2xy\mathsf{M}\xi\eta + y^2\mathsf{M}\eta^2$ неотрицательно определена, а следовательно, ее дискриминант неположителен: $(\mathsf{M}\xi\eta)^2 - \mathsf{M}\xi^2 \cdot \mathsf{M}\eta^2 \leqslant 0$.

Статистическое истолкование математического ожидания. Пусть в некоторой лотерее имеется один выигрыш, размер которого случаен и равен или x_1, или x_2, ..., или x_k. Если лотерея проводится N раз, причем N_i раз выпадает выигрыш x_i, $N = N_1 + N_2 + \ldots + N_k$, то N_i/N есть относительная частота выигрыша x_i, а $\bar{x} = \frac{1}{N}\sum_{i=1}^{k} x_i N_i$ — средний выигрыш на одну лотерею. Если ξ — случайная величина, равная размеру выигрыша в одной лотерее, то из статистической устойчивости частот следует $\frac{N_i}{N} \approx \mathsf{P}\{\xi = x_i\}$, поэтому средний выигрыш $\bar{x}$ колеблется около $\mathsf{M}\xi$:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{k} x_i N_i \approx \sum_{i=1}^{k} x_i \mathsf{P}\{\xi = x_i\} = \mathsf{M}\xi.$$

Механическая интерпретация $\mathsf{M}\xi$ и $\mathsf{D}\xi$. Интерпретируем наглядно закон распределения как расположение на прямой в точках $x_1 < x_2 < \ldots < x_k$ точечных масс $p_1, p_2, \ldots, p_k$, $\sum_{i=1}^{k} p_i = 1$. В этом случае $\mathsf{M}\xi = \sum_{i=1}^{k} x_i p_i$ есть центр тяжести, $\mathsf{D}\xi = \sum_{i=1}^{k} p_i (x_i - \mathsf{M}\xi)^2$ — момент инерции масс p_i относительно центра тяжести. Таким образом, $\mathsf{M}\xi$ характеризует место, вокруг которого группируются массы p_i, а $\mathsf{D}\xi$ — степень разбросанности масс p_i около $\mathsf{M}\xi$.

Вероятность суммы событий. Вычислим от обеих частей равенства (2) математическое ожидание и воспользуемся его аддитивностью. Получаем

$$\mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right)=\sum_{k=1}^{n}\mathsf{P}(A_k)-\sum_{1\leqslant k_1<k_2\leqslant n}\mathsf{P}(A_{k_1}A_{k_2})+$$
$$+\sum_{1\leqslant k_1<k_2<k_3\leqslant n}\mathsf{P}(A_{k_1}A_{k_2}A_{k_3})-\ldots+(-1)^{n-1}\mathsf{P}(A_1A_2\ldots A_n). \tag{18}$$

С помощью (18) можно вычислять $\mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right)$.

Пример 3. *Размещение частиц по ячейкам.* Пусть имеется N ячеек, в которые независимо друг от друга размещаются n частиц. Каждая частица с вероятностью $1/N$ может попасть в любую фиксированную ячейку. Обозначим через μ_0 число пустых ячеек после такого размещения. Вычислим вероятность $\mathsf{P}\{\mu_0=0\}$. Введем случайные события A_i, полагая, что A_i произошло тогда и только тогда, когда i-я ячейка пустая. Тогда $\{\mu_0>0\}=$ $=\bigcup_{i=1}^{N} A_i$, и мы можем применить (18). Поскольку

$$\mathsf{P}(A_i)=\left(1-\frac{1}{N}\right)^n,\quad \mathsf{P}(A_iA_j)=\left(1-\frac{2}{N}\right)^n$$

и, вообще,

$$\mathsf{P}(A_{i_1}A_{i_2}\ldots A_{i_k})=\left(1-\frac{k}{N}\right)^n,$$

то из (18) следует

$$\mathsf{P}\{\mu_0>0\}=\sum_{k=1}^{N}C_N^k(-1)^{k-1}\left(1-\frac{k}{N}\right)^n$$

или

$$\mathsf{P}\{\mu_0=0\}=1-\mathsf{P}\{\mu_0>0\}=\sum_{k=0}^{N}C_N^k(-1)^k\left(1-\frac{k}{N}\right)^n.$$

§ 14. Многомерные законы распределения

Пусть на конечном вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$ заданы случайные величины $\xi=\xi(\omega)$, $\eta=$ $=\eta(\omega)$. Пусть $x_1, \ldots, x_k$ — все возможные значения ξ,

$y_1, \ldots, y_m$ — все возможные значения η. Как мы уже знаем, с помощью вероятностей $\mathsf{P}\{\xi = x_i\}$ и $\mathsf{P}\{\eta = y_j\}$ определяются законы распределения случайных величин ξ и η:

$$\begin{aligned} \mathsf{P}_\xi(B) &= \mathsf{P}\{\xi \in B\} = \sum_{x_i \in B} \mathsf{P}\{\xi = x_i\}, \\ \mathsf{P}_\eta(B) &= \mathsf{P}\{\eta \in B\} = \sum_{y_j \in B} \mathsf{P}\{\eta = y_j\}, \end{aligned} \tag{19}$$

где B — любое числовое множество. ***Совместным распределением*** случайных величин ξ, η, или ***законом их совместного распределения***, мы будем называть вероятность $\mathsf{P}\{(\xi, \eta) \in B\}$, определенную для всех множеств

Таблица 4

Двумерный закон распределения

	y_1	y_2		y_m
x_1	p_{11}	p_{12}	...	p_{1m}
x_2	p_{21}	p_{22}	...	p_{2m}
	$\vdots$	$\vdots$		
x_k	p_{k1}	p_{k2}	...	p_{km}

B точек плоскости (x, y) и обозначаемую $\mathsf{P}_{\xi\eta}(B)$. Совместное распределение можно задать с помощью набора вероятностей

$$\mathsf{P}\{\xi = x_i,\ \eta = y_j\}, \qquad i = 1, \ldots, k; \qquad j = 1, \ldots, m,$$

полагая $\mathsf{P}\{(\xi, \eta) \in B\} = \sum_{(x_i, y_j) \in B} \mathsf{P}\{\xi = x_i,\ \eta = y_j\}$. Если обозначить $p_{ij} = \mathsf{P}(\xi = x_i,\ \eta = y_j\}$, то совместное распределение ξ, η можно задать с помощью табл. 4, в которой все $p_{ij} \geqslant 0$ и $\sum_{i=1}^{k} \sum_{j=1}^{m} p_{ij} = 1$. Любая таблица такого вида задает некоторый закон совместного распределения пары случайных величин, который мы иногда будем называть *двумерным законом распределения*, или *двумерным распределением*. Иногда двумерным законом распределения мы будем называть просто табл. 4. Законы распределения (19) отдельных случайных величин ξ и η будем называть *одномерными*.

Пара случайных величин ξ, η порождает разбиение $\alpha_{\xi\eta}$, состоящее из событий

$$A_{ij} = \{\omega\colon \ \xi(\omega) = x_i, \ \eta(\omega) = y_j\}, \ i = 1, \ldots, k; \ j = 1, \ldots, m.$$

Это разбиение, а также порожденную им алгебру $\mathscr{A}_{\xi\eta}$ будем называть *порожденными* парой ξ, η. Любое событие $A \in \mathscr{A}_{\xi\eta}$ представимо в виде $A = \{\omega\colon (\xi(\omega), \eta(\omega)) \in \in B\}$, где B — некоторое множество точек плоскости. И, наоборот, любое событие этого вида принадлежит $\mathscr{A}_{\xi\eta}$. Нетрудно видеть, что алгебры $\mathscr{A}_\xi$ и $\mathscr{A}_\eta$, порожденные случайными величинами ξ и η соответственно, есть подалгебры $\mathscr{A}_{\xi\eta}$, причем алгебра $\mathscr{A}_{\xi\eta}$ порождена объединением алгебр $\mathscr{A}_\xi$ и $\mathscr{A}_\eta$. Если $\{A_{ij}\}$ составляют разбиение $\alpha_{\xi\eta}$ и

$$A_{i\cdot} = \sum_{j=1}^{m} A_{ij}, \quad A_{\cdot j} = \sum_{i=1}^{k} A_{ij},$$

то $\{A_{i\cdot}\}$ образуют разбиение α_ξ, а $\{A_{\cdot j}\}$ — разбиение α_η. Из двумерного закона распределения можно получить одномерные законы распределения для ξ

$$\mathsf{P}\{\xi = x_i\} = p_{i\cdot} = \sum_{j=1}^{m} p_{ij}$$

и для η

$$\mathsf{P}\{\eta = y_j\} = p_{\cdot j} = \sum_{i=1}^{k} p_{ij},$$

которые иногда называют *маргинальными законами* первоначального двумерного распределения.

Аналогично для n случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ определяется n-мерный закон распределения $\mathsf{P}_{\xi_1 \ldots \xi_n}(B) = \mathsf{P}\{(\xi_1, \ldots, \xi_n) \in B\}$, где B — множество точек n-мерного пространства R^n. Этот закон можно задать вероятностями

$$\mathsf{P}\{\xi_i = x_{ij_i}, \ i = 1, \ldots, n\} = p_{j_1 j_2 \ldots j_n}, \qquad 1 \leqslant j_i \leqslant k_i, \qquad (20)$$

где $p_{l_1 l_2 \ldots l_n} \geqslant 0$, $\sum\limits_{l_1, l_2, \ldots, l_n} p_{l_1 l_2 \ldots l_n} = 1$ и $x_{i1} < x_{i2} < \ldots$ $\ldots < x_{ik_i}$ — значения, которые принимает случайная величина ξ_i. Совокупность случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ порождает разбиение $\alpha_{\xi_1 \xi_2 \ldots \xi_n}$, состоящее из событий вида $A_{l_1 l_2 \ldots l_n} = \{\omega: \xi_i(\omega) = x_{il_i}, i = 1, \ldots, n\}$, и алгебру $\mathscr{A}_{\xi_1 \ldots \xi_n}$, состоящую из событий вида $\{(\xi_1, \ldots, \xi_n) \in B\}$, где B — подмножество n-мерного пространства R^n.

Так же, как в двумерном случае, по n-мерному закону (20) определяются маргинальные одномерные, двумерные и т. п. законы распределения, например,

$$\mathsf{P}\{\xi_1 = x_{1l}\} = \sum_{l_2, \ldots, l_n} p_{l l_2 \ldots l_n},$$

$$\mathsf{P}\{\xi_1 = x_{1i}, \xi_2 = x_{2l}\} = \sum_{l_3, \ldots, l_n} p_{i l l_3 \ldots l_n}.$$

Так же, как и в случае одной случайной величины, мы часто будем считать, что случайные величины $\xi_1, \xi_2, \ldots$ $\ldots, \xi_n$ заданы, если задан их n-мерный закон распределения (20). В этом случае всегда можно построить такое конечное вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P})$, на котором можно определить случайные величины $\xi_1, \xi_2, \ldots$ $\ldots, \xi_n$ так, чтобы их n-мерное распределение совпадало с (20). Например, мы можем положить $\Omega = \{\omega\}$, где $\omega = (x_{1l_1}, \ldots, x_{nl_n})$, $1 \leqslant l_i \leqslant k_i$ и $p(\omega) = p_{l_1 l_2 \ldots l_n}$.

Иногда n случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ мы будем трактовать как компоненты случайного вектора $\xi =$ $= (\xi_1, \xi_2, \ldots, \xi_n)$. Распределением случайного вектора ξ будет n-мерное распределение $\mathsf{P}\{\xi \in B\} = \sum\limits_{x \in B} \mathsf{P}\{\xi = x\}$, где B — множество точек n-мерного пространства, $x =$ $= (x_1, \ldots, x_n)$ — возможные векторные значения случайного вектора ξ.

§ 15. Независимость случайных величин

В общем случае одномерные законы распределения не определяют многомерного закона. Однако в важном случае независимых случайных величин по одномерным законам распределения однозначно восстанавливаются многомерные распределения.

Определение 1. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называются *независимыми*, если порожденные ими алгебры

$$\mathscr{A}_{\xi_1}, \mathscr{A}_{\xi_2}, \ldots, \mathscr{A}_{\xi_n}$$

независимы.

Поскольку каждая из алгебр $\mathscr{A}_{\xi_i}$ состоит из событий вида $\{\xi_i \in B\}$, где $B \subseteq R^1$, то данное выше определение эквивалентно следующему: случайные величины $\xi_1, \ldots, \xi_n$ независимы, если для любых числовых множеств B_i

$$\mathsf{P}\{\xi_1 \in B_1, \ldots, \xi_n \in B_n\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i \in B_i\}. \qquad (21)$$

Из теоремы 6 в § 10 следует, что независимость алгебр $\mathscr{A}_{\xi_1}, \ldots, \mathscr{A}_{\xi_n}$ равносильна независимости порождающих их разбиений $\alpha_{\xi_1}, \ldots, \alpha_{\xi_n}$. Это приводит еще к одному эквивалентному определению независимости: случайные величины $\xi_1, \ldots, \xi_n$ независимы, если для любых $x_{1j_1}, \ldots, x_{nj_n}$

$$\mathsf{P}\{\xi_1 = x_{1j_1}, \ldots, \xi_n = x_{nj_n}\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i = x_{ij_i}\}.$$

Теорема 1. *Если случайные величины* $\xi_1, \xi_2, \ldots, \xi_n$ *независимы, а* $g_i(x)$ — *числовые функции, то случайные величины* $\eta_1 = g_1(\xi_1), \eta_2 = g_2(\xi_2), \ldots, \eta_n = g_n(\xi_n)$ *также независимы.*

Доказательство. Так как имеет место включение $\mathscr{A}_{g_i(\xi_i)} \subseteq \mathscr{A}_{\xi_i}$, то утверждение сразу следует из определения 1.

Определение независимости (21) можно распространить на случайные векторы $\xi_i = (\xi_{i1}, \ldots, \xi_{ir_i})$, компоненты которых являются случайными величинами. Для этого надо потребовать, чтобы равенство (21) выполнялось для любых множеств $B_i \subseteq R^{r_i}$ из r_i-мерного евклидова пространства. Для таких независимых случайных векторов тоже будет справедлива теорема, аналогичная теореме 1, если $g_i(x)$ — функции, отображающие R^{r_i} и R^{s_i}, η_i — s_i-мерные случайные векторы.

Мультипликативное свойство математических ожиданий.

Теорема 2. *Если случайные величины* $\xi_1, \xi_2, \ldots, \xi_n$ *независимы, то*

$$\mathsf{M}\xi_1\xi_2 \ldots \xi_n = \prod_{i=1}^{n} \mathsf{M}\xi_i. \tag{22}$$

Доказательство. Докажем (22) сначала для двух случайных величин. Пусть ξ, η независимы, и

$$\xi = \sum_{i=1}^{k} x_i I_{A_i}, \quad \eta = \sum_{j=1}^{m} y_j I_{B_j},$$

где $x_1 < x_2 < \ldots < x_k$, $y_1 < y_2 < \ldots < y_m$. Отсюда получаем, в силу аддитивности математического ожидания:

$$\xi\eta = \sum_{i=1}^{k} \sum_{j=1}^{m} x_i y_j I_{A_i B_j}, \quad \mathsf{M}\xi\eta = \sum_{i=1}^{k} \sum_{j=1}^{m} x_i y_j \mathsf{P}(A_i B_j).$$

Из независимости ξ, η следует $\mathsf{P}(A_i B_j) = \mathsf{P}(A_i)\mathsf{P}(B_j)$, поэтому

$$\mathsf{M}\xi\eta = \sum_{i=1}^{k} x_i \mathsf{P}(A_i) \sum_{j=1}^{m} y_j \mathsf{P}(B_j) = \mathsf{M}\xi \cdot \mathsf{M}\eta.$$

Общий случай можно доказать по индукции, если положить $\xi = \xi_1 \ldots \xi_{n-1}$, $\eta = \xi_n$ и воспользоваться независимостью ξ и η.

Из мультипликативного свойства (22) следует аддитивное свойство дисперсии.

Теорема 3. *Если случайные величины* $\xi_1, \xi_2, \ldots, \xi_n$ *независимы, то*

$$\mathsf{D}(\xi_1 + \ldots + \xi_n) = \mathsf{D}\xi_1 + \ldots + \mathsf{D}\xi_n. \tag{23}$$

Доказательство. Докажем (23) для двух независимых случайных величин ξ и η. Общий случай получается по индукции. Имеем $\mathsf{D}(\xi + \eta) = \mathsf{M}[(\xi + \eta) - \mathsf{M}(\xi + \eta)]^2 = \mathsf{M}[(\xi - \mathsf{M}\xi) + (\eta - \mathsf{M}\eta)]^2 = \mathsf{M}(\xi - \mathsf{M}\xi)^2 + \mathsf{M}(\eta - \mathsf{M}\eta)^2 + 2\mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta)$. Так как ξ, η независимы, то $\xi - \mathsf{M}\xi$ и $\eta - \mathsf{M}\eta$ также независимы. Поэтому $\mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) = \mathsf{M}(\xi - \mathsf{M}\xi) \cdot \mathsf{M}(\eta - \mathsf{M}\eta)$. Отсюда следует утверждение, так как $\mathsf{M}(\xi - \mathsf{M}\xi) = \mathsf{M}\xi - \mathsf{M}\xi = 0$.

§ 16. Евклидово пространство случайных величин

Геометрическая интерпретация. Пусть пространство элементарных событий $\Omega=\{\omega\}$ состоит из n элементов $\omega_1, \omega_2, \ldots, \omega_n$. Тогда каждой случайной величине $\xi = \xi(\omega)$ можно поставить в соответствие n-мерный вектор $\xi=(\xi(\omega_1), \ldots, \xi(\omega_n))$. Если ввести скалярное произведение

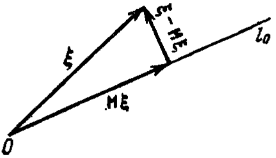

Рис. 9. Проекция ξ на прямую констант l_0.

$$(\xi, \eta)=\sum_{\omega} \xi(\omega)\eta(\omega)p(\omega)=\mathsf{M}\xi\eta,$$

норму $\|\xi\|=\sqrt{(\xi, \xi)}$ и расстояние

$$d(\xi, \eta)=\sqrt{\mathsf{M}(\xi-\eta)^2}=\|\xi-\eta\|,$$

то множество всех случайных величин, определенных на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$, можно рассматривать как n-мерное евклидово пространство.

Определим в этом пространстве прямую констант $l_0=\{\xi\colon \xi(\omega_1)=\xi(\omega_2)= \ldots =\xi(\omega_n)\}$. Спроектируем ξ на прямую l_0, т. е. найдем такую константу $m_\xi \in l_0$, что

$$d(\xi, m_\xi)=\min_{c\in l_0} d(\xi, c).$$

Так как при любой константе $c \in l_0$

$$\mathsf{M}(\xi-c)^2=\mathsf{M}(\xi-\mathsf{M}\xi)^2+(\mathsf{M}\xi-c)^2 \geqslant \mathsf{D}\xi,$$

то $m_\xi=\mathsf{M}\xi$ и $d(\xi, m_\xi)=\sqrt{\mathsf{D}\xi}$. Таким образом, проекция ξ на прямую констант l_0 — это математическое ожидание $\mathsf{M}\xi$, и $\xi-\mathsf{M}\xi$ ортогонально l_0 (ортогональность мы будем обозначать знаком $\perp$, так что в нашем случае $\xi-\mathsf{M}\xi \perp l_0$), поскольку $(\xi-\mathsf{M}\xi, 1)=0$. Расстояние ξ от l_0 равно среднему квадратическому отклонению $\sqrt{\mathsf{D}\xi}$ (см. рис. 9).

Рассмотрим две случайные величины ξ и η. Полагая $\xi=\mathsf{M}\xi+\xi_1$, $\eta=\mathsf{M}\eta+\eta_1$, найдем косинус угла $\varphi_{\xi_1\eta_1}$ между ξ_1 и η_1:

$$\cos\varphi_{\xi_1\eta_1}=\frac{(\xi_1, \eta_1)}{\|\xi_1\|\cdot\|\eta_1\|}=\frac{\mathsf{M}(\xi-\mathsf{M}\xi)(\eta-\mathsf{M}\eta)}{\sqrt{\mathsf{D}\xi\cdot\mathsf{D}\eta}}. \qquad (24)$$

Этот косинус носит название *коэффициента корреляции* между ξ и η и обозначается $\rho(\xi, \eta)$. Числитель справа в (24) носит название *ковариации* между ξ и η и обозначается

$$\mathrm{Cov}(\xi, \eta) = \mathrm{M}(\xi - \mathrm{M}\xi)(\eta - \mathrm{M}\eta). \tag{25}$$

Из (24) и (25) имеем

$$\rho(\xi, \eta) = \frac{\mathrm{Cov}(\xi, \eta)}{\sqrt{\mathrm{D}\xi \cdot \mathrm{D}\eta}}.$$

Из неравенства Коши—Буняковского $(\mathrm{M}\xi_1\eta_1)^2 \leqslant \mathrm{M}\xi_1^2 \cdot \mathrm{M}\eta_1^2$ следует, что всегда $|\rho(\xi, \eta)| \leqslant 1$. Если случайные величины ξ и η независимы, то $\mathrm{Cov}(\xi, \eta) = 0$ (так как $\mathrm{Cov}(\xi, \eta) = \mathrm{M}(\xi - \mathrm{M}\xi)(\eta - \mathrm{M}\eta) = \mathrm{M}(\xi - \mathrm{M}\xi) \cdot \mathrm{M}(\eta - \mathrm{M}\eta) = 0$), следовательно, и $\rho(\xi, \eta) = 0$. Если $\rho(\xi, \eta) = 0$, то $\xi_1 \perp \eta_1$ и случайные величины ξ и η называются *некоррелированными*. Из определения коэффициента корреляции вытекает, что при $\alpha_1\alpha_2 \neq 0$

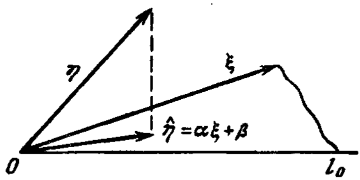

Рис. 10. Проекция η на плоскость (ξ, l_0).

$$\rho(\alpha_1\xi + \beta_1, \alpha_2\eta + \beta_2) = \frac{\alpha_1}{|\alpha_1|} \cdot \frac{\alpha_2}{|\alpha_2|} \rho(\xi, \eta).$$

Спроектируем вектор η на плоскость, в которой лежат l_0 и ξ. Проекция $\hat{\eta} = \alpha\xi + \beta$ определяется константами α и β (см. рис. 10), при которых $\eta - \alpha\xi - \beta \perp 1$ и $\eta - \alpha\xi - \beta \perp \xi$, т. е.

$$\mathrm{M}(\eta - \alpha\xi - \beta) \cdot 1 = 0,$$
$$\mathrm{M}(\eta - \alpha\xi - \beta) \cdot \xi = 0.$$

Это приводит к системе линейных уравнений относительно α и β:

$$\alpha \cdot \mathrm{M}\xi + \beta = \mathrm{M}\eta,$$
$$\alpha \cdot \mathrm{M}\xi^2 + \beta \cdot \mathrm{M}\xi = \mathrm{M}\xi\eta.$$

Решая эту систему, получаем

$$\alpha = \frac{M\xi\eta - M\xi \cdot M\eta}{M\xi^2 - (M\xi)^2} = \rho \frac{\sigma_\xi \sigma_\eta}{\sigma_\xi^2} = \rho \frac{\sigma_\eta}{\sigma_\xi},$$

$$\beta = \frac{M\xi^2 \cdot M\eta - M\xi\eta \cdot M\xi}{\sigma_\xi^2} = M\eta - \rho \frac{M\xi}{\sigma_\xi} \sigma_\eta,$$

где $\sigma_\xi^2 = D\xi$, $\sigma_\eta^2 = D\eta$, $\rho = \rho(\xi, \eta)$. Таким образом, для проекции $\hat{\eta}$ получаем выражение

$$\hat{\eta} = M\eta + \rho \frac{\sigma_\eta}{\sigma_\xi} (\xi - M\xi), \tag{26}$$

называемое *уравнением регрессии* η *на* ξ. Формула (26) дает линейное относительно ξ выражение $\hat{\eta}$, для которого $M(\eta - \hat{\eta})^2 = \min$. Вычислим это расстояние:

$$d^2(\eta, \hat{\eta}) = M(\eta - \hat{\eta})^2 = M\left(\eta - M\eta - \rho \frac{\sigma_\eta}{\sigma_\xi}(\xi - M\xi)\right)^2 =$$

$$= M(\eta - M\eta)^2 + \rho^2 \frac{\sigma_\eta^2}{\sigma_\xi^2} M(\xi - M\xi)^2 -$$

$$- 2\rho \frac{\sigma_\eta}{\sigma_\xi} M(\xi - M\xi)(\eta - M\eta) =$$

$$= \sigma_\eta^2 + \rho^2\sigma_\eta^2 - 2\rho\sigma_\eta^2 = \sigma_\eta^2(1 - \rho^2).$$

Полученное выражение $\sigma_\eta^2(1 - \rho^2)$ носит название *остаточной дисперсии*. Если $\rho^2 = 1$, то $M(\eta - \hat{\eta})^2 = 0$ и $\eta = \hat{\eta}$ с вероятностью 1, т. е. с вероятностью 1 в этом случае ξ и η линейно связаны:

$$\frac{\eta - M\eta}{\sigma_\eta} = \rho \frac{\xi - M\xi}{\sigma_\xi}.$$

Таким образом, коэффициент корреляции $\rho = \rho(\xi, \eta)$ является мерой зависимости между ξ и η. Если ξ и η независимы, то $\rho = 0$; если же $\rho^2 = 1$, то ξ и η зависимы друг от друга линейно, причем при $\rho = 1$ η монотонно возрастает вместе с ξ, а при $\rho = -1$ — убывает.

Если случайные величины $\xi_1, \ldots, \xi_n$ зависимы, то при вычислении дисперсии их суммы можно пользоваться следующей теоремой.

Теорема 4. *Имеет место формула*

$$\mathsf{D}(\xi_1+\ldots+\xi_n)=\sum_{k=1}^{n}\mathsf{D}\xi_k+2\sum_{1\leqslant k<l\leqslant n}\operatorname{Cov}(\xi_k,\xi_l).$$

Доказательство. Докажем теорему для суммы $\xi+\eta$. Общий случай доказывается аналогично. Имеем

$$\mathsf{D}(\xi+\eta)=\mathsf{M}((\xi-\mathsf{M}\xi)+(\eta-\mathsf{M}\eta))^2=$$
$$=\mathsf{M}(\xi-\mathsf{M}\xi)^2+\mathsf{M}(\eta-\mathsf{M}\eta)^2+2\mathsf{M}(\xi-\mathsf{M}\xi)(\eta-\mathsf{M}\eta)=$$
$$=\mathsf{D}\xi+\mathsf{D}\eta+2\operatorname{Cov}(\xi,\eta).$$

§ 17. Условные математические ожидания

Вернемся к понятию условной вероятности. Пусть дано разбиение

$$\alpha:\ A_1+\ldots+A_n=\Omega, \tag{27}$$

причем $\mathsf{P}(A_k)>0$ для всех k. Относительно каждого события A_k из разбиения и любого события $B\in\mathscr{A}$ можно образовать условную вероятность $\mathsf{P}(B|A_k)=\frac{\mathsf{P}(BA_k)}{\mathsf{P}(A_k)}$. Пусть $\mathscr{A}(\alpha)$ — алгебра событий, порожденная разбиением (27). Определим *условную вероятность* $\mathsf{P}(B|\mathscr{A}(\alpha))$ *относительно* $\mathscr{A}(\alpha)$ как случайную величину,

Таблица 5

Закон распределения условной вероятности

Значения $\mathsf{P}(B\mid\mathscr{A}(\alpha))$	$\mathsf{P}(B\mid A_1)$	$\mathsf{P}(B\mid A_2)$	...	$\mathsf{P}(B\mid A_n)$
Вероятности	$\mathsf{P}(A_1)$	$\mathsf{P}(A_2)$	...	$\mathsf{P}(A_n)$

которая принимает значение $\mathsf{P}(B\mid A_k)$ при $\omega\in A_k$. Закон распределения этой случайной величины $\mathsf{P}(B\mid\mathscr{A}(\alpha))$ определяется таблицей 5.

Правую часть формулы полной вероятности

$$\mathsf{P}(B)=\sum_{k=1}^{n}\mathsf{P}(A_k)\mathsf{P}(B\mid A_k)$$

можно теперь трактовать как математическое ожидание $\mathsf{MP}(B|\mathscr{A}(\alpha))$ случайной величины $\mathsf{P}(B|\mathscr{A}(\alpha))$.

Пусть разбиение α_ξ определяется случайной величиной ξ: $A_k=\{\xi=x_k\}$. Обозначим $\mathscr{A}_\xi$ алгебру, порожденную ξ. Условная вероятность $\mathsf{P}(B|\mathscr{A}_\xi)$ в этом случае есть функция от значений ξ, и мы обозначаем ее $\mathsf{P}(B|\xi)$, а ее значение — через $\mathsf{P}(B|\xi=x_k)$.

Предположим теперь, что $B_l=\{\eta=y_l\}$, $l=1, \ldots, m$, образуют разбиение, порожденное случайной величиной η. *Условным законом распределения* η *при заданном значении* $\xi=x_k$ назовем набор условных вероятностей

$$\mathsf{P}\{\eta=y_l|\xi=x_k\}=\frac{\mathsf{P}\{\eta=y_l,\ \xi=x_k\}}{\mathsf{P}\{\xi=x_k\}},\qquad l=1, \ldots, m;$$

условным математическим ожиданием η *при заданном значении* $\xi=x_k$ будет тогда сумма

$$\mathsf{M}\{\eta|\xi=x_k\}=\sum_{l=1}^{m} y_l\mathsf{P}\{\eta=y_l|\xi=x_k\}=\frac{\sum\limits_{l=1}^{m} y_l\mathsf{P}\{\eta=y_l,\ \xi=x_k\}}{\mathsf{P}\{\xi=x_k\}}. \tag{28}$$

Мы можем считать $\mathsf{M}\{\eta|\xi=x_k\}$ значениями случайной величины $\mathsf{M}(\eta|\xi)$, являющейся функцией от ξ и равной $\mathsf{M}\{\eta|\xi=x_k\}$ при $\xi=x_k$. Случайную величину $\mathsf{M}(\eta|\xi)$ будем называть *условным математическим ожиданием при заданном* ξ. От этой случайной величины можно вычислить математическое ожидание

$$\mathsf{M}[\mathsf{M}(\eta|\xi)]=\sum_{k=1}^{n}\mathsf{P}\{\xi=x_k\}\mathsf{M}\{\eta|\xi=x_k\}. \tag{29}$$

Теорема 5. *Имеет место равенство*

$$\mathsf{M}[\mathsf{M}(\eta|\xi)]=\mathsf{M}\eta. \tag{30}$$

Доказательство. Подставляя в (29) значения условных математических ожиданий (28), имеем

$$\mathsf{M}[\mathsf{M}(\eta\,|\,\xi)] = \sum_{k=1}^{n} \mathsf{P}\{\xi = x_k\}\,\mathsf{M}\{\eta\,|\,\xi = x_k\} =$$

$$= \sum_{k=l}^{n} \sum_{l=1}^{m} y_l \mathsf{P}\{\eta = y_l,\ \xi = x_k\} = \sum_{l=1}^{m} y_l \mathsf{P}\{\eta = y_l\} = \mathsf{M}\eta.$$

Теорема доказана.

Пример 4. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — случайные величины, имеющие одинаковые математические ожидания $\mathsf{M}\xi_i$, а ν — независимая от них случайная величина, принимающая целые неотрицательные значения. Определим сумму случайного числа случайных величин $\eta_\nu = \xi_1 + \ldots + \xi_\nu$ при $\nu \geqslant 1$, $\eta_\nu = 0$ при $\nu = 0$. Тогда $\mathsf{M}\eta_\nu = \mathsf{M}\xi_i \cdot \mathsf{M}\nu$. Эта формула доказывается с помощью (30). Имеем при любом $\nu = n$:

$$\mathsf{M}\{\eta_\nu\,|\,\nu = n\} = \mathsf{M}(\xi_1 + \ldots + \xi_n) = n \cdot \mathsf{M}\xi_i,$$

т. е. $\mathsf{M}\{\eta_\nu\,|\,\nu\} = \nu \cdot \mathsf{M}\xi_i$. Отсюда получаем

$$\mathsf{M}\eta_\nu = \mathsf{M}[\mathsf{M}(\eta_\nu\,|\,\nu)] = \mathsf{M}\nu \cdot \mathsf{M}\xi_i.$$

§ 18. Неравенство Чебышева. Закон больших чисел

Следующие два неравенства носят название *неравенства Чебышева.*

Теорема 6. *Для любого $x > 0$ имеют место неравенства*:

$$\mathsf{P}\{|\xi| \geqslant x\} \leqslant \frac{\mathsf{M}|\xi|}{x}, \tag{31}$$

$$\mathsf{P}\{|\xi - \mathsf{M}\xi| \geqslant x\} \leqslant \frac{\mathsf{D}\xi}{x^2}. \tag{32}$$

Доказательство. Вычисляя математическое ожидание от обеих частей неравенства

$$|\xi| = |\xi|\,I_{\{|\xi| \geqslant x\}} + |\xi|\,I_{\{|\xi| < x\}} \geqslant |\xi|\,I_{\{|\xi| \geqslant x\}} \geqslant x I_{\{|\xi| \geqslant x\}},$$

получаем

$$\mathsf{M}|\xi| \geqslant x\mathsf{M}I_{\{|\xi| \geqslant x\}} = x\mathsf{P}\{|\xi| \geqslant x\}.$$

Неравенство (32) получается из первого неравенства, если его применить к случайной величине $\eta=(\xi-\mathsf{M}\xi)^2$ и воспользоваться тем, что $\mathsf{M}\eta=\mathsf{D}\xi$. Теорема доказана.

Неравенство Чебышева (32) показывает, что при малой дисперсии $\mathsf{D}\xi$ с вероятностью, близкой к 1, случайная величина ξ концентрируется около математического ожидания $\mathsf{M}\xi$:

$$\mathsf{P}\{|\xi-\mathsf{M}\xi|<x\}\geqslant 1-\frac{\mathsf{D}\xi}{x^2}. \tag{33}$$

Неравенство Чебышева позволяет просто доказывать некоторые предельные соотношения, в которых участвуют последовательности независимых случайных величин ξ_n. В рассматриваемой в этой главе конечной схеме мы имеем право лишь утверждать, что любое конечное множество случайных величин может быть определено на одном вероятностном пространстве. В доказываемых ниже теоремах 7 и 8 мы полагаем, что при каждом n случайные величины $\xi_1, \ldots, \xi_n$ определены на некотором конечном вероятностном пространстве $(\Omega_n, \mathscr{A}_n, \mathsf{P}_n)$, причем при каждом фиксированном $k<n$ случайная величина ξ_k имеет распределение вероятностей, не зависящее от n. Вообще говоря, любую последовательность независимых случайных величин ξ_n можно определить на одном бесконечном вероятностном пространстве. Данные ниже доказательства теорем 7 и 8 имеют общий характер и не зависят от конечности рассматриваемой схемы.

Теорема 7. (Теорема Чебышева.) *Если $\xi_1, \xi_2, \ldots$ независимы и существует такая константа $c>0$, что $\mathsf{D}\xi_n\leqslant c$, $n=1, 2, \ldots$, то при любом $\varepsilon>0$*

$$\lim_{n\to\infty}\mathsf{P}\left\{\left|\frac{\xi_1+\ldots+\xi_n}{n}-\frac{\mathsf{M}\xi_1+\ldots+\mathsf{M}\xi_n}{n}\right|>\varepsilon\right\}=1. \tag{34}$$

Доказательство. Обозначим $\zeta_n=\xi_1+\ldots+\xi_n$ и применим к ζ_n/n неравенство (33). Имеем при любом $x>0$:

$$1\geqslant\mathsf{P}\left\{\left|\frac{\zeta_n}{n}-\frac{\mathsf{M}\zeta_n}{n}\right|<x\right\}\geqslant 1-\frac{\mathsf{D}\zeta_n}{x^2n^2}\geqslant 1-\frac{c}{x^2n}, \tag{35}$$

так как $D\zeta_n = \sum_{k=1}^{n} D\xi_k \leqslant nc$ (см. теорему 4). Из (35) при $n \to \infty$ имеем (34).

Следствие. *Если* $\xi_1, \xi_2, \ldots$ *независимы и одинаково распределены*, $M\xi_n = a$, $D\xi_n = \sigma^2 < \infty$, *то при любом* $x > 0$

$$\lim_{n\to\infty} \mathsf{P}\left\{\left|\frac{\xi_1 + \ldots + \xi_n}{n} - a\right| < x\right\} = 1. \tag{36}$$

Предельные утверждения типа (34) и (36) носят название *закона больших чисел.* Закон больших чисел утверждает, что с вероятностью, приближающейся при $n \to \infty$ к 1, среднее арифметическое сумм независимых слагаемых при определенных условиях становится близким к константе.

Из (36) получаем закон больших чисел в схеме Бернулли.

Теорема 8. (Теорема Бернулли.) *Пусть* μ_n — *число успехов при* n *испытаниях в схеме Бернулли с вероятностью* $0 < p < 1$ *в каждом испытании. Тогда при любом* $x > 0$

$$\lim_{n\to\infty} \mathsf{P}\left\{\left|\frac{\mu_n}{n} - p\right| \leqslant x\right\} = 1. \tag{37}$$

Доказательство. Мы можем представить μ_n в виде суммы независимых слагаемых $\xi_1 + \ldots + \xi_n$, где $\xi_i = 1$, если при i-м испытании произошел успех, и $\xi_i = 0$ в противоположном случае. Поскольку $M\xi_i = p$, $D\xi_i = p(1-p)$, то к $\mu_n = \xi_1 + \ldots + \xi_n$ применимо следствие (36). Теорема доказана.

Соотношение (37) показывает, что при больших n разность между относительной частотой μ_n/n и вероятностью успеха мала с вероятностью, близкой к 1. В условиях, когда справедливо свойство устойчивости частот, можно применять следующий принцип: *при единичном испытании маловероятное событие практически невозможно.* Считая серию в n испытаний в схеме Бернулли за единичное испытание и выбирая x таким, чтобы $\frac{D\mu_n}{x^2 n} = \frac{pq}{x^2 n}$ было мало, мы можем утверждать, что неравенство $|\mu_n/n - p| > x$ практически невозможно. Вопрос о том, какие вероятности считать малыми, зависит от конкретной прикладной задачи.

Задачи

1. Из 28 костей домино случайно выбирается одна. Найти закон распределения суммы очков на половинках этой кости. (Кость домино — это прямоугольник, разделенный на две части. Каждая из частей помечена одной из цифр 0, 1, 2, 3, 4, 5, 6. Все 28 комбинаций пар (i, j), $0 \leqslant i \leqslant j \leqslant 6$, составляют набор костей домино.)

2. Найти закон распределения случайной величины $\eta = \sin \frac{\pi}{3} \xi$, где ξ — число очков, выпадающее при бросании игральной кости.

3. Найти математическое ожидание $\mathsf{M}\xi$ и дисперсию $\mathsf{D}\xi$:

а) биномиального распределения $\mathsf{P}\{\xi = m\} = C_n^m p^m (1-p)^{n-m}$, $m = 0, \ldots, n$;

б) гипергеометрического распределения

$$\mathsf{P}\{\xi = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \qquad m = 0, 1, \ldots, \min\{n, M\};$$

в) равномерного распределения на $\{1, 2, \ldots, N\}$

$$\mathsf{P}\{\xi = m\} = \frac{1}{N}, \qquad m = 1, \ldots, N.$$

4. Из множества всех $n!$ подстановок $\begin{pmatrix} 1 & 2 & \ldots & n \\ x_1 & x_2 & \ldots & x_n \end{pmatrix}$ случайно и равновероятно выбирается одна. Найти вероятность того, что $x_k \neq k$ при всех $1 \leqslant k \leqslant n$.

5. Если $\rho(\xi, \eta) = 0$, то ξ и η не обязательно независимы. Построить пример.

6. Найти $\mathsf{M}\xi_i$, $\mathsf{D}\xi_i$ и $\mathrm{Cov}(\xi_i, \xi_j)$ в полиномиальном распределении

$$\mathsf{P}\{\xi_1 = m_1, \ldots, \xi_r = m_r\} = \frac{n!}{m_1! \ldots m_r!} p_1^{m_1} \ldots p_r^{m_r},$$

при целых неотрицательных $m_1 + \ldots + m_r = n$ и $\mathsf{P}\{\xi_1 = m_1, \ldots, \xi_r = m_r\} = 0$ в остальных случаях.

7. Из чисел 0000, 0001, ..., 9999 случайно и равновероятно выбирают число $\xi_1\xi_2\xi_3\xi_4$. Доказать, что ξ_i независимы в совокупности.

8. Найти коэффициент корреляции между ξ и ξ^2, если $\mathsf{P}\{\xi = 0\} = 1/3$, $\mathsf{P}\{\xi = 1\} = 1/2$, $\mathsf{P}\{\xi = -1\} = 1/6$.

9. Бросается игральная кость. Пусть на ней выпало ν очков. После этого та же игральная кость бросается ν раз. Обозначим η сумму выпавшего числа очков в этих ν бросаниях кости. Найти $\mathsf{M}\eta$.

10. а) Показать, что при любом $x > 0$ найдется такая случайная величина ξ, для которой $\mathsf{M}|\xi| \leqslant x$ и неравенство (31) превращается в равенство. б) Аналогичное утверждение справедливо и для неравенства (32), но в этом случае надо полагать $\mathsf{D}\xi \leqslant x^2$.

11. Пользуясь неравенством Чебышева, оценить вероятность того, что при 1000 бросаниях монеты число выпадений герба μ будет заключено между 450 и 550.

Глава 4. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ В СХЕМЕ БЕРНУЛЛИ

§ 19. Биномиальное распределение

Биномиальное распределение числа успехов μ при n независимых испытаниях в схеме Бернулли с вероятностью успеха p в каждом испытании задается вероятностями

$$\mathsf{P}\{\mu=m\}=C_n^m p^m q^{n-m}, \quad m=0, 1, \ldots, n; \ q=1-p. \quad (1)$$

Формула (1) записывается достаточно компактно и просто, однако использование ее для вычисления вероятностей $\mathsf{P}\{\mu=m\}$ при больших значениях n и m вызывает значительные трудности. При очень больших значениях n и m можно производить вычисления на быстродействующей ЭВМ. Но и в этом случае при составлении программы надо учитывать то обстоятельство, что очень большие числа, возникающие при вычислении C_n^m, приходится множить на очень малые числа $p^m q^{n-m}$. При этом надо следить, чтобы промежуточные численные результаты не выходили за диапазон допустимых значений.

Таблицы для вероятностей $C_n^m p^m q^{n-m}$ громоздки и очень неудобны для пользования, так как содержат три входа (n, p и m). Еще хуже дело обстоит с вычислением вероятностей

$$\mathsf{P}\{m_1 \leqslant \mu \leqslant m_2\}=\sum_{m=m_1}^{m_2} C_n^m p^m q^{n-m}, \quad (2)$$

которые зависят уже от четырех параметров: n, p, m_1 и m_2.

Поскольку схема независимых испытаний служит вероятностной моделью многих реальных случайных явлений, представляет значительный интерес задача о нахождении асимптотических формул, позволяющих приближенно вычислять вероятности (1) и (2) при больших значениях n, m, m_1, m_2. Такие формулы дают нам предельные теоремы.

§ 20. Теорема Пуассона

Рассмотрим сначала случай больших n и малых p.

Теорема 1. (Теорема Пуассона.) *Если $n \to \infty$ и $p \to 0$ так, что $np \to a$, то для любого фиксированного $m = 0, 1, \ldots$*

$$\mathrm{P}\{\mu = m\} = C_n^m p^m q^{n-m} \to \frac{a^m}{m!} e^{-a}. \tag{3}$$

Доказательство. Утверждение (3) сразу вытекает из равенства

$$\mathrm{P}(\mu = m) = \\ = \frac{(np)^m}{m!} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \ldots \left(1 - \frac{m-1}{n}\right)(1-p)^{n-m},$$

если учесть, что при $n \to \infty$, $np \to a$ предел $(1-p)^n$ равен e^{-a}.

Можно показать, что в предельном соотношении (3) имеет место следующая оценка:

$$\left|\mathrm{P}\{\mu = m\} - \frac{a^m}{m!} e^{-a}\right| \leqslant \frac{a^2}{n}, \qquad a = np. \tag{4}$$

Мы докажем более сильное утверждение в более общей схеме независимых испытаний. Рассмотрим n независимых испытаний с разными вероятностями успеха в разных испытаниях. Обозначим p_i вероятность успеха, $q_i = 1 - p_i$ — вероятность неуспеха в i-м испытании. Обозначим распределение μ — числа успехов при n испытаниях — через

$$\mathrm{P}\{\mu = m\} = P_n(m) = P_n(m;\ p_1, \ldots, p_n). \tag{5}$$

Такую схему независимых испытаний с разными p_i называют *схемой Пуассона*. Схема Пуассона при $p_i \equiv p$ превращается в схему Бернулли. Вероятности $\mathrm{P}\{\mu = m\}$

в схеме Пуассона не записываются в компактном виде, аналогичном (1). Например,

$$\mathsf{P}\{\mu=0\}=q_1q_2\ldots q_n,$$

$$\mathsf{P}\{\mu=1\}=p_1q_2\ldots q_n+q_1p_2q_3\ldots q_n+\ldots+q_1q_2\ldots q_{n-1}p_n,$$

. . . .

$$\mathsf{P}\{\mu=n\}=p_1p_2\ldots p_n,$$

$$\mathsf{P}\{\mu=m\}=0 \text{ при } m<0 \text{ и } m>n.$$

Обозначим

$$\Pi(m,\ a)=\frac{a^m}{m!}e^{-a}. \tag{6}$$

Имеет место

Теорема 2. *В схеме Пуассона для любого натурального n, любых вероятностей $p_1, p_2, \ldots, p_n$ и любого числового множества B имеет место неравенство*

$$\left|\mathsf{P}\{\mu\in B\}-\sum_{m\in B}\Pi(m,\ a)\right|\leqslant\sum_{i=1}^{n}p_i^2, \tag{7}$$

где $a=p_1+p_2+\ldots+p_n$.

Доказательство. Формула (6) задает вероятности $\Pi(m, a)$ более общего, чем мы рассматривали до сих пор, *распределения Пуассона*, имеющего положительные вероятности при $m=0,\ 1,\ 2,\ \ldots$ такие, что $\sum_{m=0}^{\infty}\Pi(m,\ a)=1$. Докажем, что левая часть неравенства (7) не превосходит V_n, где

$$V_n=\frac{1}{2}\sum_{m=0}^{\infty}|P_n(m;\ p_1,\ \ldots,\ p_n)-\Pi(m,\ a)|.$$

Разобьем все неотрицательные целые числа m на два множества. Положим $m\in B^+$, если $P_n(m)>\Pi(m, a)$, и $m\in B^-$ в остальных случаях. Обозначим

$$\sum\nolimits^{+}=\sum_{m\in B^+}(P_n(m)-\Pi(m,\ a)),$$

$$\sum\nolimits^{-}=\sum_{m\in B^-}(P_n(m)-\Pi(m,\ a)).$$

Так как $\sum\limits_m P_n(m) = \sum\limits_m \Pi(m, a) = 1$, то

$$0 = \sum_m (P_n(m) - \Pi(m, a)) = \sum{}^+ + \sum{}^-$$

и

$$V_n = \frac{1}{2}\sum_m |P_n(m) - \Pi(m, a)| = \sum{}^+.$$

С другой стороны, для любого числового множества B имеем

$$\Big|\sum_{m\in B}(P_n(m) - \Pi(m, a))\Big| \leqslant$$
$$\leqslant \max\Big\{\sum_{m\in B\cap B^+}(P_n(m) - \Pi(m, a)), \Big|\sum_{m\in B\cap B^-}(P_n(m) - - \Pi(m, a))\Big|\Big\} \leqslant \sum{}^+ = V_n.$$

Итак, для доказательства (7) нам достаточно доказать, что

$$V_n \leqslant \sum_{i=1}^n p_i^2. \tag{8}$$

Проведем доказательство по индукции. При $n = 1$ и $p_1 = p$ имеем

$$|P_1(0; p) - \Pi(0; p)| = e^{-p} + p - 1,$$
$$|P_1(1; p) - \Pi(1; p)| = p - pe^{-p},$$
$$|P_1(m; p) - \Pi(m; p)| = \frac{p^m}{m!}e^{-p}, \qquad m \geqslant 2,$$

откуда получаем

$$V_1 = \frac{1}{2}\sum_{m=0}^\infty |P_1(m; p) - \Pi(m, p)| =$$
$$= \frac{1}{2}\Big(-1 + p + e^{-p} + p - pe^{-p} + \sum_{m=2}^\infty \frac{p^m}{m!}e^{-p}\Big) =$$
$$= p(1 - e^{-p}) \leqslant p^2, \tag{9}$$

так как $0 \leqslant 1 - e^{-x} \leqslant x$ при $x \geqslant 0$.

Далее воспользуемся тем, что по формуле полной вероятности

$$P_n(m;\ p_1, \ldots, p_n) = P_{n-1}(m;\ p_1, \ldots, p_{n-1})\,P_1(0;\ p_n) + \\ + P_{n-1}(m-1;\ p_1, \ldots, p_{n-1})\,P(1;\ p_n) = \\ = \sum_{k=0}^{m} P_{n-1}(m-k)\,P_1(k;\ p_n), \qquad (10)$$

и при любых $a_1 \geqslant 0$, $a_2 \geqslant 0$

$$\Pi(m,\ a_1 + a_2) = \sum_{k=0}^{n} \Pi(m-k,\ a_1)\,\Pi(k,\ a_2). \qquad (11)$$

Обозначим $a_n = \sum_{k=1}^{n} p_k$, $A_n = \sum_{k=1}^{n} p_k^2$. Предположим, что

$$V_{n-1} \leqslant A_{n-1}. \qquad (12)$$

Применяя формулы (9) — (12), оценим V_n:

$$V_n = \frac{1}{2}\sum_{m=0}^{\infty} |P_n(m) - \Pi(m,\ a_n)| = \\ = \frac{1}{2}\sum_{m}\Big|\sum_{k} P_{n-1}(m-k)\,P_1(k;\ p_n) - \\ - \sum_{k}\Pi(m-k,\ a_{n-1})\,\Pi(k,\ p_n)\Big| \leqslant \\ \leqslant \frac{1}{2}\sum_{m=0}^{\infty}\sum_{k=0}^{m} |P_{n-1}(m-k) - \Pi(m-k,\ a_{n-1})|\,P_1(k;\ p_n) + \\ + \frac{1}{2}\sum_{m=0}^{\infty}\sum_{k=0}^{m}\Pi(m-k,\ a_{n-1})\,|P_1(k;\ p_n) - \Pi(k,\ p_n)| \leqslant \\ \leqslant V_{n-1} + p_n^2 \leqslant A_n.$$

Теорема доказана.

Следствие. *В схеме Бернулли при любых n и p*

$$\Big|\mathsf{P}\{\mu \in B\} - \sum_{m \in B}\Pi(m,\ a)\Big| \leqslant np^2 = \frac{a^2}{n},$$

где $a = np$.

§ 21. Локальная предельная теорема Муавра — Лапласа

Биномиальное распределение (1) случайной величины μ имеет $\mathsf{M}\mu = np$ и $\mathsf{D}\mu = npq$ (см. задачу 3 в гл. 3). Обозначим $\sigma = \sqrt{npq}$ среднее квадратическое отклонение. Доказываемая ниже теорема дает асимптотическую формулу для биномиальной вероятности (1) при p, не близких к 0 или 1.

Теорема 3. (Локальная предельная теорема Муавра — Лапласа.) *Если в схеме Бернулли $\sigma = \sqrt{npq} \to \infty$, то для любого $C > 0$ равномерно по всем $|x| \leqslant C$ вида $x = \frac{m - np}{\sigma}$, где m — целые неотрицательные числа,*

$$\mathsf{P}\left\{\frac{\mu - np}{\sqrt{npq}} = x\right\} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2}(1 + o(1)). \qquad (13)$$

Доказательство. Пусть $m = np + x\sigma$. Оценим логарифм вероятности

$$\mathsf{P}\{\mu = m\} = \mathsf{P}\left\{\frac{\mu - np}{\sigma} = x\right\} = \frac{n!}{m!\,(n-m)!} p^m q^{n-m},$$

равный

$$\log \mathsf{P}\{\mu = m\} = \log n! - \log m! - \log (n-m)! + \\ + m \log p + (n-m) \log q.$$

Воспользуемся асимптотической при $n \to \infty$ формулой Стирлинга

$$\log n! = n \log n + \log \sqrt{2\pi n} - n + \theta_n,$$

где $\theta_n = O\left(\frac{1}{n}\right)$. Обозначим $k = n - m = nq - x\sigma$. Из условия теоремы следует, что $m = np\left(1 + \frac{xq}{\sigma}\right) \to \infty$, $k = nq\left(1 - \frac{xp}{\sigma}\right) \to \infty$, поэтому можно применить формулу Стирлинга для оценки $\log n!$, $\log m!$, $\log k!$. Имеем

$$\log \mathsf{P}\{\mu = m\} = n \log n - m \log m - k \log k + \\ + m \log p + k \log q + \frac{1}{2} \log \frac{n}{2\pi m k} + \theta_n - \theta_m - \theta_k. \qquad (14)$$

Так как $\log \frac{n}{mk} = \log \frac{1}{npq} - \log\left(1 + \frac{xq}{\sigma}\right) - \log\left(1 - \frac{xp}{\sigma}\right) = 2\log\frac{1}{\sigma} + O\left(\frac{1}{\sigma}\right)$, $\frac{1}{n} = O\left(\frac{1}{\sigma^2}\right)$, $\frac{1}{m} = \frac{1}{n\left(1+\frac{xq}{\sigma}\right)} = O\left(\frac{1}{\sigma^2}\right)$, $\frac{1}{k} = O\left(\frac{1}{\sigma^2}\right)$, $\log(1+\varepsilon) = O(\varepsilon)$, $\varepsilon \to 0$, то из (14) получаем

$$\log \mathsf{P}\{\mu = m\} = -m\log\frac{m}{np} - k\log\frac{k}{nq} + \log\frac{1}{\sigma\sqrt{2\pi}} + O\left(\frac{1}{\sigma}\right). \tag{15}$$

Далее, из (15) следует

$$\log \mathsf{P}\{\mu = m\} = -(np + x\sigma)\log\left(1 + \frac{xq}{\sigma}\right) - (nq - x\sigma)\log\left(1 - \frac{xp}{\sigma}\right) + \log\frac{1}{\sigma\sqrt{2\pi}} + O\left(\frac{1}{\sigma}\right) =$$
$$= \log\frac{1}{\sigma\sqrt{2\pi}} - (np + x\sigma)\left(\frac{xq}{\sigma} - \frac{x^2q^2}{2\sigma^2} + O\left(\frac{1}{\sigma^3}\right)\right) - (nq - x\sigma)\left(-\frac{xp}{\sigma} - \frac{x^2p^2}{2\sigma^2} + O\left(\frac{1}{\sigma^3}\right)\right) + O\left(\frac{1}{\sigma}\right) =$$
$$= \log\frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2} + O\left(\frac{1}{\sigma}\right),$$

что и доказывает асимптотическую формулу (13).

§ 22. Интегральная предельная теорема Муавра — Лапласа

Для приближенного вычисления вероятностей $\mathsf{P}\{m_1 \leqslant \mu \leqslant m_2\}$ можно применять следующую теорему.

Теорема 4. (Интегральная предельная теорема Муавра — Лапласа.) *При* $\sigma = \sqrt{npq} \to \infty$ *равномерно по* $-\infty \leqslant a < b \leqslant \infty$

$$\mathsf{P}\left\{a \leqslant \frac{\mu - np}{\sqrt{npq}} \leqslant b\right\} - \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx \to 0. \tag{16}$$

Доказательство. Предположим сначала, что $|a| \leqslant C$, $|b| \leqslant C$. Пусть $m_1 =]np + a\sqrt{npq}[$, $m_2 = [np + b\sqrt{npq}]$, где $]x[$ — такое наименьшее целое число, что $x \leqslant]x[$, а $[x]$ — такое наибольшее целое число, что $[x] \leqslant x$. Тогда

$$\mathsf{P}\left\{a \leqslant \frac{\mu - np}{\sqrt{npq}} \leqslant b\right\} = \sum_{m=m_1}^{m_2} \mathsf{P}\{\mu = m\}. \tag{17}$$

Обозначим $m = np + x_m\sqrt{npq}$, тогда $\Delta x_m = x_{m+1} - x_m = 1/\sigma$. По локальной предельной теореме запишем (17) в виде

$$\mathsf{P}\left\{a \leqslant \frac{\mu - np}{\sqrt{npq}} \leqslant b\right\} = \sum_{m=m_1}^{m_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}} \Delta x_m \left(1 + O\left(\frac{1}{\sigma}\right)\right). \tag{18}$$

Справа в (18) стоит интегральная сумма, сходящаяся равномерно по a и b при $\sigma \to \infty$ к интегралу $\frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{x^2}{2}} dx$; отсюда получаем утверждение (16), когда $|a| \leqslant C$, $|b| \leqslant C$.

Снимем теперь ограничение $|a| \leqslant C$, $|b| \leqslant C$. Обозначим $\xi_n = \frac{\mu - np}{\sigma}$. Имеем равенство

$$\mathsf{P}\{|\xi_n| > C\} = 1 - \mathsf{P}\{|\xi_n| \leqslant C\}. \tag{19}$$

Как известно из анализа,

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

поэтому

$$\frac{1}{\sqrt{2\pi}} \int\limits_{-C}^{C} e^{-\frac{x^2}{2}} dx = 1 - \frac{1}{\sqrt{2\pi}} \int\limits_{|x|>C} e^{-\frac{x^2}{2}} dx. \tag{20}$$

Из (19) и (20) получаем

$$\left|\mathsf{P}\{|\xi_n| > C\} - \frac{1}{\sqrt{2\pi}} \int\limits_{|x|>C} e^{-\frac{x^2}{2}} dx\right| = \left|\mathsf{P}\{|\xi_n| \leqslant C\} - \frac{1}{\sqrt{2\pi}} \int\limits_{-C}^{C} e^{-\frac{x^2}{2}} dx\right|. \tag{21}$$

Пусть задано $\varepsilon > 0$. Тогда найдется такое C, что

$$\frac{1}{\sqrt{2\pi}} \int\limits_{|x|>C} e^{-\frac{x^2}{2}}\, dx < \frac{\varepsilon}{8}. \tag{22}$$

Зафиксируем его. По только что доказанному найдется такое n_1, что для всех $n \geqslant n_1$

$$\left| \mathsf{P}\{|\xi_n| \leqslant C\} - \frac{1}{\sqrt{2\pi}} \int\limits_{-C}^{C} e^{-\frac{x^2}{2}}\, dx \right| < \frac{\varepsilon}{8},$$

откуда, в силу (21) и (22), для тех же $n \geqslant n_1$ имеем

$$\mathsf{P}\{|\xi_n| > C\} \leqslant \frac{\varepsilon}{4}. \tag{23}$$

Берем теперь любой интервал $[a, b]$ и обозначим $[A, B] = [a, b] \cap [-C, C]$. Так как $-C \leqslant A \leqslant B \leqslant C$, то, как мы уже доказали, существует такое n_2, что для всех $n \geqslant n_2$ имеет место неравенство

$$\left| \mathsf{P}\{\xi_n \in [A, B]\} - \frac{1}{\sqrt{2\pi}} \int\limits_{A}^{B} e^{-\frac{x^2}{2}}\, dx \right| < \frac{\varepsilon}{2}. \tag{24}$$

Из неравенства

$$\left| \mathsf{P}\{\xi_n \in [a, b]\} - \frac{1}{\sqrt{2\pi}} \int\limits_{a}^{b} e^{-\frac{x^2}{2}}\, dx \right| \leqslant \mathsf{P}\{|\xi_n| > C\} +$$

$$+ \frac{1}{\sqrt{2\pi}} \int\limits_{|x|>C} e^{-\frac{x^2}{2}}\, dx + \left| \mathsf{P}\{\xi_n \in [A, B]\} - \frac{1}{\sqrt{2\pi}} \int\limits_{A}^{B} e^{-\frac{x^2}{2}}\, dx \right|$$

получаем, в силу (22)—(24), что при $n \geqslant n_0 = \max(n_1, n_2)$

$$\left| \mathsf{P}\{\xi_n \in [a, b]\} - \frac{1}{\sqrt{2\pi}} \int\limits_{a}^{b} e^{-\frac{x^2}{2}}\, dx \right| \leqslant \varepsilon$$

равномерно по всем $a \leqslant b$. Теорема доказана.

§ 23. Применения предельных теорем

Предельные теоремы Пуассона и Муавра — Лапласа применяются для приближенного вычисления вероятностей $\mathsf{P}\{\mu = m\}$ и $\mathsf{P}\{m_1 \leqslant \mu \leqslant m_2\}$ в схеме Бернулли при больших n. Приближение, даваемое теоремой Пуассона, называется *пуассоновским*. Приближение, получаемое с помощью теорем Муавра — Лапласа, называется *нормальным*, так как функция $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ есть плотность нормального распределения (см. § 31). Для распределения Пуассона $\frac{a^m}{m!}e^{-a}$ и интеграла

$$\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{u^2}{2}}\,du, \tag{25}$$

называемого *интегралом Лапласа*, имеются таблицы.

Пусть, например, нам нужно вычислить вероятность $\mathsf{P}\{m_1\leqslant\mu\leqslant m_2\}$ в схеме Бернулли с n независимыми испытаниями и с вероятностью успеха p. Вычислим $x_{m_1}=\frac{m_1-np}{\sqrt{npq}}$, $x_{m_2}=\frac{m_2-np}{\sqrt{npq}}$ и положим

$$\mathsf{P}\{m_1\leqslant\mu\leqslant m_2\}\approx\frac{1}{\sqrt{2\pi}}\int_{x_{m_1}}^{x_{m_2}} e^{-\frac{x^2}{2}}\,dx=\Phi_0(x_{m_2})-\Phi_0(x_{m_1}). \tag{26}$$

При этом мы допускаем некоторую погрешность. Можно оценить эту погрешность, но точная ее оценка очень сложна, а более простые оценки слишком грубы. Эту погрешность можно значительно уменьшить, если в правой части приближенного равенства немного изменить пределы интегрирования, полагая

$$\mathsf{P}\{m_1\leqslant\mu\leqslant m_2\}\approx\Phi_0\left(x_{m_2+1/2}\right)-\Phi_0\left(x_{m_1-1/2}\right), \tag{27}$$

где $x_{m_2+1/2}=\frac{m_2+1/2-np}{\sqrt{npq}}$, $x_{m_1-1/2}=\frac{m_1-1/2-np}{\sqrt{npq}}$.

Из табл. 6 видно, что приближенная формула (27) дает значения вероятностей $\mathsf{P}\{m_1\leqslant\mu\leqslant m_2\}$ с точностью до трех-четырех знаков после запятой даже при n порядка нескольких сотен. Обычно применяемая формула (26) такой точности не дает.

Таблица 6

m_1	m_2	Точное значение по формуле (2)	Нормальное приближение по формуле (26)	Уточненное нормальное приближение по формуле (27)
		$n=100;\ p=0,5$		
40	60	0,9648	0,9545	0,9643
45	55	0,7287	0,6827	0,7287
55	65	0,1832	0,1573	0,1831
		$n=300;\ p=0,5$		
135	165	0,9267	0,9167	0,9265
140	160	0,7747	0,7518	0,7747
160	180	0,1361	0,1238	0,1361
		$n=500;\ p=0,5$		
230	270	0,9334	0,9264	0,9333
240	260	0,6523	0,6289	0,6523
260	280	0,1950	0,1819	0,1946
		$n=1000;\ p=0,5$		
470	530	0,9463	0,9422	0,9463
530	560	0,03095	0,02882	0,03097
		$n=100;\ p=0,25$		
15	35	0,9852	0,9791	0,9845
20	30	0,7967	0,7518	0,7960
30	40	0,1492	0,1238	0,1492
		$n=300;\ p=0,25$		
60	70	0,9615	0,9545	0,9612
70	80	0,5366	0,4950	0,5366
80	90	0,2510	0,2297	0,2545
		$n=500\cdot\ p=0,25$		
105	145	0,9659	0,9611	0,9658
115	135	0,7219	0,6983	0,7218
135	155	0,1621	0,1499	0,1624

Значения вероятностей $\mathsf{P}\{m_1 \leqslant \mu \leqslant m_2\} = \sum_{m=m_1}^{m_2} C_n^m p^m (1-p)^{n-m}$ в схеме Бернулли.

Задачи

1. В большом городе в год рождается 20 000 детей. Считая вероятность рождения мальчика $p = 0{,}51$, найти такое число t, чтобы с вероятностью 0,99 можно было утверждать, что среди рожденных в течение года в этом городе детей число мальчиков превышает число девочек не менее чем на t.

2. Сколько надо произвести бросаний правильной монеты, чтобы с вероятностью 0,99 относительная частота выпадения герба отличалась от 1/2 не более чем на 0,01?

3. В таблице случайных чисел каждая цифра появляется независимо от других с вероятностью 1/10. Сколько надо набрать таких случайных чисел, чтобы с вероятностью 0,999 среди них появилось не менее 100 нулей?

4. В большом городе в среднем в течение одного дневного часа поступает один вызов на скорую помощь. С какой вероятностью за 3 дневных часа поступит более 10 вызовов?

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

Глава 5. ЦЕПИ МАРКОВА

§ 24. Марковская зависимость испытаний

Очень часто реальные случайные явления можно изучать с помощью следующей модели. Пусть состояние некоторой системы описывается точкой фазового пространства $E=\{e_1, e_2, \ldots, e_r\}$. В дальнейшем точки из E будем обозначать просто числами $1, 2, \ldots, r$. Предположим, что время t дискретно и принимает значения $t=0, 1, 2, \ldots, T$. Эволюция изучаемой системы описывается траекторией $\omega=(\omega_0, \omega_1, \ldots, \omega_T)$, где $\omega_t=i$, если в момент t система находится в состоянии i. В описываемом случае вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P})$ определяется пространством траекторий $\Omega=\{\omega\}$, алгеброй $\mathscr{A}$ всевозможных подмножеств Ω и вероятностью P, задаваемой элементарными вероятностями $p(\omega)$. События $A_i(t)=\{\omega\colon \omega_t=i\}$, $i=1, \ldots, r$, при каждом t определяют разбиение α_t, которое порождает алгебру событий $\mathscr{A}_t$. Исходя из принятой нами в § 11 терминологии, мы будем говорить, что

$$\mathscr{A}_0, \mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_T \tag{1}$$

есть последовательность случайных испытаний.

В § 11 мы описали модель последовательности независимых испытаний. В 1907 г. А. А. Марков ввел такой класс зависимых испытаний (1), который может служить моделью многих случайных явлений. Этот класс впоследствии изучался очень интенсивно и привел к сильно продвинутой в настоящее время теории марковских процессов. Простейшая модель марковского процесса — *цепь Маркова* — определяется следующим образом. В последовательности испытаний (1) зафиксируем какой-нибудь момент времени t. Алгебру событий $\mathscr{A}_t$ назовем *настоящим*, алгебру $\mathscr{A}_0^{t-1}$, порожденную алгебрами $\mathscr{A}_0, \mathscr{A}_1, \ldots, \mathscr{A}_{t-1}$, назовем *прошлым*, алгебру событий $\mathscr{A}_{t+1}^T$, порожденную алгебрами $\mathscr{A}_{t+1}, \ldots, \mathscr{A}_T$,— *будущим.* Любое событие из $\mathscr{A}_0^{t-1}$ также назовем *прошлым*, из $\mathscr{A}_{t+1}^T$ — *будущим*, из $\mathscr{A}_t$ — *настоящим.* Например, событие $\{\omega$: найдется такое k, что $t < k < T$ и $\omega_k = \omega_{k+1}\}$ принадлежит будущему, а событие $\{\omega$: для всех k, $0 \leqslant k < t$, $\omega_k \neq r\}$ — прошлому.

Определение 1. Последовательность испытаний (1) мы будем называть *цепью Маркова*, если при любом фиксированном настоящем $\omega_t = k$ прошлое $\mathscr{A}_0^{t-1}$ и будущее $\mathscr{A}_{t+1}^T$ независимы, т. е. для любых $1 \leqslant k \leqslant r$, $t = 1, 2, \ldots, T-1$, $A \in \mathscr{A}_0^{t-1}$, $B \in \mathscr{A}_{t+1}^T$

$$\mathsf{P}\{AB \mid \omega_t = k\} = \mathsf{P}\{A \mid \omega_t = k\}\, \mathsf{P}\{B \mid \omega_t = k\}. \qquad (2)$$

Поскольку из определения условных вероятностей следует, что для любых событий A, B, C с $\mathsf{P}(AC) > 0$

$$\frac{\mathsf{P}(AB \mid C)}{\mathsf{P}(A \mid C)} = \mathsf{P}(B \mid AC),$$

то условие (2) равносильно условию

$$\mathsf{P}\{B \mid \omega_t = k,\ A\} = \mathsf{P}\{B \mid \omega_t = k\}. \qquad (3)$$

§ 25. Переходные вероятности

Вычислим вероятность того, что траектория $\omega = (\omega_0, \omega_1, \ldots, \omega_T)$ равна $(i_0, i_1, \ldots, i_T)$. Для этого воспользуемся введенными выше обозначениями $A_i(t) = \{\omega: \omega_t = i\}$ и теоремой умножения из § 6. Имеем

$$\mathsf{P}\{\omega = (i_0, i_1, \ldots, i_T)\} = \\ = \mathsf{P}(A_{i_0}(0)) \prod_{t=1}^{T} \mathsf{P}\left(A_{i_t}(t) \mid A_{i_0}(0) A_{i_1}(1) \ldots A_{i_{t-1}}(t-1)\right). \quad (4)$$

Из условия (3) получаем для цепей Маркова

$$\mathsf{P}\left(A_{i_t}(t) \mid A_{i_0}(0) A_{i_1}(1) \ldots A_{i_{t-1}}(t-1)\right) = \\ = \mathsf{P}\left(A_{i_t}(t) \mid A_{i_{t-1}}(t-1)\right),$$

поэтому (4) запишется проще:

$$\mathsf{P}\{\omega = (i_0, i_1, \ldots, i_T)\} = \mathsf{P}(A_{i_0}(0)) \prod_{t=1}^{T} \mathsf{P}\left(A_{i_t}(t) \mid A_{i_{t-1}}(t-1)\right). \qquad (5)$$

В дальнейшем мы будем рассматривать *однородные* цепи Маркова, в которых условные вероятности

$$\mathbf{P}(A_j(t) \mid A_i(t-1)) = p_{ij},$$

называемые *переходными вероятностями*, не зависят от t. Таким образом, чтобы вычислить вероятность любой траектории ω в цепи Маркова, достаточно задать начальное распределение $p_i(0) = \mathbf{P}(A_i(0))$ и *матрицу переходных вероятностей*

$$\mathbf{P} = \begin{Vmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \cdot & \cdot & \cdot & \cdot \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{Vmatrix}. \tag{6}$$

Вероятность (5) записывается тогда так:

$$\mathbf{P}\{\omega = (i_0, i_1, \dots, i_T)\} = p_{i_0}(0) \prod_{t=1}^{T} p_{i_{t-1} i_t}. \tag{7}$$

Элементы матрицы переходных вероятностей обладают следующими свойствами:

$$p_{ij} \geqslant 0, \quad \sum_{j=1}^{r} p_{ij} = 1. \tag{8}$$

Любая квадратная матрица (6), элементы которой удовлетворяют условиям (8), называется *стохастической*.

Введем переходные вероятности за t шагов:

$$p_{ij}(t) = \mathbf{P}\{A_j(t+s) \mid A_i(s)\}.$$

Матрица

$$\mathbf{P}(t) = \| p_{ij}(t) \|$$

также будет стохастической. Переходные вероятности удовлетворяют при любых целых $t > 0$, $s > 0$ уравнению

$$p_{ij}(t+s) = \sum_{k=1}^{r} p_{ik}(t) p_{kj}(s). \tag{9}$$

Это уравнение выводится с помощью формулы полной вероятности

$$p_{ij}(t+s) = \mathbf{P}(A_j(t+s) \mid A_i(0)) =$$

$$= \sum_{k=1}^{r} \mathbf{P}(A_j(t+s) \mid A_i(0) A_k(s)) \, \mathbf{P}(A_k(s) \mid A_i(0)).$$

Так как $\mathbf{P}(A_j(t+s)\mid A_i(0)A_k(s))=\mathbf{P}(A_j(t+s)\mid A_k(s))$ в силу марковости и $\mathbf{P}(A_j(t+s)\mid A_k(s))=\mathbf{P}(A_j(t)\mid A_k(0))$ в силу однородности, то отсюда получаем (9).

Уравнения (9) можно записать в матричной форме

$$P(t+s)=P(t)P(s),$$

откуда имеем $P(t)=P^t$, где $P(1)=P$ — матрица (6). Предполагая $p_{ij}(0)=\delta_{ij}$ ($\delta_{ij}=0$, если $i\neq j$, $\delta_{ii}=1$), мы распространяем уравнения (9) на случай $t\geqslant 0$, $s\geqslant 0$.

Через начальные вероятности $p_i(0)$ и переходные вероятности $p_{ij}(t)$ мы можем выразить с помощью формулы полной вероятности распределение вероятностей $p_i(t)=\mathbf{P}(A_i(t))$ при любом t:

$$p_i(t)=\sum_{k=1}^{r} p_k(0)\,p_{ki}(t). \tag{10}$$

Пример 1. *Блуждание с поглощением.* Пусть по точкам 0, 1, 2, ..., N прямой блуждает частица. Время t дискретно. Если в момент t частица была в точке i, то в следующий момент $t+1$ она независимо от ее положений в более ранние моменты времени с вероятностью p_{ij} попадает в точку j. Если $\|p_{ij}\|$ задается равенствами $p_{00}=p_{NN}=1$, $p_{i,\,i+1}=p$, $p_{i,\,i-1}=1-p$, если $1\leqslant i\leqslant N-1$ и $p_{ij}=0$ при $|i-j|>1$, то мы получаем цепь Маркова, которая описывает блуждание частицы по целым точкам отрезка $[0,N]$ с поглощением на концах.

Пример 2. *Блуждание с отражением.* Пусть переходные вероятности $p_{i,\,i+1}$, $p_{i,\,i-1}$ для $1\leqslant i\leqslant N-1$ и p_{ij} для $|i-j|>1$ остаются теми же самыми. Если определить еще $p_{00}=1-p$, $p_{01}=p$, $p_{NN}=p$, $p_{N,\,N-1}=1-p$, то полученная цепь Маркова моделирует блуждание частицы по целым точкам отрезка $\left(-\frac{1}{2},\ N+\frac{1}{2}\right)$ с отражением на концах.

§ 26. Теорема о предельных вероятностях

Теорема 1. *Если при некотором t_0 все элементы $p_{ij}(t_0)$ матрицы $\mathbf{P}^{t_0}$ положительны, то существуют пределы*

$$\lim_{t\to\infty} p_{ij}(t)=p_j, \qquad j=1,\ \ldots,\ r. \tag{11}$$

Предельные вероятности p_j *не зависят от начального состояния* i *и являются единственным решением системы*

$$\sum_{k=1}^{r} x_k p_{kj} = x_j, \quad j = 1, \ldots, r, \quad \sum_{j=1}^{r} x_j = 1. \tag{12}$$

Доказательство. Обозначим

$$M_j(t) = \max_i p_{ij}(t), \quad m_j(t) = \min_i p_{ij}(t).$$

Так как $m_j(t) \leqslant p_{kj}(t) \leqslant M_j(t)$ при любом k, то из равенства

$$p_{ij}(t+1) = \sum_k p_{ik} p_{kj}(t)$$

следует, что при всех i

$$m_j(t) \leqslant p_{ij}(t+1) \leqslant M_j(t).$$

Отсюда вытекает

$$m_j(t) \leqslant m_j(t+1) \leqslant M_j(t+1) \leqslant M_j(t).$$

Таким образом, при $t \to \infty$ имеются пределы у последовательностей $m_j(t)$ и $M_j(t)$. Докажем, что эти пределы совпадают. Пусть i и j таковы, что $p_{ik}(t+t_0) = M_k(t+t_0)$, $p_{jk}(t+t_0) = m_k(t+t_0)$. Вычитая друг из друга равенства

$$M_k(t+t_0) = p_{ik}(t+t_0) = \sum_{l=1}^{r} p_{il}(t_0) p_{lk}(t),$$

$$m_k(t+t_0) = p_{jk}(t+t_0) = \sum_{l=1}^{r} p_{jl}(t_0) p_{lk}(t),$$

получаем

$$M_k(t+t_0) - m_k(t+t_0) = \sum_{l=1}^{r} (p_{il}(t_0) - p_{jl}(t_0)) p_{lk}(t).$$

Разобьем сумму справа $\sum$ на сумму $\sum^{+}$ положительных слагаемых и сумму $\sum^{-}$ отрицательных слагаемых. Тогда

$$M_k(t+t_0) - m_k(t+t_0) \leqslant M_k(t) \sum_l^{+} (p_{il}(t_0) - p_{jl}(t_0)) + \\ + m_k(t) \sum_l^{-} (p_{il}(t_0) - p_{jl}(t_0)). \tag{13}$$

Так как $0=\sum_{l}(p_{il}(t_0)-p_{jl}(t_0))=\sum^{+}+\sum^{-}$, то $\sum^{-}=-\sum^{+}$. Обозначим $\sum_{l}^{+}(p_{il}(t_0)-p_{jl}(t_0))=d_{ij}$. Из условий теоремы следует, что все $d_{ij}<1$, поэтому $d=\max\limits_{i,j} d_{ij}<1$. Теперь из (13) имеем

$$M_k(t+t_0)-m_k(t+t_0)\leqslant d\cdot(M_k(t)-m_k(t))$$

и

$$0\leqslant M_k(t)-m_k(t)\leqslant d^{\left[\frac{t}{t_0}\right]}\to 0$$

при $t\to\infty$. Так как $m_k(t)\leqslant p_{ik}(t)\leqslant M_k(t)$, то отсюда следует утверждение (11). Перейдем в уравнениях

$$p_{ij}(t+1)=\sum_{k=1}^{r}p_{ik}(t)\,p_{kj}$$

к пределу по $t\to\infty$. Получаем

$$p_j=\sum_{k=1}^{r}p_kp_{kj}.$$

Кроме того, $\sum\limits_{j=1}^{r}p_j=1$, т. е. предельные вероятности p_j удовлетворяют системе (12). Предположим, что какие-либо $x_1, \ldots, x_r$ удовлетворяют (12). Тогда они при любом t удовлетворяют системе

$$x_j=\sum_{k=1}^{r}x_kp_{kj}(t). \tag{14}$$

Это доказывается по индукции:

$$\sum_{k=1}^{r}x_kp_{kj}(t+1)=\sum_{k=1}^{r}x_k\sum_{l=1}^{r}p_{kl}p_{lj}(t)=$$
$$=\sum_{l=1}^{r}\left(\sum_{k=1}^{r}x_kp_{kl}\right)p_{lj}(t)=\sum_{l=1}^{r}x_lp_{lj}(t)=x_j.$$

Переходя в (14) к пределу по $t\to\infty$, получаем $x_j=\sum\limits_{k=1}^{r}x_kp_j=p_j$. Теорема доказана.

Из формулы (10) следует, что в условиях теоремы I $p_i(t) \to p_i$ при $t \to \infty$, причем предел не зависит от первоначального распределения $p_i(0)$.

Можно проверить, что цепь Маркова, описывающая блуждание с отражением в примере 2, удовлетворяет условиям теоремы. Предельные вероятности в этом случае можно найти с помощью системы уравнений (12).

Задачи

1. В урне содержится 5 шаров, белые и черные. Испытание состоит в том, что каждый раз из урны случайно вынимается один шар и взамен в урну возвращается шар, но другого цвета (вместо белого — черный и наоборот). Найти матрицу переходных вероятностей $\| p_{ij} \|$ для цепи Маркова, состояниями которой является количество белых шаров в урне. Найти вероятности перехода за два шага $p_{ij}(2)$. Найти предельные вероятности $\lim\limits_{t \to \infty} p_{ij}(t) = p_j$.

2. *Модель перемешивания колоды карт.* Пусть имеется три карточки с номерами 1, 2, 3. Состоянием системы назовем последовательность номеров этих карточек $i_1 i_2 i_3$. Предположим, что перемешивание происходит следующим образом: с вероятностями 1/2 состояние $i_1 i_2 i_3$ переходит в $i_3 i_1 i_2$ или в $i_1 i_3 i_2$. Найти матрицу вероятностей перехода. Найти предельные вероятности $p(i_1 i_2 i_3)$.

3. Полагая в примере 1 § 25 $N = 3$, вычислить вероятности перехода $p_{ij}(t)$ за t шагов и пределы $\lim\limits_{t \to \infty} p_{ij}(t) = p^*_{ij}$.

4. Пусть случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы и $\mathrm{P}\{\xi_k = 1\} = p$, $\mathrm{P}\{\xi_k = 0\} = q$, $p + q = 1$. Доказать, что пары (ξ_1, ξ_2), (ξ_2, ξ_3), ..., (ξ_{n-1}, ξ_n) образуют цепь Маркова. Найти переходные вероятности этой цепи за t шагов, $t = 1, 2, \ldots$

5. Образуют ли цепь Маркова значения случайных величин $\eta_t = \xi_{t-1}\xi_t$, где ξ_t взяты из задачи 4, если $0 < p < 1$?

Глава 6. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ (ОБЩИЙ СЛУЧАЙ)

§ 27. Случайные величины и их распределения

Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — произвольное вероятностное пространство. Теперь уже случайной величиной мы будем называть не любую числовую функцию $\xi = \xi(\omega)$.

Определение 1. Числовая функция $\xi = \xi(\omega)$ от элементарного события $\omega \in \Omega$ называется *случайной величиной*, если для любого числа x

$$\{\xi \leqslant x\} = \{\omega\colon \ \xi(\omega) \leqslant x\} \in \mathscr{A}. \tag{1}$$

Смысл этого определения состоит в следующем. Поскольку не любое подмножество Ω является событием и все события составляют σ-алгебру подмножеств $\mathscr{A}$, то естественно рассматривать такие функции $\xi = \xi(\omega)$, для которых имеет смысл говорить о вероятностях попадания ξ в достаточно простые числовые множества, в частности, в $\{\xi \leqslant x\}$. Свойство (1) гарантирует, что при любом x неравенство $\{\xi \leqslant x\}$ есть событие и, следовательно, имеет смысл говорить о его вероятности.

Определение 2. Функцию

$$F(x) = F_\xi(x) = \mathsf{P}\{\xi \leqslant x\}, \tag{2}$$

определенную при всех $x \in R$, назовем *функцией распределения* случайной величины ξ.

С помощью распределения (2) можно выразить вероятности попадания ξ в различные интервалы вида

$$x_1 \leqslant x \leqslant x_2, \quad x_1 < x < x_2, \quad x_1 < x \leqslant x_2, \quad x_1 \leqslant x < x_2. \tag{3}$$

Пусть $x_1 < x_2$. Тогда из разложения события $\{\xi \leqslant x_2\}$ на сумму несовместных событий $\{\xi \leqslant x_1\} + \{x_1 < \xi \leqslant x_2\}$ следует $\mathsf{P}(\xi \leqslant x_2) = \mathsf{P}(\xi \leqslant x_1) + \mathsf{P}\{x_1 < \xi \leqslant x_2\}$ и

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

Событие $\{\xi < x\}$ можно представить как счетную сумму несовместных событий

$$\sum_{n=1}^{\infty}\left\{x-\frac{1}{n-1}<\xi\leqslant x-\frac{1}{n}\right\},$$

откуда с помощью (4) получаем:

$$\mathsf{P}\{\xi<x\}=\sum_{n=1}^{\infty}\mathsf{P}\left\{x-\frac{1}{n-1}<\xi\leqslant x-\frac{1}{n}\right\}=F(x-1)+$$
$$+\lim_{N\to\infty}\sum_{n=2}^{N}\left(F\left(x-\frac{1}{n}\right)-F\left(x-\frac{1}{n-1}\right)\right)=$$
$$=\lim_{N\to\infty}F\left(x-\frac{1}{N}\right)=F(x-0). \quad (5)$$

Здесь и далее мы будем пользоваться обозначениями

$$F(x-0)=\lim_{y\uparrow x}F(y), \qquad F(x+0)=\lim_{y\downarrow x}F(y),$$
$$F(+\infty)=\lim_{y\to\infty}F(y), \qquad F(-\infty)=\lim_{y\to-\infty}F(y).$$

С помощью (3), (4) и (5) нетрудно уже получить остальные случаи:

$$\begin{gathered}
\mathsf{P}\{\xi=x\}=F(x)-F(x-0),\\
\mathsf{P}(x_1\leqslant\xi\leqslant x_2)=F(x_2)-F(x_1-0),\\
\mathsf{P}(x_1<\xi<x_2)=F(x_2-0)-F(x_1),\\
\mathsf{P}(x_1\leqslant\xi<x_2)=F(x_2-0)-F(x_1-0).
\end{gathered} \quad (6)$$

Теорема 1. *Функция распределения $F(x)$ обладает следующими свойствами:*

1) $F(x)$ *не убывает,*
2) $F(x)$ *непрерывна справа,*
3) $$F(+\infty)=1,$$
4) $$F(-\infty)=0. \quad (7)$$

Доказательство. Свойство 1) следует из (4). Свойство 2) следует из аксиомы непрерывности 4°. Так как события $B_n=\left\{x<\xi\leqslant x+\frac{1}{n}\right\}\downarrow\varnothing$, то $\mathsf{P}(B_n)=$ $=F\left(x+\frac{1}{n}\right)-F(x)\to 0$, т. е. $F(x+0)=F(x)$. Свойства

3) и 4) вытекают из аксиомы счетной аддитивности. Так как $\Omega = \sum_{n=-\infty}^{\infty} A_n$, где $A_n = \{\omega: n-1 < \xi(\omega) \leqslant n\}$, то

$$1 = \mathsf{P}(\Omega) = \sum_{n=-\infty}^{\infty} \mathsf{P}(A_n) = \lim_{N\to\infty} \sum_{n=-N+1}^{N} \mathsf{P}(A_n) = \lim_{N\to\infty} [F(N) - F(-N)]$$

и, следовательно, $F(\infty) = \lim_{N\to\infty} F(N) = 1$, $F(-\infty) = \lim_{N\to\infty} F(-N) = 0$. Теорема доказана.

Определение 3. σ-алгебра $\mathscr{B}$ числовых множеств, порожденная всевозможными интервалами вида $x_1 < x \leqslant x_2$, называется *борелевской*; множества B, входящие в $\mathscr{B}$, также называются *борелевскими*.

σ-алгебра борелевских множеств $\mathscr{B}$ содержит всевозможные интервалы вида (3) с конечными и бесконечными концами, их конечные и счетные суммы, все открытые и замкнутые множества. Таким образом, σ-алгебра множеств $\mathscr{B}$ достаточно богата и содержит все числовые множества, которые нам будут необходимы.

Назовем *полным прообразом* при отображении ξ числового множества B множество тех ω, для которых $\xi(\omega) \in B$. Обозначая полный прообраз B через $\xi^{-1}(B)$, имеем $\xi^{-1}(B) = \{\omega: \xi(\omega) \in B\}$. Из свойств полных прообразов

$$\xi^{-1}(\varnothing) = \varnothing, \quad \xi^{-1}(R) = \Omega \ (R \text{ — прямая}),$$

$$\xi^{-1}\Big(\bigcap_n B_n\Big) = \bigcap_n \xi^{-1}(B_n), \quad \xi^{-1}\Big(\bigcup_n B_n\Big) = \bigcup_n \xi^{-1}(B_n),$$

$$\xi^{-1}(B_1 \setminus B_2) = \xi^{-1}(B_1) - \xi^{-1}(B_2)$$

следует, что совокупность $\xi^{-1}(B)$ для всех борелевских множеств $B \in \mathscr{B}$ является σ-алгеброй событий $\mathscr{A}_\xi \subseteq \mathscr{A}$. Мы будем называть $\mathscr{A}_\xi$ σ-алгеброй, *порожденной* случайной величиной ξ. Можно установить, что A_ξ порождается множествами вида $\{\omega: \xi(\omega) \leqslant x\}$ и состоит из событий A вида

$$A = \xi^{-1}(B) = \{\omega: \xi(\omega) \in B\},$$

где $B \in \mathscr{B}$. Ниже мы покажем, что для каждого $B \in \mathscr{B}$ определена вероятность $\mathsf{P}\{\xi \in B\}$, которую мы будем обозначать $\mathsf{P}_\xi(B)$.

Определение 4. Функция $\mathsf{P}_\xi(B)$, определенная для всех $B \in \mathscr{B}$, называется *распределением вероятностей* случайной величины ξ.

С помощью (4)—(6) можно выразить вероятность события $\{\xi \in B\}$ для борелевских множеств B, представимых в виде конечной суммы интервалов вида (3).

В теории меры доказывается следующая

Теорема Каратеодори. *Если на алгебре $\mathscr{A}_0$ подмножеств Ω определена вероятность* P, *удовлетворяющая аксиомам* 1°, 2°, 3* (причем аксиома счетной аддитивности 3* формулируется так: если попарно несовместные $A_n \in \mathscr{A}_0$ и $A = \sum\limits_{n=1}^{\infty} A_n \in \mathscr{A}_0$, то $\mathsf{P}(A) = \sum\limits_{n=1}^{\infty} \mathsf{P}(A_n)$), *то эту вероятность можно однозначно продолжить на все множества из σ-алгебры $\mathscr{A}$, порожденной* $\mathscr{A}_0$[1]).

Нетрудно видеть, что числовые множества, составленные из конечных сумм полуинтервалов $(x_1, x_2]$, образуют алгебру $\mathscr{B}_0$. Эта алгебра порождает σ-алгебру борелевских множеств $\mathscr{B}$. Если задана функция распределения $F_\xi(x)$, то она удовлетворяет условиям (7). С помощью формулы (4) и аксиомы аддитивности мы можем по этой функции распределения определить значения вероятностей $\mathsf{P}_\xi(B) = \mathsf{P}\{\xi \in B\}$ для всех $B \in \mathscr{B}_0$. Можно доказать, что распределение вероятностей $\mathsf{P}_\xi(B)$ σ-аддитивно на алгебре множеств $\mathscr{B}_0$. Отсюда и из теоремы Каратеодори следует, что с помощью функции распределения (2) мы можем получить вероятность события $\{\xi \in B\}$ для любого борелевского множества $B \in \mathscr{B}$. Итак, распределение вероятностей P_ξ случайной величины ξ однозначно определяется функцией распределения F_ξ.

Таким образом, каждая случайная величина ξ дает такое отображение $\xi = \xi(\omega)$ множества Ω в числовую прямую R, которое порождает новое вероятностное пространство $(R, \mathscr{B}, \mathsf{P}_\xi)$.

---

[1]) См., например, Халмош П., Теория меры. — М.: ИЛ, 1953.

Из равенства $\mathsf{P}\{\xi=x\}=F(x)-F(x-0)$ следует, что в точках разрыва функции $F(x)$ имеет место $\mathsf{P}\{\xi=x\}>0$. Так как при каждом целом n может быть не более n точек x с $\mathsf{P}(\xi=x)\geqslant 1/n$, то у функции $F(x)$ имеется не более счетного числа точек разрыва.

Обозначим $x_1, x_2, \ldots$ все точки разрыва $F_\xi(x)$. Если вероятности $\mathsf{P}\{\xi=x_k\}=p_k$ таковы, что $\sum_{k=1}^{\infty} p_k=1$, то мы говорим, что случайная величина ξ имеет *дискретное распределение*. Примерами дискретных распределений служат:

1) *биномиальное*

$$\mathsf{P}\{\xi=k\}=C_n^k p^k(1-p)^{n-k}, \quad k=0,\ 1,\ \ldots,\ n; \quad 0<p<1;$$

2) *пуассоновское*

$$\mathsf{P}\{\xi=k\}=\frac{a^k}{k!}e^{-a}, \qquad k=0,\ 1,\ 2,\ \ldots; \quad 0<a;$$

3) *геометрическое*

$$\mathsf{P}\{\xi=k\}=p(1-p)^k, \quad k=0,\ 1,\ 2,\ \ldots,\ 0<p<1.$$

Мы будем говорить, что функция $p(x)=p_\xi(x)$ есть плотность распределения случайной величины ξ, если для любых $x_1<x_2$

$$\mathsf{P}\{x_1<\xi<x_2\}=\int_{x_1}^{x_2} p_\xi(x)\,dx. \tag{8}$$

Из определения (8) следует:

1. $F'_\xi(x)=p_\xi(x)$ в точках непрерывности $p_\xi(x)$;

2. $F_\xi(x)=\int_{-\infty}^{x} p_\xi(u)\,du$;

3. $F_\xi(x_2)-F_\xi(x_1)=\int_{x_1}^{x_2} p_\xi(u)\,du$ для любых $x_1<x_2$.

Если распределение имеет плотность $p_\xi(x)$, то мы будем говорить, что случайная величина ξ имеет *абсолютно непрерывное распределение*. Через плотность $p_\xi(x)$ можно выразить любую вероятность $\mathsf{P}\{\xi\in B\}$, если мы умеем вычислять интеграл по области B в следующей формуле:

$$\mathsf{P}\{\xi \in B\} = \int_B p_\xi(x)\,dx. \tag{9}$$

Для множеств B, равных сумме интервалов, интеграл (9) вычисляется обычным способом. Для того чтобы равенство (9) имело смысл при любом борелевском множестве B, нам нужно обобщить понятие интеграла, перейдя от интеграла Римана к интегралу Лебега (см. гл. 7).

Отметим, что существуют непрерывные функции распределения $F(x)$, не имеющие плотностей. Примером такой функции служит канторова функция $F(x)$, которую можно определить равенствами $F(x)=0$ при $x \leqslant 0$, $F(x)=1$ при $x \geqslant 1$ и

$$F(x) = \begin{cases} \frac{1}{2}F(3x) & \text{при} \quad 0 \leqslant x \leqslant \frac{1}{3}, \\ \frac{1}{2} & \text{при} \quad \frac{1}{3} \leqslant x \leqslant \frac{2}{3}, \\ \frac{1}{2} + \frac{1}{2}F(3x-2) & \text{при} \quad \frac{2}{3} \leqslant x \leqslant 1. \end{cases}$$

Непрерывные функции распределения, не имеющие плотностей, называются *сингулярными*. В общем случае любая функция распределения $F(x)$ представима в виде

$$F(x) = a_1F_1(x) + a_2F_2(x) + a_3F_3(x),$$

где $a_i \geqslant 0$, $a_1+a_2+a_3=1$, $F_1(x)$ — дискретная функция распределения, $F_2(x)$ — функция распределения, имеющая плотность (такие функции называются *абсолютно непрерывными*), $F_3(x)$ — сингулярная функция распределения.

Плотность распределения $p(x)$ обладает следующими двумя свойствами:

$$p(x) \geqslant 0, \quad \int_{-\infty}^{\infty} p(x)\,dx = 1, \tag{10}$$

которые легко устанавливаются из определения (8).

Функции от случайных величин. Пусть $g(x)$ отображает действительную прямую R в себя. Для любого

$B \subseteq R$ полный прообраз $g^{-1}(B)$ определяется как множество тех точек $x \in R$, для которых $g(x) \in B$.

Определение 5. Функцию $g(x)$ назовем *борелевской*, если для любого борелевского множества $B \in \mathscr{B}$ полный прообраз $g^{-1}(B) \in \mathscr{B}$, т. е. тоже борелевский.

К множеству борелевских функций принадлежат, в частности, непрерывные и кусочно непрерывные функции.

Теорема 2. *Если ξ — случайная величина, а $g(x)$ — борелевская функция, то $\eta = g(\xi)$ есть случайная величина.*

Доказательство. Рассмотрим $\eta = \eta(\omega)$ как сложную функцию $\eta = g(\xi(\omega))$. Пусть $B \in \mathscr{B}$. Так как $g(x)$ — борелевская функция, то $g^{-1}(B) = B_1 \in \mathscr{B}$. Так как $\eta^{-1}(B) = \xi^{-1}(B_1) \in \mathscr{A}$, то η — случайная величина.

Рассмотрим два примера вычисления функции распределения $F_\eta(x)$ и плотности $p_\eta(x)$ случайной величины $\eta = g(\xi)$ по функции распределения $F_\xi(x)$ и плотности $p_\xi(x)$.

Пример 1. Пусть функция $\eta = g(\xi)$ монотонно возрастает, $g^{-1}(x)$ — обратная функция. Тогда

$$F_\eta(x) = \mathsf{P}\{\eta \leqslant x\} = \mathsf{P}\{g(\xi) \leqslant x\} = \\ = \mathsf{P}\{\xi \leqslant g^{-1}(x)\} = F_\xi(g^{-1}(x)). \quad (11)$$

Дифференцируя (11) по x, имеем (если $g(x)$ дифференцируема и имеется плотность $p_\xi(x)$)

$$F'_\eta(x) = F'_\xi(g^{-1}(x)) \frac{dg^{-1}(x)}{dx},$$

откуда получаем соотношение между плотностями:

$$p_\eta(x) = \frac{1}{g'(g^{-1}(x))} p_\xi(g^{-1}(x)).$$

В частности, при $g(x) = x^3$ имеем

$$p_\eta(x) = \frac{1}{3x^{2/3}} p_\xi(\sqrt[3]{x}).$$

Пример 2. Пусть $g(x) = x^2$, $F_\xi(x)$ — непрерывная функция распределения с плотностью $p_\xi(x)$, $\eta = \xi^2$. При $x \geqslant 0$ из равенств

$$F_\eta(x) = \mathsf{P}\{\eta \leqslant x\} = \mathsf{P}\{\xi^2 \leqslant x\} = \\ = \mathsf{P}\{-\sqrt{x} \leqslant \xi \leqslant \sqrt{x}\} = F_\xi(\sqrt{x}) - F_\xi(-\sqrt{x})$$

получаем

$$p_\eta(x) = \frac{1}{2\sqrt{x}}\left(p_\xi(\sqrt{x}) + p_\xi(-\sqrt{x})\right).$$

Рассмотрим несколько примеров абсолютно непрерывных распределений.

1. *Нормальное* (или *гауссовское*) *распределение.*

Мы говорим, что случайная величина ξ имеет нормальное распределение с параметрами (a, σ), $-\infty < a < \infty$, $\sigma > 0$, если она имеет плотность

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Нормальное распределение с параметрами $(0, 1)$ с плотностью

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

называется *стандартным*. Плотность $p(x)$ удовлетворяет условию

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

2. *Равномерное распределение.*

Мы говорим, что случайная величина ξ имеет равномерное распределение на отрезке $[a, b]$, если ее плотность имеет вид

$$p_\xi(x) = \begin{cases} C & \text{при} \quad a \leqslant x \leqslant b, \\ 0 & \text{при} \quad x < a \quad \text{или} \quad x > b. \end{cases}$$

Из условия $\int_{-\infty}^{\infty} p(x)\,dx = C\int_a^b dx = C(b-a) = 1$ следует $C = \frac{1}{b-a}$.

3. *Гамма-распределение.*

Распределение с плотностью

$$p_\xi(x) = \begin{cases} 0, & x < 0, \\ \dfrac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, & x \geqslant 0, \end{cases}$$

где $\alpha > 0$ и $\lambda > 0$ — параметры, $\Gamma(\alpha)$ — гамма-функция, называется *гамма-распределением.*

Плотность $p_\xi(x)$ с $\alpha = 1$ называется плотностью *показательного распределения.*

§ 28. Многомерные распределения

Часто приходится рассматривать на одном и том же вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$ несколько случайных величин $\xi_1, \xi_2, \ldots, \xi_n$. Так как множества $\{\xi_k \leqslant \leqslant x_k\} \in \mathscr{A}$, т. е. являются событиями, то и их пересечение $\bigcap\limits_{k=1}^{n} \{\xi_k \leqslant x_k\} \in \mathscr{A}$. Поэтому существует вероятность этого события, которая называется *многомерной функцией распределения*

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

Многомерную функцию распределения мы будем иногда записывать просто $F(x_1, \ldots, x_n)$, не указывая индексами $\xi_1, \ldots, \xi_n$.

Обозначим $\Delta_{h_1 \ldots h_n} F(x_1, \ldots, x_n)$ разность n-го порядка по аргументам $x_1, \ldots, x_n$ с приращениями $h_1, \ldots \ldots, h_n$. Последовательно эти разности можно определить следующим образом:

$$\Delta_{h_1} F(x_1, \ldots, x_n) = \\ = F(x_1 + h_1, x_2, \ldots, x_n) - F(x_1, x_2, \ldots, x_n),$$

$$\Delta_{h_1 h_2} F(x_1, x_2, \ldots, x_n) = \Delta_{h_2}(\Delta_{h_1} F(x_1, \ldots, x_n)) = \\ = F(x_1 + h_1, x_2 + h_2, x_3, \ldots, x_n) - F(x_1 + h_1, x_2, \ldots, x_n) - \\ - F(x_1, x_2 + h_2, x_3, \ldots, x_n) + F(x_1, x_2, \ldots, x_n),$$

и т. д.

В общем случае имеет место равенство

$$\Delta_{h_1 \ldots h_n} F(x_1, \ldots, x_n) = \\ = \sum_{\theta_1, \ldots, \theta_n = 0}^{1} (-1)^{n+\theta_1+\ldots+\theta_n} F(x_1 + \theta_1 h_1, \ldots, x_n + \theta_n h_n),$$

где суммирование ведется по всем $\theta_i = 0$ и 1.

С помощью $F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)$ можно вычислить вероятность попадания в любой прямоугольник вида

$x_i < \xi_i \leqslant x_i + h_i,\ i = 1, \ldots, n$:

$$\mathsf{P}\{x_i < \xi \leqslant x_i + h_i,\ i = 1, \ldots, n\} = \Delta_{h_1 \ldots h_n} F(x_1, \ldots, x_n). \tag{12}$$

Доказательство формулы (12) можно провести последовательно:

$$\mathsf{P}\{x_1 < \xi_1 \leqslant x_1 + h_1,\ \xi_2 \leqslant x_2, \ldots, \xi_n \leqslant x_n\} = \\ = \mathsf{P}\{\xi_1 \leqslant x_1 + h_1,\ \xi_2 \leqslant x_2, \ldots, \xi_n \leqslant x_n\} - \\ - \mathsf{P}\{\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\} = \Delta_{h_1} F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n);$$

$$\mathsf{P}\{x_1 < \xi_1 \leqslant x_1 + h_1,\ x_2 < \xi_2 \leqslant x_2 + h_2,\ \xi_3 \leqslant x_3, \ldots, \xi_n \leqslant x_n\} = \\ = \mathsf{P}\{x_1 < \xi_1 \leqslant x_1 + h_1,\ \xi_2 \leqslant x_2 + h_2,\ \xi_3 \leqslant x_3, \ldots, \xi_n \leqslant x_n\} - \\ - \mathsf{P}\{x_1 < \xi_1 \leqslant x_1 + h_1,\ \xi_2 \leqslant x_2, \ldots, \xi_n \leqslant x_n\} = \\ = \Delta_{h_2}\left(\Delta_{h_1} F_{\xi_1 \ldots \xi_n}(x_1, x_2, \ldots, x_n)\right) = \\ = \Delta_{h_1 h_2} F_{\xi_1 \ldots \xi_n}(x_1, x_2, \ldots, x_n);$$

и т. д.

Из формулы (12) и из определения многомерной функции распределения $F(x_1, \ldots x_n)$ вытекают следующие свойства (которые доказываются аналогично одномерному случаю):

1) $F(x_1, x_2, \ldots, x_n)$ *по каждому аргументу не убывает и непрерывна справа*;

2) $F(-\infty, x_2, \ldots, x_n) = F(x_1, -\infty, x_3, \ldots, x_n) = \\ = F(x_1, \ldots, x_{n-1}, -\infty) = 0$;

3) $F(+\infty, +\infty, \ldots, +\infty) = 1$;

4) *при любых* $h_1 \geqslant 0, \ldots, h_n \geqslant 0$

$$\Delta_{h_1 \ldots h_n} F(x_1, \ldots, x_n) \geqslant 0.$$

Здесь, как и ранее,

$$F(-\infty, x_2, \ldots, x_n) = \lim_{x_1 \to -\infty} F(x_1, x_2, \ldots, x_n),$$

$$F(+\infty, x_2, \ldots, x_n) = \lim_{x_1 \to +\infty} F(x_1, x_2, \ldots, x_n).$$

Любая функция $F(x_1, \ldots, x_n)$, удовлетворяющая свойствам 1)—4), есть многомерная функция распределения некоторых $\xi_1, \ldots, \xi_n$. Пример функции

$$F(x_1, x_2) = \begin{cases} 0 & \text{при} \quad x_1 + x_2 < 1, \\ 1 & \text{при} \quad x_1 + x_2 \geqslant 1, \end{cases}$$

для которой выполнены свойства 1)—3) и не выполнено свойство 4) (так как $\Delta_{11}F(0,0)=F(1,1)-F(1,0)-F(0,1)+F(0,0)=-1$), показывает, что свойство 4) не вытекает из первых трех.

Из формулы (12) и свойства счетной аддитивности вероятности следует, что

$$F_{\xi_1\ldots\xi_m}(x_1,\ldots,x_m)=F_{\xi_1\ldots\xi_n}(x_1,\ldots,x_m,+\infty,\ldots,+\infty).$$

Назовем σ-алгебру множеств n-мерного пространства R^n, порожденную всевозможными n-мерными прямоугольниками вида $a_i<x_i\leqslant b_i$, $i=1,\ldots,n$, *борелевской* и будем ее обозначать $\mathscr{B}^n$. Множества из $\mathscr{B}^n$ также будем называть *борелевскими*. Как это было и в одномерном случае, многомерная функция распределения $F_{\xi_1\ldots\xi_n}(x_1,\ldots,x_n)$ позволяет нам при помощи формулы (12) вычислять вероятности событий вида $\xi\in B$, где $\xi=(\xi_1,\ldots,\xi_n)$, B — прямоугольники и конечные их суммы. Аналогично тому, как мы это делали в § 27, доказывается, что с помощью таких вероятностей однозначно определяется вероятность события $\xi\in B$ для всех $B\in\mathscr{B}^n$. События $\{\omega:\ \xi(\omega)\in B\}$, где $B\in\mathscr{B}^n$, образуют σ-подалгебру $\mathscr{A}_{\xi_1\ldots\xi_n}$ σ-алгебры $\mathscr{A}$. Мы будем называть $\mathscr{A}_{\xi_1\ldots\xi_n}$ σ-алгеброй, *порожденной* случайными величинами $\xi_1,\ldots,\xi_n$. Функция $\mathsf{P}_\xi(B)=\mathsf{P}\{\xi\in B\}$, определенная для всех $B\in\mathscr{B}^n$, называется *n-мерным распределением вероятностей* случайного вектора $\xi=(\xi_1,\ldots,\xi_n)$.

Дискретное многомерное распределение задается конечным или счетным набором значений $x=(x_1,\ldots,x_n)$ и неотрицательных $p(x)$ с $\sum\limits_x p(x)=1$. Вероятность $\mathsf{P}_\xi(B)=\mathsf{P}(\xi\in B)$ определяются в этом случае как $\sum\limits_{x\in B}p(x)$.

Другой частный случай дают распределения с плотностью. *Многомерной плотностью распределения* $p_\xi(x)$, $x=(x_1,\ldots,x_n)$ называется такая функция, что

$$\mathsf{P}_\xi(B)=\mathsf{P}\{\xi\in B\}=\int\limits_B p_\xi(x)\,dx, \tag{13}$$

где справа стоит n-мерный интеграл по области B. Интегралу справа можно придать смысл при любом $B\in$

$\in \mathscr{B}^n$, поэтому формула (13), вообще говоря, действует при всех $B \in \mathscr{B}^n$. Из определения плотности $p(x)$ следуют ее свойства:

$$p(x) \geqslant 0, \quad \int p(x)\,dx = \int_{-\infty}^{\infty} \cdots \int p(x)\,dx = 1. \tag{14}$$

Функция $p(x)$, удовлетворяющая (14), может быть плотностью некоторого распределения.

Из определения плотности вытекают следующие ее связи с функцией распределения:

$$F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\xi_1 \ldots \xi_n}(u_1, \ldots, u_n)\,du_1 \ldots du_n,$$

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

В точках непрерывности $x = (x_1, \ldots, x_n)$ плотности $p(x_1, x_2, \ldots, x_n)$ имеет место равенство

$$\mathsf{P}\{x_i < \xi_i < x_i + \Delta x_i, \quad i = 1, \ldots, n\} = p(x_1, \ldots, x_n)\,\Delta x_1 \ldots \Delta x_n + o(\Delta x_1 \ldots \Delta x_n),$$
$$\max_i \Delta x_i \to 0.$$

Примером многомерной плотности служит плотность $p(x)$ равномерного распределения на области $S \subseteq R^n$ конечного n-мерного объема $|S|$, задаваемая равенствами

$$p(x) = \begin{cases} \dfrac{1}{|S|} & \text{при} \quad x \in S, \\ 0 & \text{при} \quad x \notin S. \end{cases}$$

Вероятность $\mathsf{P}\{\xi \in B\}$ в этом случае определяется отношением объемов $B \cap S$ и S:

$$\mathsf{P}(\xi \in B) = \frac{|B \cap S|}{|S|}.$$

По этой формуле вычисляются так называемые геометрические вероятности (см. § 5).

§ 29. Независимость случайных величин

Случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называются *независимыми*, если независимы порожденные ими σ-алгебры

$$\mathscr{A}_{\xi_1},\ \mathscr{A}_{\xi_2},\ \ldots,\ \mathscr{A}_{\xi_n}.$$

Это определение эквивалентно тому, что для любых $B_i \in \mathscr{B}$

$$\mathbf{P}\{\xi_1 \in B_1, \ldots, \xi_n \in B_n\} = \prod_{i=1}^{n} \mathbf{P}\{\xi_i \in B_i\}. \qquad (15)$$

Частным случаем (15) является равенство

$$F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \ldots F_{\xi_n}(x_n), \qquad (16)$$

справедливое при всех x_i. Из (16) нетрудно установить, что при всех x_i и $h_i > 0$

$$\Delta_{h_1 \ldots h_n} F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \Delta_{h_i} F_{\xi_i}(x_i), \qquad (17)$$

что эквивалентно (15) для $B_i = (x_i, x_i + h_i]$. Как уже отмечалось выше, значения вероятности на всех интервалах однозначно определяют ее на борелевских множествах, поэтому из справедливости (17) вытекает справедливость (15) для любых $B_i \in \mathscr{B}$. Таким образом, равенства (16) или (17) можно взять за определение независимости случайных величин $\xi_1, \ldots, \xi_n$.

Если случайные величины дискретны, то из (17) следует, что за определение независимости в этом случае можно принять равенство

$$\mathbf{P}\{\xi_i = x_i,\ i = 1, \ldots, n\} = \prod_{i=1}^{n} \mathbf{P}\{\xi_i = x_i\}, \qquad (18)$$

справедливые при всех возможных x_i. Для распределений с плотностью $p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)$ за определение независимости можно взять равенство

$$p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = p_{\xi_1}(x_1) \ldots p_{\xi_n}(x_n), \qquad (19)$$

так как из (19), в силу (13), вытекает (15), а из (17) следует (19).

Если $\xi_1, \ldots, \xi_n$ независимы и $g_i(x)$ — борелевские функции, то случайные величины $g_1(\xi_1), \ldots, g_n(\xi_n)$ также независимы, так как $\mathscr{A}_{g_i(\xi_i)} \subseteq \mathscr{A}_{\xi_i}$.

Аналогично можно определить независимость векторов $\xi_i = \left(\xi_{i1}, \ldots, \xi_{ir_i}\right)$ как независимость порожденных ими σ-алгебр

$$\mathscr{A}_{\xi_1}, \mathscr{A}_{\xi_2}, \ldots, \mathscr{A}_{\xi_n},$$

где $\mathscr{A}_{\xi_i} = \mathscr{A}_{\xi_{i1}, \ldots, \xi_{ir_i}}$. Иначе это определение можно записать в виде равенства

$$\mathsf{P}\{\xi_i \in B_i, \quad i = 1, \ldots, n\} = \prod_{i=1}^{n} \mathsf{P}\{\xi_i \in B_i\},$$

справедливого для любых борелевских $B_i \in \mathscr{B}^{r_i}$. Аналогично, если $\xi_1, \ldots, \xi_n$ независимы, $g_i(x)$ — борелевские функции, отображающие R^{r_i} в R^{s_i}, то векторы $g_1(\xi_1)$, $g_2(\xi_2), \ldots, g_n(\xi_n)$ также независимы.

Формула композиции. Пусть ξ и η — независимые случайные величины, p_ξ, p_η — их плотности. Плотность совместного их распределения равна $p_{\xi\eta}(x, y) = p_\xi(x) p_\eta(y)$. Функция распределения суммы $\xi + \eta$ равна следующему интегралу:

$$F_{\xi+\eta}(z) = \mathsf{P}\{\xi + \eta \leqslant z\} = \int\limits_{x+y \leqslant z} p_\xi(x) p_\eta(y)\, dx\, dy. \tag{20}$$

Интеграл в (20) можно вычислять как повторный (для непрерывных плотностей — это факт из анализа, в общем случае — следствие теоремы Фубини, доказываемой в теории интеграла Лебега), поэтому

$$F_{\xi+\eta}(z) = \int_{-\infty}^{\infty} p_\xi(x)\, dx \int_{-\infty}^{z-x} p_\eta(y)\, dy = \int_{-\infty}^{\infty} F_\eta(z-x) p_\xi(x)\, dx =$$

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

Формулы

$$F_{\xi+\eta}(z) = \int_{-\infty}^{\infty} F_\eta(z-x) p_\xi(x)\, dx$$

и

$$p_{\xi+\eta}(z)=\int\limits_{-\infty}^{\infty} p_\xi(x)\,p_\eta(z-x)\,dx$$

носят название формул *композиции* или *свертки.* С помощью их мы выражаем плотность $p_{\xi+\eta}(z)$ и функцию распределения $F_{\xi+\eta}(z)$ суммы независимых случайных величин через плотности и функции распределения слагаемых.

Пример 3. Пусть ξ и η независимы, $F_\xi(x)$ — функция распределения ξ, а η имеет плотность

$$p_\eta(x)=\begin{cases}\dfrac{1}{b-a} & \text{для} \quad a\leqslant x\leqslant b,\\ 0 & \text{в остальных случаях.}\end{cases}$$

Применяя формулу композиции, имеем

$$F_{\xi+\eta}(z)=\int\limits_{-\infty}^{\infty} F_\xi(z-x)\,p_\eta(x)\,dx=\frac{1}{b-a}\int\limits_a^b F_\xi(z-x)\,dx,$$

откуда получаем, делая в интеграле замену $z-x=u$:

$$F_{\xi+\eta}(z)=\frac{1}{b-a}\int\limits_{z-b}^{z-a} F_\xi(u)\,du.$$

Отсюда следует существование плотности

$$p_{\xi+\eta}(z)=\frac{F_\xi(z-a)-F_\xi(z-b)}{b-a}.$$

Задачи

1. На прямоугольнике $0\leqslant x\leqslant a$, $0\leqslant y\leqslant b$ с равномерным распределением случайно берется точка (x, y). Найти функцию распределения и плотность площади ξ прямоугольника с вершинами $(0, 0)$, $(0, y)$, $(x, 0)$, (x, y).

2. На отрезке $[0, l]$ независимо друг от друга берутся две случайные точки с равномерным распределением. Найти функцию распределения $F(x)$ и плотность $p(x)$ расстояния между ними.

3. Пусть случайные величины ξ_i с функциями распределения $F_i(x)$ независимы. Найти функции распределения а) $\max\{\xi_1, \ldots, \xi_n\}$; б) $\min\{\xi_1, \ldots, \xi_n\}$.

4. Пусть случайные величины $\xi_1, \ldots, \xi_n$ независимы и одинаково распределены с функцией распределения $F(x)$ и плотностью $p(x)$. Упорядочив их по возрастанию, образуем «вариационный» ряд

$\xi_{(1)} \leqslant \xi_{(2)} \leqslant \ldots \leqslant \xi_{(n)}$. Найти плотность распределения $\xi_{(k)}$ и двумерную плотность распределения $\xi_{(k)}$ и $\xi_{(l)}$, $k < l$.

5. На прямоугольнике $0 \leqslant x \leqslant a$, $0 \leqslant y \leqslant b$ случайно с равномерным распределением берется точка. Доказать, что ее координаты (ξ, η) независимы.

6. На круге $x^2 + y^2 \leqslant R^2$ с равномерным распределением случайно берется точка. Показать, что ее координаты (ξ, η) зависимы.

7. Найти плотность распределения суммы $\xi_1 + \xi_2$ независимых случайных величин, если их плотности $p_{\xi_i}(x) = \lambda_i e^{-\lambda_i x}$, $x \geqslant 0$, $p_{\xi_i}(x) = 0$, $x < 0$.

8. Найти плотность распределения $p_n(x)$ суммы $\xi_1 + \ldots + \xi_n$ независимых случайных величин, каждая из которых имеет плотность $\lambda e^{-\lambda x}$, $x \geqslant 0$.

9. Случайные величины ξ_1, ξ_2 независимы и имеют плотность e^{-x}, $x \geqslant 0$. Найти функцию распределения $\eta = \dfrac{\xi_1}{\xi_1 + \xi_2}$.

Глава 7. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

§ 30. Определение математического ожидания

Математическое ожидание $\mathsf{M}\xi$ случайной величины $\xi = \xi(\omega)$, заданной на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$, определяется последовательно сначала для простых случайных величин, затем для неотрицательных случайных величин и, наконец, в общем случае.

I. Мы будем называть случайную величину ξ *простой*, если она представима в виде

$$\xi = \xi(\omega) = \sum_{j=1}^{m} x_j I_{A_j}(\omega), \tag{1}$$

где события $A_1, A_2, \ldots, A_m$ составляют разбиение, т. е. $A_i A_j = \varnothing$ при $i \neq j$ и $\sum_{i=1}^{m} A_i = \Omega$. Для простой случайной величины (1) $\mathsf{M}\xi$ определяется равенством

$$\mathsf{M}\xi = \sum_{j=1}^{m} x_j \mathsf{P}(A_j).$$

II. Для неотрицательной случайной величины ξ математическое ожидание определяется как предел

$$\mathsf{M}\xi = \lim_{n \to \infty} \mathsf{M}\xi_n \tag{2}$$

(конечный или бесконечный), где $\xi_n(\omega) \uparrow \xi(\omega)$ для каждого $\omega \in \Omega$, ξ_n — последовательность простых случайных величин.

III. В общем случае любая случайная величина ξ однозначно представима в виде

$$\xi = \xi^+ - \xi^-,$$

где $\xi^+ = \xi I_{\{\xi \geqslant 0\}}$, $\xi^- = |\xi| I_{\{\xi < 0\}}$. Полагаем

$$\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^-, \tag{3}$$

если правая часть равенства (3) имеет смысл, т. е. если $\mathsf{M}\xi^+$ и $\mathsf{M}\xi^-$ не равны ∞ одновременно. Если $\mathsf{M}\xi^+ = \mathsf{M}\xi^- = \infty$, то мы говорим, что $\mathsf{M}\xi$ не существует. Если $\mathsf{M}\xi^+ = \infty$, $\mathsf{M}\xi^- < \infty$, то полагаем $\mathsf{M}\xi = \infty$. Если $\mathsf{M}\xi^- = \infty$, $\mathsf{M}\xi^+ < \infty$, то полагаем $\mathsf{M}\xi = -\infty$.

Определенное выше математическое ожидание $\mathsf{M}\xi$ обладает следующими свойствами.

1°. Свойство линейности.

Пусть $\mathsf{M}\xi$, $\mathsf{M}\eta$ *и* $\mathsf{M}\xi + \mathsf{M}\eta$ *существуют и* c *— константа. Тогда*

$$\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta, \quad \mathsf{M}(c\xi) = c\mathsf{M}\xi.$$

2. Свойство положительности.

Если $\xi \geqslant 0$, *то и* $\mathsf{M}\xi \geqslant 0$. *Если* $\mathsf{M}\xi$ *и* $\mathsf{M}\eta$ *существуют и* $\xi \geqslant \eta$, *то* $\mathsf{M}\xi \geqslant \mathsf{M}\eta$.

3°. Свойство конечности.

Если $\mathsf{M}\xi$ *конечно, то и* $\mathsf{M}|\xi|$ *конечно. Если* $|\xi| \leqslant \eta$ *и* $\mathsf{M}\eta$ *конечно, то* $\mathsf{M}\xi$ *конечно. Если* $\mathsf{M}\xi$ *и* $\mathsf{M}\eta$ *конечны, то* $\mathsf{M}(\xi + \eta)$ *конечно.*

Эти свойства мы докажем ниже параллельно с доказательством корректности определення математического ожидания. Здесь лишь заметим, что $\mathsf{M}\xi$ всегда существует и конечно, когда ξ — простая, и $\mathsf{M}\xi$ существует для всех неотрицательных ξ. И, наконец, заметим, что свойство 3° вытекает из определения $\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^-$, $\mathsf{M}|\xi| = \mathsf{M}\xi^+ + \mathsf{M}\xi^-$ и из свойств 1° и 2°.

Корректность определения $\mathsf{M}\xi$. Для того чтобы данное выше определение $\mathsf{M}\xi$ было настоящим определением, нам надо убедиться в его корректности, т. е. независимости $\mathsf{M}\xi$ от представления (1) простой случайной величины ξ и независимости предела (2) от выбора последовательности простых случайных величин $\xi_n \uparrow \xi$.

I. *Простые случайные величины.*

Пусть имеется два представления одной и той же случайной величины

$$\xi = \sum_{j=1}^{m} x_j I_{A_j} = \sum_{k=1}^{n} y_k I_{B_k}, \tag{4}$$

где $\{A_j\}$ и $\{B_k\}$ — разбиения. Поскольку $A_j = \sum_{k=1}^{n} A_j B_k$ при каждом j и $B_k = \sum_{j=1}^{m} A_j B_k$ при каждом k и для

$\omega \in A_j \cap B_k$ $\xi(\omega) = x_j = y_k$, то

$$\mathsf{M}\xi = \sum_{j=1}^{m} x_j \mathsf{P}(A_j) = \sum_{j=1}^{m} \sum_{k=1}^{n} x_j \mathsf{P}(A_j B_k) =$$
$$= \sum_{k=1}^{n} \sum_{j=1}^{m} y_k \mathsf{P}(A_j B_k) = \sum_{k=1}^{n} y_k \mathsf{P}(B_k).$$

Доказательство свойств.

1°. Пусть случайные величины ξ и η представимы в виде (4). Так как $\{A_j B_k\}$, $j = 1, \ldots, m$; $k = 1, \ldots, n$, — разбиение и для $\omega \in A_j B_k$ $\xi(\omega) + \eta(\omega) = x_j + y_k$, то

$$\xi + \eta = \sum_{j=1}^{m} \sum_{k=1}^{n} (x_j + y_k) I_{A_j B_k},$$

откуда следует

$$M(\xi + \eta) = \sum_{j=1}^{m} \sum_{k=1}^{n} (x_j + y_k) \mathsf{P}(A_j B_k) = \sum_{j=1}^{m} x_j \sum_{k=1}^{n} \mathsf{P}(A_j B_k) +$$
$$+ \sum_{k=1}^{n} y_k \sum_{j=1}^{m} \mathsf{P}(A_j B_k) = \sum_{j=1}^{m} x_j \mathsf{P}(A_j) + \sum_{k=1}^{n} y_k \mathsf{P}(B_k) = \mathsf{M}\xi + \mathsf{M}\eta.$$

Если ξ представимо в виде (1), то $c\xi = \sum_{i=1}^{m} c x_i I_{A_i}$ и $\mathsf{M}(c\xi) = c\mathsf{M}\xi$.

2°. Если $\xi \geqslant 0$, то в (1) все $x_i \geqslant 0$, поэтому $\mathsf{M}\xi \geqslant 0$. Если $\xi \geqslant \eta$, то $\xi = \eta + (\xi - \eta)$ и $\mathsf{M}\xi = \mathsf{M}\eta + \mathsf{M}(\xi - \eta) \geqslant \mathsf{M}\eta$, так как из $\xi - \eta \geqslant 0$ следует $\mathsf{M}(\xi - \eta) \geqslant 0$.

II. *Неотрицательные случайные величины.*

Если $\xi \geqslant 0$, то всегда существует последовательность неотрицательных простых ξ_n таких, что $\xi_n(\omega) \uparrow \xi(\omega)$ при любом $\omega \in \Omega$. За такую последовательность можно взять, например,

$$\xi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{\left\{\frac{k-1}{2^n} < \xi \leqslant \frac{k}{2^n}\right\}}. \tag{5}$$

Нетрудно видеть, что $0 \leqslant \xi_n \leqslant \xi_{n+1} \leqslant \xi$ и при $\xi(\omega) \leqslant n$

$$\xi(\omega) \leqslant \xi_n(\omega) + \frac{1}{2^n},$$

следовательно, $\xi_n(\omega) \uparrow \xi(\omega)$ для любого $\omega \in \Omega$.

Покажем теперь, что для любых двух последовательностей $0 \leqslant \xi_n \uparrow \xi$, $0 \leqslant \eta_n \uparrow \eta$ простых случайных величин

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \lim_{n\to\infty} \mathbf{M}\eta_n. \tag{6}$$

Докажем сначала лемму.

Лемма 1. *Пусть η и ξ_n — простые неотрицательные случайные величины и $\xi_n \uparrow \xi \geqslant \eta$. Тогда*

$$\lim_{n\to\infty} \mathbf{M}\xi_n \geqslant \mathbf{M}\eta.$$

Доказательство. Пусть $\varepsilon > 0$. Обозначим $A_n = \{\omega\colon \xi_n(\omega) \geqslant \eta(\omega) - \varepsilon\}$. Тогда $\bar{A}_n \downarrow \varnothing$ при $n \to \infty$, следовательно, $\mathbf{P}(\bar{A}_n) \downarrow 0$. Далее, из очевидных неравенств

$$\xi_n \geqslant \xi_n I_{A_n} \geqslant (\eta - \varepsilon) I_{A_n} = \eta - \varepsilon I_{A_n} - \eta I_{\bar{A}_n}$$

и свойства 2° математического ожидания

$$\mathbf{M}\xi_n \geqslant \mathbf{M}\eta - \varepsilon \mathbf{P}(A_n) - c\mathbf{P}(\bar{A}_n),$$

где число c выбрано так, чтобы $\eta(\omega) \leqslant c$ при любом $\omega \in \Omega$. Имеем

$$\mathbf{M}\xi_n \geqslant \mathbf{M}\eta - \varepsilon - c\mathbf{P}(\bar{A}_n),$$

а так как $\mathbf{P}(\bar{A}_n) \downarrow 0$, то

$$\lim_{n\to\infty} \mathbf{M}\xi_n \geqslant \mathbf{M}\eta - \varepsilon$$

при любом $\varepsilon > 0$. Поскольку ε произвольно, то отсюда получаем доказываемое неравенство.

Используем теперь это неравенство для доказательства (6). Пусть $\xi_n \uparrow \xi$, $\eta_n \uparrow \xi$ — две последовательности простых случайных величин. Зафиксируем m и применим к $\xi_n \uparrow \xi \geqslant \eta_m$ лемму. Получаем $\lim\limits_{n\to\infty} \mathbf{M}\xi_n \geqslant \mathbf{M}\eta_m$, откуда следует $\lim\limits_{n\to\infty} \mathbf{M}\xi_n \geqslant \lim\limits_{n\to\infty} \mathbf{M}\eta_n$. Меняя местами ξ_n и η_n, приходим к равенству (6).

Доказательство свойств.

1°. Пусть $0 \leqslant \xi_n \uparrow \xi$, $0 \leqslant \eta_n \uparrow \eta$. Тогда $\xi_n + \eta_n \uparrow \xi + \eta$ и по определению

$$\mathbf{M}(\xi + \eta) = \lim_{n\to\infty} \mathbf{M}(\xi_n + \eta_n) = \lim_{n\to\infty} \mathbf{M}\xi_n + \lim_{n\to\infty} \mathbf{M}\eta_n = \mathbf{M}\xi + \mathbf{M}\eta.$$

Если $\xi \geqslant 0$ и $c \geqslant 0$, то из $\xi_n \uparrow \xi$ следует $c\xi_n \uparrow c\xi$ и $\mathsf{M}(c\xi) = \lim\limits_{n\to\infty} \mathsf{M}(c\xi_n) = c \lim\limits_{n\to\infty} \mathsf{M}\xi_n = c\mathsf{M}\xi$.

2°. Из $0 \leqslant \xi_n \uparrow \xi$ следует $0 \leqslant \mathsf{M}\xi_n \uparrow \mathsf{M}\xi$. Если $\xi \geqslant \eta$, то из $\xi = \eta + (\xi - \eta)$ вытекает $\mathsf{M}\xi = \mathsf{M}\eta + \mathsf{M}(\xi - \eta) \geqslant \mathsf{M}\eta$.

3°. При $\xi \geqslant 0$ имеем $|\xi| = \xi$ и $\mathsf{M}|\xi| = \mathsf{M}\xi$. Если $0 \leqslant \xi \leqslant \eta$ и $\mathsf{M}\eta < \infty$, то из $\mathsf{M}\xi \leqslant \mathsf{M}\eta$ следует $\mathsf{M}\xi < \infty$.

III. *Общий случай.*

Так как разложение $\xi = \xi^+ - \xi^-$ единственно, то математическое ожидание $\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^-$ определяется однозначно, если оно существует.

Доказательство свойств.

1°. Из $\xi = \xi^+ - \xi^-$ следует $c\xi = c\xi^+ - c\xi^-$ для $c > 0$ и $c\xi = |c|\xi^- - |c|\xi^+$ для $c < 0$. Отсюда $\mathsf{M}(c\xi) = c\mathsf{M}\xi$. Докажем теперь свойство аддитивности $\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta$. Заметим прежде всего, что из равенства $\xi = \xi_1 - \xi_2$, где $\xi_1 \geqslant 0$, $\xi_2 \geqslant 0$, следует $\xi_1 = \xi^+ + \delta$, $\xi_2 = \xi^- + \delta$, где $\delta \geqslant 0$. В самом деле, из равенства $\xi = \xi^+ - \xi^- = \xi_1 - \xi_2$ вытекает, что $\xi_1 - \xi^+ = \xi_2 - \xi^- \geqslant 0$; обозначая $\xi_1 - \xi^+ = \delta$, получаем $\xi_2 = \xi^- + \delta$. Далее, из $\xi = \xi_1 - \xi_2$ нетрудно получить $\mathsf{M}\xi = \mathsf{M}\xi_1 - \mathsf{M}\xi_2$, если $\mathsf{M}\xi_1$, $\mathsf{M}\xi_2$ конечны. Поскольку $\xi + \eta = (\xi^+ + \eta^+) - (\xi^- + \eta^-)$, то из только что доказанного равенства имеем

$$\mathsf{M}(\xi + \eta) = \mathsf{M}(\xi^+ + \eta^+) - \mathsf{M}(\xi^- + \eta^-),$$

откуда уже легко следует $\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta$. Этот вывод справедлив, когда $\mathsf{M}\xi$ и $\mathsf{M}\eta$ конечны. Случай бесконечных $\mathsf{M}\xi$ или $\mathsf{M}\eta$ легко анализируется отдельно.

2°. Докажем, что из $\xi \geqslant \eta$ и существования $\mathsf{M}\xi$ и $\mathsf{M}\eta$ следует $\mathsf{M}\xi \geqslant \mathsf{M}\eta$. Случай $\mathsf{M}\eta = -\infty$ тривиален. Предположим, что $\mathsf{M}\eta > -\infty$. Тогда в разложении $\xi = \eta + (\xi - \eta)$ можно воспользоваться аддитивностью математического ожидания $\mathsf{M}\xi = \mathsf{M}\eta + \mathsf{M}(\xi - \eta)$ и неравенством $\mathsf{M}(\xi - \eta) \geqslant 0$. Получаем $\mathsf{M}\xi \geqslant \mathsf{M}\eta$.

3°. Так как из $\xi = \xi^+ - \xi^-$ следует $|\xi| = \xi^+ + \xi^-$, то из конечности $\mathsf{M}\xi$ следует конечность $\mathsf{M}\xi^+$ и $\mathsf{M}\xi^-$. Все остальные свойства 3° проверяются просто.

Мультипликативное свойство.

Теорема 1. *Если ξ и η независимы и имеют конечные математические ожидания $\mathsf{M}\xi$ и $\mathsf{M}\eta$, то*

$$\mathsf{M}\xi\eta = \mathsf{M}\xi \cdot \mathsf{M}\eta. \tag{7}$$

Доказательство. Пусть ξ и η независимы. Если ξ и η простые и представимы в виде $\xi = \sum_{k=1}^{m} x_k I_{A_k}$, $\eta = \sum_{l=1}^{n} y_l I_{B_l}$, где $x_1 < x_2 < \ldots < x_m$, $y_1 < y_2 < \ldots < y_n$, то $\mathsf{P}(A_k B_l) = \mathsf{P}(A_k)\mathsf{P}(B_l)$. Поэтому

$$\mathsf{M}\xi\eta = \mathsf{M}\sum_{k=1}^{m}\sum_{l=1}^{n} x_k y_l I_{A_k B_l} = \sum_{k=1}^{m}\sum_{l=1}^{n} x_k y_l \mathsf{P}(A_k B_l) =$$

$$= \sum_{k=1}^{m}\sum_{l=1}^{n} x_k y_l \mathsf{P}(A_k)\mathsf{P}(B_l) = \sum_{k=1}^{m} x_k \mathsf{P}(A_k) \sum_{l=1}^{n} y_l \mathsf{P}(B_l) = \mathsf{M}\xi \cdot \mathsf{M}\eta.$$

Если неотрицательные ξ, η независимы, то простые $\xi_n = g_n(\xi)$ и $\eta_n = g_n(\eta)$, построенные по формуле (5), тоже будут независимы. Поэтому $\mathsf{M}\xi_n\eta_n = \mathsf{M}\xi_n \cdot \mathsf{M}\eta_n$. Так как $\xi_n \uparrow \xi$, $\eta_n \uparrow \eta$, то $\xi_n\eta_n \uparrow \xi\eta$ и $\mathsf{M}\xi_n\eta_n \uparrow \mathsf{M}\xi\eta$. Таким образом, равенство (7) доказано для неотрицательных ξ и η. В общем случае $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$. Так как $\xi^\pm$ и $\eta^\pm$ есть функции от ξ и η, то они независимы. Поэтому

$$\mathsf{M}(\xi^+ - \xi^-)(\eta^+ - \eta^-) = \mathsf{M}\xi^+\eta^+ - \mathsf{M}\xi^+\eta^- - \mathsf{M}\xi^-\eta^+ + \mathsf{M}\xi^-\eta^- =$$
$$= \mathsf{M}\xi^+ \cdot \mathsf{M}\eta^+ - \mathsf{M}\xi^+ \cdot \mathsf{M}\eta^- - \mathsf{M}\xi^- \cdot \mathsf{M}\eta^+ + \mathsf{M}\xi^- \cdot \mathsf{M}\eta^- =$$
$$= (\mathsf{M}\xi^+ - \mathsf{M}\xi^-)(\mathsf{M}\eta^+ - \mathsf{M}\eta^-) = \mathsf{M}\xi \cdot \mathsf{M}\eta.$$

Теорема доказана.

Следствие 1. *Если $\xi_1, \ldots, \xi_n$ независимы и имеют конечные математические ожидания, то* $\mathsf{M}\xi_1 \ldots \xi_n = \mathsf{M}\xi_1 \cdot \ldots \cdot \mathsf{M}\xi_n$.

Доказывается по индукции.

Интеграл Лебега. Данное нами определение математического ожидания есть не что иное, как интеграл Лебега от функции $\xi = \xi(\omega)$ по вероятностной мере P. Для такого интеграла используют обозначения $\int_\Omega \xi(\omega)\, d\mathsf{P}(\omega)$, $\int_\Omega \xi(\omega)\, \mathsf{P}(d\omega)$, $\int_\Omega \xi(\omega)\, d\mathsf{P}$, $\int_\Omega \xi\, d\mathsf{P}$, причем при интегрировании по всему пространству Ω иногда вместо $\int_\Omega$ пишут просто $\int$. Интеграл Лебега по множеству

$A \in \mathscr{A}$ определяется как интеграл от ξI_A, т. е.

$$\int_A \xi \, d\mathsf{P} = \int_\Omega \xi I_A \, d\mathsf{P}.$$

Рассмотрим вероятностное пространство $(R, \mathscr{B}, \mathsf{P}_\xi)$, где R — прямая, $\mathscr{B}$ — σ-алгебра борелевских множеств на ней, P_ξ — распределение вероятностей случайной величины ξ. Интеграл Лебега $\int g(x)\,d\mathsf{P}_\xi(x)$ от борелевской функции $g(x)$ иногда записывается как

$$\int g(x)\,dF_\xi(x)$$

и называется интегралом Лебега — Стилтьеса. Здесь $F_\xi(x)$ — функция распределения ξ, которая порождает вероятностную меру P_ξ

Свойства сходимости. Докажем две теоремы о переходе к пределу под знаком математического ожидания.

Теорема 2. (Теорема о монотонной сходимости.) *Если* $0 \leqslant \xi_n \uparrow \xi$, *то* $\lim\limits_{n\to\infty} \mathsf{M}\xi_n = \mathsf{M}\xi$.

Доказательство. Так как $0 \leqslant \xi_n \leqslant \xi$, то $0 \leqslant \mathsf{M}\xi_n \leqslant \mathsf{M}\xi$ и

$$\lim_{n\to\infty} \mathsf{M}\xi_n \leqslant \mathsf{M}\xi. \tag{8}$$

Введем простые случайные величины ξ_{nk}, так что $0 \leqslant \xi_{nk} \uparrow \xi_n$ при $k \to \infty$. Случайные величины $\eta_k = \max\limits_{1\leqslant n\leqslant k} \xi_{nk}$ также будут простыми. Так как

$$0 \leqslant \eta_k = \max_{1\leqslant n\leqslant k} \xi_{nk} \leqslant \max_{1\leqslant n\leqslant k+1} \xi_{n,k+1} = \eta_{k+1},$$

то последовательность η_k монотонно возрастает. Обозначим η предел $\lim\limits_{k\to\infty} \eta_k$. При каждом k $\eta_k \leqslant \xi_k$, поэтому

$$\lim_{k\to\infty} \mathsf{M}\eta_k = \mathsf{M}\eta \leqslant \lim_{k\to\infty} \mathsf{M}\xi_k. \tag{9}$$

Далее, при $n \leqslant k$ $\xi_{nk} \leqslant \eta_k \leqslant \eta$; полагая $k \to \infty$, имеем $\xi_n \leqslant \eta$ при всех n, откуда $\xi \leqslant \eta$ и $\mathsf{M}\xi \leqslant \mathsf{M}\eta$, что вместе с (8) и (9) доказывает теорему.

Следствие 2. *Если ряд* $\sum_{n=1}^{\infty} \xi_n$ *состоит из неотрицательных случайных величин, то*

$$M\left(\sum_{n=1}^{\infty} \xi_n\right) = \sum_{n=1}^{\infty} M\xi_n. \tag{10}$$

Доказательство. Последовательность $\eta_n = \sum_{k=1}^{n} \xi_k$ частных сумм удовлетворяет условиям теоремы 2, поэтому $\lim_{n\to\infty} M\eta_n = M \lim_{n\to\infty} \eta_n$, а это — другая запись равенства (10).

Следствие 3. *Если* $M\eta$ *конечно и события* $A_n \downarrow \varnothing$, *то*

$$\lim_{n\to\infty} M\eta I_{A_n} = 0. \tag{11}$$

Доказательство. Если $|M\eta| < \infty$, то $M|\eta| < \infty$. Разложим $|\eta|$ на сумму $\eta_n + \eta_n'$, где $\eta_n' = |\eta| I_{\bar{A}_n}$, $\eta_n = |\eta| I_{A_n}$. Тогда $M|\eta| = M\eta_n + M\eta_n'$ и $0 \leqslant \eta_n' \uparrow |\eta|$. По теореме 2 $\lim_{n\to\infty} M\eta_n' = M|\eta|$, поэтому $\lim_{n\to\infty} M\eta_n = 0$. Из $|M\eta I_{A_n}| \leqslant M|\eta| I_{A_n} \to 0$ вытекает (11).

Теорема 3. (Теорема Лебега о мажорируемой сходимости.) *Если* $\lim_{n\to\infty} \xi_n = \xi$ *(в каждой точке* $\omega \in \Omega$*) и* $|\xi_n| \leqslant \eta$, *где* $M\eta < \infty$, *то*

$$\lim_{n\to\infty} M\xi_n = M\xi. \tag{12}$$

Доказательство. При любом $\varepsilon > 0$ последовательность событий $A_n = \{\omega: \sup_{m \geqslant n} |\xi_m(\omega) - \xi(\omega)| < \varepsilon\}$ такова, что $\bar{A}_n \downarrow \varnothing$. В сумме $\xi_n = \xi_n I_{A_n} + \xi_n I_{\bar{A}_n}$ слагаемые оцениваются так:

$$\xi I_{A_n} - \varepsilon \leqslant \xi_n I_{A_n} \leqslant \xi I_{A_n} + \varepsilon, \quad -\eta I_{\bar{A}_n} \leqslant \xi_n I_{\bar{A}_n} \leqslant \eta I_{\bar{A}_n},$$

откуда

$$\xi - \varepsilon - \xi I_{\bar{A}_n} - \eta I_{\bar{A}_n} \leqslant \xi_n \leqslant \xi + \varepsilon + \eta I_{\bar{A}_n} - \xi I_{\bar{A}_n},$$

$$M\xi - \varepsilon - 2M\eta I_{\bar{A}_n} \leqslant M\xi_n \leqslant M\xi + \varepsilon + 2M\eta I_{\bar{A}_n}. \tag{13}$$

Переходя в (13) к пределу по $n\to\infty$ и применяя следствие 3, имеем

$$\mathsf{M}\xi-\varepsilon\leqslant\varliminf_{n\to\infty}\mathsf{M}\xi_n\leqslant\varlimsup_{n\to\infty}\mathsf{M}\xi_n\leqslant\mathsf{M}\xi+\varepsilon.$$

Поскольку $\varepsilon>0$ произвольно, отсюда получаем (12).

§ 31. Формулы для вычисления математического ожидания

Как мы уже отмечали в § 27, случайная величина $\xi=\xi(\omega)$, заданная на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$, с точки зрения ее вероятностных свойств вполне характеризуется своим распределением вероятностей P_ξ, поэтому ее можно рассматривать определенной на вероятностном пространстве $(R, \mathscr{B}, \mathsf{P}_\xi)$ функцией $\xi=\xi(x)=x$, $x\in R$. Отсюда можно сделать вывод, что математическое ожидание $\mathsf{M}\xi=\int\limits_\Omega\xi(\omega)\,\mathsf{P}(d\omega)$ на самом деле не зависит от вида функции $\xi(\omega)$, $\omega\in\Omega$, а зависит только от распределения вероятностей P_ξ. В самом деле, для неотрицательных случайных величин ξ имеем $\mathsf{M}\xi=\lim\limits_{n\to\infty}\mathsf{M}\xi_n$, где

$$\mathsf{M}\xi_n=\sum_{k=1}^{n\cdot 2^n}\frac{k-1}{2^n}\mathsf{P}\left\{\omega:\ \frac{k-1}{2^n}<\xi(\omega)\leqslant\frac{k}{2^n}\right\}. \quad (14)$$

Эту сумму можно выразить через закон распределения P_ξ:

$$\mathsf{M}\xi_n=\sum_{k=1}^{n\cdot 2^n}\frac{k-1}{2^n}\mathsf{P}_\xi\left\{\left(\frac{k-1}{2^n},\ \frac{k}{2^n}\right]\right\}. \quad (15)$$

Предел $\lim\limits_{n\to\infty}\mathsf{M}\xi_n$ в (14) мы обозначали как интеграл Лебега $\int\limits_\Omega\xi(\omega)\,d\mathsf{P}(\omega)$; тот же предел в (15) будет интегралом Лебега $\int\limits_0^\infty x\,d\mathsf{P}_\xi(x)$, который также называют интегралом Лебега — Стилтьеса и обозначают $\int\limits_0^\infty x\,dF_\xi(x)$.

Применяя то же рассуждение к ξ^+ и ξ^-, мы получаем выражение для $\xi=\xi^+-\xi^-$:

$$\mathsf{M}\xi=\int\limits_{-\infty}^{\infty} x\,dF_\xi(x), \tag{16}$$

зависящее только от распределения случайной величины ξ.

Если $\mathsf{M}\xi$ конечно, то в формуле (16) мы можем понимать правую часть как несобственный интеграл Римана — Стилтьеса (в этом случае он сходится абсолютно).

Интеграл Римана — Стилтьеса от $g(x)$ на конечном отрезке $(a, b]$ по неубывающей функции $F(x)$ с конечным изменением $F(b)-F(a)$ определяется как предел

$$\int\limits_a^b g(x)\,dF(x)=\lim_{n\to\infty}\sum_{k=0}^{n-1} g\left(x_k'\right)\left[F(x_{k+1})-F(x_k)\right],$$

где $x_k=a+\frac{b-a}{n}\cdot k$, $k=0, 1, \ldots, n$, $x_k<x_k'\leqslant x_{k+1}$. Несобственный интеграл $\int\limits_{-\infty}^{\infty} g(x)\,dF(x)$ определяется как предел $\lim\limits_{\substack{a\to-\infty\\ b\to\infty}}\int\limits_a^b g(x)\,dF(x)$. Если $F(x)$ имеет производную $p(x)$ и $F(x'')-F(x')=\int\limits_{x'}^{x''} p(u)\,du$ для всех $a\leqslant x'<x''\leqslant b$, то $\int g(x)\,dF(x)=\int g(x)\,p(x)\,dx$.

Выведем формулы, по которым вычисляются $\mathsf{M}\xi$ и $\mathsf{M}g(\xi)$ для непрерывных случайных величин.

Теорема 4. *Если случайная величина* ξ *имеет плотность* $p_\xi(x)$ *и* $\int\limits_{-\infty}^{\infty} |x| p_\xi(x)\,dx < \infty$, *то*

$$\mathsf{M}\xi = \int\limits_{-\infty}^{\infty} x p_\xi(x)\,dx. \qquad (17)$$

Доказательство. Мы будем предполагать, что плотность $p_\xi(x) = p(x)$ интегрируема по Риману и справа в (17) стоит несобственный интеграл Римана (доказательство остается справедливым и для интеграла Лебега).

Рассмотрим сначала неотрицательную случайную величину ξ с функцией распределения

$$\mathsf{P}\{\xi \leqslant x\} = F_\xi(x) = \begin{cases} 0 & \text{при} \quad x < 0, \\ F(0) & \text{при} \quad x = 0, \\ F(0) + \int\limits_0^x p(u)\,du & \text{при} \quad x > 0. \end{cases} \qquad (18)$$

Обозначим $A_k = \left\{ \frac{k-1}{2^n} < \xi \leqslant \frac{k}{2^n} \right\}$ и введем последовательность простых случайных величин

$$\xi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{A_k}.$$

Тогда $\mathsf{M}\xi = \lim\limits_{n\to\infty} \mathsf{M}\xi_n$. Имеем

$$\mathsf{M}\xi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \int\limits_{(k-1)/2^n}^{k/2^n} p(u)\,du$$

и

$$\int\limits_0^n x p(x)\,dx - \frac{1}{2^n} \leqslant \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \int\limits_{(k-1)/2^n}^{k/2^n} p(x)\,dx \leqslant \int\limits_0^\infty x p(x)\,dx.$$

Переходя в неравенствах

$$\int_0^n xp(x)\,dx - \frac{1}{2^n} \leqslant \mathsf{M}\xi_n \leqslant \int_0^\infty xp(x)\,dx$$

к пределу по $n \to \infty$, устанавливаем справедливость (17) для неотрицательных случайных величин. В общем случае $\xi = \xi^+ - \xi^-$ и ξ^+ и ξ^- имеют распределения вида (18) с плотностями (при $x > 0$) $p_{\xi_+}(x) = p(x)$ и $p_{\xi_-}(x) = p(-x)$. Имеем

$$\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^- = \int_0^\infty xp(x)\,dx - \int_0^\infty xp(-x)\,dx = \int_{-\infty}^\infty xp(x)\,dx.$$

Теорема 5. *Если ξ имеет плотность $p_\xi(x)$, функция $g(x)$ непрерывна и интеграл $\int_{-\infty}^\infty |g(x)|p_\xi(x)\,dx$ сходится, то*

$$\mathsf{M}g(\xi) = \int_{-\infty}^\infty g(x)\,p_\xi(x)\,dx. \tag{19}$$

Доказательство. Сначала рассмотрим непрерывные функции $g(x)$, равные нулю вне интервала $[a, b]$. Для каждого $n = 1, 2, \ldots$ положим $x_{nk} = a + \frac{b-a}{n}k$,

$$g_n(x) = \begin{cases} 0 & \text{при } x \leqslant a \text{ или } x > b, \\ g(x_{nk}) & \text{при } x_{n,k-1} < x \leqslant x_{nk}. \end{cases}$$

Пусть $\varepsilon > 0$. Тогда найдется такое n_0, что для всех $n \geqslant n_0$ и всех $x \in [a, b]$ справедливо неравенство $|g_n(x) - g(x)| < \varepsilon$, т. е. $g_n(x) \to g(x)$ при $n \to \infty$ равномерно по x. Кроме того, при $n \geqslant n_0$

$$|g_n(x)| \leqslant |g(x)| + \varepsilon,$$

и $g(x)$ ограничена. Применяя теорему Лебега о мажорируемой сходимости, имеем

$$\lim_{n \to \infty} \mathsf{M}g_n(\xi) = \mathsf{M}g(\xi). \tag{20}$$

С другой стороны,

$$\mathsf{M}g_n(\xi) = \sum_{k=1}^{n} g(x_{nk}) \int_{x_{n,k-1}}^{x_{nk}} p(x)\,dx = \int_a^b g_n(x)\,p(x)\,dx.$$

Отсюда и из неравенства $|g_n(x) - g(x)| \leqslant \varepsilon$ имеем при $n \geqslant n_0$

$$\left| \int_a^b g(x)\,p(x)\,dx - \mathsf{M}g_n(\xi) \right| \leqslant \varepsilon.$$

Отсюда и из (20) получаем равенство (19). Рассмотрим теперь неотрицательные $g(x) \geqslant 0$. Положим

$$g_n(x) = \begin{cases} g(x) & \text{при} \quad |x| \leqslant n, \\ 0 & \text{при} \quad |x| > n. \end{cases}$$

Случайные величины $\eta_n = g_n(\xi)$ монотонно сходятся к $\eta = g(\xi)$, поэтому по теореме о монотонной сходимости $\mathsf{M}g_n(\xi) \uparrow \mathsf{M}g(\xi)$. Отсюда и из

$$\mathsf{M}g_n(\xi) = \int_{-n}^{n} g(x)\,p(x)\,dx \to \int_{-\infty}^{\infty} g(x)\,p(x)\,dx$$

следует (19) для неотрицательных $g(x)$. В общем случае $g(x) = g^+(x) - g^-(x)$, где $g^+(x) = \max\{g(x), 0\}$, $g^-(x) = -\min\{g(x), 0\}$. Имеем

$$\mathsf{M}g(\xi) = \mathsf{M}g^+(\xi) - \mathsf{M}g^-(\xi) =$$

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

Теорема доказана.

Теоремы 4 и 5 почти так же доказываются и в случае произвольного распределения $F_\xi(x)$ с заменой (17) и (19) на интегралы Стилтьеса (Римана — Стилтьеса):

$$\mathsf{M}\xi = \int_{-\infty}^{\infty} x\,dF_\xi(x), \tag{21}$$

$$\mathsf{M}g(\xi) = \int_{-\infty}^{\infty} g(x)\,dF_\xi(x). \tag{22}$$

Равенства (21) и (22) имеют место, когда интегралы сходятся абсолютно. Для дискретных случайных величин (21) и (22) переходят в ряды

$$\mathsf{M}\xi = \sum_{k=1}^{\infty} x_k \mathsf{P}\{\xi = x_k\}, \tag{23}$$

$$\mathsf{M}g(\xi) = \sum_{k=1}^{\infty} g(x_k) \mathsf{P}\{\xi = x_k\}, \tag{24}$$

причем равенства (23) и (24) имеют место, когда ряды сходятся абсолютно.

Замечание 1. Формулы (19), (22), (24) справедливы и в более общем случае, когда борелевская функция $g(x_1, \ldots, x_m)$ отображает R^m в R^1. Пусть случайный вектор $\xi = (\xi_1, \ldots, \xi_m)$ имеет функцию распределения $F_{\xi_1 \ldots \xi_m}(x_1, \ldots, x_m)$ и плотность $p_{\xi_1 \ldots \xi_m}(x_1, \ldots, x_m)$ (если она существует). Тогда имеют место следующие формулы для вычисления математических ожиданий:

$$\mathsf{M}g(\xi_1, \ldots, \xi_m) = \int_{-\infty}^{\infty} \cdots \int g(x_1, \ldots, x_m)\, dF_{\xi_1 \ldots \xi_m}(x_1, \ldots, x_m),$$

$$\mathsf{M}g(\xi_1, \ldots, \xi_m) = \int_{-\infty}^{\infty} \cdots \int g(x_1, \ldots, x_m)\, p_{\xi_1 \ldots \xi_m}(x_1, \ldots, x_m)\, dx_1 \ldots dx_m.$$

Доказательство аналогично тому, которое проводится в случае формул (19), (22), (24).

Замечание 2. При вычислении математических ожиданий $\mathsf{M}\xi$, $\mathsf{M}g(\xi)$ очень часто используются приемы, позволяющие обходить формулы (16), (17), (19), (22)—(24), тем более, что нередки случаи, когда закон распределения либо очень сложен, либо вообще не выписывается в явном виде. Один из таких приемов состоит в том, что случайная величина ξ, математическое ожидание которой мы собираемся вычислять, представляется в виде суммы более простых случайных величин (например, индикаторов): $\xi = \theta_1 + \theta_2 + \ldots + \theta_m$ и далее используется аддитивное свойство $\mathsf{M}\xi = \mathsf{M}\theta_1 + \mathsf{M}\theta_2 + \ldots + \mathsf{M}\theta_m$. Другой прием вычисления математических ожиданий связан с использованием производящих и характеристических функций (см. гл. 8 и 9).

В § 13 мы изучали некоторые свойства математических ожиданий в конечной схеме. В этой главе мы установили, что математическое ожидание $\mathsf{M}\xi$ в общем случае обладает теми же самыми свойствами, если только предполагать в соответствующих местах существование или конечность $\mathsf{M}\xi$. Так же, как в гл. 3, в общем случае определяются моменты k-го порядка, центральные, абсолютные и абсолютные центральные моменты, в частности дисперсия, ковариация, коэффициент корреляции. Доказательства неравенств Иенсена, Коши — Буняковского, Ляпунова, Чебышева, данные в гл. 3, легко переносятся и на общий случай. Аналогично доказательство теоремы Чебышева (закон больших чисел) в § 18 дано в такой форме, которая годится и для общего случая. Мы будем в дальнейшем пользоваться этими результатами, не проводя здесь еще раз доказательств, которые были даны в гл. 3 в конечной схеме.

Вычислим $\mathsf{M}\eta$ и $\mathsf{D}\eta$ случайной величины η, распределенной нормально с параметрами $(0, 1)$:

$$\mathsf{M}\eta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}}\,dx = 0,$$

$$\mathsf{D}\eta = \mathsf{M}\eta^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\,dx = -\left.\frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\right|_{-\infty}^{\infty} + $$
$$+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = 1.$$

При вычислении $\mathsf{D}\eta$ мы воспользовались методом интегрирования по частям, полагая $v = x$, $u = -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $dv = dx$, $du = \frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Если η распределена нормально с параметрами $(0, 1)$, то $\xi = \sigma\eta + a$ имеет нормальное распределение с параметрами (a, σ) и $M\xi = a$, $D\xi = \sigma^2$. Таким образом, параметры нормального распределения a и σ равны математическому ожиданию и среднему квадратическому отклонению.

Вычислим $M\xi$ и $D\xi$ равномерного на $[a, b]$ распределения. Имеем

$$M\xi = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2},$$

$$M\xi^2 = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + b^2}{3}$$

и

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{(a-b)^2}{12}.$$

Вычислим $M\xi$ и $D\xi$ гамма-распределения:

$$M\xi = \int_0^\infty \frac{\lambda^\alpha x^\alpha}{\Gamma(\alpha)} e^{-\lambda x}\,dx = \frac{1}{\lambda\Gamma(\alpha)}\int_0^\infty u^\alpha e^{-u}\,du = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda},$$

$$M\xi^2 = \int_0^\infty \frac{\lambda^\alpha x^{\alpha+1}}{\Gamma(\alpha)} e^{-\lambda x}\,dx = \frac{1}{\lambda^2\Gamma(\alpha)}\int_0^\infty x^{\alpha+1}e^{-x}\,dx = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\lambda^2},$$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

Задачи

1. Случайная величина ξ имеет нормальное распределение с параметрами $(0, \sigma)$. Найти ее моменты $M\xi^n$.

2. Найти $Me^{-\xi}$ для случайной величины ξ в задаче 1.

3. Вычислить $M\xi^n$ при натуральном n, если ξ имеет нормальное распределение с параметрами (a, σ).

4. Случайные величины ξ_i, $i = 1, \ldots, n$, независимы, $M\xi_i = a_i$, $D\xi_i = \sigma_i^2$. Найти дисперсию $D\eta_n$, где $\eta_n = \xi_1\xi_2 \ldots \xi_n$.

5. Неотрицательные случайные величины $\xi_1, \ldots, \xi_n$ независимы и одинаково распределены. Найти математическое ожидание $M\eta_k$

случайной величины

$$\eta_k = \frac{\xi_k + \alpha}{\sum_{i=1}^{n} \xi_i + n\alpha},$$

где $\alpha > 0$ — константа.

6. Случайная величина ξ имеет Г-распределение с плотностью $\frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}$, $x \geqslant 0$. Найти $M\xi^\beta$. При каких β это математическсе ожидание конечно?

7. Случайные величины (ξ, η) — это координаты равномерно распределенной точки в круге $x^2 + y^2 \leqslant R^2$. Найти их математические ожидания и дисперсии.

Глава 8. ПРОИЗВОДЯЩИЕ ФУНКЦИИ

§ 32. Целочисленные случайные величины и их производящие функции

Дискретную случайную величину ξ, принимающую только целые неотрицательные значения, будем называть *целочисленной* случайной величиной. Закон распределения целочисленной случайной величины определяется вероятностями

$$p_n = \mathsf{P}\{\xi = n\}, \qquad n = 0, 1, 2, \ldots, \tag{1}$$

для которых

$$\sum_{n=0}^{\infty} p_n = 1. \tag{2}$$

Закон распределения (1) удобно изучать с помощью *производящей функции*, которая определяется как следующее математическое ожидание:

$$\varphi_\xi(s) = \mathsf{M}s^\xi.$$

Через закон распределения (1) производящая функция выражается суммой ряда

$$\varphi_\xi(s) = \sum_{n=0}^{\infty} p_n s^n, \tag{3}$$

который абсолютно сходится при $|s| \leqslant 1$. Поскольку

$$p_n = \frac{1}{n!}\varphi^{(n)}(0), \qquad n = 0, 1, 2, \ldots, \tag{4}$$

то между законами распределения $\{p_n\}$ и производящими функциями равенства (3) и (4) устанавливают взаимно однозначное соответствие. Определенная рядом (3) производящая функция называется иногда *вероятностной производящей функцией. Производящей функцией* любой числовой последовательности $a_0, a_1, a_2, \ldots$ называется сумма ряда

$$a_0 + a_1 s + a_2 s^2 + \ldots,$$

если он имеет ненулевой радиус сходимости. Из (2) следует, что вероятностная производящая функция $\varphi_\xi(s)$ в точке $s = 1$ равна 1.

Вычислим производящие функции распределений некоторых целочисленных случайных величин.

1) *Биномиальное распределение.*

$$\mathrm{P}\{\xi = m\} = C_n^m p^m q^{n-m}, \qquad m = 0,\ 1,\ 2,\ \ldots,\ n,\ p + q = 1,$$

$$\varphi(s) = \sum_{m=0}^{n} C_n^m p^m q^{n-m} s^m = (ps + q)^n.$$

2) *Пуассоновское распределение.*

$$\mathrm{P}\{\xi = n\} = \frac{a^n}{n!} e^{-a}, \qquad n = 0,\ 1,\ 2,\ \ldots,$$

$$\varphi(s) = \sum_{n=0}^{\infty} \frac{a^n s^n}{n!} e^{-a} = e^{a(s-1)}.$$

3) *Геометрическое распределение.*

$$\mathrm{P}\{\xi = n\} = q^n p, \qquad n = 0,\ 1,\ 2,\ \ldots,\ p + q = 1,$$

$$\varphi(s) = \sum_{n=0}^{\infty} q^n p s^n = \frac{p}{1 - qs}.$$

§ 33. Факториальные моменты

Вместо моментов $\mathrm{M}\xi^r$ в случае целочисленных случайных величин удобнее иметь дело с *факториальными моментами* $\mathrm{M}\xi^{[r]}$, где $\xi^{[r]} = \xi(\xi - 1) \ldots (\xi - r + 1)$, $\xi^{[0]} = 1$. Через факториальные моменты $\mathrm{M}\xi^{[r]}$ можно выразить моменты $\mathrm{M}\xi^r$ и наоборот. Например, первый факториальный момент есть просто математическое ожидание, а $\mathrm{M}\xi^2 = \mathrm{M}\xi^{[2]} + \mathrm{M}\xi$ и, следовательно, $\mathrm{D}\xi = \mathrm{M}\xi^{[2]} + \mathrm{M}\xi - (\mathrm{M}\xi)^2$.

Факториальные моменты легко вычисляются через производные производящих функций в точке $s = 1$.

Имеет место равенство

$$\mathsf{M}\xi^{[r]} = \varphi_\xi^{(r)}(1), \tag{5}$$

справедливое при любом целом неотрицательном r. Если ряд (3), определяющий $\varphi_\xi(s)$, сходится в какой-либо точке $s > 1$, то его можно дифференцировать почленно в $s = 1$, и мы получаем

$$\varphi_\xi^{(r)}(1) = \sum_n n^{[r]} p_n. \tag{6}$$

В противном случае мы определяем $\varphi_\xi^{(r)}(1)$ либо как $\lim\limits_{s \uparrow 1} \varphi^{(r)}(s)$, либо как левую производную в $s = 1$, определяемую предельным переходом $\varphi^{(k)}(1) = \lim\limits_{h \downarrow 0} \frac{\varphi^{(k-1)}(1) - \varphi^{(k-1)}(1-h)}{h}$ последовательно при $k = 1, \ldots, r$, $\varphi^{(0)}(s) = \varphi(s)$. В обоих случаях получаем (6). Поскольку

$$\mathsf{M}\xi^{[r]} = \sum_n n^{[r]} p_n, \tag{7}$$

то (6) и (7) доказывают (5). Заметим, что в равенстве (5) обе части могут быть бесконечными.

Таким образом, $\mathsf{M}\xi$ и $\mathsf{D}\xi$ можно следующим образом выразить через производные $\varphi_\xi(s)$:

$$\mathsf{M}\xi = \varphi_\xi'(1), \tag{8}$$

$$\mathsf{D}\xi = \varphi_\xi''(1) + \varphi_\xi'(1) - [\varphi_\xi'(1)]^2. \tag{9}$$

Вычислим с помощью (8) и (9) $\mathsf{M}\xi$ и $\mathsf{D}\xi$ биномиального, пуассоновского и геометрического распределений.

1) *Биномиальное распределение.*

$$\varphi_\xi'(s) = np(ps+q)^{n-1}, \quad \varphi_\xi''(s) = n(n-1)p^2(ps+q)^{n-2},$$

$$\mathsf{M}\xi = np, \quad \mathsf{D}\xi = n(n-1)p^2 + np - n^2p^2 = npq.$$

2) *Пуассоновское распределение.*

$$\varphi_\xi'(s) = ae^{a(s-1)}, \quad \varphi_\xi''(s) = a^2e^{a(s-1)},$$

$$\mathsf{M}\xi = a, \quad \mathsf{D}\xi = a^2 + a - a^2 = a.$$

3) *Геометрическое распределение.*

$$\varphi'_\xi(s)=\frac{pq}{(1-qs)^2},\quad \varphi''_\xi(s)=\frac{2pq}{(1-qs)^3},$$

$$\mathsf{M}\xi=\frac{q}{p},\quad \mathsf{D}\xi=\frac{2q^2}{p^2}+\frac{q}{p}-\frac{q^2}{p^2}=\frac{q}{p^2}.$$

Многомерные производящие функции. Аналогичным образом можно определить многомерные производящие функции. Пусть $\xi=(\xi_1, \ldots, \xi_r)$ — случайный вектор с целочисленными неотрицательными компонентами ξ_i. Обозначим

$$p_\alpha=\mathsf{P}\{\xi=\alpha\},$$

где $\alpha=(\alpha_1, \ldots, \alpha_r)$ — возможные значения вектора ξ. *Многомерной производящей функцией* называется

$$\varphi_\xi(s_1, \ldots, s_r)=\mathsf{M}s_1^{\xi_1}s_2^{\xi_2}\ldots s_r^{\xi_r}=\sum_\alpha p_\alpha s_1^{\alpha_1}\ldots s_r^{\alpha_r}.$$

Она обладает свойствами, аналогичными свойствам одномерных производящих функций. В частности, с помощью производных $\varphi_\xi(s_1, \ldots, s_r)$ вычисляются смешанные факториальные моменты

$$\mathsf{M}\xi_1^{[k_1]}\xi_2^{[k_2]}\ldots\xi_r^{[k_r]}=$$
$$=\left.\frac{\partial^{k_1+k_2+\ldots+k_r}\varphi_\xi(s_1, s_2, \ldots, s_r)}{\partial s_1^{k_1}\partial s_2^{k_2}\ldots\partial s_r^{k_r}}\right|_{s_1=s_2=\ldots=s_r=1}.$$

Пример. Полиномиальное распределение

$$\mathsf{P}\{\xi=\alpha\}=\frac{n!}{\alpha_1!\ldots\alpha_r!}p_1^{\alpha_1}\ldots p_r^{\alpha_r},\quad \alpha_1+\ldots+\alpha_r=n,$$
$$p_1+\ldots+p_r=1,$$

имеет производящую функцию

$$\varphi_\xi(s_1, \ldots, s_r)=(p_1s_1+\ldots+p_rs_r)^n.$$

§ 34. Мультипликативное свойство

Теорема 1. *Если $\xi_1, \xi_2, \ldots, \xi_n$ — независимые целочисленные случайные величины, $\varphi_{\xi_k}(s)$, $k=1, \ldots, n$, — их производящие функции, то*

$$\varphi_{\xi_1+\ldots+\xi_n}(s)=\prod_{k=1}^n\varphi_{\xi_k}(s). \tag{10}$$

Доказательство. Из независимости $\xi_1, \xi_2, \ldots$ $\ldots, \xi_n$ следует независимость $s^{\xi_1}, s^{\xi_2}, \ldots, s^{\xi_n}$. Из мультипликативного свойства математического ожидания имеем равенство

$$\mathsf{M}s^{\xi_1+\ldots+\xi_n} = \mathsf{M}s^{\xi_1} \ldots s^{\xi_n} = \prod_{k=1}^{n} \mathsf{M}s^{\xi_k},$$

равносильное (10).

Если целочисленные ξ и η независимы и $p_n = \mathsf{P}\{\xi = n\}$, $q_n = \mathsf{P}\{\eta = n\}$, то распределение их суммы $r_n = \mathsf{P}\{\xi + \eta = n\}$ по формуле полной вероятности определяется равенством

$$r_n = \sum_{k=0}^{n} \mathsf{P}\{\xi = k\} \mathsf{P}\{\eta = n - k\} = \sum_{k=0}^{n} p_k q_{n-k}. \quad (11)$$

Распределение $\{r_n\}$ называется *композицией* или *сверткой* распределений $\{p_n\}$ и $\{q_n\}$. Теорема 1 позволяет нам иногда с помощью производящих функций находить свертку распределений, не прибегая к формулам (11). Например, из равенства

$$(ps+q)^{n_1}(ps+q)^{n_2} = (ps+q)^{n_1+n_2}$$

вытекает, что свертка двух биномиальных распределений с одинаковыми p и разными числами испытаний n_1 и n_2 дает опять биномиальное распределение с тем же самым p и числом испытаний $n_1 + n_2$. Аналогично, из равенства

$$e^{a_1(s-1)} \cdot e^{a_2(s-1)} = e^{(a_1+a_2)(s-1)}$$

следует, что композиция двух пуассоновских законов с параметрами a_1 и a_2 дает опять пуассоновский закон с параметром $a_1 + a_2$. Этим свойством пуассоновских распределений мы пользовались в § 20.

Распределение с производящей функцией $\frac{ps}{1-qs}$ можно интерпретировать как число испытаний в схеме Бернулли до первого успеха. Обозначим в этой схеме ξ_r число испытаний до r-го успеха включительно. Случайная величина ξ_r представима в виде суммы $\xi_r = \tau_1 + \tau_2 + \ldots + \tau_r$, где τ_i независимы, одинаково распределены и имеют производящие функции $\varphi_{\tau_i}(s) = \frac{ps}{1-qs}$ (τ_1 — число испытаний до первого успеха включительно,

τ_2 — число испытаний от первого успеха до второго успеха и т. д.). По свойству мультипликативности имеем

$$\varphi_{\xi_r}(s) = \frac{p^r s^r}{(1-qs)^r}. \tag{12}$$

Разлагая (12) в ряд, получаем

$$\varphi_{\xi_r}(s) = p^r s^r \sum_{\alpha=0}^{\infty} \binom{-r}{\alpha} \cdot (-1)^\alpha s^\alpha q^\alpha =$$

$$= p^r s^r \sum_{\alpha=0}^{\infty} \frac{-r(-r-1)\dots(-r-\alpha+1)\cdot(-1)^\alpha}{\alpha!} q^\alpha s^\alpha =$$

$$= p^r s^r \sum_{\alpha=0}^{\infty} \frac{(r+\alpha-1)(r+\alpha-2)\dots r}{\alpha!} s^\alpha q^\alpha = p^r s^r \sum_{\alpha=0}^{\infty} C_{r+\alpha-1}^{\alpha} s^\alpha q^\alpha,$$

откуда

$$\mathsf{P}\{\xi_r = n\} = C_{n-1}^{r-1} p^r q^{n-r}, \qquad n = r,\ r+1, \dots$$

Сумма случайного числа случайных величин. Пусть $\xi_1, \xi_2, \dots$ — последовательность целочисленных независимых одинаково распределенных случайных величин с производящей функцией $\varphi_\xi(s)$ и ν — независимая от них целочисленная случайная величина с производящей функцией $\varphi_\nu(s)$. Определим сумму случайного числа случайных величин равенствами $\zeta_\nu = \xi_1 + \xi_2 + \dots + \xi_\nu$ при $\nu \geqslant 1$, $\zeta_0 = 0$.

Теорема 2. *Производящая функция* $\varphi_{\zeta_\nu}(s)$ *равна суперпозиции*

$$\varphi_{\zeta_\nu}(s) = \varphi_\nu(\varphi_\xi(s)). \tag{13}$$

Доказательство. Вычислим $\varphi_{\zeta_\nu}(s) = \mathsf{M}s^{\zeta_\nu}$ с помощью условных математических ожиданий, используя равенство

$$\mathsf{M}\{s^{\xi_1+\dots+\xi_\nu} \mid \nu = n\} = \mathsf{M}s^{\xi_1+\dots+\xi_n} = [\varphi_\xi(s)]^n.$$

Получаем

$$\varphi_{\zeta_\nu}(s) = \mathsf{M}s^{\zeta_\nu} = \mathsf{M}[\mathsf{M}(s^{\zeta_\nu} \mid \nu)] = \mathsf{M}[\varphi_\xi(s)]^\nu = \varphi_\nu(\varphi_\xi(s)),$$

что и требовалось доказать.

С помощью (13), (8) и (9) вычислим математическое ожидание и дисперсию ζ_ν:

$$\varphi'_{\zeta_\nu}(s) = \varphi'_\nu(\varphi_\xi(s))\,\varphi'_\xi(s), \quad \varphi'_{\zeta_\nu}(1) = \varphi'_\nu(1)\,\varphi'_\xi(1),$$

$$\varphi''_{\zeta_\nu}(s) = \varphi''_\nu(\varphi_\xi(s))\,[\varphi'_\xi(s)]^2 + \varphi'_\nu(\varphi_\xi(s))\,\varphi''_\xi(s),$$

$$\varphi''_{\zeta_\nu}(1) = \varphi''_\nu(1)\cdot[\varphi'_\xi(1)]^2 + \varphi'_\nu(1)\,\varphi''_\xi(1),$$

$$\mathsf{M}\zeta_\nu = \mathsf{M}\nu\cdot\mathsf{M}\xi,$$

$$\mathsf{D}\zeta_\nu = \varphi''_\nu(1)\cdot[\varphi'_\xi(1)]^2 + \varphi'_\nu(1)\,\varphi''_\xi(1) + \mathsf{M}\zeta_\nu - (\mathsf{M}\zeta_\nu)^2 =$$
$$= (\mathsf{M}\nu^2 - \mathsf{M}\nu)\cdot(\mathsf{M}\xi)^2 + \mathsf{M}\nu\cdot(\mathsf{M}\xi^2 - \mathsf{M}\xi) + \mathsf{M}\nu\cdot\mathsf{M}\xi -$$
$$- (\mathsf{M}\nu)^2\cdot(\mathsf{M}\xi)^2 = \mathsf{D}\nu\cdot(\mathsf{M}\xi)^2 + \mathsf{M}\nu\cdot\mathsf{D}\xi.$$

§ 35. Теорема непрерывности

Докажем, что соответствие между законами распределения $\{p_n\}$ и производящими функциями (3) не только взаимно однозначно, но и взаимно непрерывно.

Теорема 3. *Пусть* $\varphi_r(s) = \sum_{n=0}^{\infty} p_n^{(r)} s^n$, $r = 1, 2, \ldots,$ — *последовательность вероятностных производящих функций,* $\varphi(s) = \sum_{n=0}^{\infty} p_n s^n$ — *производящая функция последовательности* $\{p_n\}$. *Для того чтобы при каждом n* $\lim_{r\to\infty} p_n^{(r)} = p_n$, *необходимо и достаточно, чтобы при всех* $0 \leqslant s < 1$

$$\lim_{r\to\infty} \varphi_r(s) = \varphi(s).$$

Доказательство. Предположим, что $\lim_{r\to\infty} p_n^{(r)} = p_n$. Пусть $\varepsilon > 0$ и $0 \leqslant s < 1$. В правой части неравенства

$$|\varphi_r(s) - \varphi(s)| \leqslant \sum_{k=0}^{\infty} |p_k^{(r)} - p_k|\cdot s^k \leqslant$$
$$\leqslant \sum_{k=0}^{N-1} |p_k^{(r)} - p_k| + \sum_{k=N}^{\infty} s^k = \sum_{k=0}^{N-1} |p_k^{(r)} - p_k| + \frac{s^N}{1-s}$$

выберем N таким, чтобы $s^N/(1-s)<\varepsilon/2$, а затем выберем r_0 таким, чтобы $\sum_{k=0}^{N-1}|p_k^{(r)}-p_k|<\varepsilon/2$ при $r\geqslant r_0$. Тогда при тех же $r\geqslant r_0$ имеем $|\varphi_r(s)-\varphi(s)|<\varepsilon$, что и доказывает необходимость. Докажем теперь достаточность. Из ограниченной последовательности $0\leqslant p_0^{(r)}\leqslant 1$ выбираем сходящуюся подпоследовательность $p_0^{(r1)}\to p_0$. Из ограниченной последовательности $0\leqslant p_1^{(r1)}\leqslant 1$ выбираем сходящуюся подпоследовательность $p_1^{(r2)}\to p_1$ и т. д. Из последовательностей

$$\begin{array}{l} p_n^{(1,1)},\ p^{(2,1)},\ p_n^{(3,1)},\ \ldots \\ p_n^{(1,2)},\ p_n^{(2,2)},\ p_n^{(3,2)},\ \ldots \\ p_n^{(1,3)},\ p_n^{(2,3)},\ p_n^{(3,3)},\ \ldots \\ \ldots\ldots\ldots\ldots\ldots \end{array}$$

выбираем диагональную сходящуюся подпоследовательность $p_n^{(r,r)}$, которая сходится к p_n при любом n. Предположим, что хотя бы при одном n последовательность $p_n^{(r)}$ не сходится к p_n. Тогда можно выбрать две сходящиеся к разным пределам подпоследовательности $p_n^{(r')}\to p_n^*$, $p_n^{(r'')}\to p_n^{**}$. По первой части теоремы $\varphi_{r'}(s)\to\varphi^*(s)=\sum_n p_n^* s^n$ и $\varphi_{r''}(s)\to\varphi^{**}(s)=\sum_n p_n^{**}s^n$. Так как по условию $\varphi_n(s)\to\varphi(s)$, то $\varphi^*(s)=\varphi^{**}(s)=\varphi(s)$ и $p_n^*=p_n^{**}=p_n$, т. е. $\lim_{r\to\infty}p_n^{(r)}=p_n$.

Замечание. Как показывает пример $\varphi_r(s)=s^r\to 0\equiv\varphi(s)$, $0\leqslant s<1$, предельные величины p_n могут не образовывать распределение вероятностей, так как, вообще говоря, $\sum_{n=0}^{\infty}p_n\leqslant 1$. Если потребовать, чтобы $\lim_{s\uparrow 1}\varphi(s)=1$, то $\sum_{n=0}^{\infty}p_n=1$ и в пределе мы получаем распределение вероятностей $\{p_n\}$.

Применим теорему 3 к доказательству предельной теоремы Пуассона (см. § 20):

$$\lim_{n\to\infty}C_n^m\left(\frac{a}{n}\right)^m\left(1-\frac{a}{n}\right)^{n-m}=\frac{a^m}{m!}e^{-a}. \tag{14}$$

Производящая функция биномиального распределения для $p=\frac{a}{n}$ равна

$$\left(s\frac{a}{n}+1-\frac{a}{n}\right)^n=\left(1+\frac{a}{n}(s-1)\right)^n.$$

Из равенства

$$\lim_{n\to\infty}\left(1+\frac{a}{n}(s-1)\right)^n=e^{a(s-1)}$$

по теореме 3 вытекает (14), что и доказывает предельную теорему Пуассона.

§ 36. Ветвящиеся процессы

Проиллюстрируем применение аппарата производящих функций на примере *ветвящихся процессов*. Пусть имеются некоторые однотипные частицы, которые размножаются независимо друг от друга. Пусть p_n — вероятность того, что одна частица превращается в n частиц, $\varphi(s)=\sum_{n=0}^{\infty}p_ns^n$ — производящая функция распределения вероятностей $\{p_n\}$. Обозначим $\mu(t)$ — число частиц в t-м поколении и $\varphi_t(s)=\mathsf{M}s^{\mu(t)}$ — производящую функцию $\mu(t)$. Предположим, что $\mu(0)=1$. Тогда $\varphi_1(s)=\varphi(s)$. Пусть $\xi_{t1}, \xi_{t2}, \ldots, \xi_{tn}, \ldots$ — независимые случайные величины с распределением, определяемым производящей функцией $\varphi(s)$. Тогда число частиц $\mu(t+1)$ в $(t+1)$-м поколении, согласно нашему определению, есть сумма $\xi_{t1}+\xi_{t2}+\ldots+\xi_{t,\mu(t)}$ случайного числа независимых случайных слагаемых (ξ_{tk} — это число потомков k-й частицы t-го поколения). По теореме 2 отсюда вытекает, что

$$\varphi_{t+1}(s)=\varphi_t(\varphi(s)), \tag{15}$$

т. е. $\varphi_2(s)=\varphi(\varphi(s))$, $\varphi_3(s)=\varphi(\varphi(\varphi(s)))$ и $\varphi_t(s)$ есть t-я итерация функции $\varphi(s)$. Соотношение (15) позволяет нам вычислить $\mathsf{M}\mu(t)=A(t)$. Обозначим $\varphi'(1)=A$. Продифференцируем (15) по s в точке 1. Получим $A(t+1)=A(t)\cdot A$, откуда

$$A(t)=A^t. \tag{16}$$

Поведение ветвящегося процесса существенно определяется значением параметра A — средним числом непосредственных потомков одной частицы. Из (16) мы видим, что при $t \to \infty$

$$A(t) \to 0, \quad \text{если} \quad A < 1,$$
$$A(t) \to \infty, \quad \text{если} \quad A > 1,$$
$$A(t) \equiv 1, \quad \text{если} \quad A = 1.$$

Назовем ветвящийся процесс *докритическим*, *надкритическим* или *критическим*, если соответственно $A < 1$, $A > 1$ или $A = 1$.

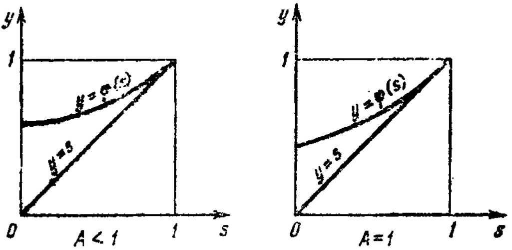

Рис. 11. Графики производящих функций $\varphi(s)$ докритического и критического ветвящихся процессов.

Если $\mu(t) = 0$, то мы будем говорить, что ветвящийся процесс *выродился к моменту времени* t. Вероятность этого события равна

$$\mathbf{P}\{\mu(t) = 0\} = \varphi_t(0).$$

Так как $\{\mu(t) = 0\} \subseteq \{\mu(t+1) = 0\}$, то $\mathbf{P}\{\mu(t) = 0\}$ не убывает и при $t \to \infty$ имеет предел

$$\lim_{t \to \infty} \mathbf{P}\{\mu(t) = 0\} = q,$$

который мы назовем *вероятностью вырождения*. Предельная вероятность q — это вероятность того, что процесс выродится в каком-либо поколении. Предположим, что $\varphi(s) \not\equiv s$. Докажем следующую теорему.

Теорема 4. *Для того чтобы $q < 1$, необходимо и достаточно, чтобы процесс был надкритическим.*

Доказательство. Соотношение (15) можно записать иначе:

$$\varphi_{t+1}(s) = \varphi(\varphi_t(s)). \tag{17}$$

Подставляя в (17) $s = 0$, имеем

$$\varphi_{t+1}(0) = \varphi(\varphi_t(0)). \qquad (18)$$

Переходя в (18) к пределу по $t \to \infty$, имеем

$$q = \varphi(q),$$

так как $q = \lim\limits_{t\to\infty} \varphi_t(0)$. Таким образом, q есть решение уравнения

$$s = \varphi(s). \qquad (19)$$

Это уравнение всегда имеет решение $s = 1$. Если других решений в $[0, 1]$ нет, то отсюда следует, что $q = 1$. При $A \leqslant 1$ других решений уравнения (19) нет, так как при всех $0 \leqslant s < 1$ выполнено неравенство $s < \varphi(s)$ (см. рис. 11). Действительно, $1 - \varphi(s) = \varphi'(\theta)(1 - s)$, где $s < \theta < 1$, поэтому из $\varphi'(\theta) < 1$ вытекает $1 - \varphi(s) < 1 - s$. При $0 < s < 1$ вторая производная $\varphi''(s) > 0$, поэтому уравнение $s = \varphi(s)$ не может иметь более двух корней в $[0, 1]$. Так как $\varphi(0) \geqslant 0$ и при $A > 1$ существуют $s_1 < 1$, для которых $\varphi(s_1) < s_1$, то найдется корень $0 \leqslant s_0 < 1$ уравнения (19) (рис. 12). Докажем, что в этом случае $q = s_0$. В самом деле, нетрудно установить, что $\varphi(s) > s$ при $0 < s < s_0$ и $\varphi(s) < s$ при $s_0 < s < 1$. Так как $\varphi_{t+1}(0) > \varphi_t(0)$, то из (18) вытекает, что

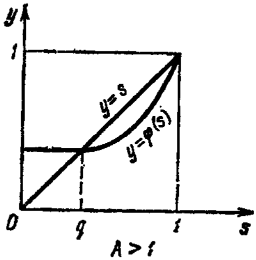

Рис. 12. График производящей функции $\varphi(s)$ надкритического ветвящегося процесса.

$$\varphi_t(0) < \varphi(\varphi_t(0))$$

при любом t, следовательно, $\varphi_t(0) < s_0$ при всех t и $q = \lim\limits_{t\to\infty} \varphi_t(0) \leqslant s_0 < 1$. Таким образом, вероятность q не может быть равна 1, а так как она есть корень уравнения (19), то $q = s_0 < 1$. Теорема доказана.

Задачи

1. Найти производящую функцию равномерного распределения, сосредоточенного в точках $0, 1, 2, \ldots, N - 1$. С помощью производных производящей функции вычислить математическое ожидание и дисперсию этого распределения.

2. Функция $1-\sqrt{1-s}$ есть вероятностная производящая функция. Найти соответствующее распределение. Что можно сказать о его математическом ожидании?

3. Дана производящая функция $\varphi(s)=\sum\limits_n \mathsf{P}\{\xi=n\}s^n$ случайной величины ξ. Найти производящую функцию $A(s)=\sum\limits_{n=0}^{\infty} a_n s^n$ для вероятностей $a_n=\mathsf{P}\{\xi>n\}$.

4. Пусть число потомков одной частицы в ветвящемся процессе определяется производящей функцией

$$\varphi(s)=1-\frac{A(1-s)}{\frac{B}{2A}(1-s)+1}.$$

Найти ее t-ю итерацию $\varphi(s)$. Найти вероятиость вырождения q.

Глава 9. ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ

§ 37. Определение и простейшие свойства характеристических функций

В предыдущей главе мы познакомились с аналитическим аппаратом производящих функций, который успешно используется в задачах исследования свойств распределений целочисленных случайных величин. В общем случае аналогичную роль играют характеристические функции. Для их определения нам нужно понятие математического ожидания распространить на комплексные случайные величины. Пусть $\zeta = \xi + i\eta$, где ξ и η — пара действительных случайных величин, у которых существуют и конечны $\mathsf{M}\xi$ и $\mathsf{M}\eta$. Тогда математическое ожидание комплексной случайной величины определим как сумму

$$\mathsf{M}\zeta = \mathsf{M}\xi + i\mathsf{M}\eta. \tag{1}$$

Основные свойства математического ожидания (например, свойства аддитивности и мультипликативности) естественным образом переносятся на случай (1). Остановимся лишь на доказательстве неравенства

$$|\mathsf{M}\zeta| \leqslant \mathsf{M}\,|\zeta|. \tag{2}$$

Если случайная величина ζ простая, т. е. принимает лишь конечное число значений $\zeta = z_k = x_k + iy_k$, причем $\mathsf{P}\{\zeta = z_k\} = p_k$, то (2) есть простое следствие свойства модуля комплексных величин:

$$|\mathsf{M}\zeta| = \left|\sum_k z_k p_k\right| \leqslant \sum_k |z_k|\,p_k = \mathsf{M}\zeta. \tag{3}$$

Пусть $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$, а $\xi_n^\pm$, $\eta_n^\pm$ — последовательности простых случайных величин, для которых $\xi_n^\pm \uparrow \xi^\pm$, $\eta_n^\pm \uparrow \eta^\pm$ и, значит, $\xi_n = \xi_n^+ - \xi_n^- \to \xi$, $\eta_n = \eta_n^+ - \eta_n^- \to \eta$. Тогда $\zeta_n \to \zeta$, и по определению $\mathsf{M}\xi$, $\mathsf{M}\eta$

имеем

$$\mathsf{M}\zeta = \lim_{n\to\infty} \mathsf{M}\zeta_n,$$

где $\zeta_n = \xi_n + i\eta_n$. В силу (3) при любом n

$$|\mathsf{M}\zeta_n| \leqslant \mathsf{M}|\zeta_n|.$$

Покажем, что $\lim\limits_{n\to\infty} \mathsf{M}|\zeta_n| = \mathsf{M}|\zeta|$. В самом деле, из

$$|\zeta_n| \leqslant |\xi_n| + |\eta_n| = \xi_n^+ + \xi_n^- + \eta_n^+ + \eta_n^- \leqslant$$
$$\leqslant \xi^+ + \xi^- + \eta^+ + \eta^- = |\xi| + |\eta|$$

и $\zeta_n \to \zeta$ по теореме Лебега о мажорируемой сходимости вытекает $\mathsf{M}|\zeta_n| \to \mathsf{M}|\zeta|$, так как мы рассматриваем лишь случай $\mathsf{M}|\xi| < \infty$, $\mathsf{M}|\eta| < \infty$.

Определение 1. *Характеристической функцией* случайной величины ξ мы будем называть функцию $f_\xi(t)$ от действительного аргумента t, равную

$$f_\xi(t) = \mathsf{M}e^{it\xi}. \tag{4}$$

Раскрывая в (4) $e^{i\varphi}$ по формуле Эйлера $e^{i\varphi} = \cos\varphi + i\sin\varphi$, мы имеем

$$f_\xi(t) = \mathsf{M}\cos\xi t + i\mathsf{M}\sin\xi t. \tag{5}$$

Иногда мы будем вместо $f_\xi(t)$ писать просто $f(t)$. Если $F_\xi(x)$ есть функция распределения ξ, а $p_\xi(x)$ — ее плотность (если она существует), то по общим формулам вычисления математического ожидания имеем:

$$f_\xi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF_\xi(x), \quad f_\xi(t) = \int_{-\infty}^{\infty} e^{itx} p_\xi(x)\, dx. \tag{6}$$

Если распределение ξ дискретно, то

$$f_\xi(t) = \sum_k e^{itx_k} \mathsf{P}\{\xi = x_k\}. \tag{7}$$

Из (6) и (7) видно, что характеристическая функция $f_\xi(t)$ вполне определяется функцией распределения $F_\xi(x)$ случайной величины ξ.

Перечислим несколько простейших свойств характеристических функций.

1) $|f(t)| \leqslant 1$ *при каждом действительном* t; $f(0) = 1$.

Доказательство просто получается из неравенства (2), так как $|e^{it\xi}| = 1$ и $|f(t)| = |\mathsf{M}e^{it\xi}| \leqslant \mathsf{M}|e^{it\xi}| = 1$.

2) $f(t)$ *равномерно непрерывна по* t.

Для доказательства этого свойства установим сначала справедливость следующей леммы, которая нам понадобится далее.

Лемма 1. *При действительных* φ *и любом целом* $n \geqslant 1$ *имеют место неравенства*

$$\left| e^{i\varphi} - \sum_{k=0}^{n-1} \frac{(i\varphi)^k}{k!} \right| \leqslant \frac{|\varphi|^n}{n!}. \tag{8}$$

Доказательство. Поскольку $|e^{i\varphi}| = 1$, то $\left| \int_0^{\varphi} e^{iu}\,du \right| = |e^{i\varphi} - 1| \leqslant \int_0^{|\varphi|} du = |\varphi|$. Далее доказываем (8) по индукции. Пусть (8) справедливо при некотором n. Тогда, так как

$$\int_0^{\varphi} \left(e^{iu} - \sum_{k=0}^{n-1} \frac{(iu)^k}{k!} \right) du = \frac{1}{i} \left(e^{i\varphi} - \sum_{k=0}^{n} \frac{(i\varphi)^k}{k!} \right),$$

то

$$\left| e^{i\varphi} - \sum_{k=0}^{n} \frac{(i\varphi)^k}{k!} \right| = \left| \int_0^{\varphi} \left(e^{iu} - \sum_{k=0}^{n-1} \frac{(iu)^k}{k!} \right) du \right| \leqslant$$
$$\leqslant \int_0^{|\varphi|} \frac{u^n}{n!}\,du = \frac{|\varphi|^{n+1}}{(n+1)!}.$$

Для доказательства 2) рассмотрим событие $A = \{|\xi| \leqslant X\}$ и в правой части неравенства

$$|f(t+h) - f(t)| = |\mathsf{M}e^{it\xi}(e^{ih\xi} - 1)| \leqslant$$
$$\leqslant \mathsf{M}|e^{ih\xi} - 1|\, I_A + \mathsf{M}|e^{ih\xi} - 1|\, I_{\bar{A}},$$

где I_A и $I_{\bar{A}}$ — индикаторы событий A и $\bar{A}$, применим неравенство (8) при $n = 1$ при оценке первого слагаемого и $|e^{it\xi} - 1| \leqslant 2$ при оценке второго слагаемого.

Тогда

$$|f(t+h)-f(t)|\leqslant|h|\cdot \mathsf{M}|\xi|I_A+2\mathsf{M}I_{\overline{A}}\leqslant$$
$$\leqslant|h|X+2\mathsf{P}\{|\xi|>X\}.$$

Пусть $\varepsilon>0$. Выбирая сначала X таким, чтобы $\mathsf{P}\{|\xi|>>X\}<\frac{\varepsilon}{4}$, а затем полагая $\delta=\frac{\varepsilon}{2X}$, получаем, что $|f(t+h)-f(t)|<\varepsilon$ при $|h|<\delta$.

3) *Если* $\eta=a\xi+b$, *где* a *и* b — *константы, то* $f_\eta(t)=e^{itb}f_\xi(at)$.

Легко вытекает из определения:

$$f_\eta(t)=\mathsf{M}e^{it\eta}=\mathsf{M}e^{it(a\xi+b)}=e^{itb}\mathsf{M}e^{ita\xi}=e^{itb}f_\xi(at).$$

4) *Если* $\xi_1, \xi_2, \ldots, \xi_n$ *независимы, то*

$$f_{\xi_1+\ldots+\xi_n}(t)=\prod_{k=1}^{n}f_{\xi_k}(t). \tag{9}$$

Из независимости $\xi_1, \xi_2, \ldots, \xi_n$ следует независимость $e^{it\xi_1}, e^{it\xi_2}, \ldots, e^{it\xi_n}$; применяя к ним свойство мультипликативности математического ожидания, получаем

$$f_{\xi_1+\ldots+\xi_n}(t)=\mathsf{M}e^{it\sum\limits_{k=1}^{n}\xi_k}=\mathsf{M}\prod_{k=1}^{n}e^{it\xi_k}=\prod_{k=1}^{n}\mathsf{M}e^{it\xi_k}.$$

5) $f_\xi(-t)=\overline{f}_\xi(t)$.

Вытекает из $\overline{e^{it\xi}}=e^{-it\xi}$ и свойства 3).

6) *Обозначим* $m_n=\mathsf{M}\xi^n$. *Если* m_n конечно, *то существуют все производные* $f^{(k)}(t)$ *с* $k\leqslant n$ *и*

$$f^{(k)}(0)=i^k\mathsf{M}\xi^k. \tag{10}$$

Кроме того, имеет место разложение

$$f(t)=\sum_{k=0}^{n}\frac{(it)^k}{k!}m_k+R_n(t), \tag{11}$$

где $R_n(t)=o(t^n)$ *при* $t\to\infty$.

Доказательство. Если мы k раз формально продифференцируем (4), то получим равенство

$$f_\xi^{(k)}(t) = i^k \mathsf{M}\xi^k e^{it\xi} = i^k \int_{-\infty}^{\infty} x^k e^{itx} dF_\xi(x). \qquad (12)$$

Полагая в (12) $t = 0$, приходим к (10). Для обоснования законности дифференцирования под знаком математического ожидания в (12) рассуждаем по индукции. Пусть формула (12) справедлива при $k < n$. Поскольку

$$\frac{f^{(k)}(t+h) - f^{(k)}(t)}{h} = i^k \mathsf{M}\xi^k \frac{e^{it\xi}(e^{ih\xi} - 1)}{h}, \qquad (13)$$

$$\left| \xi^k e^{it\xi} \frac{(e^{ih\xi} - 1)}{h} \right| \leqslant |\xi|^{k+1} \quad \text{и} \quad \mathsf{M}|\xi|^{k+1} < \infty,$$

то в правой части (13) по теореме Лебега о мажорируемой сходимости можно перейти к пределу по $h \to 0$. Таким образом мы доказываем справедливость (12) при $k+1$. Для оценки остаточного члена $R_n(t)$ в (11) применим лемму 1 к разности

$$|R_n(t)| = \left| \mathsf{M}\left[e^{it\xi} - \sum_{k=0}^{n} \frac{(it\xi)^k}{k!} \right] \right| \leqslant$$

$$\leqslant \mathsf{M}\left| e^{it\xi} - \sum_{k=0}^{n} \frac{(it\xi)^k}{k!} \right| \leqslant \frac{|t|^{n+1}}{(n+1)!} \mathsf{M}|\xi|^{n+1} I_A +$$

$$+ 2\mathsf{M} \frac{|t|^n |\xi|^n}{n!} I_{\bar{A}}, \qquad (14)$$

где A — событие, введенное в 2) (здесь в первом слагаемом мы воспользовались неравенством (8), а во втором — неравенством

$$\left| e^{i\varphi} - \sum_{k=0}^{n} \frac{(i\varphi)^k}{k!} \right| \leqslant \left| e^{i\varphi} - \sum_{k=0}^{n-1} \frac{(i\varphi)^k}{k!} \right| + \frac{|\varphi|^n}{n!} \leqslant 2\frac{|\varphi|^n}{n!} \Bigg).$$

Так как $I_A = 1$ при $|\xi| \leqslant X$, то из (14) получаем

$$|R_n(t)| \leqslant \frac{|t|^{n+1} X^{n+1}}{(n+1)!} + \frac{2|t|^n}{n!} \mathsf{M}|\xi|^n I_{\bar{A}}.$$

Пусть $\varepsilon>0$. Выбираем сначала X таким, чтобы $\mathsf{M}|\xi|^n I_{\bar{A}}<\frac{\varepsilon}{4}$, а затем $\delta=\frac{(n+1)\varepsilon}{2X}$. Тогда при $|t|<\delta$ имеем $|R_n(t)|\leqslant\frac{|t|^n}{n!}\varepsilon$, что и требовалось.

7) *Если* $\varphi_\xi(s)=\mathsf{M}s^\xi$ — *производящая функция целочисленной случайной величины, то*

$$f_\xi(t)=\varphi_\xi(e^{it}). \tag{15}$$

Следует из определения.

Вычисление характеристических функций некоторых законов распределений.

1) С помощью формулы (15) получаем характеристические функции следующих распределений целочисленных случайных величин:

а) *Биномиальное распределение*

$$\mathsf{P}\{\xi=m\}=C_n^m p^m(1-p)^{n-m},\qquad m=0,\ 1,\ \ldots,\ n,$$
$$f_\xi(t)=(pe^{it}+1-p)^n.$$

б) *Пуассоновское распределение*

$$\mathsf{P}\{\xi=n\}=\frac{a^n}{n!}e^{-a},\qquad n=0,\ 1,\ 2,\ \ldots,$$
$$f_\xi(t)=\exp(a(e^{it}-1)).$$

в) *Геометрическое распределение*

$$\mathsf{P}\{\xi=n\}=pq^n,\qquad n=0,\ 1,\ 2,\ \ldots,\ q=1-p,$$
$$f_\xi(t)=\frac{p}{1-qe^{it}}.$$

2) *Вырожденное распределение* $\mathsf{P}\{\xi=C\}=1$,

$$f_\xi(t)=e^{itC}$$

3) *Нормальное распределение.* Если случайная величина ξ имеет нормальное распределение с параметрами $(0, 1)$, то

$$f(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{itx-x^2/2}\,dx. \tag{16}$$

Дифференцируя равенство (16) по t, получаем

$$f'(t)=\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{itx-x^2/2}\,dx.$$

Интегрируя по частям, приходим к дифференциальному уравнению

$$f'(t)=\frac{i}{\sqrt{2\pi}}\left[-e^{itx-x^2/2}\Big|_{-\infty}^{\infty}+it\int_{-\infty}^{\infty} e^{itx-x^2/2}\,dx\right]=-tf(t).$$

Решая это уравнение с начальным условием $f(0)=1$, получаем

$$f(t)=e^{-t^2/2}.$$

В общем случае нормального распределения с параметрами (a, σ) имеем, согласно свойству 3):

$$f(t)=e^{ita-\sigma^2t^2/2}. \tag{17}$$

4) *Равномерное на* (a, b) *распределение*

$$f(t)=\frac{1}{b-a}\int_a^b e^{itx}\,dx=\frac{e^{itb}-e^{ita}}{it(b-a)}. \tag{18}$$

Отметим частные случаи (18). При $a=-l, b=l$ имеем

$$f(t)=\frac{e^{itl}-e^{-itl}}{2itl}=\frac{\sin tl}{tl}. \tag{19}$$

При $a=0$, $b=L$ имеем

$$f(t)=\frac{e^{itL}-1}{itL}.$$

5) *Гамма-распределение с плотностью*

$$p_\alpha(x)=\frac{x^{\alpha-1}}{\Gamma(\alpha)}e^{-x}, \qquad x\geqslant 0.$$

Обозначим $f_\alpha(t)$ характеристическую функцию, соответствующую $p_\alpha(x)$. Поскольку $p_{\alpha+\beta}(x)$ есть свертка $p_\alpha(x)$, и $p_\beta(x)$:

$$p_{\alpha+\beta}(x)=\int_0^x p_\beta(x-y)\,p_\alpha(y)\,dy=\frac{e^{-x}}{\Gamma(\alpha)\,\Gamma(\beta)}\int_0^x y^{\alpha-1}(x-y)^{\beta-1}\,dy=$$

$$=\frac{x^{\alpha+\beta-1}e^{-x}}{\Gamma(\alpha)\,\Gamma(\beta)}\int_0^1 z^{\alpha-1}(1-z)^{\beta-1}\,dz=\frac{x^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)}e^{-x},\qquad x\geqslant 0,$$

то, в силу 4), $f_{\alpha+\beta}(t)=f_\alpha(t)\,f_\beta(t)$. Вычислим сначала $f_1(x)=\int_0^\infty e^{itx}p_1(x)\,dx=\int_0^\infty e^{itx-x}\,dx$. Интегрируя по частям, получаем

$$f_1(t)=\int_0^\infty e^{itx-x}\,dx=-e^{itx-x}\Big|_0^\infty+it\int_0^\infty e^{itx-x}\,dx=1+itf_1(t)$$

и

$$f_1(t)=\frac{1}{1-it}. \tag{20}$$

Из (20) для любого целого n имеем

$$f_n(t)=\frac{1}{(1-it)^n}.$$

Из $f_1(t)=[f_{1/n}(t)]^n$ получаем $f_{1/n}(t)=(1-it)^{-1/n}$, и, далее, $f_{m/n}(t)=[f_{1/n}(t)]^m=(1-it)^{-m/n}$. Таким образом, для любого рационального $\alpha>0$

$$f_\alpha(t)=(1-it)^{-\alpha}. \tag{21}$$

Так как плотность $p_\alpha(x)$ непрерывно зависит от α и, как мы увидим в § 39, из $p_{\alpha_n}(x)\to p_\alpha(x)$ следует $f_{\alpha_n}(t)\to f_\alpha(t)$, формула (21) справедлива для всех $\alpha>0$. Заметим, что при дробных α из многозначной функции (21) выделяется однозначная ветвь, для которой $f_\alpha(0)=1$.

§ 38. Формулы обращения для характеристических функций

В § 37 мы установили, что каждой функции распределения $F_\xi(x)$ соответствует характеристическая функция $f_\xi(t)$. Пусть существует непрерывная плотность $p_\xi(x)$. Тогда характеристическая функция вычисляется по формуле

$$f_\xi(t)=\int_{-\infty}^{\infty} e^{itx}p_\xi(x)\,dx, \tag{22}$$

т. е. $f_\xi(t)$ есть преобразование Фурье функции $p_\xi(x)$. В анализе доказывается, что при $f_\xi(t)\in L_1$, т. е. при конечности интеграла $\int_{-\infty}^{\infty}|f_\xi(t)|\,dt$, имеет место формула обращения для преобразования Фурье (22):

$$p_\xi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}f_\xi(t)\,dt. \tag{23}$$

Исходя из этой формулы, мы докажем формулу обращения в общем случае. Сначала докажем несколько вспомогательных утверждений.

Лемма 2. *Пусть ξ и η независимы. Если ξ имеет функцию распределения $F(x)$, а η равномерно распределена на отрезке $[a,b]$, то существует плотность $p_{\xi+\eta}(x)$, которая выражается формулой*

$$p_{\xi+\eta}(x)=\frac{F(x-a)-F(x-b)}{b-a}.$$

Доказательство. По формуле композиции

$$F_{\xi+\eta}(x)=\int_{-\infty}^{\infty}F_\xi(x-y)\,p_\eta(y)\,dy=\frac{1}{b-a}\int_a^b F(x-y)\,dy=$$

$$=\frac{1}{b-a}\int_{x-b}^{x-a}F(z)\,dz. \tag{24}$$

Исходя из (24), мы можем для любых $x_1<x_2$ записать

$$F_{\xi+\eta}(x_2)-F_{\xi+\eta}(x_1)=\frac{1}{b-a}\left[\int_{x_2-b}^{x_2-a}F(z)\,dz-\int_{x_1-b}^{x_1-a}F(z)\,dz\right]=$$

$$=\int_{x_1}^{x_2}\frac{F(x-a)-F(x-b)}{b-a}\,dx,$$

откуда и следует утверждение.

Замечание. Если η равномерно распределена на $[-l, l]$, то

$$p_{\xi+\eta}(x) = \frac{F(x+l) - F(x-l)}{2l}.$$

Лемма 3. *Пусть ξ и η независимы, ξ имеет ограниченную плотность $p_\xi(x) = p(x)$ и η имеет плотность $p_\eta(x)$. Обозначим $p_\theta(x)$ плотность суммы $\xi + \theta\eta$, где θ — параметр. Тогда в точках непрерывности $p(x)$ имеет место равенство*

$$\lim_{\theta \to 0} p_\theta(x) = p(x).$$

Доказательство. По формуле композиции имеем

$$p_\theta(x) = \int_{-\infty}^{\infty} p(x-y)\, p_\eta\left(\frac{y}{\theta}\right) \frac{dy}{\theta},$$

откуда

$$p_\theta(x) - p(x) = \int_{-\infty}^{\infty} [p(x-y) - p(x)]\, p_\eta\left(\frac{y}{\theta}\right) \frac{dy}{\theta}$$

и

$$|p_\theta(x) - p(x)| \leqslant \int_{|y| \leqslant \delta} |p(x-y) - p(x)|\, p_\eta\left(\frac{y}{\theta}\right) \frac{dy}{\theta} + \int_{|y| > \delta} |p(x-y) - p(x)|\, p_\eta\left(\frac{y}{\theta}\right) \frac{dy}{\theta}. \qquad (25)$$

Пусть x — точка непрерывности $p(x)$. Зафиксируем любое $\varepsilon > 0$. Тогда можно выбрать такое $\delta > 0$, что при $|y| \leqslant \delta$ выполняется неравенство $|p(x-y) - p(x)| \leqslant \varepsilon/2$. Так как плотность $p(x)$ ограничена, то существует такая константа $C < \infty$, что $p(x) \leqslant C$.

Тогда из (25) следует

$$|p_\theta(x) - p(x)| < \frac{\varepsilon}{2} \int_{|y| \leqslant \delta} p_\eta\left(\frac{y}{\theta}\right) \frac{dy}{\theta} + C\mathsf{P}\left\{|\eta| \geqslant \frac{\delta}{\theta}\right\}.$$

Выберем $\theta_0 > 0$ так, чтобы $\mathsf{P}\left\{|\eta| \geqslant \frac{\delta}{\theta_0}\right\} < \frac{\varepsilon}{2C}$. Тогда при всех $|\theta| \leqslant \theta_0$ имеет место неравенство $|p_\theta(x) - p(x)| < \varepsilon$.

Формулу обращения в общем случае дает

Теорема 1. *Пусть $f_\xi(t)$ — характеристическая функция и $F_\xi(x)$ — соответствующая функция распределения. Тогда, если точки $x+l$ и $x-l$ являются точками непрерывности функции $F_\xi(x)$, то*

$$F_\xi(x+l)-F_\xi(x-l)=\lim_{\sigma\to 0}\frac{1}{\pi}\int_{-\infty}^{\infty} e^{-ixt} f_\xi(t)\frac{\sin tl}{t}e^{-\frac{\sigma^2t^2}{2}}\,dt. \tag{26}$$

Доказательство. Пусть случайные величины ξ, η, ζ независимы, ξ имеет функцию распределения $F_\xi(x)$, η имеет равномерное распределение на интервале $(-1, 1)$, ζ имеет нормальное распределение с параметрами $(0, 1)$. Тогда по лемме 2 $\xi+l\eta$ имеет плотность

$$p(x)=\frac{F_\xi(x+l)-F_\xi(x-l)}{2l},$$

а $\xi+l\eta+\sigma\zeta$ имеет характеристическую функцию

$$f_\xi(t)\frac{\sin tl}{tl}e^{-\frac{\sigma^2t^2}{2}}\in L_1,$$

поэтому ее плотность $p_\sigma(x)$ выражается по формуле обращения (23):

$$p_\sigma(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt} f_\xi(t)\frac{\sin tl}{tl}e^{-\frac{\sigma^2t^2}{2}}\,dt. \tag{27}$$

По лемме 3

$$\lim_{\sigma\to 0} p_\sigma(x)=\frac{F_\xi(x+l)-F_\xi(x-l)}{2l}, \tag{28}$$

если $x+l$ и $x-l$ — точки непрерывности $F_\xi(x)$. Переходя к пределу в (27) и пользуясь равенством (28), получаем (26).

Теорема 2. *Каждой характеристической функции $f_\xi(t)$ соответствует только одна функция распределения $F_\xi(x)$.*

Доказательство. В формуле (26) разность $F_\xi(x_2)-F_\xi(x_1)$ для точек $x_2=x+l$ и $x_1=x-l$ непрерывности $F_\xi(x)$ однозначно определяется по $f_\xi(t)$. Полагая в разности $F_\xi(x_2)-F_\xi(x_1)$ $x_1\to-\infty$ по точкам непрерывности x_1, мы однозначно определяем $F_\xi(x_2)$

в точках непрерывности x_2, а так как в любой точке

$$F_\xi(x) = \lim_{x_2 \downarrow x} F(x_2),$$

где предел берется по точкам непрерывности x_2, то $F_\xi(x)$ однозначно определяется $f_\xi(t)$. Теорема доказана.

§ 39. Теорема о непрерывном соответствии между множеством характеристических функций и множеством функций распределения

В § 38 мы установили, что между множеством функций распределения $\{F_\xi(x)\}$ и множеством их характеристических функций $\{f_\xi(t)\}$ имеется взаимно однозначное соответствие. Покажем, что это соответствие не только взаимно однозначное, но и взаимно непрерывное.

Определение 2. Мы будем говорить, что последовательность функций распределения $F_n(x)$ *слабо сходится* к $F(x)$, и писать

$$F_n(x) \Rightarrow F(x),$$

если $F_n(x) \to F(x)$ в каждой точке непрерывности предельной функции.

Если $F_n(x)$ — функция распределения ξ_n, $F(x)$ — функция распределения ξ, то мы будем также иногда говорить, что ξ_n слабо сходится к ξ, и обозначать $\xi_n \Rightarrow \xi$; иногда мы будем говорить, что ξ_n *сходится к* ξ *по распределению*. Из слабой сходимости $\xi_n \Rightarrow \xi$ следует, что

$$\mathsf{P}\{x_1 < \xi_n \leqslant x_2\} \to \mathsf{P}\{x_1 < \xi \leqslant x_2\}, \quad n \to \infty,$$

если только $\mathsf{P}\{\xi = x_1\} = \mathsf{P}\{\xi = x_2\} = 0$. Пример $\mathsf{P}\left\{\xi_n = \frac{1}{n}\right\} = 1$, $\mathsf{P}\{\xi = 0\} = 1$ показывает, что из $\xi_n \Rightarrow \xi$ не вытекает сходимость $F_{\xi_n}(x) \to F_\xi(x)$ в каждой точке, так как $F_{\xi_n}(0) = 0$ и $F_\xi(0) = 1$.

Одно из самых важных свойств характеристических функций содержится в следующих двух теоремах. Пусть $F_n(x)$, $F(x)$ — функции распределения, $f_n(t)$, $f(t)$ — соответствующие им характеристические функции.

Теорема 3. (Прямая предельная теорема.) *Если $F_n(x) \Rightarrow F(x)$, то $f_n(t) \to f(t)$ в каждой точке t.*

Теорема 4. (Обратная предельная теорема.) *Если $f_n(t)$ сходится в каждой точке t к некоторой функции $f(t)$, непрерывной в нуле, то $F_n(x) \Rightarrow F(x)$ и $f(t)$ есть характеристическая функция распределения $F(x)$.*

Доказательство этих теорем будет следовать из леммы и двух теорем Хелли.

Лемма 4. *Если $F_n(x) \to F(x)$ на всюду плотном на прямой множестве D, то $F_n(x) \Rightarrow F(x)$.*

Доказательство. Пусть x — точка непрерывности $F(x)$, $x', x'' \in D$ и $x' < x < x''$. Имеем

$$F_n(x') \leqslant F_n(x) \leqslant F_n(x''),$$

и

$$F(x') = \lim_{n\to\infty} F_n(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant$$
$$\leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant \lim_{n\to\infty} F_n(x'') = F(x''). \qquad (29)$$

Так как $F(x') \leqslant F(x) \leqslant F(x'')$ и разность $F(x'') - F(x')$ может быть сделана как угодно малой, то из (29) следует $\lim\limits_{n\to\infty} F_n(x) = F(x)$, что и требовалось доказать.

Теорема 5. (Первая теорема Хелли.) *Из всякой последовательности функций распределения $\{F_n\}$ можно выбрать слабо сходящуюся подпоследовательность.*

Доказательство. Пусть $D = \{x_k\}$ — всюду плотное на прямой счетное множество. Из ограниченной последовательности $0 \leqslant F_n(x_1) \leqslant 1$ выбираем сходящуюся подпоследовательность $F_{1n}(x_1)$, предел которой обозначим $F(x_1)$. Из ограниченной последовательности $0 \leqslant \leqslant F_{1n}(x_2) \leqslant 1$ выбираем сходящуюся подпоследовательность $F_{2n}(x_2) \to F(x_2)$ и т. д. Далее выбираем диагональную подпоследовательность $F_{nn}(x)$, для которой $F_{nn}(x_k) \to F(x_k)$ для любой точки $x_k \in D$. По лемме 4 отсюда вытекает $F_{nn}(x) \Rightarrow F(x)$.

Замечание. $F(x)$ может не быть функцией распределения. Например, если $F_n(x) = 0$ при $x < n$ и $F_n(x) = 1$ при $x \geqslant n$, то $F_n(x) \Rightarrow F(x) \equiv 0$.

Теорема 6. (Вторая теорема Хелли.) *Если $g(x)$ — непрерывная ограниченная функция на прямой и $F_n(x) \Rightarrow \Rightarrow F(x)$, $F(\infty) - F(-\infty) = 1$, то*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x)\, dF_n(x) = \int_{-\infty}^{\infty} g(x)\, dF(x). \qquad (30)$$

Доказательство. Пусть $a < b$ — точки непрерывности $F(x)$. Докажем сначала, что

$$\lim_{n\to\infty}\int_a^b g(x)\,dF_n(x)=\int_a^b g(x)\,dF(x). \qquad (31)$$

Пусть $\varepsilon > 0$. Разделим $[a, b]$ точками непрерывности $a = x_0, x_1, \ldots, x_{N-1}, x_N = b$ функции $F(x)$ на такие отрезки $[x_{k-1}, x_k]$, что $|g(x)-g(x_k)| < \varepsilon$ для точек $x \in [x_{k-1}, x_k]$. Это сделать можно, так как $g(x)$ равномерно непрерывна на $[a, b]$, а точки непрерывности $F(x)$ расположены всюду плотно. Определим ступенчатую функцию

$$g_\varepsilon(x) = g(x_k) \text{ на } x \in (x_{k-1}, x_k],$$

для которой $|g_\varepsilon(x)-g(x)| \leqslant \varepsilon$ на $x \in [a, b]$. Тогда

$$\left|\int_a^b g(x)\,dF_n(x)-\int_a^b g(x)\,dF(x)\right| \leqslant$$

$$\leqslant \int_a^b |g(x)-g_\varepsilon(x)|\,dF_n(x)+$$

$$+\left|\int_a^b g_\varepsilon dF_n-\int_a^b g_\varepsilon dF\right|+\int_a^b |g(x)-g_\varepsilon(x)|\,dF(x) \leqslant$$

$$\leqslant 2\varepsilon + M\left[\sum_{k=1}^{N}\left[F_n(x_k)-F(x_k)-(F_n(x_{k-1})-F(x_{k-1}))\right]\right],$$

где $M=\sup\limits_x |g(x)|$. При $n\to\infty$ последнее слагаемое может быть сделано как угодно малым, откуда и следует (31). Для доказательства (30) выберем $X > 0$ таким, чтобы $F(-X) < \varepsilon/4$ и $1-F(X) < \varepsilon/4$ и чтобы точки $\pm X$ были точками непрерывности $F(x)$. Тогда, так как $F_n(\pm X)\to F(\pm X)$, можно выбрать n_0 таким, что при $n \geqslant n_0$ $F_n(-X) < \varepsilon/2$ и $1-F_n(X) \leqslant \varepsilon/2$. Оценим разность

$$\left| \int_{-\infty}^{\infty} g(x)\,dF_n(x) - \int_{-\infty}^{\infty} g(x)\,dF(x) \right| \leqslant$$

$$\leqslant \left| \int_{-X}^{X} g(x)\,dF(x) - \int_{-X}^{X} g(x)\,dF_n(x) \right| +$$

$$+ \left| \int_{|x|>X} g(x)\,dF_n(x) \right| + \left| \int_{|x|>X} g(x)\,dF(x) \right| \leqslant$$

$$\leqslant \left| \int_{-X}^{X} g(x)\,dF(x) - \int_{-X}^{X} g(x)\,dF_n(x) \right| + M\varepsilon + M\varepsilon/2. \quad (32)$$

На основании (31) заключаем, что правая часть (32) может быть сделана как угодно малой, что и доказывает теорему.

Доказательство теоремы 3. По теореме 6 из $F_n(x) \Rightarrow F(x)$ вытекает $f_n(t) = \int e^{itx}\,dF_n \to \int e^{itx}\,dF = f(t)$. Можно доказать, что эта сходимость будет равномерной на каждом конечном интервале t.

Доказательство теоремы 4. По теореме 5 из последовательности $F_n(x)$ можно выбрать подпоследовательность $F_{nn}(x) \Rightarrow F^*(x)$. Докажем, что $F^*(x)$ — функция распределения, т. е. что $F^*(\infty) = 1$, $F^*(-\infty) = 0$. Для этого мы используем неравенство

$$\mathsf{P}\{|\xi| \leqslant X\} \geqslant \frac{\left| \dfrac{1}{2\tau} \displaystyle\int_{-\tau}^{\tau} f(t)\,dt \right| - \dfrac{1}{\tau X}}{1 - \dfrac{1}{\tau X}}, \quad (33)$$

где $f(t)$ — характеристическая функция ξ, $X > 0$, $\tau > 0$. В частности, при $\tau X = 2$

$$\mathsf{P}\{|\xi| \leqslant X\} \geqslant 2 \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f(t)\,dt \right| - 1. \quad (34)$$

Докажем (33). Имеем

$$\left|\frac{1}{2\tau}\int_{-\tau}^{\tau} f(t)\,dt\right| = \left|\frac{1}{2\tau}\int_{-\tau}^{\tau} \mathsf{M}e^{it\xi}\,dt\right| = \left|\frac{1}{2\tau}\mathsf{M}\int_{-\tau}^{\tau} e^{it\xi}\,dt\right| =$$
$$= \left|\mathsf{M}\,\frac{\sin\tau\xi}{\tau\xi}\left(I_{\{|\xi|\leqslant X\}} + I_{\{|\xi|>X\}}\right)\right| \leqslant \mathsf{M}I_{\{|\xi|\leqslant X\}} +$$
$$+ \frac{1}{\tau X}\mathsf{M}I_{\{|\xi|>X\}} = \mathsf{P}\{|\xi|\leqslant X\} + \frac{1}{\tau X}\left(1 - \mathsf{P}\{|\xi|\leqslant X\}\right),$$

откуда и следует (33).

По предположению $f(t)$ непрерывна в нуле, поэтому существует такое $\tau_0 > 0$, что при $0 < \tau \leqslant \tau_0$

$$\left|\frac{1}{2\tau}\int_{-\tau}^{\tau} f(t)\,dt\right| \geqslant 1 - \varepsilon/4.$$

Так как $f_n(t) \to f(t)$ в каждой точке t, то существует такое n_0, что при $n \geqslant n_0$

$$\left|\int_{-\tau}^{\tau} f_n(t)\,dt - \int_{-\tau}^{\tau} f(t)\,dt\right| < \frac{\varepsilon\tau}{2}$$

(теорема 3 § 24 о мажорируемой сходимости). Тогда при $n \geqslant n_0$

$$\left|\frac{1}{2\tau}\int_{-\tau}^{\tau} f_n(t)\,dt\right| \geqslant 1 - \varepsilon/2,$$

и по неравенству (34)

$$\mathsf{P}\{|\xi_n|\leqslant 2/\tau\} =$$
$$= F_n(2/\tau) - F_n(-2/\tau) \geqslant 2(1 - \varepsilon/2) - 1 = 1 - \varepsilon,$$

т. е. $F_n(2/\tau) - F_n(-2/\tau) \geqslant 1 - \varepsilon$, следовательно, $F^*(+\infty) = 1$, $F^*(-\infty) = 0$. Докажем теперь, что $F_n \Rightarrow F$. Предположим, что $F_n \not\Rightarrow F$. Тогда существуют две подпоследовательности $F_{n'} \Rightarrow F^*$ и $F_{n''} \Rightarrow F^{**}$. По прямой предельной теореме $f_{n'} \to f^*$, $f_{n''} \to f^{**}$, но так как $f_n \to f$, то $f^* = f^{**} = f$. Теорема доказана.

Задачи

1. Найти характеристическую функцию распределения, задаваемого плотностью $\frac{1}{2}e^{-|x|}$.

2. Плотность распределения случайной величины ξ задана формулами

$$p_\xi(x) = \begin{cases} \dfrac{x^{\alpha-1}}{2\Gamma(\alpha)}e^{-x}, & x \geqslant 0, \\ \dfrac{|x|^{\beta-1}}{2\Gamma(\beta)}e^{x} & x < 0, \end{cases}$$

с положительными α и β. Найти характеристическую функцию $f_\xi(t)$.

3. Пусть $f_1(t)$ и $f_2(t)$ — характеристические функции, $0 \leqslant p \leqslant 1$. Доказать, что $f(t) = pf_1(t) + (1-p)f_2(t)$ тоже будет характеристической функцией.

4. Если $f(t)$ — характеристическая функция, то $\operatorname{Re} f(t)$ также будет характеристической функцией. Доказать.

5. Пользуясь простейшими свойствами характеристических функций, показать, что функции: а) $\sin t$, б) $\sin t + 1$, в) $\frac{1}{1+t^4}$, г) $|\cos t|$ не могут быть характеристическими.

6. Показать, что характеристическая функция $f_\xi(t)$ случайной величины ξ вещественна при всех t тогда и только тогда, когда распределение ξ симметрично (т. е. ξ и $-\xi$ имеют одинаковые распределения).

Глава 10. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

§ 40. Центральная предельная теорема для одинаково распределенных независимых слагаемых

Ранее мы доказали, что распределение числа успехов μ в схеме Бернулли при $n \to \infty$ и постоянном $0 < p < 1$ обладает следующим предельным свойством:

$$\lim_{n\to\infty} \mathsf{P}\left\{ \frac{\mu - \mathsf{M}\mu}{\sqrt{\mathsf{D}\mu}} \leqslant x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du. \qquad (1)$$

Функцию нормального распределения будем обозначать $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$. Функция нормального распределения $\Phi(x)$ выражается через интеграл Лапласа $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-u^2/2}\, du$, введенный в гл. 4, следующим образом: $\Phi(x) = \frac{1}{2} + \Phi_0(x)$. Этот результат является очень частным случаем так называемой центральной предельной теоремы. Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — последовательность независимых случайных величин. Мы будем говорить, что для этой последовательности выполнена *центральная предельная теорема*, если при любом x справедливо следующее предельное соотношение для сумм $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$:

$$\lim_{n\to\infty} \mathsf{P}\left\{ \frac{\zeta_n - \mathsf{M}\zeta_n}{\sqrt{\mathsf{D}\zeta_n}} \leqslant x \right\} = \Phi(x). \qquad (2)$$

Так как в схеме Бернулли число успехов можно представить в виде суммы $\mu = \mu_1 + \mu_2 + \ldots + \mu_n$ независимых случайных величин с $\mathsf{P}\{\mu_i = 1\} = p$,

$\mathsf{P}\{\mu_i = 0\} = 1 - p$, то результат (1) есть частный случай центральной предельной теоремы (2).

Для справедливости центральной предельной теоремы (2) на случайные величины $\zeta_1, \zeta_2, \ldots, \zeta_n, \ldots$ надо налагать те или иные дополнительные условия.

Мы докажем центральную предельную теорему сначала для одинаково распределенных случайных величин.

Теорема 1. *Если случайные величины* $\xi_1, \xi_2, \ldots$ *независимы, одинаково распределены и имеют конечные* $\mathsf{M}\xi_i = a$ *и* $\mathsf{D}\xi_i = \sigma^2 > 0$, *то*

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{\xi_1 + \ldots + \xi_n - na}{\sigma\sqrt{n}} \leqslant x\right\} = \Phi(x).$$

Доказательство. Используем аппарат характеристических функций. Обозначим $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$, $\tilde{\xi}_k = \xi_k - a_k$, $\tilde{\zeta}_n = \dfrac{\zeta_n - na}{\sigma\sqrt{n}}$, тогда $\tilde{\zeta}_n = \dfrac{1}{\sigma\sqrt{n}}\sum\limits_{k=1}^{n}\tilde{\xi}_k$. Пусть $f(t) = f_{\tilde{\xi}_k}(t)$ — характеристическая функция $\tilde{\xi}_k$. Так как $\mathsf{M}\tilde{\xi}_k = 0$, $\mathsf{M}\tilde{\xi}_k^2 = \mathsf{D}\xi_k = \sigma^2$, то, в силу свойства 6) в § 37,

$$f(t) = 1 - \frac{t^2\sigma^2}{2} + o(t^2)$$

при $t \to 0$ и, следовательно, при любом фиксированном t имеем

$$f_{\tilde{\zeta}_n}(t) = \left[f\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{-\frac{t^2}{2}},$$
$$n \to \infty.$$

Отсюда вытекает утверждение теоремы.

§ 41. Теорема Ляпунова

Центральная предельная теорема имеет место также при некоторых условиях и для неодинаковых независимых слагаемых. Мы докажем ниже эту теорему в условиях Ляпунова.

Пусть теперь ξ_k, $k = 1, 2, \ldots$, независимы и имеют не обязательно одно и то же распределение. Обозначим

$$\mathsf{M}\xi_k = a_k, \quad \mathsf{D}\xi_k = b_k^2, \quad \mathsf{M}|\xi_k - a_k|^3 = c_k^3$$

и

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

Теорема 2. (Теорема Ляпунова.) *Если* $\xi_1, \xi_2, \ldots$ *независимы,* a_k, b_k, c_k *конечны и* $C_n/B_n \to 0$, *то*

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{\xi_1 + \ldots + \xi_n - A_n}{B_n} \leqslant x\right\} = \Phi(x). \tag{3}$$

Доказательство. Положим $\tilde{\xi}_k = \xi_k - a_k$, $f_k(t) = f_{\tilde{\xi}_k}(t)$. Так как $\mathsf{M}\tilde{\xi}_k = 0$, $\mathsf{M}\tilde{\xi}_k^2 = b_k^2$, $\mathsf{M}|\tilde{\xi}_k|^3 = c_k^3 < \infty$, то

$$f_k(t) = 1 - \frac{t^2}{2} b_k^2 + r_k(t), \tag{4}$$

где

$$|r_k(t)| \leqslant \frac{c_k^3 |t|^3}{6}, \quad \left|-\frac{t^2 b_k^2}{2} + r_k(t)\right| \leqslant \frac{|t|^2}{2} b_k^2 \tag{5}$$

(это доказывается так же, как свойство 6) из § 37). В силу независимости $\tilde{\xi}_k$, характеристическая **функция** $\tilde{\zeta}_n = \frac{1}{B_n}\sum_{k=1}^{n} \tilde{\xi}_k$ равна произведению

$$f_{\tilde{\zeta}_n}(t) = \prod_{k=1}^{n} f_k\left(\frac{t}{B_n}\right).$$

Логарифмируя, получаем

$$\log f_{\tilde{\zeta}_n}(t) = \sum_{k=1}^{n} \log f_k\left(\frac{t}{B_n}\right). \tag{6}$$

Из разложения

$$\log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1} x^k}{k}$$

следует, что

$$\log(1+x) = x + \alpha(x), \tag{7}$$

где при $|x| \leqslant 1/2$

$$|\alpha(x)| = \left|\sum_{k=2}^{\infty} \frac{(-1)^{k-1} x^k}{k}\right| \leqslant \frac{1}{2}\sum_{k=2}^{\infty} |x|^k \leqslant |x|^2. \tag{8}$$

Из условия теоремы вытекает

$$\frac{b_k^2}{B_n^2} \leqslant \frac{(\mathsf{M}\,|\,\tilde{\xi}_k\,|^3)^{2/3}}{B_n^2} = \frac{c_k^2}{B_n^2} \leqslant \left(\frac{C_n}{B_n}\right)^2 \to 0 \tag{9}$$

(мы здесь воспользовались неравенством $(\mathsf{M}\,|\,\xi\,|^r)^{1/r} \leqslant \leqslant (\mathsf{M}\,|\,\xi\,|^{r+1})^{1/(r+1)}$). Таким образом, $b_k^2/B_n^2 \to 0$ при $n \to \infty$ равномерно по $1 \leqslant k \leqslant n$. Пусть $T > 0$ и $|\,t\,| \leqslant T$. Согласно (4) и (5)

$$f_k\left(\frac{t}{B_n}\right) = 1 + \varepsilon_k\left(\frac{t}{B_n}\right),$$

где, в силу (9),

$$\left|\varepsilon_k\left(\frac{t}{B_n}\right)\right| = \left|-\frac{t^2}{2}\cdot\frac{b_k^2}{B_n^2} + r_k\left(\frac{t}{B_n}\right)\right| \leqslant \frac{T^2}{2}\cdot\frac{b_k^2}{B_n^2}. \tag{10}$$

Выберем n_0 таким, что $\left|\varepsilon_k\left(\frac{t}{B_n}\right)\right| \leqslant \frac{1}{2}$ при $n \geqslant n_0$, и применим в правой части (6) представление (7) с оценкой (8). Получаем

$$\log f_{\zeta_n}(t) = -\frac{t^2}{2} + \sum_{k=1}^{n} r_k\left(\frac{t}{B_n}\right) + \sum_{k=1}^{n} \theta_k \varepsilon_k^2\left(\frac{t}{B_n}\right), \tag{11}$$

где $|\theta_k| \leqslant 1$. Используя в (11) оценки (5) и (10), имеем при $|t| \leqslant T$:

$$\left|\log f_{\zeta_n}(t) + \frac{t^2}{2}\right| \leqslant \sum_{k=1}^{n}\left|r_k\left(\frac{t}{B_n}\right)\right| +$$

$$+ \max_{1\leqslant k\leqslant n}\left|\varepsilon_k\left(\frac{t}{B_n}\right)\right|\sum_{k=1}^{n}\left|\varepsilon_k\left(\frac{t}{B_n}\right)\right| \leqslant \frac{T^3}{6}\left(\frac{C_n}{B_n}\right)^3 +$$

$$+ \frac{T^4}{4}\cdot\frac{\max\limits_{1\leqslant k\leqslant n} b_k^2}{B_n^2}\sum_{k=1}^{n}\frac{b_k^2}{B_n^2} \to 0,$$

поэтому

$$\lim_{n\to\infty} f_{\zeta_n}(t) = e^{-\frac{t^2}{2}},$$

что равносильно утверждению (3).

§ 42. Применения центральной предельной теоремы

Доказанные выше предельные теоремы имеют большое теоретическое и прикладное значение. Основным в условиях этих теорем является то, что в сумме $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$ каждое слагаемое ξ_k дает малый в общую величину суммы ζ_n случайный вклад. В частности, это выражается тем, что $\mathsf{D}\xi_k/\mathsf{D}\zeta_n \to 0$ при $n \to \infty$ равномерно по $1 \leqslant k \leqslant n$. В приложениях часто используют предположение о том, что встречающиеся при расчетах случайные величины имеют приближенно нормальное распределение. На предположении нормальности построена так называемая теория ошибок измерения, в которой изучаются методы учета случайных ошибок при измерениях тех или иных параметров в экспериментах. В антропологии, например, обработка результатов измерения параметров человеческого тела также ведется на основе предположения нормальности распределения этих параметров. Основанием для предположения нормальности в этих случаях служит большой статистический материал, накопленный при измерениях. Центральная предельная теорема дает гипотезе нормальности некоторое теоретическое обоснование, так как часто на величину какого-либо параметра в реальном явлении влияет много случайных независимых факторов, причем влияние каждого из них невелико, а суммарно они дают некоторый ощутимый эффект. Известно ироническое высказывание одного статистика на этот счет: «Каждый уверен в справедливости нормального закона распределения, экспериментаторы — потому, что они думают, что это математическая теорема, математики — потому, что они думают, что это экспериментальный факт». Это изречение лишний раз нам напоминает, что математические теории строятся не на самих реальных явлениях, а лишь на их математических моделях. Поэтому в применениях теории вероятностей, как и вообще математики, надо никогда не забывать о здравом смысле и всегда заботиться о том, чтобы рассматривалась подходящая модель, правильно отражающая соответствующее явление.

Рассмотрим несколько примеров на применение центральной предельной теоремы. При этом мы будем придерживаться следующей терминологии. Если последовательность случайных величин ζ_n такова, что при некоторых A_n и B_n

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{\zeta_n - A_n}{B_n} \leqslant x\right\} = \Phi(x), \tag{12}$$

то мы будем говорить, что случайная величина ζ_n *асимптотически нормальна* с параметрами (A_n, B_n) или просто (A_n, B_n)*-асимптотически нормальна*, а равенство (12) будем использовать в допредельной форме для приближенной оценки вероятности, полагая

$$\mathsf{P}\left\{\frac{\zeta_n - A_n}{B_n} \leqslant x\right\} \approx \Phi(x).$$

Пример 1. *Ошибки измерения.* При измерении некоторой величины a мы получаем приближенное значение ξ. Сделанная ошибка $\delta = \xi - a$ может быть представлена в виде суммы двух ошибок

$$\delta = (\xi - \mathsf{M}\xi) + (\mathsf{M}\xi - a),$$

первая из которых $\xi - \mathsf{M}\xi$ называется *случайной ошибкой*, а вторая $\mathsf{M}\xi - a$ — *систематической ошибкой*. Хорошие методы измерения не должны иметь систематической ошибки, поэтому мы будем далее полагать $\mathsf{M}\xi = a$. Случайная ошибка δ имеет нулевое математическое ожидание $\mathsf{M}\delta = 0$. Пусть $\mathsf{D}\delta = \sigma^2$. Для уменьшения этой ошибки производят n независимых измерений $\xi_1, \xi_2, \ldots, \xi_n$ и принимают за оценку измеряемой величины a среднее арифметическое $\hat{a} = \frac{1}{n}(\xi_1 + \ldots + \xi_n)$. Какая при этом допускается погрешность? По центральной предельной теореме сумма $\xi_1 + \ldots + \xi_n$ одинаково распределенных независимых случайных величин с $\mathsf{M}\xi_i = a$, $\mathsf{D}\xi_i = \sigma^2 > 0$ $(an, \sigma\sqrt{n})$-асимптотически нормальна. Поэтому $\hat{a}$ при больших n $(a, \sigma/\sqrt{n})$-асимптотически нормальна и

$$\mathsf{P}\{|\hat{a} - a| \leqslant \varepsilon\} \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-\varepsilon\sqrt{n}/\sigma}^{\varepsilon\sqrt{n}/\sigma} e^{-\frac{u^2}{2}}\, du. \tag{13}$$

Из (13) формально можно было бы сделать вывод, что с помощью как угодно грубых методов измерения получаются при больших n как угодно точные результаты. Это противоречит здравому смыслу. В чем тут дело? По-видимому, измерение грубыми методами не подчиняется той модели, на основе которой получена формула (13). И, действительно, при крупном масштабе деления измерительного инструмента нельзя гарантировать отсутствие систематической ошибки.

Часто, отвлекаясь от ошибки округления, принимают, что каждое измерение ξ_i имеет нормальное распределение с параметрами (a, σ). Тогда (13) из приближенного равенства превращается в точное.

Пример 2. *Логарифмически-нормальное распределение.* В антропологии обычно рост или вес человека определенного возраста и пола считают случайной величиной, имеющей нормальное распределение. Однако во многих случаях с гораздо большим основанием можно считать, что логарифмы этих параметров имеют нормальное распределение. Если случайная величина η такова, что $\xi = \log \eta$ имеет нормальное распределение, то говорят, что η имеет *логарифмически-нормальное распределение*, или, короче, *лог-нормальное распределение.* Лог-нормальности роста и веса можно дать некоторое теоретическое обоснование. В самом деле, вес, например, получается в результате воздействия многих независимых причин, однако эти причины воздействуют на вес не аддитивно, а мультипликативно, т. е.

$$\eta = \eta_1 \eta_2 \ldots \eta_n,$$

где η_i — близкие к единице независимые случайные величины. В этом случае

$$\log \eta = \sum_{i=1}^{n} \log \eta_i,$$

и $\log \eta$ в силу центральной предельной теоремы имеет в пределе нормальное распределение.

Пример 3. С помощью центральной предельной теоремы можно доказывать и чисто аналитические факты. Докажем, например, что

$$\lim_{n \to \infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}.$$

В самом деле, пусть ζ_n есть случайная величина, имеющая распределение Пуассона с параметром n. Тогда

$$\mathsf{P}\{\zeta_n \leqslant n\} = e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!}.$$

Поскольку $\zeta_n = \xi_1 + \ldots + \xi_n$, где ξ_k независимы, $\mathsf{M}\xi_k = 1$ и распределены по закону Пуассона, то ζ_n асимптотически нормальна с параметрами $(n, \sqrt{n})$; поэтому $\mathsf{P}\{\zeta_n \leqslant n\} \to 1/2$.

Задачи

1. В предположении, что размер одного шага пешехода равномерно распределен в интервале от 70 см до 80 см и размеры разных шагов независимы, найти вероятность того, что за 10 000 шагов он пройдет расстояние не менее 7,49 км и не более 7,51 км.

2. Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы и одинаково распределены, $\mathsf{M}\xi_i = 0$, $\mathsf{M}\xi_i^2 = 1$. Показать, что для последовательности $\lambda_1\xi_1, \lambda_2\xi_2, \ldots, \lambda_n\xi_n, \ldots$, где λ_n — числовая последовательность, удовлетворяющая условию

$$\frac{\max\limits_{1 \leqslant k \leqslant n} \lambda_k^2}{\sum\limits_{k=1}^{n} \lambda_k^2} \to 0, \tag{14}$$

справедлива центральная предельная теорема. Построить пример, показывающий, что при нарушении условия (14) центральная предельная теорема может не выполняться.

3. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимы и имеют следующие распределения:

$$\mathsf{P}\{\xi_n = n^\alpha\} = \mathsf{P}\{\xi_n = -n^\alpha\} = \frac{1}{2n^\beta},$$

$$\mathsf{P}\{\xi_n = 0\} = 1 - \frac{1}{n^\beta}.$$

При каких α и β выполнено условие теоремы Ляпунова?

4. Случайные величины ξ_n, η_n независимы и имеют пуассоновские распределения с $\mathsf{M}\xi_n = \mathsf{M}\eta_n = \lambda n$. Найти

$$\lim_{n \to \infty} \mathsf{P}\left\{\frac{\xi_n - \eta_n}{\sqrt{n}} \leqslant x\right\}.$$

5. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимы и равномерно распределены на отрезке $[0, 1]$. Найти вероятность того, что $\prod\limits_{k=1}^{100} \xi_k \leqslant \frac{10}{2^{100}}$.

Глава 11. МНОГОМЕРНЫЕ ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ

§ 43. Определение и простейшие свойства

Пусть случайный вектор $\xi = (\xi_1, \ldots, \xi_k)$ имеет многомерную функцию распределения $F_{\xi_1 \ldots \xi_k}(x_1, \ldots, x_k) = \mathbf{P}\{\xi_1 \leqslant x_1, \ldots, \xi_k \leqslant x_k\}$, которую мы иногда будем обозначать кратко $F_\xi(x)$, полагая $x = (x_1, \ldots, x_k)$. Аналогично, плотность $p_{\xi_1 \ldots \xi_k}(x_1, \ldots, x_k)$, если она существует, будем обозначать $p_\xi(x)$. *Многомерной характеристической функцией* случайного вектора ξ назовем

$$f_\xi(t) = f_{\xi_1 \ldots \xi_k}(t_1, \ldots, t_k) = \mathbf{M}e^{i(t, \xi)}, \tag{1}$$

где $t = (t_1, \ldots, t_k)$, $(t, \xi) = \sum_{\alpha=1}^{k} t_\alpha \xi_\alpha$. Характеристическая функция определена для всех t с действительными компонентами t_α. Характеристическая функция (1) определяется с помощью $F_\xi(x)$ и $p_\xi(x)$ следующим образом:

$$f_\xi(t) = \int e^{i(t, x)} dF_\xi(x), \quad f_\xi(t) = \int e^{i(t, x)} p_\xi(x)\, dx,$$

где интегралы берутся по всему k-мерному пространству R^k.

Свойства характеристических функций.

1) *При всех* $t \in R^k$ $|f(t)| \leqslant 1$, $f(0) = 1$.

Очевидное следствие из $|e^{i(t, \xi)}| = 1$.

2) $f(t)$ *равномерно непрерывна по* t.

Доказательство. Обозначим событие $A = \{|\xi_\alpha| \leqslant X,\ \alpha = 1, \ldots, k\}$ и напишем неравенство

$$|f(t+h) - f(t)| = |\mathbf{M}e^{i(t, \xi)}(e^{i(h, \xi)} - 1)| \leqslant$$
$$\leqslant \mathbf{M}|e^{i(h, \xi)} - 1| = \mathbf{M}|e^{i(h, \xi)} - 1|I_A +$$
$$+ \mathbf{M}|e^{i(h, \xi)} - 1|I_{\bar{A}} \leqslant \mathbf{M}|(h, \xi)|I_A + 2\mathbf{M}I_{\bar{A}} \leqslant$$
$$\leqslant X|h| + 2\mathbf{P}\{\xi \notin [-X, X]^k\},$$

где $|h| = \sum_{\alpha=1}^{k} |h_\alpha|$ и $[-X, X]^k$ — прямоугольник

$$\{x: |x_\alpha| \leqslant X, \ \alpha = 1, \ldots, k\}.$$

Пусть дано $\varepsilon > 0$. Выбираем сначала X так, чтобы $\mathsf{P}\{\xi \notin [-X, X]^k\} < \varepsilon/4$. Тогда при всех $|h| < \varepsilon/(2X)$ $|f(t+h) - f(t)| < \varepsilon$, что и требовалось доказать.

3) *Если* $\xi(1), \xi(2), \ldots, \xi(n)$ — *независимые случайные векторы и* $\zeta = \xi(1) + \xi(2) + \ldots + \xi(n)$, *то*

$$f_\zeta(t) = \prod_{\alpha=1}^{n} f_{\xi(\alpha)}(t).$$

Доказывается с помощью мультипликативного свойства математического ожидания.

4) *Характеристическая функция для вектора* $(\xi_1, \ldots, \xi_m)$, $m < k$, *получается из характеристической функции* $f_{\xi_1 \ldots \xi_k}(t_1, \ldots, t_k)$ *следующим образом:*

$$f_{\xi_1 \ldots \xi_m}(t_1, \ldots, t_m) = f_{\xi_1 \ldots \xi_k}(t_1, \ldots, t_m, 0, \ldots, 0).$$

Очевидно.

5) $f_{\xi_1 + \ldots + \xi_k}(t) = f_{\xi_1 \ldots \xi_k}(t, t, \ldots, t).$

Вытекает из $\sum_{\alpha=1}^{k} \xi_\alpha \cdot t = t \sum_{\alpha=1}^{k} \xi_\alpha$.

6) *Для независимости* $\xi_1, \ldots, \xi_k$ *необходимо и достаточно, чтобы*

$$f_{\xi_1 \ldots \xi_k}(t_1, \ldots, t_k) = \prod_{\alpha=1}^{k} f_{\xi_\alpha}(t_\alpha).$$

Необходимость следует из мультипликативного свойства математического ожидания. Достаточность будет следовать из доказываемой ниже формулы обращения.

7) *Если* $\eta = C\xi$ — *линейное преобразование*

$$\eta_\alpha = \sum_{\beta=1}^{k} c_{\alpha\beta} \xi_\beta, \qquad \alpha = 1, \ldots, m,$$

с матрицей $C = \|c_{\alpha\beta}\|$, $\alpha = 1, \ldots, m$; $\beta = 1, \ldots, k$, *то*

$$f_\eta(t) = f_\xi(C^* t),$$

где C^* — сопряженная C матрица, преобразующая вектор $t=(t_1, \ldots, t_m)$ по формулам

$$\sum_{\alpha=1}^{m} c_{\alpha\beta} t_\alpha, \quad \beta=1, \ldots, k.$$

Доказательство.

$$\mathsf{M}e^{i(t,\eta)} = \mathsf{M}e^{i\sum_{\alpha=1}^{m} t_\alpha \eta_\alpha} = \mathsf{M}e^{i(t, C\xi)} = \mathsf{M}e^{i\sum_{\alpha=1}^{m} t_\alpha \sum_{\beta=1}^{k} c_{\alpha\beta}\xi_\beta} =$$
$$= \mathsf{M}e^{i\sum_{\beta=1}^{k} \xi_\beta \sum_{\alpha=1}^{m} c_{\alpha\beta} t_\alpha} = \mathsf{M}e^{i(C^*t, \xi)} = f_\xi(C^*t).$$

Замечание 1. Если $m=k$, детерминант $|C| \neq 0$ и имеется плотность $p_\xi(x)$, то $\eta = C\xi$ также имеет плотность $p_\eta(y)$, которая связана с $p_\xi(x)$ формулой

$$p_\eta(y) = \frac{1}{|C|} p_\xi(C^{-1}y). \tag{2}$$

В самом деле, для любого $A \in \mathscr{B}^k$ имеем $\mathsf{P}\{\xi \in A\} = \int_A p_\xi(x)\,dx$. Делая в интеграле замену $x = C^{-1}y$, получаем

$$\mathsf{P}\{\xi \in A\} = \int_{CA} p_\xi(C^{-1}y)\,|C^{-1}|\,dy =$$
$$= \mathsf{P}\{\eta \in CA\} = \int_{CA} p_\eta(y)\,dy,$$

откуда следует (2).

Замечание 2. Из 3) и 7) следует, что при преобразовании $\eta = C\xi + b$ имеется следующая связь между характеристическими функциями f_ξ и f_η:

$$f_\eta(t) = e^{i(t,b)} f(C^*t).$$

8) $f_\xi(-t) = \bar{f}_\xi(t) = f_{-\xi}(t)$.

Очевидно.

Обозначим моменты

$$m_{\alpha_1 \ldots \alpha_k} = \mathsf{M}\xi_1^{\alpha_1}\xi_2^{\alpha_2} \ldots \xi_k^{\alpha_k}.$$

9) *Если конечны все $m_{\alpha_1 \ldots \alpha_k}$ с $\alpha_1 + \ldots + \alpha_k = r$, то*

$$m_{\alpha_1 \ldots \alpha_k} = i^{\alpha} \frac{\partial^\alpha f_\xi(0, \ldots, 0)}{\partial t_1^{\alpha_1} \ldots \partial t_k^{\alpha_k}}, \qquad \alpha = \alpha_1 + \ldots + \alpha_k \leqslant r, \tag{3}$$

и

$$f(t)=\sum_{\alpha=0}^{r} i^{\alpha} \sum_{\alpha_1+\ldots+\alpha_r=\alpha} \frac{t_1^{\alpha_1}\ldots t_k^{\alpha_k}}{\alpha_1!\ldots\alpha_k!} m_{\alpha_1\ldots\alpha_k}+R_r(t), \qquad (4)$$

где $R_r(t)=o(|t|^r)$ *при* $|t|=|t_1|+\ldots+|t_k|\to 0$.

Доказательство. Равенство (3) доказывается так же, как в одномерном случае. Для доказательства (4) опять введем событие $A=\{|\xi_\alpha|\leqslant X,\ \alpha=1,\ldots,k\}$ и в правой части неравенства

$$|R_r(t)|=\left|\mathbf{M}\left(e^{i(t,\xi)}-\sum_{\alpha=0}^{r}\frac{i^\alpha (t,\xi)^\alpha}{\alpha!}\right)\right|\leqslant$$

$$\leqslant \mathbf{M}\left|e^{i(t,\xi)}-\sum_{\alpha=0}^{r} i^\alpha\frac{(t,\xi)^\alpha}{\alpha!}\right|(I_A+I_{\overline{A}})$$

воспользуемся неравенством $\left|e^{i\varphi}-\sum_{\alpha=0}^{l}\frac{(i\varphi)^\alpha}{\alpha!}\right|\leqslant\frac{|\varphi|^{l+1}}{(l+1)!}$ при $l=r$ и $l=r-1$. Получаем

$$|R_r(t)|\leqslant \mathbf{M}\frac{|(t,\xi)|^{r+1}}{(r+1)!}I_A+2\frac{\mathbf{M}|(t,\xi)|^r I_{\overline{A}}}{r!}\leqslant$$

$$\leqslant\frac{X^{r+1}|t|^{r+1}}{(r+1)!}+2\frac{|t|^r}{r!}\mathbf{M}(|\xi_1|+\ldots+|\xi_k|)^r I_{\overline{A}}.$$

Для каждого $\varepsilon>0$ выбираем сначала X так, чтобы второй член был $<\varepsilon|t|^r/2$. Тогда при $|t|\leqslant t_0=\varepsilon(r+1)!/(2X^{r+1})$ получаем $|R_r(t)|<\varepsilon|t|^r$.

Примеры.

1. Если $\mathbf{P}\{\xi=c\}=1$, то

$$f_\xi(t)=e^{i(t,c)}.$$

2. Пусть $\varphi_{\xi_1\ldots\xi_k}(s_1,\ldots,s_k)=Ms_1^{\xi_1}\ldots s_k^{\xi_k}$ — многомерная производящая функция случайного вектора $\xi=(\xi_1,\ldots,\xi_k)$. Тогда $f_\xi(t_1,\ldots,t_k)=\varphi_\xi(e^{it_1},\ldots,e^{it_k})$. В частности, для полиномиального распределения

$$\varphi_\xi(s_1,\ldots,s_k)=(p_1s_1+\ldots+p_ks_k)^n$$

и

$$f_\xi(t_1,\ldots,t_k)=(p_1e^{it_1}+\ldots+p_ke^{it_k})^n.$$

3. Пусть $\xi_1, \ldots, \xi_k$ — нормально распределенные независимые случайные величины с $M\xi_\alpha = a_\alpha$ и $D\xi_\alpha = b_\alpha$. Тогда

$$f_{\xi_1 \ldots \xi_k}(t) = e^{i \sum\limits_{\alpha=1}^{k} t_\alpha a_\alpha - \frac{1}{2} \sum\limits_{\alpha=1}^{k} b_\alpha t_\alpha^2}.$$

§ 44. Формула обращения

Мы будем исходить из следующей формулы обращения для преобразования Фурье, доказываемой в анализе. Пусть случайный вектор $\xi = (\xi_1, \ldots, \xi_k)$ имеет непрерывную плотность $p_\xi(x)$ и характеристическую функцию $f_\xi(t) \in L_1$ (т. е. $\int\limits_{R^k} |f_\xi(t)|\,dt < \infty$). Тогда

$$p_\xi(x) = \frac{1}{(2\pi)^k} \int\limits_{R^k} e^{-i(t,x)} f_\xi(t)\,dt. \tag{5}$$

Основываясь на (5), докажем формулу обращения в общем случае. Пусть $\eta = (\eta_1, \ldots, \eta_k)$ имеет независимые компоненты, причем η_α имеет равномерное в $(-l_\alpha, l_\alpha)$ распределение, а $\theta = (\theta_1, \ldots, \theta_k)$ — случайный вектор с независимыми нормально распределенными компонентами с $M\theta_\alpha = 0$, $D\theta_\alpha = 1$. Образуем вектор $\zeta = \xi + \eta + \sigma \cdot \theta$. Его характеристическая функция, если ξ, η, θ независимы, равна

$$f_\zeta(t) = f_\xi(t) \prod_{\alpha=1}^{k} \frac{\sin l_\alpha t_\alpha}{l_\alpha t_\alpha} \cdot e^{-\frac{\sigma^2}{2} \sum\limits_{\alpha=1}^{k} t_\alpha^2} \in L_1.$$

Поэтому по формуле (5)

$$p_\zeta(x) = \frac{1}{(2\pi)^k} \int\limits_{R^k} e^{-i(t,x)} f_\zeta(t)\,dt. \tag{6}$$

Обозначим $\Delta(x, l)$ прямоугольник с вершинами $x_\alpha \pm l_\alpha$, $\alpha = 1, \ldots, k$. Так же, как в одномерном случае, доказывается, что при $\sigma \to 0$

$$p_\zeta(x) \to p_{\xi+\eta}(x) = \frac{1}{2^k l_1 \ldots l_k} \mathsf{P}\{\xi \in \Delta(x, l)\}$$

в точках непрерывности x предельной плотности. Поэтому из (6) получаем общую формулу обращения

$$\mathsf{P}\{\xi \in \Delta(x, l)\} = \frac{1}{\pi^k} \lim_{\sigma \to 0} \int_{R^k} e^{-i(t, x)} f_\xi(t) \prod_{\alpha=1}^{k} \frac{\sin l_\alpha t_\alpha}{t_\alpha} e^{-\frac{\sigma^2}{2} \sum_{\alpha=1}^{k} t_\alpha^2} dt, \quad (7)$$

справедливую для всех тех прямоугольников $\Delta(x, l)$, для которых вероятность попадания ξ на границу равна нулю. Поскольку в (7) $\Delta(x, l)$ можно выбирать так, что x_α и l_α образуют всюду плотное множество, то мы получаем из нее следующую теорему единственности.

Теорема 1. *По характеристической функции $f_\xi(t)$ функция распределения восстанавливается однозначно.*

§ 45. Предельные теоремы для характеристических функций

Пусть случайный вектор $\xi_1 = (\xi_1, \ldots, \xi_k)$ имеет функцию распределения $F_{\xi_1 \ldots \xi_k}(x_1, \ldots, x_k) = F(x_1, \ldots, x_k)$. По этой функции мы определяем одномерные функции распределения $F_{\xi_\alpha}(x)$. Обозначим D_α множество точек разрыва $F_{\xi_\alpha}(x)$. Как известно, D_α не более чем счетно. Множество $D = D_1 \cup \ldots \cup D_k$ также не более чем счетно. $F(x_1, \ldots, x_k)$ непрерывна во всех точках $x = (x_1, \ldots, x_k)$, если никакое $x_\alpha \notin D$, так как при $h = (h_1, \ldots, h_k)$ с $h_\alpha \geqslant 0$

$$0 \leqslant F(x+h) - F(x) = \mathsf{P}\{\xi_\alpha \leqslant x_\alpha + h_\alpha, \ \alpha = 1, \ldots, k\} - \mathsf{P}\{\xi_\alpha \leqslant x_\alpha, \ \alpha = 1, \ldots, k\} \leqslant \sum_{\alpha=1}^{k} [F_{\xi_\alpha}(x_\alpha + h_\alpha) - F_{\xi_\alpha}(x_\alpha)],$$

и аналогичное неравенство можно написать при $h_\alpha < 0$. k-мерный прямоугольник $x_\alpha < \xi_\alpha \leqslant y_\alpha$, $\alpha = 1, \ldots, k$, назовем *прямоугольником непрерывности*, если никакое x_α или y_α не принадлежит D. Для прямоугольника непрерывности вероятность

$$\mathsf{P}\{x_\alpha < \xi_\alpha \leqslant y_\alpha, \quad \alpha = 1, \ldots, k\} = \Delta_{h_1 \ldots h_k} F(x_1, \ldots, x_k),$$
$$h_\alpha = y_\alpha - x_\alpha,$$

непрерывна по всем своим аргументам x_α, y_α.

Определение. Мы будем говорить, что последовательность $F_n(x)$ *слабо сходится к* $F(x)$, и писать

$$F_n(x) \Rightarrow F(x), \tag{8}$$

если $F_n(x) \to F(x)$ в каждой точке непрерывности предельной функции.

Если $F_n(x)$ — функция распределения ξ_n, $F(x)$ — функция распределения ξ, то при $F_n(x) \Rightarrow F(x)$ мы будем также говорить, что ξ_n *слабо сходится к* ξ, и обозначать $\xi_n \Rightarrow \xi$; иногда мы будем говорить, что ξ_n сходится к ξ *по распределению*. Из слабой сходимости следует, в частности, что $\mathbf{P}\{\xi_n \in \Delta\} \to \mathbf{P}\{\xi \in \Delta\}$ для любого прямоугольника непрерывности Δ по предельному распределению. Если ξ_n сходится к ξ по распределению, то это значит, что распределения ξ_n и ξ близки друг к другу. Требовать в (8) сходимости в каждой точке было бы неразумно, так как, например, при $k = 1$, $\xi_n = 1/n$, $\xi = 0$ $F_{\xi_n}(x) \Rightarrow F_\xi(x)$, но $F_{\xi_n}(0) \not\to F_\xi(0)$, в то же время ξ_n и ξ близки друг к другу. Нетрудно доказать, что из $F_n(x) \Rightarrow F(x)$ и непрерывности $F(x)$ во всех точках следует равномерная сходимость $F_n(x) \to F(x)$.

Одно из самых важных свойств характеристических функций содержится в следующих предельных теоремах. Пусть $F_n(x)$, $F(x)$ — функции распределения, $f_n(t)$, $f(t)$ — соответствующие им характеристические функции.

Теорема 2. (Прямая предельная теорема.) *Если $F_n(x) \Rightarrow F(x)$, то $f_n(t) \to f(t)$ в каждой точке $t \in R^k$.*

Теорема 3. (Обратная предельная теорема.) *Если $f_n(t)$ сходится в каждой точке $t \in R^k$ к некоторой функции $f(t)$, непрерывной в нуле, то $F_n(x) \Rightarrow F(x)$ и $f(t)$ есть характеристическая функция $F(x)$.*

Доказательство этих теорем вытекает из следующей леммы и двух теорем Хелли.

Лемма. *Если D — всюду плотное множество R^k и $F_n(x) \to F(x)$ для всех x из D, то $F_n(x) \Rightarrow F(x)$.*

Доказательство. Будем писать $x \leqslant y$, $x < y$, если при всех α $x_\alpha \leqslant y_\alpha$ или $x_\alpha < y_\alpha$ соответственно. Пусть x — точка непрерывности $F(x)$. Тогда для любых

$x', x'' \in D, x' < x < x''$, имеем

$$F_n(x') \leqslant F_n(x) \leqslant F_n(x''),$$

$$F(x') = \lim_{n\to\infty} F_n(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant$$
$$\leqslant \lim_{n\to\infty} F_n(x'') = F(x'').$$

Поскольку $F(x') \leqslant F(x) \leqslant F(x'')$ и разность $F(x'') - F(x')$ может быть сделана как угодно малой, имеем $\lim\limits_{n\to\infty} F_n(x) = F(x)$.

Теорема 4. (Первая теорема Хелли.) *Из всякой последовательности функций распределения $\{F_n\}$ можно выбрать слабо сходящуюся подпоследовательность.*

Доказательство. Пусть $D = \{x_n\}$ — всюду плотное в R^k счетное множество. Из ограниченной последовательности $0 \leqslant F_n(x_1) \leqslant 1$ выбираем сходящуюся подпоследовательность $F_{1n}(x_1) \to F(x_1)$ (так мы обозначаем предел). Из ограниченной последовательности $0 \leqslant F_{1n}(x_2) \leqslant 1$ выбираем сходящуюся подпоследовательность $F_{2n}(x_2) \to F(x_2)$ и т. д. Далее выбираем диагональную подпоследовательность $F_{nn}(x)$, для которой

$$F_{nn}(x_k) \to F(x_k)$$

для любого $x_k \in D$. По лемме отсюда вытекает $F_{nn}(x) \Rightarrow F(x)$.

Замечание 3. $F(x)$ может не быть функцией распределения. Построить пример.

Теорема 5. (Вторая теорема Хелли.) *Если $g(x)$ — непрерывная ограниченная функция на R^k и $F_n(x) \Rightarrow F(x)$, то*

$$\lim_{n\to\infty} \int_{R^k} g(x)\,dF_n(x) = \int_{R^k} g(x)\,dF(x). \tag{9}$$

Доказательство. Пусть Δ — прямоугольник непрерывности $F(x)$. Докажем сначала, что

$$\lim_{n\to\infty} \int_{\Delta} g(x)\,dF_n(x) = \int_{\Delta} g(x)\,dF(x). \tag{10}$$

Пусть $|g(x)| \leqslant M$. Выберем $\varepsilon > 0$. Разобьем Δ на прямоугольники непрерывности Δ_α с центрами x_α и положим $g_\varepsilon(x) = g(x_\alpha)$, если $x \in \Delta_\alpha$. Выберем разбиение Δ столь мелким, чтобы для любого $x \in \Delta$

$$|g(x) - g_\varepsilon(x)| < \varepsilon$$

(это можно сделать из-за равномерной непрерывности $g(x)$ на Δ). Тогда

$$\left|\int_\Delta g(x)\,dF_n(x) - \int_\Delta g(x)\,dF(x)\right| \leqslant \left|\int_\Delta g_\varepsilon\,dF_n - \int_\Delta g_\varepsilon\,dF\right| + \\ + \int_\Delta |g(x) - g_\varepsilon(x)|\,dF_n + \int_\Delta |g_\varepsilon(x) - g(x)|\,dF(x) \leqslant \\ \leqslant 2\varepsilon + M \cdot \sum_{\alpha=1}^{N} |\mathbf{P}\{\xi_n \in \Delta_\alpha\} - \mathbf{P}\{\xi \in \Delta_\alpha\}|,$$

где N — число прямоугольников разбиения. При $n \to \infty$ последнее слагаемое может быть сделано как угодно малым, что и доказывает (10). Для доказательства (9) выберем столь большой прямоугольник непрерывности Δ, что $F(\Delta) \geqslant 1 - \varepsilon$ (через $F(A)$ мы иногда будем обозначать $\mathbf{P}\{\xi \in A\}$ для случайного вектора ξ с функцией распределения $F(x)$). Тогда существует такое n_0, что для всякого $n \geqslant n_0$ $F_n(\Delta) \geqslant 1 - 2\varepsilon$, следовательно, $F(\bar\Delta) \leqslant \varepsilon$, $F_n(\bar\Delta) \leqslant 2\varepsilon$. Далее, (9) вытекает из (10) и

$$\left|\int_{R^k} g\,dF_n - \int_{R^k} g\,dF\right| \leqslant \left|\int_\Delta g\,dF_n - \int_\Delta g\,dF\right| + \\ + \left|\int_{\bar\Delta} g\,dF_n\right| + \left|\int_{\bar\Delta} g\,dF\right| \leqslant 3\varepsilon M + \left|\int_\Delta g\,dF_n - \int_\Delta g\,dF\right|.$$

Доказательство теоремы 2. По второй теореме Хелли из $F_n(x) \Rightarrow F(x)$ вытекает $f_n(t) = \int_{R^k} e^{i(t,x)}\,dF_n \to$ $\to f(t) = \int_{R^k} e^{i(t,x)}\,dF$ в каждой точке $t \in R^k$. Нетрудно доказать, что сходимость равномерна на любом ограниченном множестве t.

Доказательство теоремы 3. По первой теореме Хелли из $F_n(x)$ можно выбрать подпоследовательность $F_{nn}(x) \Rightarrow F^*(x)$. Надо доказать, что $F^*(x)$ — функция распределения. Это вытекает из неравенства

$$\mathsf{P}\{|\xi_\alpha| \leqslant X,\ \alpha = 1, \ldots, k\} \geqslant$$

$$\geqslant \frac{\left| \frac{1}{2^k \tau^k} \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} f(t)\, dt \right| - \frac{1}{\tau X}}{1 - \frac{1}{\tau X}}. \quad (11)$$

В частности, при $\tau X = 2$

$$\mathsf{P}\{|\xi_\alpha| \leqslant X, \quad \alpha = 1, \ldots, k\} \geqslant$$

$$\geqslant 2 \left| \frac{1}{2^k \tau^k} \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} f(t)\, dt \right| - 1. \quad (12)$$

Докажем (11). Пусть $A = \{|\xi_\alpha| \leqslant X, \quad \alpha = 1, \ldots, k\}$. Имеем

$$\left| \frac{1}{2^k \tau^k} \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} f(t)\, dt \right| = \left| \frac{1}{2^k \tau^k} \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} \mathsf{M} e^{i(t, \xi)}\, dt \right| =$$

$$= \left| \frac{\mathsf{M}}{2^k \tau^k} \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} e^{i(t, \xi)}\, dt \right| = \left| \mathsf{M} \prod_{\alpha=1}^{k} \frac{\sin \tau\xi_\alpha}{\tau\xi_\alpha} (I_A + I_{\bar{A}}) \right| \leqslant$$

$$\leqslant \mathsf{P}(A) + \frac{1}{\tau X}(1 - \mathsf{P}(A)),$$

откуда вытекает (11).

По предположению $f(t)$ непрерывна в нуле, следовательно, для любого $\varepsilon > 0$ существует такое τ_0, что при $0 < \tau \leqslant \tau_0$

$$\left| \frac{1}{2^k \tau^k} \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} f(t)\, dt \right| \geqslant 1 - \frac{\varepsilon}{4}.$$

Так как $f_n(t) \to f(t)$ в каждой точке t, то существует такое n_0, что при $n \geqslant n_0$

$$\left| \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} f_n(t)\, dt - \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} f(t)\, dt \right| \leqslant 2^{k-2} \tau^k \varepsilon$$

(теорема 3 § 30 о мажорируемой сходимости). Тогда при $n \geqslant n_0$ $\left| \frac{1}{2^k \tau^k} \int_{-\tau}^{\tau} \ldots \int_{-\tau}^{\tau} f_n(t)\,dt \right| \geqslant 1 - \frac{\varepsilon}{2}$ и по неравенству (12) $\mathsf{P}\left\{ |\xi_{n\alpha}| \leqslant \frac{2}{\tau},\ \alpha = 1, \ldots, k \right\} \geqslant 2\left(1 - \frac{\varepsilon}{2}\right) - 1 = 1 - \varepsilon$, следовательно, $F(R^k) = 1$.

Докажем теперь, что $F_n \Rightarrow F$. Предположим, что $F_n \not\Rightarrow F$. Тогда существуют две подпоследовательности $F_{n'} \Rightarrow F^*$ и $F_{n''} \Rightarrow F^{**}$. По прямой предельной теореме $f_{n'} \to f^*$ и $f_{n''} \to f^{**}$, но так как по условию теоремы $f_n \to f$, то $f^* = f^{**} = f$ и по теореме 1 $F^* = F^{**}$. Теорема доказана.

§ 46. Многомерное нормальное распределение и связанные с ним распределения

Мы будем говорить, что случайный вектор $\xi = (\xi_1, \ldots, \xi_k)$ имеет *нормальное* (или *гауссовское*) распределение, если его характеристическая функция имеет вид

$$f_\xi(t) = e^{i(t,\,a) - \frac{1}{2}(Bt,\,t)}, \tag{13}$$

где $a = (a_1, \ldots, a_k)$ — вектор, а $B = \|b_{\alpha\beta}\|$ — симметричная $k \times k$-матрица неотрицательно определенной квадратичной формы $(Bt, t) = \sum_{\alpha,\beta=1}^{k} b_{\alpha\beta} t_\alpha t_\beta \geqslant 0$ [1]). Мы будем также говорить, что случайный вектор с характеристической функцией (13) *(a, B)-нормален.*

Из (13) следует, что каждая компонента ξ_α имеет характеристическую функцию

$$f_{\xi_\alpha}(t) = e^{ita_\alpha - \frac{b_{\alpha\alpha}}{2} t^2},$$

т. е. нормально распределена с $\mathsf{M}\xi_\alpha = a_\alpha$, $\mathsf{D}\xi_\alpha = b_{\alpha\alpha}$. Далее нам удобно будет перейти к центрированному

---

[1]) В случае $B = 0$ распределение (13) вырождается в константу a. В этом вырожденном случае распределение также удобно причислять к нормальному.

вектору $\tilde{\xi} = \xi - a$, для которого

$$f_{\tilde{\xi}}(t) = e^{-\frac{1}{2}(Bt,\, t)}.$$

Поскольку конечны все $M\tilde{\xi}_\alpha^2$, то конечны и смешанные моменты $M\tilde{\xi}_\alpha\tilde{\xi}_\beta$, поэтому их можно вычислять с помощью производных характеристической функции в $t = 0$. Получаем

$$\mathrm{Cov}(\xi_\alpha, \xi_\beta) = M\tilde{\xi}_\alpha\tilde{\xi}_\beta = -\frac{\partial^2 f_{\tilde{\xi}}(0)}{\partial t_\alpha\, \partial t_\beta} = b_{\alpha\beta};$$

таким образом, $B = \|b_{\alpha\beta}\|$ — это *ковариационная матрица* $(\xi_1, \ldots, \xi_k)$.

Далее мы будем $\tilde{\xi}$ обозначать просто ξ. Одно из важных свойств нормального распределения состоит в том, что любое линейное преобразование $\eta = C\xi$ нормально распределенного вектора ξ с $M\xi = 0$ и $\|\mathrm{Cov}(\xi_\alpha, \xi_\beta)\| = B$ приводит к нормально распределенному вектору η с $M\eta = 0$ и $\|\mathrm{Cov}(\eta_\alpha, \eta_\beta)\| = CBC^*$. Это следует из свойства 7) § 43, по которому

$$f_\eta(t) = f_\xi(C^*t) = e^{-\frac{1}{2}(BC^*t,\, C^*t)} = e^{-\frac{1}{2}(CBC^*t,\, t)}.$$

Пусть C — такая ортогональная матрица, что

$$CBC^* = \begin{Vmatrix} d_{11} & 0 & \ldots & 0 \\ 0 & d_{22} & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_{kk} \end{Vmatrix} = D, \quad d_{\alpha\alpha} \geqslant 0.$$

Тогда

$$f_\eta(t) = e^{-\frac{1}{2}\sum\limits_{\alpha=1}^{k} d_{\alpha\alpha} t_\alpha^2}, \tag{14}$$

т. е. $\eta_1, \ldots, \eta_k$ независимо распределены, причем при $d_{\alpha\alpha} > 0$ η_α имеет нормальное распределение с параметрами $(0, \sqrt{d_{\alpha\alpha}})$, а при $d_{\alpha\alpha} = 0$ с вероятностью 1 $\eta_\alpha = 0$. Если матрица B имеет ранг k, то матрица D также имеет ранг k, т. е. все $d_{\alpha\alpha} > 0$. В этом случае η имеет

k-мерную плотность

$$p_\eta(y_1, \ldots, y_k) = \prod_{\alpha=1}^{k} \frac{1}{\sqrt{2\pi d_{\alpha\alpha}}} e^{-\frac{y_\alpha^2}{2d_{\alpha\alpha}}} = \frac{1}{(2\pi)^{k/2}\sqrt{|D|}} e^{-\frac{1}{2}(D^{-1}y,\, y)}.$$

Поскольку $\xi = C^{-1}\eta$ и $|C| = 1$, то по формуле (2) из § 43

$$p_\xi(x) = p_\eta(Cx) = \frac{1}{(2\pi)^{k/2}\sqrt{|D|}} e^{-\frac{1}{2}(D^{-1}Cx,\, Cx)} =$$

$$= \frac{1}{(2\pi)^{k/2}\sqrt{|B|}} e^{-\frac{1}{2}(C^*D^{-1}Cx,\, x)} = \frac{1}{(2\pi)^{k/2}\sqrt{B}} e^{-\frac{1}{2}(B^{-1}x,\, x)}, \quad (15)$$

так как $B^{-1} = C^*D^{-1}C$, $|B| = |D|$. Нормальное распределение $(\xi_1, \ldots, \xi_k)$ с плотностью (15) называется *невырожденным*. Если B диагональна с одинаковыми диагональными элементами, то нормальное распределение называется *сферическим*. В этом случае плотность (15) зависит лишь от расстояния точки x от начала координат. Если же ранг r матрицы B меньше k, то $d_{\alpha\alpha} > 0$ для $\alpha = 1, \ldots, r$, $d_{r+1,\, r+1} = \ldots = d_{kk} = 0$ (при соответствующем преобразовании C). В этом случае, как уже говорилось выше, $\mathsf{P}\{\eta_{r+1} = \ldots = \eta_k = 0\} = 1$, т. е. все распределение сосредоточено на пространстве меньшего числа измерений, определяемого равенствами

$$\sum_{\beta=1}^{k} c_{\alpha\beta}\xi_\beta = 0, \qquad \alpha = r+1, \ldots, k.$$

Выбирая на этом подпространстве координаты $\eta_\alpha = \sum_{\beta=1}^{k} c_{\alpha\beta}\xi_\beta$, $\alpha = 1, \ldots, r$, мы получаем, в силу (14), на этом подпространстве плотность

$$f_{\eta_1 \ldots \eta_r}(x_1, \ldots, x_r) = \frac{1}{(2\pi)^{r/2}\sqrt{d_{11} \ldots d_{rr}}} e^{-\frac{1}{2}\sum_{\alpha=1}^{r} \frac{x_\alpha^2}{d_{\alpha\alpha}}}.$$

В этом случае нормальное распределение называется *вырожденным*.

Если мы в общем случае $r \leqslant k$ к случайным величинам $\eta_1, \ldots, \eta_r$ применим еще линейное преобразование $\theta_\alpha = \eta_\alpha / \sqrt{d_{\alpha\alpha}}$, так что $(\theta_1, \ldots, \theta_r)$ будут независимы н $(0, 1)$-нормальны, то мы приходим к следующему утверждению.

Теорема 6. *Для того чтобы случайный вектор $\xi = (\xi_1, \ldots, \xi_k)$ был нормально распределен, необходимо и достаточно, чтобы имело место представление*

$$\xi_\alpha = \sum_{\beta=1}^{r} g_{\alpha\beta}\theta_\beta + a_\alpha,$$

где $\|g_{\alpha\beta}\|$ — некоторая матрица, $\mathsf{M}\xi_\alpha = a_\alpha$, а $\theta_1, \ldots, \theta_r$ — независимые нормально распределенные случайные величины с параметрами $(0, 1)$.

Одно из самых важных свойств нормального распределения состоит в том, что оно выступает в роли предельного распределения для достаточно общей схемы сумм независимых случайных векторов. Мы докажем здесь методом характеристических функций следующую предельную теорему.

Теорема 7. *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — последовательность независимых одинаково распределенных случайных векторов $\xi_n = (\xi_{n1}, \ldots, \xi_{nk})$ с $\mathsf{M}\xi_{nk} = 0$ и конечными $\mathrm{Cov}(\xi_{n\alpha}, \xi_{n\beta}) = b_{\alpha\beta}$. Обозначим $\zeta_n = \xi_1 + \ldots + \xi_n$. Тогда функция распределения случайного вектора $\zeta'_n = \dfrac{\zeta_n - na}{\sqrt{n}}$ слабо сходится к нормальной функции распределения с нулевыми математическими ожиданиями и матрицей ковариации $B = \| b_{\alpha\beta} \|$.*

Доказательство. Обозначим $f(t)$ характеристическую функцию $\tilde{\xi}_n = \xi_n - a$. Поскольку $\mathsf{M}\tilde{\xi}_n = 0$ и $\mathsf{M}\tilde{\xi}_{n\alpha}\tilde{\xi}_{n\beta} = b_{\alpha\beta}$, то по свойству 9) § 43

$$f\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2n} \sum_{\alpha, \beta=1}^{k} b_{\alpha\beta} t_\alpha t_\beta + o\left(\frac{1}{n}\right).$$

Поэтому при любом t

$$f_{\zeta'_n}(t) = \left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n \to e^{-\frac{1}{2}\sum\limits_{\alpha, \beta=1}^{k} b_{\alpha\beta} t_\alpha t_\beta},$$

откуда, в силу теоремы 3 § 45, следует доказываемое утверждение.

Частным случаем этой теоремы является

Теорема 8. *Пусть* $\xi = (\xi_1, \ldots, \xi_k)$ *имеет полиномиальное распределение с вероятностями исходов* $p = (p_1, \ldots, p_k)$ *и* n *испытаниями. Распределение вектора* $(\xi - np)/\sqrt{n}$ *при* $n \to \infty$ *слабо сходится к нормальному с нулевым средним и матрицей ковариации* $\|\delta_{\alpha\beta} p_\alpha - p_\alpha p_\beta\|$, *где* $\delta_{\alpha\beta}$ — *символ Кронекера.*

Доказательство. Случайный вектор ξ представим в виде суммы $\eta_1 + \eta_2 + \ldots + \eta_n$ независимых векторов $\eta_\alpha = (\eta_{\alpha 1}, \eta_{\alpha 2}, \ldots, \eta_{\alpha k})$, где $\eta_{\alpha\beta} = 1$, если при α-м испытании произошел исход β и $\eta_{\alpha\beta} = 0$ в противоположном случае. Поскольку $\mathsf{M}\eta_{\alpha\beta} = p_\beta$ и $\mathrm{Cov}(\eta_{\alpha\beta}, \eta_{\alpha\gamma}) = p_\beta \delta_{\beta\gamma} - p_\beta p_\gamma$, то применима теорема 7, откуда и следует утверждение.

Замечание 4. Из слабой сходимости ζ_n к предельному вектору ζ следует, что для любого прямоугольника непрерывности Δ предельного распределения

$$\mathsf{P}\{\zeta_n \in \Delta\} \to \mathsf{P}\{\zeta \in \Delta\}. \tag{16}$$

Ясно, что из (16) вытекает справедливость аналогичного утверждения для конечных сумм таких прямоугольников и для множеств, которые можно приблизить этими суммами. Другими словами, для любого измеримого по Жордану множества A с $\mathsf{P}\{\zeta \in \partial A\} = 0$, где ∂A — граница A, при $\zeta_n \Rightarrow \zeta$

$$\mathsf{P}\{\zeta_n \in A\} \to \mathsf{P}\{\zeta \in A\}. \tag{17}$$

Можно доказать, что (17) справедливо для любого борелевского A с $\mathsf{P}\{\zeta \in \partial A\} = 0$. Так же, как в одномерном случае, мы используем обычно предельное соотношение (17) в допредельной форме, считая, что при достаточно больших n левая часть (17) приближенно равна правой.

Сферическое нормальное распределение. Как уже говорилось выше, распределение $\xi = (\xi_1, \ldots, \xi_k)$ с плотностью

$$p_\xi(x) = \frac{1}{(2\pi)^{k/2}\sigma^k} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{k} x_i^2} \tag{18}$$

называется *сферическим нормальным распределением.* Это распределение инвариантно относительно любого ортогонального преобразования $\eta = C\xi$, так как

$$f_\eta(t) = f_\xi(C^*t) = e^{-\frac{\sigma^2}{2}(C^*t,\ C^*t)} = e^{-\frac{\sigma^2}{2}(t,\ t)},$$

т. е. $f_\eta(t) = f_\xi(t)$. Из сферического нормального распределения мы выведем несколько стандартных распределений, имеющих большое значение в математической статистике и других приложениях теории вероятностей.

χ^2-распределение. Рассмотрим сферическое распределение (18) с $\sigma = 1$. Найдем распределение случайной величины

$$\chi_k^2 = \xi_1^2 + \ldots + \xi_k^2.$$

Найдем сначала плотность $p_{\chi_k}(x)$ случайной величины χ_k (она нам понадобится дальше). Вероятность события $x < \chi_k < x + dx$ можно получить из k-мерной нормальной плотности (18) с $\sigma = 1$, интегрируя ее по k-мерному сферическому слою радиуса x и толщины dx. В результате, поскольку $(k-1)$-мерный объем $(k-1)$-мерной сферы радиуса x пропорционален x^{k-1}, получаем

$$p_{\chi_k}(x) = C_k x^{k-1} e^{-\frac{x^2}{2}}.$$

Для определения C_k воспользуемся тем, что по свойству плотности $\int\limits_0^\infty p_{\chi_k}(x)\,dx = 1$, откуда получаем

$$C_k \int\limits_0^\infty x^{k-1} e^{-\frac{x^2}{2}}\,dx = 2^{\frac{k}{2}-1}\,\Gamma\left(\frac{k}{2}\right) C_k = 1$$

и

$$p_{\chi_k}(x) = \frac{x^{k-1}}{2^{\frac{k}{2}-1}\,\Gamma\left(\frac{k}{2}\right)}\, e^{-\frac{x^2}{2}}, \qquad x \geqslant 0. \tag{19}$$

Используя связь между плотностями p_{χ_k} и $p_{\chi_k^2}$, имеем

$$p_{\chi_k^2}(x)=\frac{1}{2\sqrt{x}}\,p_{\chi_k}(\sqrt{x})=\frac{x^{\frac{k}{2}-1}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}e^{-\frac{x}{2}},\quad x\geqslant 0. \tag{20}$$

Распределение с плотностью (20) называется *χ^2-распределением с k степенями свободы.* При $k=2$ χ_2^2 имеет показательную плотность $\frac{1}{2}e^{-\frac{x}{2}}$, $x\geqslant 0$. Плотность (19) при $k=3$ называется *плотностью распределения Максвелла* и дает в кинетической теории газов распределение абсолютной величины скорости частиц.

Распределение Стьюдента. Пусть случайные величины $\xi_0, \xi_1, \ldots, \xi_k$ независимы и нормально распределены с параметрами $(0, 1)$. В статистике мы часто будем использовать случайную величину

$$\tau_k=\frac{\xi_0}{\sqrt{\frac{1}{k}\sum\limits_{\alpha=1}^{k}\xi_\alpha^2}},$$

называемую *отношением Стьюдента.* Распределение случайной величины τ_k, называемое *распределением Стьюдента с k степенями свободы*, имеет плотность

$$s_k(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}},\quad -\infty<x<\infty. \tag{21}$$

Плотность (21) можно вывести следующим образом. Обозначая $\chi_k=\sqrt{\sum\limits_{\alpha=1}^{k}\xi_\alpha^2}$, представим τ_k в виде отношения двух независимых случайных величин

$$\tau_k=\frac{\xi_0\sqrt{k}}{\chi_k},$$

распределение которых известно (числитель имеет нормальное распределение $(0, \sqrt{k})$, а знаменатель — распределение (19)). Функция распределения $S_k(x)$ случайной величины τ_k равна интегралу

$$S_k(x) = \mathsf{P}\{\tau_k \leqslant x\} = \iint\limits_{\frac{u}{v} \leqslant x,\ v>0} p_{\xi_0\sqrt{k}}(u)\, p_{\chi_k}(v)\, du\, dv =$$

$$= a_k \iint\limits_{\frac{u}{v} \leqslant x,\ v>0} v^{k-1} e^{-\frac{1}{2}\left(\frac{u^2}{k} + v^2\right)}\, du\, dv,$$

где

$$a_k = \frac{1}{\sqrt{2\pi k}} \cdot \frac{1}{2^{\frac{k}{2}-1}\Gamma\left(\frac{k}{2}\right)}.$$

От переменных (u, v) перейдем к новым переменным (y, z) по формулам

$$u = yz, \quad v = z. \tag{22}$$

Якобиан преобразования равен $\frac{\partial(u, v)}{\partial(y, z)} = z$, поэтому

$$\iint\limits_{\frac{u}{v} \leqslant x,\ v>0} v^{k-1} e^{-\frac{1}{2}\left(\frac{u^2}{k} + v^2\right)}\, du\, dv = \int\limits_{-\infty}^{x} dy \int\limits_{0}^{\infty} z^k e^{-\frac{z^2}{2}\left(\frac{y^2}{k}+1\right)}\, dz =$$

$$= \int\limits_{-\infty}^{x} \left(\frac{y^2}{k} + 1\right)^{-\frac{k+1}{2}} dy \int\limits_{0}^{\infty} w^k e^{-\frac{w^2}{2}}\, dw =$$

$$= 2^{\frac{k-1}{2}}\Gamma\left(\frac{k+1}{2}\right) \int\limits_{-\infty}^{x} \left(\frac{y^2}{k} + 1\right)^{-\frac{k+1}{2}} dy,$$

откуда следует (21). Заметим, что плотность (21) при $k \to \infty$ сходится к нормальной плотности $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

***F*-распределение.** Пусть $\xi_1, \ldots, \xi_p, \eta_1, \ldots, \eta_q$ — независимые $(0, 1)$-нормальные случайные величины.

Обозначим

$$F_{pq} = \frac{\frac{1}{p}\sum_{\alpha=1}^{p} \xi_\alpha^2}{\frac{1}{q}\sum_{\beta=1}^{q} \eta_\beta^2}.$$

Распределение F_{pq} имеет плотность

$$p^{\frac{p}{2}} q^{\frac{q}{2}} \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \cdot \frac{x^{\frac{p}{2}-1}}{(q+xp)^{\frac{p+q}{2}}}, \qquad x \geqslant 0, \tag{23}$$

и называется *F-распределением Фишера*. Для вывода (23) воспользуемся тем, что $\frac{p}{q} F_{pq}$ есть отношение двух независимых случайных величин, имеющих распределение χ^2 с p и q степенями свободы соответственно. Поэтому

$$\mathbf{P}\left\{\frac{p}{q} F_{pq} \leqslant x\right\} =$$

$$= \frac{1}{2^{\frac{p+q}{2}} \Gamma\left(\frac{p}{2}\right) \Gamma\left(\frac{q}{2}\right)} \iint\limits_{\substack{\frac{u}{v} \leqslant x \\ u \geqslant 0,\ v \geqslant 0}} u^{\frac{p}{2}-1} v^{\frac{q}{2}-1} e^{-\frac{u+v}{2}} \, du \, dv.$$

Делая опять под интегралом замену переменных (22), получаем

$$\iint\limits_{\substack{\frac{u}{v} \leqslant x \\ u \geqslant 0,\ v \geqslant 0}} u^{\frac{p}{2}-1} v^{\frac{q}{2}-1} e^{-\frac{u+v}{2}} \, du \, dv =$$

$$= \int_0^x y^{\frac{p}{2}-1} \, dy \int_0^\infty z^{\frac{p+q}{2}-1} e^{-\frac{z(1+y)}{2}} \, dz =$$

$$= 2^{\frac{p+q}{2}} \Gamma\left(\frac{p+q}{2}\right) \int_0^x \frac{y^{\frac{p}{2}-1}}{(1+y)^{\frac{p+q}{2}}} \, dy$$

и

$$\mathsf{P}\left\{\frac{p}{q}F_{pq}\leqslant x\right\}=\frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)}\int_0^x\frac{y^{\frac{p}{2}-1}}{(1+y)^{\frac{p+q}{2}}}\,dy,$$

откуда уже нетрудно вывести (23). Просто связанная с F_{pq} случайная величина

$$\frac{\xi_1^2+\ldots+\xi_p^2}{\xi_1^2+\ldots+\xi_p^2+\eta_1^2+\ldots+\eta_q^2} \tag{24}$$

имеет более симметричное β-распределение с плотностью

$$\frac{1}{\mathrm{B}\left(\frac{p}{2},\ \frac{q}{2}\right)}\,x^{\frac{p}{2}-1}(1-x)^{\frac{q}{2}-1},\qquad 0\leqslant x\leqslant 1. \tag{25}$$

Функции χ^2-распределения, распределения Стьюдента и F-распределения табулированы.

Задачи

1. Случайные величины ξ_1, ξ_2 — координаты точки, равномерно распределенной в треугольнике $\xi_1\geqslant 0$, $\xi_2\geqslant 0$, $\xi_1+\xi_2\leqslant 1$. Найти их двумерную характеристическую функцию.

2. Пусть $f(t)$ — характеристическая функция случайной величины ξ_1. Найти характеристическую функцию ξ_1, ξ_2, если $\xi_2=1-\xi_1$.

3. Случайные величины ξ_1, ξ_2 имеют сферическое нормальное распределение с плотностью

$$\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}.$$

Найти вероятности $\mathsf{P}\{|\xi|\leqslant 1,\ |\eta|\leqslant 1\}$ и $\mathsf{P}\{\xi^2+\eta^2\leqslant 1\}$.

4. Случайные величины ξ_0', ξ_0'', $\xi_1, \ldots, \xi_n$ независимы и имеют нормальное распределение с параметрами (0, 1). Выразить через распределение Стьюдента распределение случайной величины

$$\eta=\frac{\xi_0'-\xi_0''}{\sqrt{\sum\limits_{\alpha=1}^{n}\xi_\alpha^2}}.$$

5. Доказать, что случайная величина (24) имеет плотность β-распределения (25).

Глава 12. УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

§ 47. Лемма Бореля — Кантелли. Закон «0 или 1» Колмогорова

Пусть на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$ определена последовательность событий $A_n \in \mathscr{A}$. С каждой такой последовательностью можно связать события

$A^* = \{\omega\colon\ \omega \in A_n$ для бесконечно многих $n\}$,

$A_* = \{\omega\colon\ \omega \in A_n$ для всех, кроме конечного числа $n\}$,

которые называются соответственно *верхним* и *нижним пределами* последовательности $\{A_n\}$. Мы будем обозначать

$$A^* = \limsup_{n\to\infty} A_n, \quad A_* = \liminf_{n\to\infty} A_n.$$

Нетрудно видеть, что

$$A^* = \bigcap_{n=1}^{\infty} \bigcup_{m \geqslant n} A_m, \quad A_* = \bigcup_{n=1}^{\infty} \bigcap_{m \geqslant n} A_m,$$

поэтому A^* и A_* принадлежат $\mathscr{A}$, т. е. являются событиями. Если $A^* = A_* = A$, то мы будем говорить, что A есть *предел* A_n, и будем писать $A = \lim\limits_{n\to\infty} A_n$. Если ввести индикаторы I_{A_n}, то легко видеть, что

$$I_{A^*} = \limsup_{n\to\infty} I_{A_n} \Leftrightarrow A^* = \limsup_{n\to\infty} A_n,$$

$$I_{A_*} = \liminf_{n\to\infty} I_{A_n} \Leftrightarrow A_* = \liminf_{n\to\infty} A_n,$$

$$I_A = \lim_{n\to\infty} I_{A_n} \Leftrightarrow A = \lim_{n\to\infty} A_n.$$

Монотонные последовательности A_n всегда имеют предел. Если $A_1 \subseteq A_2 \subseteq \ldots$, то $A_n \uparrow A_* = A^* = \bigcup\limits_n A_n$, а если

$A_1 \supseteq A_2 \supseteq \ldots$, то $A_n \downarrow A^* = A_* = \bigcap\limits_n A_n$. В этих случаях из аксиомы непрерывности легко получить $\mathbf{P}(A_n) \uparrow \mathbf{P}\left(\bigcup\limits_n A_n\right)$ и $\mathbf{P}(A_n) \downarrow \mathbf{P}\left(\bigcap\limits_n A_n\right)$. Поскольку для любой последовательности $\{A_n\}$

$$B_n = \bigcup_{m \geqslant n} A_m \downarrow A^* = \limsup_{n\to\infty} A_n \quad \text{и} \quad C_n = \bigcap_{m \geqslant n} A_m \uparrow A_* = \liminf_{n\to\infty} A_n,$$

то

$$\mathbf{P}\{\limsup_{n\to\infty} A_n\} = \lim_{n\to\infty} \mathbf{P}(B_n), \qquad \mathbf{P}\{\liminf_{n\to\infty} A_n\} = \lim_{n\to\infty} \mathbf{P}\{C_n\}.$$

Условия, при которых вероятность события $A^* = \limsup\limits_{n\to\infty} A_n$ равна нулю или единице, дает нижеследующая

Лемма 1. (Лемма Бореля — Кантелли.) *Если*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty, \tag{1}$$

то $\mathbf{P}(A^*) = 0$. *Если* $A_1, A_2, \ldots$ *независимы и*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty, \tag{2}$$

то $\mathbf{P}(A^*) = 1$.

Доказательство. Рассмотрим случайную величину

$$\xi = \sum_{n=1}^{\infty} I_{A_n},$$

равную числу тех номеров n, при которых происходит A_n (т. е. $\xi(\omega) = k$, если ровно для k номеров n $\omega \in A_n$). По теореме о монотонной сходимости

$$\mathbf{M}\xi = \sum_{n=1}^{\infty} \mathbf{M} I_{A_n} = \sum_{n=1}^{\infty} \mathbf{P}(A_n),$$

поэтому из (1) следует $\mathbf{M}\xi < \infty$, т. е. случайная величина с вероятностью 1 конечна, а так как $\mathbf{P}(A^*) = \mathbf{P}\{\xi = \infty\}$, то первая часть леммы доказана. Для доказательства второй части воспользуемся независимостью $A_1, A_2, \ldots$ в соотношении

$$\mathsf{P}\{A^*\} = \lim_{n\to\infty} \mathsf{P}\left\{\bigcup_{m=n}^{\infty} A_m\right\} = 1 - \lim_{n\to\infty} \mathsf{P}\left\{\bigcap_{m=n}^{\infty} \bar{A}_m\right\} =$$

$$= 1 - \lim_{n\to\infty} \lim_{k\to\infty} \mathsf{P}\left\{\bigcap_{m=n}^{k} \bar{A}_m\right\} = 1 - \lim_{n\to\infty} \lim_{k\to\infty} \prod_{m=n}^{k} (1 - \mathsf{P}(A_m)) =$$

$$= 1 - \lim_{n\to\infty} \prod_{m=n}^{\infty} (1 - \mathsf{P}(A_m)) = 1,$$

так как ряд (2) расходится.

Следствие. *Если $A_1, A_2, \ldots$ независимы, то $\mathsf{P}(A^*)$ равно 0 или 1 в зависимости от того, сходится или расходится ряд $\sum_n \mathsf{P}(A_n)$.*

Это следствие является частным случаем более общего закона «0 или 1» А. Н. Колмогорова.

Пусть на вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$ определена последовательность $\xi_1, \xi_2, \ldots$ независимых случайных величин. (Это означает, что любая конечная их совокупность $\xi_1, \xi_2, \ldots, \xi_N$ независима.) Ранее мы уже определяли σ-алгебру $\mathscr{A}_{\xi_1 \ldots \xi_n}$, порожденную случайными величинами $\xi_1, \ldots, \xi_n$, как σ-алгебру всех событий A, представимых в виде $A = \{\omega: (\xi_1(\omega), \ldots, \xi_n(\omega))\} \in B$, где $B \in \mathscr{B}^n$ — борелевские множества из пространства R^n. Аналогично определяются $\mathscr{A}_{\xi_n} \subseteq \mathscr{A}_{\xi_n \xi_{n+1}} \subseteq \ldots$ Объединение всех $\mathscr{A}_{\xi_n}, \mathscr{A}_{\xi_n \xi_{n+1}}, \ldots$ есть алгебра событий; минимальную σ-алгебру, порожденную этой алгеброй, обозначим $\mathscr{A}_{\xi_n \xi_{n+1} \ldots}$. Последовательность $\mathscr{A}_{\xi_n \xi_{+1} \ldots}, \mathscr{A}_{\xi_{n+1} \xi_{n+2} \ldots}, \ldots$ есть последовательность невозрастающих σ-алгебр. Назовем σ-алгебру $\mathscr{C} = \bigcap_{n=1}^{\infty} \mathscr{A}_{\xi_n \xi_{n+1} \ldots}$ *остаточной σ-алгеброй* последовательности $\{\xi_n\}$; события $A \in \mathscr{C}$ также будем называть *остаточными*. Это название отражает тот факт, что любое $A \in \mathscr{C}$ не зависит от любого конечного числа случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ и определяется лишь «бесконечно далекими» значениями последовательности $\xi_1, \xi_2, \ldots$ Примерами остаточных событий являются $\left\{\sum_n \xi_n \text{ сходится}\right\}$, $\left\{\sum_n \xi_n \text{ расходится}\right\}$.

Теорема 1. (Закон «0 или 1» Колмогорова.) *Если $\xi_1, \xi_2, \ldots$ — независимые случайные величины, то всякое остаточное событие $A \in \mathscr{C}$ имеет вероятность* $\mathsf{P}(A)$, *равную* 0 *или* 1.

Доказательство. Если $A \in \mathscr{C}$, то $A \in \mathscr{A}_{\xi_n \xi_{n+1} \ldots}$ при любом n. Так как $\mathscr{A}_{\xi_1 \ldots \xi_{n-1}}$ и $\mathscr{A}_{\xi_n \xi_{n+1} \ldots}$ независимы, то $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$ для любого $B \in \mathscr{A}_{\xi_1 \ldots \xi_n}$ при каждом n. Следовательно, A не зависит от любого $B \in \mathscr{A}_{\xi_1 \xi_2 \ldots}$, а так как $\mathscr{C} \subseteq \mathscr{A}_{\xi_1 \xi_2 \ldots}$, то A не зависит от самого себя, т. е. $\mathsf{P}(AA) = \mathsf{P}(A) = \mathsf{P}^2(A)$, откуда и следует утверждение.

Следствие. *Если $\xi_1, \xi_2, \ldots$ независимы, то ξ_n либо сходится с вероятностью* 1, *либо расходится с вероятностью* 1; *то же самое справедливо для ряда* $\sum_n \xi_n$.

§ 48. Различные виды сходимости случайных величин

Сходимость почти наверное. Мы будем говорить, что последовательность $\xi_1, \xi_2, \ldots$ *сходится почти наверное* (п. н.) к случайной величине ξ, и писать $\xi_n \xrightarrow{\text{п. н.}} \xi$, если

$$\mathsf{P}\{\lim_{n\to\infty} \xi_n = \xi\} = 1,$$

т. е. вероятность события

$$\{\omega\colon \lim_{n\to\infty} \xi_n(\omega) \neq \xi(\omega)\}$$

равна нулю.

Покажем, что сходимость $\xi_n \xrightarrow{\text{п.н.}} \xi$ эквивалентна тому, что для каждого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}\{\omega\colon \sup_{m \geqslant n} |\xi_m - \xi| > \varepsilon\} = 0. \tag{3}$$

В самом деле, событие $\{\xi_n \to \xi\}$ можно записать так:

$$\{\xi_n \to \xi\} = \bigcap_{r=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m \geqslant n} \left\{|\xi_m - \xi| \leqslant \frac{1}{r}\right\},$$

а противоположное событие представимо в виде

$$\{\xi_n \not\to \xi\} = \bigcup_{r=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m \geqslant n} \left\{ |\xi_m - \xi| > \frac{1}{r} \right\}.$$

Для того чтобы $\mathsf{P}\{\xi_n \not\to \xi\}$, необходимо и достаточно, чтобы при всех r

$$\mathsf{P}\left\{ \bigcap_{n=1}^{\infty} \bigcup_{m \geqslant n} \left\{ |\xi_m - \xi| > \frac{1}{r} \right\} \right\} = 0, \tag{4}$$

а так как

$$\bigcup_{m \geqslant n} \left\{ |\xi_m - \xi| > \frac{1}{r} \right\} = \left\{ \sup_{m \geqslant n} |\xi_m - \xi| > \frac{1}{r} \right\},$$

то из (4) следует, что при любом $r \geqslant 1$

$$\lim_{n \to \infty} \mathsf{P}\left\{ \sup_{m \geqslant n} |\xi_m - \xi| > \frac{1}{r} \right\} = 0,$$

что равносильно (3).

Сходимость по вероятности. Мы говорим, что ξ_n *сходится по вероятности* к ξ (и обозначаем $\xi_n \xrightarrow{\mathsf{P}} \xi$), если для любого $\varepsilon > 0$

$$\mathsf{P}\{|\xi_n - \xi| > \varepsilon\} \to 0, \quad n \to \infty.$$

Доказанный ранее закон больших чисел для сумм $\zeta_n = \xi_1 + \ldots + \xi_n$ независимых одинаково распределенных случайных величин с $\mathsf{M}\xi_i = a$ и $\mathsf{D}\xi_i = \sigma^2 < \infty$ дает пример сходимости по вероятности $\frac{\zeta_n}{n} \xrightarrow{\mathsf{P}} a$, так как $\forall \varepsilon > 0$

$$\mathsf{P}\left\{ \left| \frac{\zeta_n}{n} - a \right| > \varepsilon \right\} \to 0, \quad n \to \infty.$$

Поскольку $\{|\xi_n - \xi| > \varepsilon\} \subseteq \{\sup_{m \geqslant n} |\xi_m - \xi| > \varepsilon\}$, то из условия (3) вытекает, что $\xi_n \xrightarrow{\text{п. н}} \xi$ влечет за собой $\xi_n \xrightarrow{\mathsf{P}} \xi$.

Мы будем говорить, что последовательность случайных величин ξ_n *фундаментальна по вероятности*, если $\forall \varepsilon > 0$

$$\mathsf{P}\{|\xi_n - \xi_m| > \varepsilon\} \to 0, \quad n,\ m \to \infty.$$

Теорема 2. *Для того чтобы $\xi_n \xrightarrow{\mathsf{P}} \xi$, необходимо и достаточно, чтобы последовательность ξ_n была фундаментальна по вероятности.*

Доказательство. Если $\xi_n \xrightarrow{\mathsf{P}} \xi$, то из неравенства

$$\mathsf{P}\{|\xi_n - \xi_m| > \varepsilon\} \leqslant \mathsf{P}\left\{|\xi_n - \xi| > \frac{\varepsilon}{2}\right\} + \mathsf{P}\left\{|\xi_m - \xi| > \frac{\varepsilon}{2}\right\}$$

вытекает фундаментальность $\{\xi_n\}$. Для доказательства достаточности воспользуемся следующей леммой.

Лемма 2. *Если последовательность ξ_n фундаментальна по вероятности, то из нее можно выбрать подпоследовательность, сходящуюся п. н.*

Доказательство. Положим $n_1 = 1$ и по индукции определим n_k как наименьшее $N > n_{k-1}$, для которого

$$\mathsf{P}\left\{|\xi_r - \xi_s| > \frac{1}{2^k}\right\} < \frac{1}{2^k}$$

при всех $r, s \geqslant N$. Тогда

$$\sum_k \mathsf{P}\left\{|\xi_{n_{k+1}} - \xi_{n_k}| > \frac{1}{2^k}\right\} < \sum_k \frac{1}{2^k} < \infty,$$

поэтому по лемме Бореля — Кантелли с вероятностью 1 осуществляется лишь конечное число событий $|\xi_{n_{k+1}} - \xi_{n_k}| > \frac{1}{2^k}$. Поэтому ряд $\xi_1 + \sum_{k=1}^{\infty} (\xi_{n_{k+1}} - \xi_{n_k})$ сходится с вероятностью 1. Полагая $\xi = \xi_1 + \sum_{k=1}^{\infty} (\xi_{n_{k+1}} - \xi_{n_k})$ для тех ω, для которых ряд сходится, и нулю в остальных точках, получаем $\xi_{n_k} \xrightarrow{\text{п. н.}} \xi$. Лемма доказана.

Докажем теперь достаточность в теореме 2. Если ξ_n фундаментальна по вероятности, то в силу леммы существует случайная величина ξ и подпоследовательность ξ_{n_k}, $\xi_{n_k} \xrightarrow{\text{п. н.}} \xi$. Но в этом случае $\xi_n \xrightarrow{\mathsf{P}} \xi$, так как

$$\mathsf{P}\{|\xi_n - \xi| > \varepsilon\} \leqslant \mathsf{P}\left\{|\xi_n - \xi_{n_k}| > \frac{\varepsilon}{2}\right\} + \mathsf{P}\left\{|\xi_{n_k} - \xi| > \frac{\varepsilon}{2}\right\} \to 0.$$

Докажем еще одно следствие сходимости по вероятности.

Теорема 3. *Если* $\xi_n \xrightarrow{\mathsf{P}} \xi$, *то функция распределения* $F_{\xi_n}(x)$ *слабо сходится к функции распределения* $F_\xi(x)$.

Доказательство. Обозначим событие $\{|\xi_n - \xi| \leqslant \varepsilon\} = A_n$. Так как при $\omega \in A_n$

$$\xi - \varepsilon \leqslant \xi_n \leqslant \xi + \varepsilon,$$

то при любом x мы имеем

$$\{\xi_n \leqslant x\} \subseteq \{\xi \leqslant x + \varepsilon\} \cup \bar{A}_n,$$
$$\{\xi \leqslant x - \varepsilon\} \subseteq \{\xi_n \leqslant x\} \cup \bar{A}_n,$$

откуда следует

$$\begin{aligned} &\mathsf{P}\{\xi \leqslant x - \varepsilon\} - \mathsf{P}(\bar{A}_n) \leqslant \mathsf{P}\{\xi_n \leqslant x\} \leqslant \\ &\qquad\qquad \leqslant \mathsf{P}\{\xi \leqslant x + \varepsilon\} + \mathsf{P}(\bar{A}_n), \\ &\mathsf{P}\{\xi \leqslant x - \varepsilon\} \leqslant \varliminf_{n\to\infty} \mathsf{P}\{\xi_n \leqslant x\} \leqslant \varlimsup_{n\to\infty} \mathsf{P}\{\xi_n \leqslant x\} \leqslant \\ &\qquad\qquad \leqslant \mathsf{P}\{\xi \leqslant x + \varepsilon\}. \end{aligned} \tag{5}$$

Если x — точка непрерывности $F_\xi(x)$, то из (5) получаем $\lim\limits_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$, что и требовалось доказать.

Если $F_{\xi_n}(x)$ слабо сходится к вырожденному распределению, то имеет место обратное утверждение.

Теорема 4. *Если* $F_{\xi_n} \Rightarrow F_\xi$ *и* F_ξ *вырождено в точке* c, *то* $\xi_n \xrightarrow{\mathsf{P}} c$.

Доказательство. Так как $F_{\xi_n}(c + \varepsilon) \to 1$ и $F_{\xi_n}(c - \varepsilon) \to 0$, то $\mathsf{P}\{c - \varepsilon < \xi_n \leqslant c + \varepsilon\} \to 1$, т. е. $\mathsf{P}\{|\xi_n - c| > \varepsilon\} \to 0$, что и требовалось доказать.

Следующий пример показывает, что сходимость п. н. сильнее сходимости по вероятности. Пусть пространство элементарных событий — это отрезок $\Omega = [0, 1]$, события — это борелевские множества на нем, вероятность — мера Лебега. Для $2^k \leqslant n < 2^{k+1}$ определим

$$\xi_n(\omega) = \begin{cases} 1, & \dfrac{n - 2^k}{2^k} < \omega < \dfrac{n + 1 - 2^k}{2^k}, \\ 0 & \text{в остальных точках.} \end{cases}$$

Так как при любом $1 > \varepsilon > 0$ $\mathsf{P}\{|\xi_n| > \varepsilon\} = \frac{1}{2^k}$, то $\xi_n \xrightarrow{\mathsf{P}} 0$, но в то же время $\mathsf{P}\{\xi_n \xrightarrow{\text{п. н.}}\not\to 0\} = 1$.

Сходимость в среднем. Мы будем говорить, что последовательность ξ_n ***сходится в среднем порядка*** $r > 0$, если

$$\mathsf{M}|\xi_n - \xi|^r \to 0, \quad n \to 0. \tag{6}$$

Если $r = 2$, то сходимость (6) называется сходимостью *в среднем квадратическом*. Сходимость в среднем порядка r будем обозначать $\xi_n \xrightarrow{r} \xi$. Из неравенства Чебышева

$$\mathsf{P}\{|\xi_n - \xi| > \varepsilon\} \leqslant \frac{\mathsf{M}|\xi_n - \xi|^r}{\varepsilon^r}$$

вытекает, что сходимость $\xi_n \xrightarrow{r} \xi$ влечет $\xi_n \xrightarrow{\mathsf{P}} \xi$.

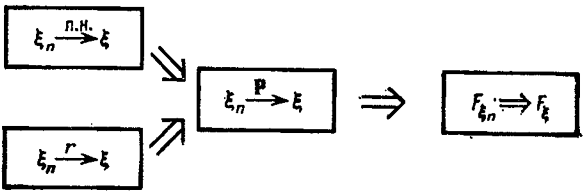

Рис. 13. Связь между различными видами сходимости случайных величин.

Таким образом, мы установили соотношения между разными видами сходимости случайных величин (см. рис. 13).

§ 49. Усиленный закон больших чисел

Исходя из неравенства Чебышева

$$\mathsf{P}\{|\zeta_n - \mathsf{M}\zeta_n| \geqslant x\} \leqslant \frac{\mathsf{D}\zeta_n}{x^2},$$

примененного к суммам $\zeta_n = \xi_1 + \ldots + \xi_n$ независимых случайных величин $\xi_1, \ldots, \xi_n$, мы доказали ранее закон больших чисел (теорема Чебышева), который можно сформулировать так: если $\xi_1, \xi_2, \ldots$ независимы и $\mathsf{D}\xi_k$ ограничены, то

$$\frac{\zeta_n - \mathsf{M}\zeta_n}{n} \xrightarrow{\mathsf{P}} 0. \qquad (7)$$

В случае, когда $\xi_1, \xi_2, \ldots$ независимы и одинаково распределены, сходимость по вероятности (7) имеет место при более слабом условии конечности $\mathsf{M}\xi_n = a$.

Теорема 5 (Хинчин). *Если $\xi_1, \xi_2, \ldots$ независимы, одинаково распределены и $\mathsf{M}\xi_n = a$ конечно, то имеет место закон больших чисел:*

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\mathsf{P}} a.$$

Доказательство. Характеристическая функция $f(t)$ случайной величины $\xi_k - a$ представима в окрестности нуля в виде $f(t) = 1 + o(t)$, поэтому, обозначая $\zeta'_n = \xi_1 + \ldots + \xi_n - na$, имеем

$$f_{\zeta'_n/n}(t) = \left[f\left(\frac{t}{n}\right)\right]^n \to 1,$$

откуда следует слабая сходимость ζ'_n/n к нулю, что равносильно (см. теоремы 3 и 4 § 48) $\zeta'_n/n \xrightarrow{\mathsf{P}} 0$.

Оказывается, можно доказать в условиях теоремы 1 более сильное утверждение, принадлежащее А. Н. Колмогорову. Это так называемый *усиленный закон больших чисел*, утверждающий, что

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{п. н}} a.$$

Далее нам понадобится неравенство Колмогорова, усиливающее известное неравенство Чебышева.

Теорема 6. (Неравенство Колмогорова.) *Пусть $\xi_1, \ldots, \xi_n$ независимы и имеют конечные $\mathsf{M}\xi_k$ и $\mathsf{D}\xi_k$. Тогда*

$$\mathsf{P}\{\max_{1 \leqslant k \leqslant n} |\zeta_k - \mathsf{M}\zeta_k| \geqslant x\} \leqslant \frac{\mathsf{D}\zeta_n}{x^2}, \qquad (8)$$

где $\zeta_k = \xi_1 + \ldots + \xi_k$.

Доказательство. Далее будем считать $\mathsf{M}\xi_k = 0$. Это не ограничивает общности, так как всегда можно перейти от ξ_k к $\xi_k - \mathsf{M}\xi_k$. Введем случайную величину

$\nu = \min\{k: |\zeta_k| \geqslant x\}$. Если $\max\limits_{1 \leqslant k \leqslant n} |\zeta_k| < x$, то положим $\nu = n+1$. Так как $\zeta_n^2 \geqslant \zeta_n^2 \sum\limits_{k=1}^{n} I_{\{\nu=k\}}$, то

$$\mathsf{M}\zeta_n^2 \geqslant \sum_{k=1}^{n} \mathsf{M}\zeta_n^2 I_{\{\nu=k\}} =$$

$$= \sum_{k=1}^{n} \mathsf{M}(\xi_1 + \ldots + \xi_k + \xi_{k+1} + \ldots + \xi_n)^2 I_{\{\nu=k\}} \geqslant$$

$$\geqslant \sum_{k=1}^{n} \mathsf{M}(\xi_1 + \ldots + \xi_k)^2 I_{\{\nu=k\}} +$$

$$+ 2\sum_{k=1}^{n-1} \mathsf{M}(\xi_1 + \ldots + \xi_k) I_{\{\nu=k\}} (\xi_{k+1} + \ldots + \xi_n).$$

Случайная величина $I_{\{\nu=k\}}$ зависит лишь от $\xi_1, \ldots, \xi_k$, поэтому $(\xi_1 + \ldots + \xi_k) I_{\{\nu=k\}}$ не зависит от $\xi_{k+1}, \ldots, \xi_n$ и

$$\mathsf{M}(\xi_1 + \ldots + \xi_k) I_{\{\nu=k\}} (\xi_{k+1} + \ldots + \xi_n) =$$

$$= \mathsf{M}(\xi_1 + \ldots + \xi_k) I_{\{\nu=k\}} \cdot \mathsf{M}(\xi_{k+1} + \ldots + \xi_n) = 0.$$

Так как для $\omega \in \{\nu = k\}$ имеем $\zeta_k \geqslant x$ и $\mathsf{P}\{\nu \leqslant n\} = \mathsf{P}\{\max\limits_{1 \leqslant k \leqslant n} |\zeta_k| \geqslant x\}$, то

$$\mathsf{M}\zeta_n^2 \geqslant \sum_{k=1}^{n} \mathsf{M}\zeta_k^2 I_{\{\nu=k\}} \geqslant x^2 \mathsf{P}\{\nu \leqslant n\} = x^2 \mathsf{P}\{\max_{1 \leqslant k \leqslant n} |\zeta_k| \geqslant x\}.$$

Неравенство (8) доказано.

Докажем теорему об усиленном законе больших чисел для независимых разно распределенных случайных величин.

Теорема 7. *Пусть* $\xi_1, \xi_2, \ldots$ *независимы*, $\mathsf{M}\xi_n = 0$, $\mathsf{D}\xi_n = \sigma_n^2$ *и*

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty.$$

Тогда

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{п. н.}} 0. \qquad (9)$$

Доказательство. Обозначим $\zeta_n = \xi_1 + \ldots + \xi_n$. По критерию (3) из § 48 сходимость (9) равносильна условию

$$\mathsf{P}\left\{\sup_{k\geqslant n}\left|\frac{\zeta_k}{k}\right|>\varepsilon\right\}\to 0,\quad n\to\infty, \tag{10}$$

при любом $\varepsilon>0$. Обозначим A_n событие

$$A_n=\left\{\max_{2^{n-1}\leqslant k<2^n}\left|\frac{\zeta_k}{k}\right|>\varepsilon\right\}.$$

Тогда (10) равносильно

$$\mathsf{P}\left\{\bigcup_{k=n}^{\infty}A_k\right\}\to 0,\quad n\to\infty. \tag{11}$$

По неравенству Колмогорова

$$\mathsf{P}(A_n)\leqslant\mathsf{P}\left\{\max_{2^{n-1}\leqslant k<2^n}|\zeta_n|>\varepsilon\cdot 2^{n-1}\right\}\leqslant$$
$$\leqslant\mathsf{P}\left\{\max_{1\leqslant k\leqslant 2^n}|\zeta_k|>\varepsilon\cdot 2^{n-1}\right\}\leqslant 4\,\frac{\mathsf{D}\zeta_{2^n}}{\varepsilon^2\cdot 2^{2n}}.$$

Далее

$$\sum_{k=1}^{\infty}\mathsf{P}(A_k)\leqslant 4\varepsilon^{-2}\sum_{k=1}^{\infty}2^{-2k}\sum_{n=1}^{2^k}\sigma_n^2\leqslant$$
$$\leqslant 4\varepsilon^{-2}\sum_{n=1}^{\infty}\sigma_n^2\sum_{\{k:\,2^k\geqslant n\}}2^{-2k}\leqslant 8\varepsilon^{-2}\sum_{n=1}^{\infty}\frac{\sigma_n^2}{n^2}<\infty,$$

так как $\sum_{k=k_0}^{\infty}2^{-2k}\leqslant 2\cdot 2^{-2k_0}$. Из сходимости ряда $\sum_k\mathsf{P}(A_k)$ следует (11), так как

$$\mathsf{P}\left\{\bigcup_{k=n}^{\infty}A_k\right\}\leqslant\sum_{k=n}^{\infty}\mathsf{P}(A_k)\to 0,\quad n\to\infty.$$

Докажем следующую вспомогательную лемму.

Лемма 3. *Математическое ожидание ξ конечно тогда и только тогда, когда* $\sum_{n=1}^{\infty}\mathsf{P}\{|\xi|>n\}<\infty$.

Доказательство. Если $\mathsf{M}\xi$ конечно, то и $M|\xi|$ конечно, и наоборот. Из очевидных неравенств

$$\sum_{n=1}^{\infty}(n-1)\mathsf{P}\{n-1<|\xi|\leqslant n\}\leqslant \mathsf{M}|\xi|\leqslant \leqslant\sum_{n=1}^{\infty}n\mathsf{P}\{n-1<|\xi|\leqslant n\}$$

и соотношений

$$\sum_{n=1}^{\infty}n\mathsf{P}\{n-1<|\xi|\leqslant n\}= =\sum_{n=0}^{\infty}\mathsf{P}\{|\xi|>n\}\leqslant 1+\sum_{n=1}^{\infty}\mathsf{P}\{|\xi|>n\},$$

$$\sum_{n=1}^{\infty}(n-1)\mathsf{P}\{n-1<|\xi|\leqslant n\}= =\sum_{n=1}^{\infty}n\mathsf{P}\{n-1<|\xi|\leqslant n\}-\mathsf{P}\{\xi>0\}=\sum_{n=1}^{\infty}\mathsf{P}\{|\xi|>n\}$$

вытекают неравенства

$$\sum_{n=1}^{\infty}\mathsf{P}\{|\xi|>n\}\leqslant \mathsf{M}|\xi|\leqslant 1+\sum_{n=1}^{\infty}\mathsf{P}\{|\xi|>n\},$$

откуда следует утверждение леммы.

Для независимых одинаково распределенных случайных величин справедливо более сильное утверждение, дающее необходимое и достаточное условие усиленного закона больших чисел.

Теорема 8. (Усиленный закон больших чисел Колмогорова.) *Пусть* $\xi_1, \xi_2, \ldots$ *независимы и одинаково распределены. Для того чтобы*

$$\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\xrightarrow{\text{п. н.}}a$$

необходимо и достаточно, чтобы существовало конечное $\mathsf{M}\xi_n=a$.

Доказательство. *Достаточность.* Введем случайные величины

$$\tilde{\xi}_n=\begin{cases}\xi_n, & \text{если } |\xi_n|\leqslant n,\\ 0, & \text{если } |\xi_n|>n,\end{cases}$$

и $\tilde{\zeta}_n = \tilde{\xi}_1 + \ldots + \tilde{\xi}_n$. Случайные величины $\tilde{\xi}_1, \tilde{\xi}_2, \ldots$ также независимы, так как $\tilde{\xi}_n$ есть функция ξ_n. Из равенства

$$\frac{\zeta_n - na}{n} = \frac{\zeta_n - \tilde{\zeta}_n}{n} + \frac{\tilde{\zeta}_n - \mathsf{M}\tilde{\zeta}_n}{n} + \left(\frac{\mathsf{M}\tilde{\zeta}_n}{n} - a\right) \qquad (12)$$

заключаем, что теорема будет доказана, если мы покажем, что справа все три слагаемых с вероятностью 1 сходятся к нулю. Третье слагаемое неслучайно и бесконечно мало, так как оно равно среднему арифметическому

$$-\frac{1}{n}\sum_{k=1}^{n} \mathsf{M}\xi_k I_{\{|\xi_k|>k\}}$$

сходящихся к нулю $\mathsf{M}\xi_k I_{\{|\xi_k|>k\}} \to 0$, $k \to \infty$, членов. Обозначим $A_n = \{\xi_n \neq \tilde{\xi}_n\}$. Имеем

$$\sum_{n=1}^{\infty} \mathsf{P}(A_n) = \sum_{n=1}^{\infty} \mathsf{P}\{|\xi_n| > n\} = \sum_{n=1}^{\infty} \mathsf{P}\{|\xi_1| > n\},$$

где последний ряд сходится в силу конечности $\mathsf{M}\xi_1$ по только что доказанной лемме. Поэтому по лемме Бореля — Кантелли лишь для конечного числа номеров n $\tilde{\xi}_n \neq \xi_n$. Следовательно, в (12)

$$\frac{\zeta_n - \tilde{\zeta}_n}{n} \xrightarrow{\text{п. н.}} 0.$$

Осталось доказать $\dfrac{\tilde{\zeta}_n - M\tilde{\zeta}_n}{n} \xrightarrow{\text{п. н.}} 0$. Применим теорему 3. Для этого докажем, что

$$\sum_{n=1}^{\infty} \frac{\mathsf{D}\tilde{\xi}_n}{n^2} < \infty.$$

Поскольку $\mathsf{D}\tilde{\xi}_n \leqslant \mathsf{M}\tilde{\xi}_n^2 \leqslant \sum_{k=1}^{\infty} k^2 \mathsf{P}\{k-1 < |\xi| \leqslant k\}$, то

$$\sum_{n=1}^{\infty} \frac{\mathsf{D}\tilde{\xi}_n}{n^2} \leqslant \sum_{n=1}^{\infty}\sum_{k=1}^{n} \frac{k^2}{n^2} \mathsf{P}\{k-1 < |\xi_1| \leqslant k\} =$$

$$= \sum_{k=1}^{\infty} k^2 \mathsf{P}\{k-1 < |\xi_1| \leqslant k\} \sum_{n \geqslant k} \frac{1}{n^2}.$$

Так как

$$\sum_{n>k} \frac{1}{n^2} \leqslant \int_k^\infty \frac{dx}{x^2} = \frac{1}{k}, \text{ то } \sum_{n\geqslant k} \frac{1}{n^2} \leqslant \frac{1}{k} + \frac{1}{k^2} = \frac{k+1}{k^2}$$

и

$$\sum_{n=1}^\infty \frac{\mathsf{D}\tilde{\xi}_n}{n^2} \leqslant \sum_{k=1}^\infty \frac{k^2(k+1)}{k^2} \mathsf{P}\{k-1 < |\xi_1| \leqslant k\} \leqslant$$

$$\leqslant \sum_{k=1}^\infty (k+1)\mathsf{P}\{k-1 < |\xi_1| \leqslant k\} \leqslant$$

$$\leqslant 2 + \sum_{k=1}^\infty (k-1)\mathsf{P}\{k-1 < |\xi_1| \leqslant k\} \leqslant 2 + \mathsf{M}|\xi_1| < \infty.$$

Необходимость. Если $\frac{\zeta_n}{n} \xrightarrow{\text{п. н.}} a$, то

$$\frac{\xi_n}{n} = \frac{\zeta_n}{n} - \frac{n-1}{n} \cdot \frac{\zeta_{n-1}}{n-1} \xrightarrow{\text{п. н.}} 0,$$

т. е. с вероятностью 1 осуществляется лишь конечное число событий $\left|\frac{\xi_n}{n}\right| > 1$. По лемме Бореля — Кантелли это влечет за собой

$$\sum_{n=1}^\infty \mathsf{P}\{|\xi_n| > n\} = \sum_{n=1}^\infty \mathsf{P}\{|\xi_1| > n\} < \infty.$$

Следовательно, по лемме этого параграфа конечно $\mathsf{M}\xi_1$.

Следствие. *В схеме Бернулли для числа успехов имеет место не только закон больших чисел*

$$\frac{\mu_n}{n} \xrightarrow{\mathsf{P}} p,$$

но и усиленный закон больших чисел

$$\frac{\mu_n}{n} \xrightarrow{\text{п. н}} p.$$

Следствие вытекает из теоремы 8, так как $\mu_n = \xi_1 + \ldots + \xi_n$, где $\xi_1, \ldots, \xi_n$ независимы и $\mathsf{P}\{\xi_i = 1\} = p$, $\mathsf{P}\{\xi_i = 0\} = 1 - p$.

Задачи

1. Случайные величины ξ_n, $n = 1, 2, \ldots$, независимы и одинаково распределены. Доказать, что с вероятностью 1 произойдет лишь конечное число событий $A_n = \{|\xi_n| \geqslant \sqrt{n}\}$ тогда и только тогда, когда $\mathsf{D}\xi_n$ конечна.

2. Доказать, что сходимость ξ_n к ξ почти наверное или по вероятности влечет за собой сходимость в том же смысле $f(\xi_n)$ к $f(\xi)$, если $f(x)$ — непрерывная функция.

3. Если $f(x)$ — непрерывная ограниченная функция, то из $\xi_n \xrightarrow{\mathsf{P}} \xi$ следует сходимость $f(\xi_n)$ к $f(\xi)$ в среднем r-го порядка при любом $r > 0$. Доказать.

4. Показать, что в условиях теоремы 7 можно получить более сильное утверждение о сходимости

$$\frac{\lambda_1\xi_1 + \ldots + \lambda_n\xi_n}{n} \xrightarrow{\text{п. н}} 0,$$

где λ_n — некоторая последовательность, стремящаяся к бесконечности.

5. Случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимы, одинаково распределены, $\mathsf{M}\xi_1$ конечно. Независимые от них случайные величины $\theta_1, \theta_2, \ldots, \theta_n, \ldots$ независимы между собой и удовлетворяют условию $|\theta_n| \leqslant 1$ и $\mathsf{M}\theta_n = 0$, $n = 1, 2, \ldots$ Справедливо ли утверждение:

$$\frac{\theta_1\xi_1 + \ldots + \theta_n\xi_n}{n} \xrightarrow{\text{п. н}} 0?$$

Глава 13. СТАТИСТИЧЕСКИЕ ДАННЫЕ

§ 50. Основные задачи математической статистики

В гл. 1 говорилось, что теория вероятностей занимается изучением математических моделей случайных явлений. Имея подходящую математическую модель какого-либо случайного явления, мы можем рассчитывать вероятности тех или иных событий и по этим вероятностям мы можем, пользуясь статистической устойчивостью частот, предсказывать частоты этих событий. Если вероятностная модель выбрана правильно, то такие предсказания будут выполняться со случайными ошибками, которые также можно рассчитывать в рамках выбранной модели.

Математическая статистика выделяется из теории вероятностей в самостоятельную область, хотя основные методы и приемы рассуждений в ней остаются теми же самыми. Причиной этого является специфичность задач математической статистики, являющихся в известной мере обратными к задачам теории вероятностей. Если в теории вероятностей мы считаем заданной модель явления и производим расчет возможного реального течения этого явления, то в математической статистике мы исходим из известных реализаций каких-либо случайных событий, из так называемых *статистических данных*, которые обычно носят числовой характер. Математическая статистика разрабатывает различные методы, которые позволяют по этим статистическим данным подобрать подходящую теоретико-вероятностную модель. Например, пусть имеется n независимых наблюдений в схеме Бернулли и пусть в m из них произошло событие A Поскольку модель в схеме Бернулли определяется числом испытаний n и вероятностью $p = \mathrm{P}(A)$, то на этом примере мы сталкиваемся с одной из задач математической статистики: как по m осуществлениям события A в n независимых испытаниях определить вероятность $p = \mathrm{P}(A)$?

Перечислим те основные задачи, которые решает математическая статистика, на примере схемы Бернулли.

а) *Проверка статистических гипотез.* Из каких-либо априорных соображений мы можем предполагать, что $p = p_0$, где p_0 — некоторое фиксированное значение. По относительной частоте $\frac{m}{n}$ мы должны решить, справедлива гипотеза $p = p_0$ или нет. Поскольку при больших n относительная частота $\frac{m}{n}$ близка к p, то статистический критерий по проверке гипотезы $p = p_0$ должен основываться на разности $\left|\frac{m}{n} - p_0\right|$. Если она большая, то, по-видимому, гипотеза неверна, если же она мала, то у нас нет основания отвергать гипотезу $p = p_0$.

б) *Статистическое оценивание неизвестных параметров.* Иногда нам требуется по наблюденному m указать то число $\hat{p}$, которое можно принять за вероятность p в схеме Бернулли. В нашем примере естественно взять $\hat{p} = \frac{m}{n}$. Оценка должна быть в том или ином смысле близкой к оцениваемому параметру.

в) *Доверительные интервалы.* Иногда нас интересует не точное значение неизвестного параметра p, а требуется указать тот интервал $\underline{p} \leqslant p \leqslant \bar{p}$, в котором с вероятностью, близкой к единице, лежит параметр p. Такой интервал $(\underline{p}(m), \bar{p}(m))$, концы которого случайны и зависят лишь от наблюдаемого значения m, называется *доверительным интервалом.*

В последующих главах мы уточним понятия, связанные с этими основными задачами, и рассмотрим эти задачи применительно к некоторым вероятностным моделям.

§ 51. Выборочный метод

Терминология многих статистических задач связана со следующей урновой схемой. Пусть имеется урна с карточками, на которых нанесены числа $X_1, X_2, \ldots, X_N$. Из урны случайно выбираются n карточек с числами $x_1, x_2, \ldots, x_n$. Полученный набор чисел

$$x_1, x_2, \ldots, x_n \tag{1}$$

называется *выборкой объема* n из *генеральной совокупности*

$$X_1, X_2, \ldots, X_N. \tag{2}$$

Как известно, выборка может быть *без возвращения*, когда каждое подмножество $\{X_{i_1}, \ldots, X_{i_n}\}$ мощности n из всего множества (2) появляется с вероятностью $1/C_N^n$, и *с возвращением*, когда каждый упорядоченный набор $(X_{i_1}, \ldots, X_{i_n})$, где могут быть повторения, появляется с вероятностью $1/N^n$. Нетрудно видеть, что в случае выборки с возвращением $x_1, \ldots, x_n$ являются независимыми случайными величинами с законом распределения случайной величины ξ, которая с одной и той же вероятностью $1/N$ принимает каждое из значений (2), если все X_j различны:

$$\mathsf{P}\{\xi = X_j\} = \frac{1}{N}, \qquad j = 1, \ldots, N.$$

В этом случае мы говорим, что (1) есть *независимая выборка объема* n, или *независимая реализация* объема n случайной величины ξ.

Упорядочивая выборку (1) по возрастанию, мы получаем *вариационный ряд*

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}.$$

С любой выборкой (1) можно связать так называемое *эмпирическое*, или *выборочное*, распределение, приписывая каждому значению x_i вероятность $1/n$. *Эмпирической* (или *выборочной*) *функцией распределения* будет

$$\hat{F}(x) = \frac{1}{n} \sum_{k=1}^{n} I_{\{x_k \leqslant x\}}.$$

Поскольку выборка (1) случайна, то эмпирическая функция распределения при каждом x есть случайная величина. Математическое ожидание (среднее), дисперсия, моменты эмпирического распределения также будут случайными величинами и будут называться соответственно *эмпирическими* (или *выборочными*) *математическим ожиданием* (*средним*), *дисперсией*, *моментами*.

Таким образом, выборочное среднее есть среднее арифметическое элементов выборки

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k, \tag{3}$$

и выборочная дисперсия равна

$$s^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2. \tag{4}$$

Выборочные моменты и центральные моменты порядка r определяются выражениями

$$\frac{1}{n} \sum_{i=1}^{n} x_i^r, \quad \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^r.$$

В прикладных курсах математической статистики большое место занимает так называемая *описательная статистика*, в которой излагаются рациональные способы задания статистических данных и вычисления сводных характеристик типа (3) и (4). Например, если $x_i = a + y_i$, то $\bar{x} = a + \bar{y}$, где $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$, и $s^2 = \frac{1}{n} \sum_{i=1}^{n} y_i^2 - (a - \bar{y})^2$. Эти формулы облегчают вычисления в случае, когда числа x_i большие. Подбирая подходящее a, мы сводим все вычисления к арифметическим действиям над числами y_i с небольшим числом знаков.

«Выборочная» терминология сохраняется и в том случае, когда генеральная совокупность (2) не состоит из конечного числа элементов N, а просто есть некий генератор независимых случайных величин x_i с каким-либо распределением[1]). Такой идеализацией в статистике пользуются или при очень больших N (например, при статистических обследованиях в демографии, экономике, социологии), или в том случае, когда элементы выборки (1) можно получать какой-либо однородной

---

[1]) В математической статистике случайные величины обозначаются часто буквами x_i, y_j и т. д., являющимися элементами выборки.

процедурой любое число раз (например, результаты измерений, размер деталей при массовом их изготовлении и т. д.).

В дальнейшем мы будем в основном заниматься независимыми выборками. Относительно бесповторной выборки докажем лишь следующую теорему. Обозначим

$$\overline{X}=\frac{1}{N}\sum_{i=1}^{N}X_i,\quad S^2=\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{X})^2$$

среднее и дисперсию генеральной совокупности (2).

Теорема 1. *Эмпирическое среднее $\bar{x}$ бесповторной выборки* (1) *имеет следующее математическое ожидание и дисперсию*:

$$\mathsf{M}\bar{x}=\overline{X},\quad \mathsf{D}\bar{x}=\frac{S^2}{n}\cdot\frac{N-n}{N-1}. \tag{5}$$

Доказательство. Воспользуемся формулами

$$\mathsf{M}\bar{x}=\frac{1}{n}\sum_{i=1}^{n}\mathsf{M}x_i,\quad \mathsf{D}x_i=\frac{1}{n^2}\left(\sum_{i=1}^{n}\mathsf{D}x_i+2\sum_{i<j}\mathrm{Cov}(x_i,x_j)\right). \tag{6}$$

Вычислим $\mathsf{M}x_i$, $\mathsf{D}x_i$ $\mathrm{Cov}(x_i,x_j)$. Поскольку для вычисления нам нужны лишь двумерные распределения x_i, x_j, рассмотрим конечное вероятностное пространство $(\Omega,\mathcal{A},\mathsf{P})$, где элементарные события $\omega=(k,l)$, $1\leqslant k\neq l\leqslant N$, и элементарные вероятности $p(\omega)=\frac{1}{N(N-1)}$. Случайные величины x_i, x_j определим равенствами

$$x_i(k,l)=X_k,\quad x_j(k,l)=X_l.$$

Тогда

$$\mathsf{M}x_i=\frac{1}{N}\sum_{k=1}^{N}X_k=\overline{X},\quad \mathsf{D}x_i=\frac{1}{N}\sum_{k=1}^{N}(X_k-\overline{X})^2=S^2$$

и при $i\neq j$

$$\mathrm{Cov}(x_i,x_j)=\frac{1}{N(N-1)}\sum_{k\neq l}(X_k-\overline{X})(X_l-\overline{X})=$$

$$=\frac{1}{N(N-1)}\left[-\sum_{k=1}^{N}(X_k-\overline{X})^2+\left[\sum_{k=1}^{N}(X_k-\overline{X})\right]^2\right]=-\frac{S^2}{N-1}.$$

Подставляя полученные значения в (6), получаем формулы (5).

Замечание. Для выборки с возвращением дисперсия $\bar{x}$ равна S^2/n. По неравенству Чебышева при $N \geqslant n \to \infty$ мы получим $\bar{x} \xrightarrow{\mathsf{P}} \bar{\bar{X}}$ как в случае выборки с возвращением, так и в случае выборки без возвращения.

Задачи

1. Из конечной генеральной совокупности $\{X_1, X_2, \ldots, X_N\}$ берутся последовательно две бесповторные выборки $(x_1, \ldots, x_{n_1})$ и $(y_1, \ldots, y_{n_2})$, $n_1 + n_2 \leqslant N$. Найти ковариацию и коэффициент корреляции между средними $\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i$ и $\bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i$.

2. Найти математическое ожидание $\mathsf{M}s^2$ выборочной дисперсии $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$, где $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, для бесповторной выборки $x_1, \ldots, x_n$ из конечной генеральной совокупности $(X_1, X_2, \ldots, X_N)$.

3. Найти математическое ожидание $\mathsf{M}s^2$ выборочной дисперсии $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$, если $x_1, \ldots, x_n$ — независимая выборка из распределения с дисперсией $\mathsf{D}x_i = \sigma^2$.

4. Вычислить $\mathsf{M}x_{(k)}$ и $\mathsf{D}x_{(k)}$, если вариационный ряд $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$ получен из независимой выборки $x_1, \ldots, x_n$ с равномерным распределением в $(0, a)$.

Глава 14. СТАТИСТИЧЕСКИЕ КРИТЕРИИ

§ 52. Статистические гипотезы

Пусть случайная величина ξ или случайный вектор $\xi=(\xi_1, \ldots, \xi_n)$ имеет плотность $p(x; \theta)$, зависящую от параметра θ, одномерного или многомерного, принимающего значения из некоторого множества Θ. В частности, если $p(x; \theta)$ — одномерная плотность и независимая выборка

$$x_1, x_2, \ldots, x_n \tag{1}$$

получена из распределения с этой плотностью, то n-мерная плотность, соответствующая выборке (1), равна произведению

$$p(x_1, \ldots, x_n; \theta)=\prod_{k=1}^{n} p(x_k; \theta).$$

Хотя мы будем далее говорить о $p(x; \theta)$ как о плотности, все сказанное с очевидными видоизменениями будет применимо и к дискретным случайным величинам с законом распределения

$$p(x; \theta)=\mathbf{P}\{\xi=x\},$$

где x принимает счетное или конечное число значений.

Значение параметра θ вполне определяет плотность $p(x; \theta)$. Те или иные предположения о значениях параметра θ мы будем называть *статистическими гипотезами*. Статистическая гипотеза называется *простой*, если она состоит в том, что $\theta=\theta_0$, где θ_0 — некоторое фиксированное значение. Если же наше предположение заключается в том, что $\theta \in \Theta_0$, где Θ_0 — подмножество множества параметров Θ, состоящее более чем из одной точки, то мы говорим о *сложной гипотезе*. Рассмотрим примеры.

Пример 1. Пусть $p(x; a, \sigma)=\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$ — плотность нормального распределения, зависящая от двумерного параметра (a, σ). Гипотеза $(a, \sigma) = (0, 1)$ является простой, а гипотеза $a = a_0$, где a_0 фиксировано, — сложной.

Пример 2. Пусть $p(x; \theta) = C_n^x \theta^x (1-\theta)^{n-x}$ — вероятность x успехов в схеме Бернулли с n независимыми испытаниями. Примером простой гипотезы служит $\theta = 1/2$, а примером сложной — $\theta > 1/2$.

Задача проверки статистических гипотез ставится следующим образом. Известно, что выборка (1) получена из распределения, имеющего плотность вида $p(x; \theta)$. Относительно параметра θ имеется некоторая *основная*, или *проверяемая*, гипотеза H_0: $\theta \in \Theta_0$. Мы должны построить такой статистический критерий, который позволяет нам заключить, согласуется ли выборка (1) с гипотезой H_0 или нет. Обычно критерий строится с помощью критического множества. Из множества X всех возможных значений $x = (x_1, \ldots, x_n)$ выборки (1) выделяется такое подмножество S, называемое *критическим*, что при $x \in S$ гипотеза H_0 отвергается, а в остальных случаях она принимается. Критическое множество S выбирается таким, чтобы вероятность $\mathbf{P}_\theta(S) = \int\limits_S p(x; \theta)\, dx$ выборке x попасть в S при гипотезе H_0 была мала. Получаемый с помощью критического множества S статистический критерий называют иногда *S-критерием*. Естественно, что множество S, удовлетворяющее этому требованию, можно выбрать многими способами. Более определенный выбор возникает в том случае, когда нам задана *конкурирующая*, или *альтернативная*, гипотеза H_1: $\theta \in \Theta_1$. Мы будем рассматривать главным образом случай двух простых гипотез: проверяемой гипотезы H_0: $p_0(x) = p(x; \theta_0)$ и конкурирующей гипотезы H_1: $p_1(x) = p(x; \theta_1)$. Есть задачи, в которых гипотезы H_0 и H_1 равноправны. Так обстоит дело при разбиении множества каких-либо объектов на два вида по значениям определенных параметров. Однако очень часто в реальных задачах гипотезы H_0 и H_1 выступают неравноправно. Например, размер годной детали, изготавливаемой на заводе, есть случайная величина, имеющая нормальное распределение с параметрами (a_0, σ_0). Предположим, что дефектная деталь имеет соответствующий размер также нормально распределенным, но уже с параметрами (a, σ_0), где $a \neq a_0$. Технический контроль, на который поступают изготовленные детали, исходит из того, что детали должны быть годными, и поэтому проверяет гипотезу H_0, т. е. их годность. В этом случае H_0 — основная гипотеза, и на контроле надо уловить те детали, которые изготовлены в условиях конкурирующей гипотезы H_1.

§ 53. Уровень значимости и мощность критерия

Рассмотрим две простые гипотезы: проверяемую H_0: $\theta = \theta_0$, и конкурирующую H_1: $\theta = \theta_1$. С каждым S-критерием связаны ошибки двух родов. *Ошибка первого рода* — отвержение гипотезы H_0, когда она верна; принимая гипотезу H_0 в случае, когда верна конкурирующая гипотеза H_1, мы делаем *ошибку второго рода*. Обозначим

$$\mathsf{P}_i(B) = \int_B p(x; \theta_i)\, dx, \qquad i = 0, 1. \tag{1}$$

Тогда вероятность ошибки первого рода S-критерия равна

$$\alpha = \mathsf{P}_0(S), \tag{2}$$

а вероятность ошибки второго рода равна

$$\beta = \mathsf{P}_1(\bar{S}), \tag{3}$$

где $\bar{S} = X \setminus S$. Иногда мы кратко вероятности ошибок первого и второго родов будем называть просто ошибками первого и второго рода.

Задача построения S-критерия для проверки простой гипотезы H_0 при конкурирующей гипотезе H_1 ставится следующим образом. Вероятность ошибки первого рода α называется *уровнем значимости* S-критерия. Функцией мощности $W = W(S; \theta)$ S-критерия называется следующая функция от θ:

$$W(S; \theta) = \int_S p(x; \theta)\, dx, \tag{4}$$

т. е. вероятность отвергнуть гипотезу H_0, когда истинное значение параметра равно θ. Как видно из (2), (3) и (4), вероятности ошибок первого и второго рода следующим образом выражаются через функцию мощности:

$$\alpha = W(S; \theta), \quad 1 - \beta = W(S; \theta).$$

Итак, сначала задается уровень значимости α и рассматривается множество $\mathcal{S}_\alpha$ всех S-критериев с уровнем значимости α. Среди этих критериев выбирается критерий S^*, для которого мощность при $\theta = \theta_1$ принимает наибольшее значение, т. е.

$$W(S^*; \theta_0) = \alpha, \quad W(S^*; \theta_1) = \max_{S \in \mathcal{S}_\alpha} W(S; \theta_1). \tag{5}$$

Критерий S^*, удовлетворяющий условиям (5), называется *оптимальным*, или *наиболее мощным*, критерием. Оптимальный критерий, удовлетворяющий (5), не всегда существует, поэтому нам удобно будет обобщить понятие статистического критерия. Для этого опишем S-критерий с помощью функции $\varphi(x)$, определенной следующим образом:

$$\varphi(x) = \begin{cases} 1, & \text{если } x \in S, \\ 0, & \text{если } x \notin S. \end{cases} \tag{6}$$

Мы можем истолковывать $\varphi(x)$ как вероятность отвергнуть гипотезу H_0, когда выборка приобретает значение x. Критерии, описываемые функцией вида (6), называются *нерандомизированными*. Введем понятие *рандомизированного* критерия (от англ. random — случайный). Пусть задана функция $\varphi(x)$, такая, что $0 \leqslant \varphi(x) \leqslant 1$ для всех x. Мы предполагаем, что с каждым значением выборки x связывается некий случайный эксперимент (*рандомизация*) с двумя исходами 1 и 0, причем вероятность 1 равна $\varphi(x)$, а вероятность 0 равна $1 - \varphi(x)$. В зависимости от исхода этой рандомизации действует и наш рандомизированный критерий. Если выпала 1, то H_0 отвергается; если выпал 0, то H_0 принимается. Функцию мощности этого критерия, который можно назвать *φ-критерием*, обозначим $W(\varphi, \theta)$. Она равна

$$W(\varphi, \theta) = \int \varphi(x)\, p(x; \theta)\, dx = \mathsf{M}_\theta \varphi(\xi),$$

где M_θ означает математическое ожидание по распределению $p(x; \theta)$, а ξ — случайная величина, плотность которой равна $p(x;\theta)$. Уровень значимости φ-критерия равен

$$\alpha = W(\varphi;\theta_0) = \mathsf{M}_{\theta_0}\varphi(\xi),$$

а вероятность ошибки второго рода равна

$$\beta = 1 - W(\varphi;\theta_1) = 1 - \mathsf{M}_{\theta_1}\varphi(\xi).$$

Рассмотрим множество $\mathcal{S}_\alpha$ всех φ-критериев с фиксированным уровнем значимости α. Мы будем называть φ^*-критерий *оптимальным*, или *наиболее мощным*, если

$$W(\varphi^*;\theta_0) = \alpha, \quad W(\varphi^*;\theta_1) = \max_{\varphi\in\mathcal{S}_\alpha} W(\varphi;\theta_1). \tag{7}$$

Задача (7) всегда допускает решение.

§ 54. Оптимальный критерий Неймана — Пирсона

Обозначим $p_0(x) = p(x;\theta_0)$, $p_1(x) = p(x;\theta_1)$, $\mathsf{M}_0\varphi = \int\varphi(x)\,p_0(x)\,dx$, $\mathsf{M}_1\varphi = \int\varphi(x)\,p_1(x)\,dx$. Оптимальный критерий (7) можно искать среди критериев, которые определяются *отношением правдоподобия* $p_1(x)/p_0(x)$.

Теорема 1. (Теорема Неймана — Пирсона.) *Для любого $0\leqslant\alpha\leqslant 1$ существуют такие числа $c\geqslant 0$ и $0\leqslant\varepsilon\leqslant 1$, что φ^*-критерий с функцией*

$$\varphi^*(x) = \begin{cases} 1, & \text{если } p_1(x) > cp_0(x), \\ \varepsilon, & \text{если } p_1(x) = cp_0(x), \\ 0, & \text{если } p_1(x) < cp_0(x), \end{cases} \tag{8}$$

определяет оптимальный критерий с уровнем значимости α, удовлетворяющий (7).

Доказательство. Пусть $0<\alpha<1$. Случаи $\alpha=0$ и $\alpha=1$ проверяются отдельно, и мы не будем здесь этим заниматься. Рассмотрим функцию от c

$$g(c) = \mathsf{P}\{p_1(\xi) > cp_0(\xi)\mid H_0\}$$

в предположении, что верна гипотеза H_0. Функция

$$1 - g(c) = \mathsf{P}\left\{\frac{p_1(\xi)}{p_0(\xi)}\leqslant c\mid H_0\right\}$$

есть функция распределения случайной величины $p_1(\xi)/p_0(\xi)$, поэтому она непрерывна справа и $g(\infty) =$

$= 0$, $g(0-) = 1$. Определим c_α из условия

$$g(c_\alpha) \leqslant \alpha < g(c_\alpha - 0).$$

Если $g(c_\alpha) < g(c_\alpha - 0)$, то выбираем

$$\varepsilon_\alpha = \frac{\alpha - g(c_\alpha)}{g(c_\alpha - 0) - g(c_\alpha)}.$$

Если $g(c_\alpha) = g(c_\alpha - 0)$, то полагаем $\varepsilon_\alpha = 0$. В случае, когда $g(c) \equiv \alpha$ для целого отрезка $c_1 \leqslant c \leqslant c_2$, принимаем за c_α любую точку этого отрезка, например, самую левую. Полагая c и ε в (8) равными найденным c_α и ε_α, строим функцию φ^*. Докажем, что полученный φ^*-критерий имеет уровень значимости α и обладает свойством оптимальности (7). Докажем сначала, что уровень значимости φ^*-критерия равен α. Имеем

$$\mathbf{M}_0\varphi^* = \int\limits_{p_1(x) > c_\alpha p_0(x)} p_0(x)\,dx + \frac{\alpha - g(c_\alpha)}{g(c_\alpha - 0) - g(c_\alpha)} \int\limits_{p_1(x) = c_\alpha p_0(x)} p_0(x)\,dx =$$

$$= g(c_\alpha) + \frac{\alpha - g(c_\alpha)}{g(c_\alpha - 0) - g(c_\alpha)} \cdot (g(c_\alpha - 0) - g(c_\alpha)) = \alpha.$$

Пусть φ — любой другой критерий с $\mathbf{M}_0\varphi \leqslant \alpha$. Покажем, что тогда $\mathbf{M}_1\varphi^* \geqslant \mathbf{M}_1\varphi$. Рассмотрим интеграл

$$\int (\varphi^*(x) - \varphi(x))(p_1(x) - c_\alpha p_0(x))\,dx. \qquad (9)$$

Разобьем его на два слагаемых

$$\int\limits_{\varphi^* > \varphi} (\varphi^* - \varphi)(p_1 - c_\alpha p_0)\,dx + \int\limits_{\varphi^* < \varphi} (\varphi^* - \varphi)(p_1 - c_\alpha p_0)\,dx. \qquad (10)$$

В первом слагаемом интегрирование производится по точкам x, для которых $\varphi^*(x) > \varphi(x) \geqslant 0$, поэтому в этом интеграле $p_1(x) \geqslant c_\alpha p_0(x)$, т. е. подынтегральная функция неотрицательна. Аналогично, во втором интеграле (10) $\varphi^*(x) < \varphi(x) \leqslant 1$, поэтому $p_1(x) \leqslant c_\alpha p_0(x)$, и подынтегральная функция также неотрицательна. Отсюда заключаем, что интеграл (9) неотрицателен, т. е.

$$\int (\varphi^* - \varphi)\, p_1\,dx \geqslant c_\alpha \int (\varphi^* - \varphi) p_0\,dx,$$

или

$$\mathbf{M}_1\varphi^* - \mathbf{M}_1\varphi \geqslant c_\alpha(\alpha - \mathbf{M}_0\varphi) \geqslant 0,$$

что и требовалось доказать.

Замечание. Теорема справедлива и для дискретных распределений $p_0(x)$ и $p_1(x)$. В доказательстве в этом случае надо везде интегралы заменять суммами.

§ 55. Оптимальные критерии для проверки гипотез о параметрах нормального и биномиального распределений

Пусть (1) есть независимая выборка из нормального распределения с параметрами (a, σ). Пусть σ известно, а относительно a имеются две гипотезы:

$$\text{гипотеза } H_0\colon\ a = a_0,$$

$$\text{гипотеза } H_1\colon\ a = a_1 > a_0.$$

Построим оптимальный критерий Неймана — Пирсона. В этом случае

$$p_j(x) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a_j)^2}, \quad j = 0,\ 1,$$

и

$$\frac{p_1(x)}{p_0(x)} = \exp\left\{ n\bar{x}(a_1 - a_0) - \frac{n}{2\sigma^2}(a_1^2 - a_0^2) \right\}, \tag{11}$$

где $\bar{x}$ — выборочное среднее. Из (11) следует, что область значений x, для которых $p_1(x)/p_0(x) > C$, определяется неравенством $\bar{x} > C_1$ при некотором C_1. Как известно, среднее $\bar{x}$ распределено нормально с параметрами $\left(a, \frac{\sigma}{\sqrt{n}}\right)$. Определим теперь ошибки первого и второго рода:

$$\alpha = \mathbf{P}\{\bar{x} > C_1 \mid H_0\} = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{C_1 - a_0}{\sigma}\sqrt{n}}^{\infty} e^{-\frac{u^2}{2}}\, du = 1 - \Phi\left(\frac{C_1 - a_0}{\sigma}\sqrt{n}\right), \tag{12}$$

$$\beta = \mathbf{P}\{\bar{x} \leqslant C_1 \mid H_1\} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{C_1 - a_1}{\sigma}\sqrt{n}} e^{-\frac{u^2}{2}}\, du = \Phi\left(\frac{C_1 - a_1}{\sigma}\sqrt{n}\right). \tag{13}$$

Обозначим u_γ то значение, для которого

$$1-\Phi(u_\gamma)=\gamma. \tag{14}$$

u_γ носит название *квантиль нормального распределения.* Тогда из (12) и (13) и $u_\gamma=-u_{1-\gamma}$ вытекает

$$\frac{C_1-a_0}{\sigma}\sqrt{n}=u_\alpha,\quad \frac{C_1-a_1}{\sigma}\sqrt{n}=-u_\beta,$$

$$C_1=a_0+u_\alpha\frac{\sigma}{\sqrt{n}}=a_1-u_\beta\frac{\sigma}{\sqrt{n}}$$

и

$$n=\frac{\sigma^2(u_\alpha+u_\beta)^2}{(a_1-a_0)^2}. \tag{15}$$

Равенство (15) дает тот объем выборки, который при оптимальном критерии обеспечивает ошибки первого и второго рода α и β (если правая часть (15) — нецелая, то за n надо брать ближайшее большее целое число).

Рассмотрим теперь следующие две гипотезы:

$$H_0:\ a=0,\quad \sigma=\sigma_0,$$
$$H_1:\ a=0,\quad \sigma=\sigma_1>\sigma_0.$$

В этом случае отношение правдоподобия

$$\frac{p_1(x)}{p_0(x)}=\frac{\sigma_0^n}{\sigma_1^n}\exp\left\{\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\sum_{i=1}^{n}x_i^2\right\}$$

приводит к критическому множеству

$$\sum_{i=1}^{n}x_i^2\geqslant C_1.$$

Поскольку случайная величина

$$\frac{\chi_n^2}{\sigma^2}=\sum_{i=1}^{n}\frac{x_i^2}{\sigma^2}$$

имеет при гипотезе $(0,\sigma)$ χ^2-распределение с n степенями свободы с функцией распределения

$$K_n(x)=\int_0^x k_n(u)\,du,\qquad x\geqslant 0,$$

и плотностью

$$k_n(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x > 0,$$

то ошибки первого и второго рода определяются из равенств

$$\alpha = \mathsf{P}\left\{\frac{1}{\sigma_0^2}\sum_{i=1}^{n} x_i^2 \geqslant \frac{C_1}{\sigma_0^2}\right\} = 1 - K_n\left(\frac{C_1}{\sigma_0^2}\right),$$

$$\beta = \mathsf{P}\left\{\frac{1}{\sigma_1^2}\sum_{i=1}^{n} x_i^2 < \frac{C_1}{\sigma_1^2}\right\} = K_n\left(\frac{C_1}{\sigma_1^2}\right).$$

Построим оптимальный критерий в схеме Бернулли. Пусть $0 < p_0 < p_1 < 1$. Рассмотрим следующие две гипотезы:

$$H_0\colon\ p_0(x) = C_n^x p_0^x (1-p_0)^{n-x}, \qquad x = 0, 1, \ldots, n,$$

$$H_1\colon\ p_1(x) = C_n^x p_1^x (1-p_1)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

Оптимальный критерий для проверки гипотезы H_0 против конкурирующей гипотезы H_1 строится, исходя из неравенства

$$\frac{p_1(x)}{p_0(x)} = \left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right]^x \frac{(1-p_1)^n}{(1-p_0)^n} \geqslant C,$$

которое равносильно неравенству $x \leqslant C_1$ при некотором C_1. Для вычисления ошибок первого и второго рода воспользуемся тем, что число положительных успехов x асимптотически нормально с параметрами $\left(np, \sqrt{np(1-p)}\right)$. Имеем

$$\alpha = \mathsf{P}\{x \geqslant C_1 \mid H_0\} = \mathsf{P}\left\{\frac{x - np_0}{\sqrt{np_0(1-p_0)}} \geqslant \frac{C_1 - np_0}{\sqrt{np_0(1-p_0)}} \mid H_0\right\},$$

$$\beta = \mathsf{P}\{x < C_1 \mid H_1\} = \mathsf{P}\left\{\frac{x - np_1}{\sqrt{np_1(1-p_1)}} < \frac{C_1 - np_1}{\sqrt{np_1(1-p_1)}} \mid H_1\right\}.$$

Отсюда, используя квантиль u_γ, определенную (14), получаем при заданных α и β границу

$$C_1 \approx np_0 + u_\alpha \sqrt{np_0(1-p_0)} \approx np_1 - u_\beta \sqrt{np_1(1-p_1)}$$

и необходимый объем выборки

$$n \approx \frac{\left(u_\alpha \sqrt{p_0(1-p_0)} + u_\beta \sqrt{p_1(1-p_1)}\right)^2}{(p_0-p_1)^2}.$$

§ 56. Критерии для проверки сложных гипотез

На примерах выборок из нормального распределения разберем те задачи, которые возникают при проверке сложных гипотез.

Пример 3. Пусть независимая выборка (1) взята из нормального распределения с параметрами (a, σ), причем σ известно. Рассмотрим простую проверяемую гипотезу H_0: $a = a_0$ и одностороннюю сложную конкурирующую гипотезу H_1: $a > a_0$. Действуя так же, как в § 55 при различении двух простых гипотез, находим, что критерий $\bar{x} > C_1 = a_0 + u_\alpha \sigma / \sqrt{n}$ будет иметь уровень значимости α и будет наиболее мощным для любой простой гипотезы $a_1 > a_0$. Функция мощности этого критерия $W(a)$ будет иметь график, изображенный на рис. 14, и ошибка второго рода $\beta(a) = 1 - W(a)$ при $a \downarrow a_0$ в пределе равна $1-\alpha$. Поэтому по критерию $\bar{x} > C_1$ мы можем лишь с малой ошибкой α отвергнуть гипотезу H_0. В случае $\bar{x} \leqslant C_1$ мы не имеем больших оснований утверждать только на основе выборки (1), что $a = a_0$, а не $a > a_0$, так как при a, близких к a_0, вероятность события $\bar{x} \leqslant C_1$ близка к единице. Поэтому при $\bar{x} \leqslant C_1$ мы говорим, что выборка (1) не противоречит гипотезе H_0, и если эта гипотеза имеет какое-либо обоснование, независимое от выборки (1), то выборка в этом случае ее подтверждает.

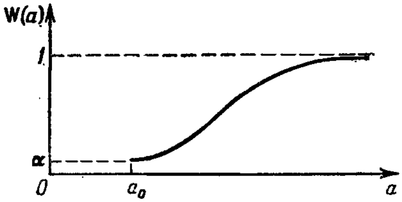

Рис. 14. Функция мощности $W(a)$ критерия $\bar{x} > a_0 + u_\alpha \dfrac{\sigma}{\sqrt{n}}$.

Пример 4. Пусть гипотеза H_0 остается прежней, а конкурирующая гипотеза будет двусторонней H_1: $a \neq a_0$. В этом случае для значений $a = a_1 < a_0$ и

$a=a_1>a_0$ теорема Неймана — Пирсона дает разные оптимальные критерии $\bar{x}<C_1$ и $\bar{x}>C_1$, т. е. не существует такого критерия с уровнем значимости α, который максимизировал бы функцию мощности $W(a)$ во

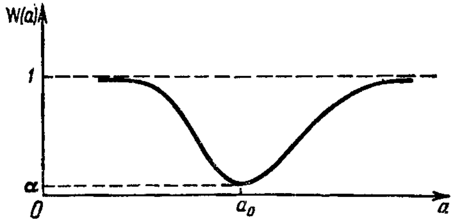

Рис. 15. Функция мощности $W(a)$ двустороннего критерия $|\bar{x}-a|>\frac{u_{\alpha/2}\sigma}{\sqrt{n}}$.

всех точках $a\neq a_0$. В этом случае применяют двусторонний критерий, по которому гипотеза H_0 отвергается, когда

$$|\bar{x}-a_0|\geqslant\frac{u_{\alpha/2}\sigma}{\sqrt{n}}.$$

Функция мощности такого критерия равна

$$W(a)=1-\frac{1}{\sqrt{2\pi}}\int\limits_{\frac{a_0-a}{\sigma}\sqrt{n}-u_{\alpha/2}}^{\frac{a_0-a}{\sigma}\sqrt{n}+u_{\alpha/2}} e^{-\frac{x^2}{2}}\,dx.$$

Уровень значимости этого критерия равен α, а график имеет вид, изображенный на рис. 15.

Пример 5. Пусть имеются две независимые выборки: выборка $x_1, x_2, \ldots, x_{n_1}$ из нормального распределения $(0, \sigma_1)$ и выборка $y_1, y_2, \ldots, y_{n_2}$ из нормального распределения $(0, \sigma_2)$. Рассмотрим основную гипотезу H_0: $\sigma_1=\sigma_2$, и конкурирующую гипотезу H_1: $\sigma_1\neq\sigma_2$.

Статистика

$$F=\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}x_i^2}{\frac{1}{n_2}\sum_{i=1}^{n_2}y_i^2} \tag{16}$$

имеет при гипотезе H_0 F-распределение Фишера. Критерий можно построить на основе статистики (16). Пусть F_γ — такая квантиль F-статистики (16), что

$$\mathsf{P}\{F\geqslant F_\gamma\mid H_0\}=\gamma.$$

Будем принимать гипотезу H_0 тогда и только тогда, когда

$$F_{1-\alpha/2}\leqslant F\leqslant F_{\alpha/2}.$$

Этот критерий имеет уровень значимости α.

§ 57. Непараметрические критерии

В математической статистике часто требуется проверить гипотезу, что независимая выборка

$$x_1,\ x_2,\ \ldots,\ x_n \tag{17}$$

взята из генеральной совокупности с функцией распределения $F(x)$. Относительно конкурирующей гипотезы, кроме независимости x_i в (17), других предположений не делается. В этом случае применяются так называемые непараметрические статистические критерии, которые строятся на основе какой-либо статистики $g(x_1, \ldots$ $\ldots, x_n, F)$, зависящей от F, причем распределение этой статистики при справедливости основной гипотезы известно точно или асимптотически при $n\to\infty$. Обычно статистика положительна, и при любой конкурирующей гипотезе ее значение возрастает. Выбирается такое g_α, чтобы $g\geqslant g_\alpha$ при основной гипотезе выполнялось с вероятностью ошибки первого рода α. Основная гипотеза принимается, если $g<g_\alpha$, и отвергается, если $g\geqslant g_\alpha$. Одним из наиболее известных таких критериев является *χ^2-критерий Пирсона.*

Выберем точки $z_0=-\infty<z_1<z_2<\ldots<z_{r-1}<$ $<z_r=\infty$. По известной функции $F(x)$ вычисляем веро-

ятности $p_k = F(z_k) - F(z_{k-1})$, $k = 1, \ldots, r$. Обозначим ν_k число тех x_i из выборки (17), которые удовлетворяют условию $z_{k-1} < x_i \leqslant z_k$. Тогда при справедливости основной гипотезы случайные величины

$$\nu_1, \nu_2, \ldots, \nu_r \tag{18}$$

имеют полиномиальное распределение

$$\mathsf{P}\{\nu_i = n_i, \ i = 1, \ldots, r\} = \frac{n!}{n_1! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r},$$
$$n_1 + \ldots + n_r = n. \tag{19}$$

Первоначальную задачу мы редуцируем теперь к проверке гипотезы о том, что частоты (18) получены из полиномиального распределения (19) с вероятностями исходов

$$p_1, p_2, \ldots, p_r.$$

Статистика, на основе которой строится критерий, называется χ^2-*статистикой Пирсона* и определяется суммой

$$\chi_n^2 = \sum_{k=1}^{r} \frac{(\nu_k - np_k)^2}{np_k}. \tag{20}$$

Теорема 2. *Распределение χ_n^2 при $n \to \infty$ слабо сходится к χ^2-распределению с $(r-1)$-й степенью свободы с функцией распределения*

$$K_{r-1}(x) = \frac{1}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r-1}{2}\right)} \int_0^x u^{\frac{r-1}{2}-1} e^{-\frac{u}{2}} du, \qquad x \geqslant 0.$$

Доказательство. Из теоремы 8 § 46 следует, что случайный вектор $\eta^{(n)} = \left(\eta_1^{(n)}, \ldots, \eta_r^{(n)}\right)$ с компонентами

$$\eta_k^{(n)} = \frac{\nu_k - np_k}{\sqrt{np_k}} \tag{21}$$

сходится при $n \to \infty$ к нормальному распределению с нулевыми средними и матрицей ковариации

$$\| \mathrm{Cov}(\eta_k, \eta_l) \| = \| \delta_{kl} - \sqrt{p_k p_l} \|$$

(будем обозначать η_k случайные величины с предельным нормальным распределением). Характеристическая функция предельного нормального распределения равна $e^{-\frac{1}{2}Q_\eta(t)}$, где

$$Q_\eta(t)=\sum_{k=1}^{r}t_k^2-\sum_{k,l}t_kt_l\sqrt{p_kp_l}=\sum_{k=1}^{r}t_k^2-\left(\sum_{k=1}^{r}\sqrt{p_k}\,t_k\right)^2. \quad (22)$$

Преобразуем вектор η ортогональным преобразованием $\xi=C\eta$. Тогда характеристические функции f_ξ и f_η связаны соотношением

$$f_\xi(u)=f_\eta(C^*u),\quad u=(u_1,\ \ldots,\ u_r),$$

где C^* — преобразование, сопряженное C (см. свойство 7) из § 43). Если $t=C^*u$, то $u=Ct$, $t=(t_1,\ \ldots$ $\ldots,\ t_r)$. Выберем C так, чтобы

$$u_1=\sqrt{p_1}\,t_1+\ \ldots\ +\sqrt{p_r}\,t_r, \quad (23)$$

а остальные $u_k=\sum_{l=1}^{r}c_{kl}t_l$ подберем так, чтобы матрица

$$C=\left\|\begin{matrix}\sqrt{p_1} & \sqrt{p_2} & \ldots & \sqrt{p_r}\\ c_{21} & c_{22} & \ldots & c_{2r}\\ \cdot & \cdot & \cdot & \cdot\\ c_{r1} & c_{r2} & \ldots & c_{rr}\end{matrix}\right\|$$

была ортогональной. Тогда в $f_\xi(u)=e^{-Q_\xi(u)/2}$ квадратичная форма, в силу (22) и (23), равна

$$Q_\xi(u)=Q_\eta(C^*u)=\sum_{k=1}^{r}t_k^2-u_1^2=\sum_{k=1}^{r}u_k^2-u_1^2=\sum_{k=2}^{r}u_k^2,$$

т. е. вектор $\xi=(\xi_1,\ \ldots,\ \xi_r)$ имеет независимые компоненты, причем $D\xi_1=0$, $\xi_2,\ \ldots,\ \xi_r$ распределены нормально с параметрами (0, 1). Из предельной теоремы и формул (20), (21) следует, что $\chi_n^2=\eta_1^{(n)^2}+\ \ldots\ +\eta_r^{(n)^2}$ сходится к распределению $\eta_1^2+\ \ldots\ +\eta_r^2$. В силу ортогональности преобразования C сумма $\eta_1^2+\ \ldots\ +\eta_r^2$ переходит в сумму $\xi_1^2+\xi_2^2+\ \ldots\ +\xi_r^2$, что и доказывает теорему (см. определение χ^2-распределения в § 46).

Полученный результат применяется следующим способом. Задаемся уровнем значимости α. Тогда, в силу теоремы 2, при больших n с вероятностью, приближенно равной α, выполняется неравенство

$$\chi_n^2 = \sum_{k=1}^{r} \frac{(\nu_k - np_k)^2}{np_k} \geqslant k_\alpha(r-1),$$

где $k_\alpha(r-1)$ — α-квантиль χ^2-распределения с $(r-1)$-й степенью свободы (см. § 64, п. в)). Мы считаем основную гипотезу принятой, если $\chi_n^2 < k_\alpha(r-1)$, и отвергнутой, если выполнено обратное неравенство. Выбор точек деления $z_1, \ldots, z_{r-1}$ должен удовлетворять двум требованиям. Во-первых, вероятности $p_1, p_2, \ldots, p_r$ должны достаточно хорошо отражать вид функции распределения $F(x)$ (для этого r должно быть больше и p_i — меньше). Во-вторых, для того чтобы можно было пользоваться предельной теоремой, np_i и, соответственно, ν_i должны быть не очень маленькими (для этого r должно быть не очень большим). Обычно на практике требуют, чтобы

$$np_i \geqslant 10, \quad \nu_i \geqslant 10.$$

Из этих противоположных требований и выбираются точки $z_1, \ldots, z_{r-1}$.

Другим примером непараметрического критерия является *критерий Колмогорова*. Этот критерий основан на статистике

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|, \tag{24}$$

где $F(x)$ — непрерывная функция распределения генеральной совокупности, $F_n(x)$ — эмпирическая функция распределения, построенная по выборке (1):

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I_{\{x_k \leqslant x\}}.$$

Докажем, что распределение случайной величины D_n инвариантно относительно $F(x)$.

Теорема 3. *Если $F(x)$ непрерывна, то распределение статистики D_n не зависит от $F(x)$.*

Доказательство. Докажем, что D_n при любой непрерывной $F(x)$ имеет такое же распределение, как и в случае, когда $F(x)$ задает равномерное распределение на отрезке $[0, 1]$.

Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины и каждая из них имеет функцию распределения $F(x)$. Предположим, что $F(a) = 0$, $F(b) = 1$ и $0 < F(x) < 1$ при $a < x < b$, причем a и b могут быть и бесконечными. Обозначим через B множество, состоящее из тех точек $a < x < b$, для которых при любом $\varepsilon > 0$ $F(x) < F(x+\varepsilon)$. Нетрудно видеть, что при любом $0 < y < 1$ существует единственная точка $x \in B$, для которой $F(x) = y$. Примем это x за значение обратной функции: $x = F^{-1}(y)$.

Введем случайные величины $\eta_k = F^{-1}(\xi_k)$, $k = 1, \ldots, n$. Они независимы, так как $\xi_1, \ldots, \xi_n$ независимы, и равномерно распределены в $(0, 1)$, так как события $\{\eta_k \leqslant y\}$ и $\{\xi_k \leqslant F^{-1}(y)\}$ равносильны и при любом $0 < y < 1$

$$\mathsf{P}\{\eta_k \leqslant y\} = \mathsf{P}\{\xi_k \leqslant F^{-1}(y)\} = F(F^{-1}(y)) = y.$$

Обозначим более подробно эмпирические функции распределения для выборок $\xi_1, \ldots, \xi_n$

$$F_n(x; \xi_1, \ldots, \xi_n) = \frac{1}{n} \sum_{k=1}^{n} I_{\{\xi_k \leqslant x\}}$$

и $\eta_1, \ldots, \eta_n$

$$F_n(y; \eta_1, \ldots, \eta_n) = \frac{1}{n} \sum_{k=1}^{n} I_{\{\eta_k \leqslant y\}}.$$

Положим $y = F(x)$, $x \in B$. Тогда из равносильности событий $\{\xi_k \leqslant x\}$ и $\{\eta_k \leqslant y\}$ следует

$$F_n(y; \eta_1, \ldots, \eta_n) = F_n(x; \xi_1, \ldots, \xi_n). \tag{25}$$

Верхнюю грань в (24) можно брать по $x \in B$, поэтому, в силу (25), с вероятностью 1

$$D_n = \sup_{x \in B} |F_n(x; \xi_1, \ldots, \xi_n) - F(x)| = \sup_{0<y<1} |F_n(y; \eta_1, \ldots, \eta_n) - y|,$$

что нам и требовалось доказать.

А. Н. Колмогоров доказал, что при $n \to \infty$ для любой непрерывной $F(x)$ имеет место следующая предельная теорема:

$$\lim_{n\to\infty} \mathsf{P}\{\sqrt{n}\, D_n \leqslant x\} = K(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2},\ x>0. \quad (26)$$

На основе предельного соотношения (26) строится непараметрический критерий Колмогорова. Пусть k_α — α-квантиль предельного распределения (26)

$$1 - K(k_\alpha) = \alpha.$$

Тогда гипотеза о том, что выборка (17) взята из распределения с функцией $F(x)$, принимается, если $\sqrt{n}\, D_n \leqslant k_\alpha$, и отвергается, если $\sqrt{n}\, D_n > k_\alpha$. Уровень значимости этого критерия равен приближенно α.

С той же самой предельной функцией $K(x)$ связан *критерий Смирнова*. Он состоит в следующем. Пусть $x_1, \ldots, x_{n_1}$ и $y_1, \ldots, y_{n_2}$ — две независимые выборки, первая имеет функцию распределения $F(x)$, вторая — $G(x)$. Обозначим

$$D_{n_1, n_2} = \sup_{-\infty<x<\infty} | F_{n_1}(x;\ x_1, \ldots, x_{n_1}) - F_{n_2}(x;\ y_1, \ldots, y_{n_2}) |.$$

Н. В. Смирнов доказал, что если $F(x) = G(x)$ и непрерывны, то при $n_1, n_2 \to \infty$, $\frac{n_1}{n_2} \to \tau$, $0 < \tau < \infty$, случайная величина $\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\, D_{n_1, n_2}$ в пределе имеет тот же закон распределения $K(x)$, определенный рядом (26). Эта предельная теорема позволяет нам строить критерий по проверке гипотезы о том, что выборки $x_1, \ldots, x_{n_1}$ и $y_1, \ldots, y_{n_2}$ взяты из одного и того же распределения.

Задачи

1. Имеется независимая выборка $x_1, \ldots, x_n$. По гипотезе H_0 все x_i равномерно распределены в $[0, 2]$, по гипотезе все H_1 равномерно распределены в $[1, 3]$. Построить критерий с наименьшей величиной $max(\alpha, \beta)$, где α и β — вероятности ошибок 1-го и 2-го рода.

2. Пусть $x_1, \ldots, x_n$ — независимая выборка из распределения с плотностью $\lambda e^{-\lambda x}$, $x \geqslant 0$. Построить оптимальный критерий проверки гипотезы H_0: $\lambda = \lambda_0$ при конкурирующей гипотезе H_1: $\lambda = \lambda_1 < \lambda_0$ с уровнем значимости α. Вычислить мощность W этого критерия.

3. По двум независимым выборкам: $x_1, \ldots, x_n$ из нормального распределения (a_1, σ_1) и $y_1, \ldots, y_{n_2}$ из нормального распределения

(a_2, σ_2) построить с уровнем значимости α оптимальный критерий проверки гипотезы H_0: $a_1 = a_2$ при конкурирующей гипотезе H_1: $a_1 < a_2$. Параметры σ_1 и σ_2 считать известными.

4. В таблице случайных чисел на 1000 знаков цифры 0, 1, ..., 9 встретились следующее число раз:

цифры	0	1	2	3	4	5	6	7	8	9
частота	90	105	112	97	108	101	93	87	103	104

С помощью χ^2-критерия Пирсона проверить гипотезу о том, что все цифры встречаются в таблице случайных чисел равновероятно. За уровень значимости принять $\alpha = 0{,}05$.

Глава 15. ОЦЕНКИ ПАРАМЕТРОВ

§ 58. Статистические оценки и их свойства

Мы будем иметь дело с независимой выборкой

$$x_1,\ x_2,\ \ldots,\ x_n \tag{1}$$

из распределения $F(x)$, принадлежащего некоторому семейству распределений $\mathcal{F}$. Пусть θ — параметр, однозначно определяемый по каждому распределению F из семейства $\mathcal{F}$. Например, $\theta = \int\limits_{-\infty}^{\infty} x\,dF(x)$ или $\theta = \int\limits_{-\infty}^{\infty} x^r\,dF(x)$ и т. п. Таким образом, $\theta = \theta(F)$ — это функционал распределения $F \in \mathcal{F}$. Очень часто мы будем предполагать, что само семейство $\mathcal{F}$ определяется одним или несколькими такими параметрами. Тогда любая $F \in \mathcal{F}$ есть функция распределения $F(x;\theta)$ или $F(x;\theta_1,\ \ldots,\ \theta_r)$, зависящая от одного или нескольких параметров. Такое семейство распределений называется *параметрическим.* В любом из этих случаев задача оценки (или, как еще говорят, оценивания) параметра θ состоит в нахождении такой функции

$$\hat{\theta} = \hat{\theta}(x_1,\ \ldots,\ x_n) \tag{2}$$

от выборки (1), которая в каком-либо смысле близка к параметру θ, если выборка взята из распределения F с $\theta(F) = \theta$. При этом предполагается, что функция (2) не зависит от значения оцениваемого параметра θ и других неизвестных параметров, от которых может зависеть F. Вообще, любая функция вида (2) от выборки носит название *статистики.* Таким образом, $\hat{\theta}$ — это статистика (2). Содержательность это определение приобретает только тогда, когда мы налагаем на эту статистику дополнительные условия, обеспечивающие ее близость к параметру θ.

Оценка $\hat\theta = \hat\theta(x_1, \ldots, x_n)$ называется *несмещенной*, если при любом возможном θ

$$\mathsf{M}\hat\theta = \theta, \tag{3}$$

т. е. среднее значение $\hat\theta$ равно θ. Значение свойства (3) можно пояснить на примере большого числа N независимых выборок объема n из одного и того же распределения. Обозначим $\hat\theta_i$ значение оценки (2) для i-й выборки. Если оценка несмещенная, то $\mathsf{M}\hat\theta_i = \theta$, $\hat\theta_1, \ldots, \hat\theta_N$, независимы и одинаково распределены. Тогда по усиленному закону больших чисел

$$\frac{\hat\theta_1 + \ldots + \hat\theta_N}{N} \xrightarrow{\text{п. н.}} \theta.$$

Если конечна дисперсия $\mathsf{D}\hat\theta_i = \sigma^2$, то по центральной предельной теореме разность

$$\frac{\hat\theta_1 + \ldots + \hat\theta_N}{N} - \theta$$

будет $(0, \sigma/\sqrt{N})$-асимптотически нормальна, т. е. при больших N неравенство

$$\left|\frac{\hat\theta_1 + \ldots + \hat\theta_N}{N} - \theta\right| \leqslant \frac{u_{\alpha/2}\sigma}{\sqrt{N}}$$

выполняется приближенно с вероятностью $1 - \alpha$ (здесь $u_{\alpha/2}$ — квантиль нормального распределения, определенная формулой (14) из § 55).

Приведем примеры несмещенных оценок. Если выборка (1) взята из семейства с конечным r-м моментом $m_r = \int x^r \, dF(x)$, то выборочный r-й момент

$$\hat m_r = \frac{1}{n} \sum_{i=1}^{n} x_i^r \tag{4}$$

будет несмещенной оценкой m_r, так как

$$\mathsf{M}\hat m_r = \frac{1}{n} \sum_{i=1}^{n} \mathsf{M}x_i^r = m_r.$$

В частности, выборочное среднее $\bar{x}$ есть несмещенная оценка математического ожидания $a = \int x\,dF(x)$. Выборочная дисперсия

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

не является несмещенной оценкой дисперсии $\sigma^2 = \int (x-a)^2\,dF(x)$, так как s^2 можно представить в виде

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

Отсюда

$$\mathsf{M}s^2 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2, \tag{5}$$

поскольку $\mathsf{M}(x_i - a)^2 = \sigma^2$, $\mathsf{M}(\bar{x} - a)^2 = \frac{\sigma^2}{n}$. Равенство (5) дает нам возможность построить несмещенную оценку дисперсии

$$s_1^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{6}$$

Заметим, что из несмещенности оценки s_1^2 для σ^2 не следует несмещенность оценки s_1 для σ. Поэтому при большом числе N выборок (1) для оценки σ предпочтительнее пользоваться оценкой $\sqrt{\frac{1}{N}\sum_{i=1}^{N}s_{1i}^2}$, а не $\frac{1}{N}\sum_{i=1}^{N}s_{1i}$, где s_{1i}^2 — значение выборочной несмещенной дисперсии (6) для i-й выборки. Заметим, что обычно вместо s_1^2 в (6) пользуются обозначением s^2. Очень часто нас интересуют асимптотические свойства оценок $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$ для выборок (1) объема $n \to \infty$. Оценка $\hat{\theta}_n$ (вернее, последовательность оценок $\hat{\theta}_n$) называется *состоятельной*, если при $n \to \infty$ она сходится по вероятности к параметру θ

$$\hat{\theta}_n \xrightarrow{P} \theta.$$

Примером состоятельной оценки может служить выборочный r-й момент $\hat{m}_r$ в (4), так как при конечности m_r по усиленному закону больших чисел $\hat{m}_r \xrightarrow{\text{п. н.}} m_r$ при $n \to \infty$, а следовательно, и $\hat{m}_r \xrightarrow{\mathsf{P}} m_r$.

Для установления состоятельности оценки $\hat{\theta}_n$ полезна следующая

Теорема 1. *Если* $\mathsf{M}\hat{\theta}_n \to \theta$ *и* $\mathsf{D}\hat{\theta}_n \to 0$ *при* $n \to \infty$, *то оценка* $\hat{\theta}_n$ *— состоятельная.*

Доказательство. По неравенству Чебышева при любом $\varepsilon > 0$

$$\mathsf{P}\{|\hat{\theta}_n - \mathsf{M}\hat{\theta}_n| > \varepsilon\} \leqslant \frac{\mathsf{D}\hat{\theta}_n}{\varepsilon^2} \to 0. \tag{7}$$

Из (7) и неравенства

$$|\hat{\theta}_n - \theta| \leqslant |\hat{\theta}_n - \mathsf{M}\hat{\theta}_n| + |\mathsf{M}\hat{\theta}_n - \theta|$$

следует, что при $n \to \infty$ вероятность события $|\hat{\theta}_n - \theta| > \varepsilon$ стремится к нулю, что и требовалось доказать.

С помощью теоремы 1 во многих случаях легко доказывается состоятельность оценок $\hat{\theta}_n$.

§ 59. Условные законы распределения

Рассмотрим сначала случай, когда вектор $\xi = (\xi_1, \ldots, \xi_n)$ имеет дискретное распределение

$$\mathsf{P}\{\xi = x\} = p_\xi(x) = p(x) = p(x_1, \ldots, x_n),$$

где $x = (x_1, \ldots, x_n)$ пробегает конечное или счетное множество возможных значений ξ, $p(x) \geqslant 0$, $\sum\limits_x p(x) = 1$. Пусть имеется функция $t(x) = t(x_1, \ldots, x_n)$. Условным распределением ξ при условии $t(\xi) = t$ назовем совокупность условных вероятностей при фиксированном t:

$$p(x \mid t) = \mathsf{P}\{\xi = x \mid t(\xi) = t\} =$$
$$= \frac{\mathsf{P}\{\xi = x,\ t(\xi) = t\}}{\mathsf{P}\{t(\xi) = t\}} = \frac{p(x)}{\sum\limits_{x' : t(x') = t} p(x')}. \tag{8}$$

Не более чем счетное число вероятностей (8) отличны от нуля; t мы выбираем такими, чтобы знаменатель в (8) не был равен нулю. Если $g(x)$ — числовая функция от векторного аргумента $x=(x_1, \ldots, x_n)$, то $\eta = g(\xi)$ будет случайной величиной. Ее математическое ожидание равно

$$\mathsf{M}\eta = \mathsf{M}g(\xi) = \sum_x g(x)\,p(x).$$

Условное математическое ожидание $\mathsf{M}\{g(\xi)\,|\,t(\xi)=t\}$ определим с помощью условного распределения (8):

$$\mathsf{M}\{g(\xi)\,|\,t(\xi)=t\} = \sum_{x:\ t(x)=t} g(x)\,p(x\,|\,t) = \frac{\sum\limits_{x:\ t(x)=t} g(x)\,p(x)}{\sum\limits_{x:\ t(x)=t} p(x)}. \tag{9}$$

Как видно из (9), условное математическое ожидание $\mathsf{M}\{g(\xi)\,|\,t(\xi)=t\}$ есть функция от t. Обозначим ее $g_1(t)$. Подставляя вместо t случайную величину $\tau = t(\xi)$, мы получаем, что условное математическое ожидание есть случайная величина $g_1(\tau)$. Вычислим математическое ожидание от $g_1(\tau)$:

$$\mathsf{M}g_1(\tau) = \sum_t g_1(t)\,\mathsf{P}\{\tau=t\} = \sum_t g_1(t) \sum_{x:\ t(x)=t} p(x) = \sum_t \sum_{x:\ t(x)=t} g(x)\,p(x) = \sum_x g(x)\,p(x).$$

Таким образом, мы показали, что

$$\mathsf{M}g(\xi) = \mathsf{M}\,[\mathsf{M}\{g(\xi)\,|\,t(\xi)=t\}], \tag{10}$$

т. е. при вычислении математического ожидания от $g(\xi)$ сначала можно вычислить условное математическое ожидание $g(\xi)$ при условии $t(\xi)=t$, а затем осреднить это условное математическое ожидание по вероятностям условий.

Формула (10) сохраняет смысл и в том случае, когда ξ имеет не дискретное распределение, а, например, имеет плотность $p(x) = p(x_1, \ldots, x_n)$. Пусть плотность $p(x)$ непрерывна в точке x. Тогда при $\Delta_i \to 0$, $i=1, \ldots, n$,

$$\mathsf{P}\{x_i < \xi_i < x_i + \Delta_i,\quad i=1, \ldots, n\} = p(x)\,\Delta_1 \ldots \Delta_n + o(\Delta_1 \ldots \Delta_n).$$

Вычислим условную вероятность

$$\mathsf{P}\{x_1 < \xi_1 < x_1 + \Delta_1, \ldots, x_m < \xi_m < x_m + \Delta_m \mid x_i < \xi_i < x_i + \Delta_i, \quad i = m+1, \ldots, n\} = \frac{p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)\Delta_1 \ldots \Delta_n + o(\Delta_1 \ldots \Delta_n)}{p_{\xi_{m+1} \ldots \xi_n}(x_{m+1}, \ldots, x_n)\Delta_{m+1} \ldots \Delta_n + o(\Delta_{m+1} \ldots \Delta_n)}.$$

Переходя к пределу по $\Delta_i \to 0$, получаем

$$\frac{1}{\Delta_1 \ldots \Delta_m}\mathsf{P}\{x_1 < \xi_1 < x_1 + \Delta_1, \ldots, x_m < \xi_m < x_m + \Delta_m \mid x_i < \xi_i < x_i + \Delta_i, \quad i = m+1, \ldots, n\} \to \frac{p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)}{p_{\xi_{m+1} \ldots \xi_n}(x_{m+1}, \ldots, x_n)}, \tag{11}$$

где

$$p_{\xi_{m+1} \ldots \xi_n}(x_{m+1}, \ldots, x_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_m, x_{m+1}, \ldots, x_n)\,dx_1 \ldots dx_m$$

Предел левой части (11) естественно назвать *условной плотностью* $\xi_1, \ldots, \xi_m$ *при заданных* $\xi_{m+1}, \ldots, \xi_n$:

$$p_{\xi_1 \ldots \xi_m \mid \xi_{m+1} \ldots \xi_n}(x_1, \ldots, x_m \mid x_{m+1}, \ldots, x_n) = \frac{p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_m, x_{m+1}, \ldots, x_n)}{p_{\xi_{m+1} \ldots \xi_n}(x_{m+1}, \ldots, x_n)}.$$

Математическое ожидание

$$\mathsf{M}g(\xi_1, \ldots, \xi_n) = \int \ldots \int g(x_1, \ldots, x_n)\,p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)\,dx_1 \ldots dx_n$$

можно вычислять по формуле (10), вычислив сначала условное математическое ожидание

$$\mathsf{M}\{g(\xi_1, \ldots, \xi_n) \mid \xi_{m+1} = x_{m+1}, \ldots, \xi_n = x_n\} = \int \ldots \int g(x_1, \ldots, x_n) \times$$

$$\times p_{\xi_1 \ldots \xi_m \mid \xi_{m+1} \ldots \xi_n}(x_1, \ldots, x_m \mid x_{m+1}, \ldots, x_n)\,dx_1 \ldots dx_m = \frac{\int \ldots \int g(x_1, \ldots, x_n)\,p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)\,dx_1 \ldots dx_m}{\int \ldots \int p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)\,dx_1 \ldots dx_m}$$

и осредняя его затем по $p_{\xi_{m+1}\ldots\xi_n}(x_{m+1}, \ldots, x_n)$:

$$\mathbf{M}\{\mathbf{M}[g(\xi_1, \ldots, \xi_n) \mid \xi_{m+1}, \ldots, \xi_n]\} =$$

$$= \int \ldots \int p_{\xi_{m+1}\ldots\xi_n}(x_{m+1}, \ldots, x_n)\, dx_{m+1} \ldots dx_n \int \ldots$$

$$\ldots \int g(x_1, \ldots, x_n)\, p_{\xi_1\ldots\xi_m \mid \xi_{m+1}\ldots\xi_n}(x_1, \ldots, x_m \mid x_{m+1}, \ldots$$

$$\ldots, x_n)\, dx_1 \ldots dx_m =$$

$$= \int \ldots \int g(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n =$$

$$= \mathbf{M} g(\xi_1, \ldots \xi_n). \quad (12)$$

Формулу (12) можно вывести и в более общем случае. Пусть имеются дифференцируемые функции $t_1 = t_1(x)$, $t_2 = t_2(x)$, $\ldots$, $t_m = t_m(x)$. Предположим, что к ним можно подобрать функции $y_j = y_j(x)$, $j = 1, \ldots, n-m$, такие, что преобразование C, задаваемое функциями

$$\begin{aligned} t_i &= t_i(x), \quad i = 1, \ldots, m, \\ y_j &= y_j(x), \quad j = 1, \ldots, n-m, \end{aligned} \quad (13)$$

взаимно однозначно в соответствующей области. Тогда плотности $p_\xi(x)$ и $p_{\tau,\eta}(t, y)$, где $\tau_i = t_i(\xi)$, $\eta_j = y_j(\xi)$, $\tau = (\tau_1, \ldots, \tau_m)$, $\eta = (\eta_1, \ldots, \eta_{n-m})$, $t = (t_1, \ldots, t_m)$, $y = (y_1, \ldots, y_{n-m})$, будут связаны равенством

$$p_\xi(x) = p_{\tau,\eta}(t, y)\, |J|, \quad (14)$$

где J — якобиан преобразования C. Пусть имеется функция $g(\xi_1, \ldots, \xi_n)$. Вычислим условное математическое ожидание $g(\xi_1, \ldots, \xi_n)$ при условии $\tau = t$. Обозначим $x_k(t, y) = x_k$, $k = 1, \ldots, n$, $x(t, y) = (x_1(t, y), \ldots, x_n(t, y))$ функции, задающие обратное преобразование C^{-1}. Тогда

$$\mathbf{M}\{g(\xi) \mid \tau = t\} =$$

$$= \int \ldots \int g(x(t, y))\, p_{\eta\mid\tau}(y \mid t)\, dy_1 \ldots dy_{n-m} =$$

$$= \frac{\int \ldots \int g(x(t, y))\, p_{\eta,\tau}(y, t)\, dy_1 \ldots dy_{n-m}}{\int \ldots \int p_{\eta,\tau}(y, t)\, dy_1 \ldots dy_{n-m}}$$

и

$$\mathsf{M}[\mathsf{M}\{g(\xi)\,|\,\tau\}] =$$

$$= \int \ldots \int \mathsf{M}\{g(\xi)\,|\,\tau = t\}\, p_\tau(t)\, dt_1 \ldots dt_m =$$

$$= \int \ldots \int g(x(t, y))\, p_{\eta,\tau}(y, t)\, dt_1 \ldots dt_m\, dy_1 \ldots dy_{n-m} =$$

$$= \int \ldots \int g(x_1, \ldots, x_n)\, p_{\xi_1 \ldots \xi_n}(x_1, \ldots x_n)\, dx_1 \ldots dx_n =$$

$$= \mathsf{M}\, g(\xi) \quad (15)$$

(здесь мы воспользовались равенством (14)).

§ 60. Достаточные статистики

Понятие достаточной статистики играет важную роль в теории оценок.

Определение 1. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — векторная случайная величина, распределение которой $p(x; \theta)$ зависит от параметра θ, и $t(x) = (t_1(x), \ldots, t_m(x))$ — векторная функция (набор m статистик) от $x = (x_1, \ldots, x_n)$. Мы будем называть $t(x)$ *достаточной статистикой*, если условное распределение $\xi = (\xi_1, \ldots, \xi_n)$ при условии $t(\xi) = t$ не зависит от параметра θ.

Мы будем далее иметь в виду два случая, разобранных в § 59: либо $p_\xi(x; \theta)$ — дискретное распределение вероятностей, либо $p_\xi(x; \theta)$ — n-мерная плотность и существует взаимно однозначное преобразование C: $x = (x_1, \ldots, x_n)$ в $(t, y) = (t_1, \ldots, t_m, y_1, \ldots, y_{n-m})$, задаваемое формулами (13).

Как мы увидим ниже, оценки, зависящие только от достаточных статистик, обладают преимуществами по сравнению с другими оценками. Во-первых, они используют не всю информацию, содержащуюся в выборке (1), а лишь ту ее часть, которая существенна для оценки параметра. Во-вторых, каждой несмещенной оценке $\hat{\theta}$ с конечной дисперсией соответствует другая несмещенная оценка $\hat{\hat{\theta}}$, зависящая от достаточной статистики, с $\mathsf{D}\hat{\hat{\theta}} \leqslant \mathsf{D}\hat{\theta}$.

Прежде всего докажем критерий факторизации, позволяющий легко находить достаточные статистики.

Теорема 2. *Если распределение $p(x;\theta)$ представимо в виде*

$$p(x;\theta)=g(t(x);\theta)h(x), \tag{16}$$

то $t(x)$ есть достаточная статистика.

Доказательство. Рассмотрим сначала дискретное распределение. Согласно формуле (8) условная вероятность $\xi=x$ при условии $t(\xi)=t$ равна

$$p_\theta(x\,|\,t)=\frac{p(x;\theta)}{\sum\limits_{x:\ t(x)=t} p(x;\theta)}. \tag{17}$$

Если выполнено (16), то из (17) получаем

$$p_\theta(x\,|\,t)=\frac{g(t;\theta)\,h(x)}{g(t;\theta)\sum\limits_{x:\ t(x)=t} h(x)}=\frac{h(x)}{\sum\limits_{x:\ t(x)=t} h(x)},$$

т. е. $t(x)$ — достаточная статистика. Если, наоборот, условная вероятность $p_\theta(x\,|\,t)=p(x\,|\,t)$ не зависит от параметра θ, то из теоремы умножения вероятностей имеем

$$p(x;\theta)=p(x\,|\,t)\,p_1(t;\theta),$$

где $p_1(t;\theta)$ — распределение t, т. е. имеет место представление (16).

Если $p(x;\theta)$ — плотность, то будем предполагать, что имеется преобразование (13) и плотности $p_\xi(x;\theta)$ и $p_{\tau,\eta}(t,y;\theta)$ связаны соотношением (14). Тогда условная плотность η при условии $\tau=t$, равная

$$p_{\eta\,|\,\tau}(y\,|\,t)=$$
$$=\frac{p_{\tau,\eta}(t_1,\ldots,t_m,y_1,\ldots,y_{n-m})}{\int\ldots\int p_{\tau,\eta}(t_1,\ldots,t_m,y_1,\ldots,y_{n-m})\,dy_1\ldots dy_{n-m}}$$

в силу (14) и (16), представима в виде

$$p_{\eta\,|\,\tau}(y\,|\,t)=$$
$$=\frac{g(t;\theta)\,h(x(t,y))\,J^{-1}(x(t,y))}{\int\ldots\int g(t;\theta)\,h(x(t,y))\,J^{-1}(x(t,y))\,dy_1\ldots dy_{n-m}}=$$
$$=\frac{h(x(t,y))\,J^{-1}(x(t,y))}{\int\ldots\int h(x(t,y))J^{-1}(x(t,y))\,dy_1\ldots dy_{n-m}}$$

и, следовательно, не зависит от θ. Так как

$$\mathsf{M}\{g(\xi)\mid \tau=t\}=\int\ldots\int g(x(t,\,y))\,p_{\eta\mid\tau}(y\mid t)\,dy_1\ldots dy_{n-m}$$

не зависит от θ, то взяв $g(x)=1$ для $x\in B$ и $g(x)=0$ для $x\notin B$, где $B\in\mathscr{B}^n$ — борелевское множество из R^n, получаем, что $\mathsf{P}\{\xi\in B\mid\tau=t\}$ не зависит от θ при любом $B\in\mathscr{B}^n$, т. е. t — достаточная статистика. Пусть, наоборот $p_{\eta\mid\tau}(y\mid t)$ не зависит от θ. Тогда из

$$p_{\eta,\,\tau}(y,\,t;\,\theta)=p_{\eta\mid\tau}(y\mid t)p_\tau(t;\,\theta)$$

и (14) имеем

$$p_\xi(x;\,\theta)=p_{\eta\mid\tau}(y\mid t)\,p_\tau(t;\,\theta)\,|J|,$$

т. е. плотность $p_\xi(x;\theta)$ представима в виде (16). Теорема доказана.

Второе из указанных выше свойств достаточных статистик вытекает из следующей теоремы.

Теорема 3. (Теорема Колмогорова — Блекуэлла.) *Пусть t — достаточная статистика семейства распределений $p(x;\theta)$, а $\hat\theta(x)$ — несмещенная оценка параметра θ с конечной дисперсией, построенная по выборке* (1). *Тогда условное математическое ожидание $\hat\theta$ при фиксированном t*

$$\hat{\hat\theta}=\mathsf{M}\{\hat\theta\mid t\}$$

будет несмещенной оценкой θ с дисперсией $\mathsf{D}\hat{\hat\theta}\leqslant\mathsf{D}\hat\theta$.

Доказательство. Из свойства (15) имеем

$$\mathsf{M}\hat{\hat\theta}=\mathsf{M}[\mathsf{M}(\hat\theta\mid t)]=\mathsf{M}\hat\theta=\theta,$$

т. е. оценка $\hat{\hat\theta}$ несмещена ($\hat{\hat\theta}$ действительно является оценкой, так как не зависит от θ, поскольку t — достаточная статистика). Вычислим $\mathsf{D}\hat\theta$:

$$\mathsf{D}\hat\theta=\mathsf{M}(\hat\theta-\theta)^2=\mathsf{M}(\hat\theta-\hat{\hat\theta}+\hat{\hat\theta}-\theta)^2=$$
$$=\mathsf{M}(\hat\theta-\hat{\hat\theta})^2+\mathsf{M}(\hat{\hat\theta}-\theta)^2+2\mathsf{M}(\hat\theta-\hat{\hat\theta})(\hat{\hat\theta}-\theta). \quad (18)$$

Так как

$$\mathsf{M}(\hat\theta-\hat{\hat\theta})(\hat{\hat\theta}-\theta)=\mathsf{M}[\mathsf{M}(\hat\theta-\hat{\hat\theta})(\hat{\hat\theta}-\theta)\mid t]=$$
$$=\mathsf{M}[(\hat{\hat\theta}-\theta)\cdot\mathsf{M}\{(\hat\theta-\hat{\hat\theta})\mid t\}],$$

а $\mathsf{M}\{(\hat{\theta} - \hat{\hat{\theta}}) \mid t\} = 0$, то из (18) получаем $\mathsf{D}\hat{\theta} \geqslant \mathsf{D}\hat{\hat{\theta}}$. Теорема доказана.

Пример 1. Пусть выборка (1) взята из схемы Бернулли ($x_i = 1$, если в i-м испытании был успех, $x_i = 0$ в противоположном случае). Параметром в этом случае служит вероятность p. Вероятность появления выборки (1) равна

$$\mathsf{P}(x;\ p) = \prod_{k=1}^{n} p^{x_k} q^{1-x_k} = p^{x_1 + \ldots + x_n} (1-p)^{n - x_1 - \ldots - x_n},$$

откуда по критерию факторизации следует, что число успехов $x_1 + \ldots + x_n$ есть достаточная статистика.

Пример 2. Пусть (1) — независимая выборка из нормального распределения с параметрами (a, σ). Тогда по критерию факторизации

$$p(x;\ a,\ \sigma) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2\right\} =$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2a\sum_{i=1}^{n} x_i + na^2\right)\right\},$$

т. е. $\sum_{i=1}^{n} x_i$ и $\sum_{i=1}^{n} x_i^2$ — достаточные статистики.

§ 61. Эффективность оценок

Как мы видели в § 60, несмещенные оценки $\hat{\theta}$ параметра θ с меньшей дисперсией предпочтительней остальных оценок. Естественно поставить вопрос о нахождении оценок с наименьшей дисперсией. Некоторый подход к решению этого вопроса дает неравенство Рао — Крамера. Пусть $p(x;\theta) = p(x_1, \ldots, x_n; \theta)$ — плотность, зависящая от параметра θ, а $\hat{\theta} = \varphi(x) = \varphi(x_1, \ldots, x_n)$ — оценка параметра θ по выборке $x_1, \ldots, x_n$, не обязательно несмещенная. Обозначим $g(\theta) = \mathsf{M}\hat{\theta} = \int \ldots \int \varphi(x) \times p(x;\theta)\, dx$. Предположим, что выполнены некоторые условия регулярности, при которых интегралы

$$\int p(x;\ \theta)\, dx \equiv 1, \qquad \int \varphi(x)\, p(x;\ \theta)\, dx \equiv g(\theta)$$

можно дифференцировать по параметру θ. В этом случае справедливы равенства

$$\int \frac{\partial p}{\partial \theta}\, dx \equiv 0, \qquad (19)$$

$$\int \varphi(x) \frac{\partial p}{\partial \theta}\, dx = g'(\theta). \qquad (20)$$

Математическое ожидание (здесь ξ имеет распределение $p(\xi, \theta)$)

$$J(\theta) = \mathsf{M}\left(\frac{\partial \log p(\xi;\ \theta)}{\partial \theta}\right)^2 = \int \left(\frac{\partial \log p(x;\ \theta)}{\partial \theta}\right)^2 p(x;\ \theta)\, dx \qquad (21)$$

носит название *информации Фишера относительно семейства* $p(x; \theta)$.

Теорема 4. (Неравенство Рао — Крамера.) *Если семейство плотностей* $p(x; \theta)$ *и оценка* $\hat{\theta} = \varphi(x)$ *таковы, что выполнены условия* (19) *и* (20), *то имеет место неравенство*

$$\mathsf{D}\hat{\theta} \geqslant \frac{[g'(\theta)]^2}{J(\theta)}. \qquad (22)$$

Доказательство. Условия (19) и (20) перепишем в эквивалентном виде:

$$\begin{aligned} &\int p(x;\ \theta) \frac{\partial \log p(x;\ \theta)}{\partial \theta}\, dx \equiv 0, \\ &\int \varphi(x)\, p(x;\ \theta) \frac{\partial \log p}{\partial \theta}\, dx \equiv g'(\theta). \end{aligned} \qquad (23)$$

Умножим первое из тождеств (23) на $g(\theta)$ и вычтем его из второго:

$$\int [\varphi(x) - g(\theta)] \frac{\partial \log p}{\partial \theta}\, p(x;\ \theta)\, dx \equiv g'(\theta). \qquad (24)$$

Полагая в (24) $\varphi_1(x) = \varphi(x) - g(\theta)$, $\varphi_2(x) = \frac{\partial p}{\partial \theta}$, применим неравенство Коши — Буняковского

$$\left[\int \varphi_1(x)\, \varphi_2(x)\, p(x;\ \theta)\, dx\right]^2 \leqslant$$
$$\leqslant \int \varphi_1^2(x)\, p(x;\ \theta)\, dx \cdot \int \varphi_2^2(x)\, p(x;\ \theta)\, dx.$$

Имеем отсюда:

$$[g'(\theta)]^2 = \left[\int (\varphi(x) - g(\theta)) \frac{\partial \log p}{\partial \theta} p(x; \theta)\, dx\right]^2 \leqslant$$
$$\leqslant (\varphi(x) - g(\theta))^2 p(x; \theta)\, dx \cdot \int \left(\frac{\partial \log p}{\partial \theta}\right)^2 p(x; \theta)\, dx,$$

а это равносильно неравенству (22).

Замечание 1. Теорема 4 остается справедливой, если под $p(x; \theta)$ понимать вероятности дискретного распределения, а под интегралами — суммы.

Замечание 2. Если тождества (19) можно еще раз дифференцировать по θ:

$$\int \frac{\partial^2 \log p}{\partial \theta^2}\, dx \equiv 0,$$

то информацию Фишера (21) можно записать в другом виде:

$$J(\theta) = -\mathsf{M}\, \frac{\partial^2 \log p(\xi; \theta)}{\partial \theta^2} = -\int \frac{\partial^2 \log p}{\partial \theta^2} p(x; \theta)\, dx. \quad (25)$$

В самом деле, обозначая $p' = \dfrac{\partial p}{\partial \theta}$, $p'' = \dfrac{\partial^2 p}{\partial \theta^2}$, имеем

$$\frac{\partial \log p}{\partial \theta} = \frac{p'}{p}, \qquad \frac{\partial^2 \log p}{\partial \theta^2} = \frac{p''}{p} - \left(\frac{p'}{p}\right)^2,$$

откуда

$$\int \frac{\partial^2 \log p}{\partial \theta^2} p\, dx = \int p''\, dx - \int \left(\frac{p'}{p}\right)^2 p\, dx = -J(\theta),$$

что и утверждалось.

Замечание 3. Из первого тождества (23) следует $\mathsf{M}\, \dfrac{\partial \log p(\xi; \theta)}{\partial \theta} = 0$, поэтому информацию Фишера (21) можно записать иначе:

$$J(\theta) = \mathsf{D}\left(\frac{\partial \log p(\xi, \theta)}{\partial \theta}\right).$$

Замечание 4. Если $x_1, \ldots, x_n$ независимы, то их совместная плотность $p_n(x_1, \ldots, x_n; \theta)$ есть произведение одномерных плотностей:

$$p_n(x_1, \ldots, x_n; \theta) = \prod_{k=1}^{n} p(x_k; \theta).$$

В этом случае информация Фишера $J_n(\theta) = \mathsf{M}\left(\frac{\partial \log p_n}{\partial\theta}\right)^2$ зависит от n линейно:

$$J_n(\theta) = nJ_1(\theta), \tag{26}$$

где $J_1(\theta) = \int \left(\frac{\partial \log p(x;\ \theta)}{\partial\theta}\right)^2 p(x;\theta)\,dx$ — информация Фишера одного наблюдения x_k, а (22) превращается в неравенство следующего вида:

$$\mathsf{D}\hat\theta \geqslant \frac{[g'(\theta)]^2}{nJ_1(\theta)}. \tag{27}$$

Формула (26) следует из

$$J_n(\theta) = \mathsf{D}\left(\sum_{i=1}^{n} \frac{\partial \log p(x_i;\ \theta)}{\partial\theta}\right) = \sum_{i=1}^{n} \mathsf{D}\left(\frac{\partial \log p(x_i;\theta)}{\partial\theta}\right).$$

Замечание 5. Если оценка $\hat\theta$ несмещенная, то $g(\theta) \equiv \theta$, и в неравенствах (22) и (27) числитель равен $g'(\theta) \equiv 1$.

В условиях теоремы 4 неравенства (22) и (27) дают оценку снизу дисперсии оценок $\hat\theta$. Ниоткуда не следует, что эта оценка достигается, однако во многих важных случаях, как мы увидим ниже, она является нижней границей дисперсии θ хотя бы в асимптотическом смысле при $n \to \infty$.

Пример 3. Пусть $x_1, \ldots, x_n$ — независимая выборка из нормального распределения с параметрами $(a,\ \sigma)$, σ — известно. Так как $\log p(x;\ a) = -\frac{(x-a)^2}{2\sigma^2} - \log\sqrt{2\pi}\,\sigma$, $\frac{\partial \log p}{\partial a} = \frac{x-a}{\sigma^2}$, то $J_n(a) = n\frac{\mathsf{M}(\xi - a)^2}{\sigma^4} = \frac{n}{\sigma^2}$. Для оценки $\hat a = \bar x$ имеем $\mathsf{D}\hat a = \frac{\sigma^2}{n} = \frac{1}{nJ_1(a)}$, т. е. в этом случае в (27) достигается равенство.

Ниже мы всегда будем предполагать, что выполнены условия теоремы 4.

Определение 2. Назовем *эффективностью* оценки $\hat\theta$ отношение

$$e(\hat\theta) = \frac{[g'(\theta)]^2}{\mathsf{D}\hat\theta \cdot J(\theta)}.$$

Оценка $\hat{\theta}$ с эффективностью $e(\hat{\theta}) = 1$ называется *эффективной.*

Эти определения обычно применяют к несмещенным оценкам.

Оценка $\bar{x}$ в примере 1 эффективна. Если неравенство в (22) или (27) для некоторой оценки превращается в равенство, то эффективность оценки $\hat{\theta}$ — это отношение минимально возможной дисперсии к дисперсии данной оценки:

$$e(\hat{\theta}) = \frac{\min\limits_{\theta_1} \mathsf{D}\hat{\theta}_1}{\mathsf{D}\hat{\theta}}.$$

Эффективность всегда удовлетворяет неравенствам $0 \leqslant e(\hat{\theta}) \leqslant 1$. Конечно, при нарушении условий теоремы 4 неравенства (22) и (27) могут не выполняться и могут существовать «сверхэффективные» оценки $\hat{\theta}$ с дисперсией $\mathsf{D}\hat{\theta}$, убывающей при $n \to \infty$ быстрее, чем $O\left(\frac{1}{n}\right)$.

Пример 4. Пусть $x_1, \ldots, x_n$ — независимая выборка из распределения с плотностью

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

В этом случае нарушается условие (19) и оценка $\hat{\theta} = \min\limits_{1 \leqslant k < n} x_k$ обладает «сверхэффективностью», так как

$$\mathsf{M}\hat{\theta} = \theta + \frac{1}{n}, \quad \mathsf{D}\hat{\theta} = \frac{1}{n^2}.$$

Важным понятием в теории статистических оценок является также асимптотическая эффективность. Будем предполагать выполненными условия теоремы 1.

Определение 3. *Асимптотической эффективностью* $e_0(\hat{\theta}_n)$ оценки $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$, построенной по независимой выборке $x_1, \ldots, x_n$, назовем предел

$$e_0(\hat{\theta}) = \lim_{n \to \infty} \frac{[g'_n(\theta)]^2}{nJ_1(\theta)\,\mathsf{D}\hat{\theta}_n},$$

если он существует. Оценка $\hat{\theta}_n$ называется *асимтотически эффективной*, если $e_0(\hat{\theta}_n) = 1$.

Таким образом, если $\hat{\theta}$ — несмещенная оценка с асимптотической эффективностью $e_0(\hat{\theta})$, то ее дисперсия $D\hat{\theta}$ при больших n асимптотически равна $[e_0(\hat{\theta})\cdot nJ_1(\theta)]^{-1}$.

Для асимптотически нормальных при $n\to\infty$ оценок $\hat{\theta}_n$ полезно другое определение асимптотической эффективности.

Определение 4. Если оценка $\hat{\theta}_n$ при $n\to\infty$ асимптотически нормальна с параметрами $\left(\theta, \frac{\sigma_\theta}{\sqrt{n}}\right)$, то ее *асимптотической эффективностью* называется отношение

$$e_0(\hat{\theta}_n)=\frac{1}{J_1(\theta)\,\sigma_\theta^2},$$

т. е. в этом случае за математическое ожидание и дисперсию оценки $\hat{\theta}$ мы принимаем математическое ожидание и дисперсию аппроксимирующего нормального закона распределения. Аналогично, если $e_0(\hat{\theta})=1$, то оценка будет называться *асимптотически эффективной.*

С понятием асимптотической эффективности асимптотически нормальных оценок мы встретимся в следующем параграфе.

§ 62. Методы нахождения оценок

До сих пор мы занимались свойствами оценок параметров, не затрагивая вопроса о способах их нахождения. Сейчас мы познакомимся с некоторыми методами нахождения оценок.

1. Метод моментов. Пусть $x_1, \ldots, x_n$ — независимая выборка из распределения с плотностью $p(x;\theta_1, \ldots, \theta_r)$, зависящей от r параметров $\theta_1, \ldots, \theta_r$. Предположим, что все моменты

$$m_k(\theta_1, \ldots, \theta_r)=\int x^k p(x;\theta_1, \ldots, \theta_r)\,dx, \qquad k=1, \ldots, r,$$

конечны и что система уравнений

$$b_k=m_k(\theta_1, \ldots, \theta_r), \qquad k=1, \ldots, r,$$

однозначно разрешима, причем ее решение

$$\theta_k = m_k^{-1}(b_1, \ldots, b_r), \qquad k = 1, \ldots, r,$$

дается непрерывными обратными функциями m_k^{-1}. При этих условиях имеет место

Теорема 5. *Оценки* $\hat{\theta}_k$, $k = 1, \ldots, r$, *получаемые как решение системы*

$$\hat{m}_k = m_k(\theta_1, \ldots, \theta_r), \qquad k = 1, \ldots, r, \tag{28}$$

где $\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ — *выборочные моменты, состоятельны.*

Доказательство. Согласно нашим предположениям система (28) имеет единственное решение $\hat{\theta}_k = m_k^{-1}(\hat{m}_1, \ldots, \hat{m}_r)$, причем m_k^{-1} — непрерывные функции. По усиленному закону больших чисел $\hat{m}_k$ сходятся п. н. к m_k, а из непрерывности функций m_k^{-1} отсюда следует, что $\hat{\theta}_k$ при $n \to \infty$ п. н. сходятся к θ_k. Теорема доказана.

Метод нахождения оценок, описанный в теореме 1, носит название *метода моментов.* Этот метод дает состоятельные оценки, но часто их эффективность и асимптотическая эффективность меньше единицы.

2. Метод наибольшего правдоподобия. Пусть $x_1, \ldots, x_n$ — независимая выборка из распределения с плотностью $p(x; \theta)$, зависящей от параметра θ. Совместную плотность

$$L(x; \theta) = p(x_1; \theta) \ldots p(x_n; \theta),$$

рассматриваемую как функцию параметра θ, называют *функцией правдоподобия.* Оценкой наибольшего правдоподобия называется оценка $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$, которая обращает в максимум функцию правдоподобия:

$$L(x; \hat{\theta}) = \max_{\theta} L(x; \theta).$$

Если функция правдоподобия $L(x; \theta)$ дифференцируема по θ, то оценку наибольшего правдоподобия $\hat{\theta}$ можно найти, решив относительно θ *уравнение правдоподобия*

$$\frac{\partial L(x; \theta)}{\partial \theta} = 0. \tag{29}$$

Установим некоторые свойства оценок наибольшего правдоподобия. Будем предполагать, что выполнены следующие условия:

1) Пусть параметр θ изменяется в интервале $\theta_1 < \theta < \theta_2$ и истинное значение параметра θ_0 **лежит** внутри этого интервала. Предположим, что в этом интервале существуют производные $\frac{\partial p}{\partial \theta}$, $\frac{\partial^2 p}{\partial \theta^2}$, $\frac{\partial^3 p}{\partial \theta^3}$.

2) Интеграл $\int p(x;\theta)\,dx$ можно два раза дифференцировать под знаком интеграла, так что

$$\int \frac{\partial p}{\partial \theta}\,dx \equiv 0, \quad \int \frac{\partial^2 p}{\partial \theta^2}\,dx \equiv 0.$$

3) $J_1(\theta_0) = \int \left(\frac{\partial \log p}{\partial \theta}\right)^2 p(x;\theta)\,dx \Big|_{\theta=\theta_0} > 0$, $\left|\frac{\partial^3 p}{\partial \theta^3}\right| \leqslant H(x)$, и $\mathrm{M}_\theta H(\xi) = \int H(x)\,p(x;\theta)\,dx \leqslant M$, где M не зависит от θ.

Теорема 6. *При выполнении условий* 1), 2), 3) *уравнение правдоподобия* (29) *имеет решение* $\hat{\theta}$, *которое при* $n \to \infty$ *сходится по вероятности к* θ_0. *Эта оценка наибольшего правдоподобия асимптотически нормальна и асимптотически эффективна.*

Доказательство. Уравнение правдоподобия (29) равносильно уравнению

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = \frac{\partial \log L}{\partial \theta} = \sum_{k=1}^{n} \frac{\partial \log p(x_k;\theta)}{\partial \theta} = 0. \tag{30}$$

Разложим $\frac{\partial \log p(x;\theta)}{\partial \theta}$ по формуле Тейлора в окрестности точки θ_0:

$$\frac{\partial \log p}{\partial \theta} = \frac{\partial \log p}{\partial \theta}\Big|_{\theta=\theta_0} + (\theta-\theta_0)\frac{\partial^2 \log p}{\partial \theta^2}\Big|_{\theta=\theta_0} + \frac{(\theta-\theta_0)^2}{2}\,\delta H(x), \tag{31}$$

где $|\delta| \leqslant 1$. Разделив (30) на n и воспользовавшись разложением (31), имеем

$$\frac{1}{n}\frac{\partial \log L}{\partial \theta} = B_0 + B_1(\theta-\theta_0) + \frac{1}{2}\delta \cdot B_2(\theta-\theta_0)^2, \tag{32}$$

где

$$B_0 = \frac{1}{n}\sum_{k=1}^{n} \frac{\partial \log p(x_k;\,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0},$$

$$B_1 = \frac{1}{n}\sum_{k=1}^{n} \frac{\partial^2 \log p(x_k;\,\theta)}{\partial\theta^2}\bigg|_{\theta=\theta_0},$$

$$B_2 = \frac{1}{n}\sum_{k=1}^{n} H(x_k).$$

По закону больших чисел $B_0 \xrightarrow{\mathsf{P}} \mathsf{M}B_0 = 0$, $B_1 \xrightarrow{\mathsf{P}} \mathsf{M}B_1 = -J_1(\theta_0)$ (см. (25) из § 61), $B_2 \xrightarrow{\mathsf{P}} \mathsf{M}B_2$, $|\mathsf{M}B_2| \leqslant M$. Пусть теперь $h > 0$ и $\varepsilon > 0$ фиксированы. Выбираем n_0 таким, чтобы при всех $n \geqslant n_0$

$$\mathsf{P}\{|B_0| \geqslant h^2\} < \frac{\varepsilon}{3}, \quad \mathsf{P}\left\{B_1 > -\frac{J_1(\theta_0)}{2}\right\} < \frac{\varepsilon}{3},$$
$$\mathsf{P}\{|B_2| > 2M\} < \frac{\varepsilon}{3}. \tag{33}$$

Обозначим через S событие, состоящее в том, что одновременно выполняются неравенства

$$|B_0| \leqslant h^2, \quad B_1 \leqslant -\frac{J_1(\theta_0)}{2}, \quad |B_2| \leqslant 2M.$$

В силу (33) $\mathsf{P}(\bar{S}) < \varepsilon$ и $\mathsf{P}(S) > 1 - \varepsilon$. При $\theta = \theta_0 \pm h$ уравнение (32) преобретает вид

$$B_0 \mp B_1 h + \frac{1}{2}\delta B_2 h^2 = 0, \tag{34}$$

В множестве S

$$\left|B_0 + \frac{1}{2}\delta B_2 h^2\right| \leqslant (M+1)h^2,$$

и при $h < \dfrac{J_1(\theta_0)}{2(M+1)}$ знак левой части (31) определяется знаком члена $\mp B_1 h$. Так как $\dfrac{1}{n}\dfrac{\partial \log L}{\partial\theta}$ непрерывно зависит от θ, то в интервале $(\theta_0 - h,\ \theta_0 + h)$ при $n \geqslant n_0$ с вероятностью $\geqslant 1 - \varepsilon$ существует корень $\hat{\theta}$. Таким образом, мы доказали первую часть теоремы. Перепишем равенство

$$B_0 + B_1(\hat{\theta} - \theta_0) + \frac{B_2\delta}{2}(\hat{\theta} - \theta_0)^2 = 0$$

в следующем виде:

$$(\hat{\theta} - \theta^0)\sqrt{J_1(\theta_0)n} = \frac{\dfrac{1}{\sqrt{J_1(\theta_0)n}}\sum_{k=1}^{n}\dfrac{\partial \log p(x_k;\theta)}{\partial\theta}\Big|_{\theta=\theta_0}}{-\dfrac{B_1}{J_1(\theta_0)} - \dfrac{1}{2}\delta B_2\dfrac{\hat{\theta}-\theta_0}{J_1(\theta_0)}}. \quad (35)$$

Числитель в (35) по центральной предельной теореме асимптотически нормален с параметрами (0, 1), а знаменатель при $n \to \infty$ сходится по вероятности к 1. Поэтому случайная величина $(\hat{\theta} - \theta_0)\sqrt{J_1(\theta_0)n}$ асимптотически нормальна с параметрами (0, 1), что и доказывает теорему.

Задачи

1. Случайная величина μ подчиняется биномиальному закону распределения $\mathsf{P}\{\mu = k\} = C_n^k p^k (1-p)^{n-k}$. Найти несмещенные оценки $\hat{a}$ и $\widehat{\sigma^2}$ математического ожидания $a = np$ и дисперсии $\sigma^2 = npq$ этого распределения, считая параметр p неизвестным.

2. Случайная величина ξ имеет геометрическое распределение $\mathsf{P}\{\xi = k\} = pq^k$, $q = 1 - p$, $k = 0, 1, 2, \ldots$ Найти выраженные через ξ несмещенные оценки $\hat{a}$ и $\hat{\sigma}^2$ тех же величин $a = np$ и $\sigma^2 = npq$, что и в задаче 1.

3. Случайная величина ξ имеет распределение Пуассона $\mathsf{P}\{\xi = k\} = \Pi(a; k) = \dfrac{a^k}{k!}e^{-a}$. Найти несмещенную оценку $\varphi_m(\xi)$ вероятности $\Pi(ma; 0) = e^{-ma}$ в законе Пуассона со значеиием параметра ma, где $m = 2, 3 \ldots$

4. Пусть $x_1, x_2, \ldots, x_n$ — независимая выборка из семейства распределений $\mathcal{F}$. Найти достаточные статистики в случае, когда $\mathcal{F}$ есть: а) семейство пуассоновских распределений $p_n = \dfrac{a^n}{h!}e^{-a}$, $n = 0, 1, 2, \ldots$; б) семейство показательных распределений с плотностью $\lambda e^{-\lambda x}$, $x \geqslant 0$; в) семейство равномерных распределений с плотностью

$$p(x) = \begin{cases} \dfrac{1}{c}, & \text{если } 0 \leqslant x \leqslant c, \\ 0 & \text{в остальных случаях.} \end{cases}$$

5. По независимой выборке $x_1, \ldots, x_n$ из нормального распределения с параметрами $(a, 1)$ построена несмещенная оценка $\hat{a} = x_1$. Найти несмещенную оценку $\hat{\hat{a}} = \mathsf{M}(x_1 \mid \bar{x})$, где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ — достаточная статистика.

6. Пусть $x_1, \ldots, x_n$ — независимая выборка из равномерного в $(0, \theta)$ распределения. Найти оценку $\hat{\theta}$ наибольшего правдоподобия параметра θ.

Глава 16. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

§ 63. Определение доверительных интервалов

Пусть

$$x_1, \ldots, x_n \tag{1}$$

— выборка (далее мы всегда будем предполагать, что она независимая) из некоторого распределения с плотностью $p(x; \theta) = p(x_1, \ldots, x_n; \theta)$, зависящей от параметра θ, который может изменяться в интервале $\theta_0 < \theta < \theta_1$. Пусть $y(x_1, \ldots, x_n)$ — некоторая статистика (т. е. функция от выборки) и $F(x; \theta) = \mathbf{P}\{\eta \leqslant x\}$ — функция распределения случайной величины $\eta = y(x_1, \ldots, x_n)$, когда выборка (1) имеет распределение с плотностью $p(x_1, \ldots, x_n; \theta)$. Предположим, что $F(x; \theta)$ есть убывающая функция от параметра θ. Обозначим $x_\gamma(\theta)$ квантиль распределения $F(x; \theta)$, т. е. корень уравнения

$$F(x; \theta) = 1 - \gamma.$$

В этом случае квантиль $x_\gamma(\theta)$ есть возрастающая функция от θ. Зададимся малым числом $\alpha > 0$, например, $\alpha = 0{,}05$ или $\alpha = 0{,}01$. Пусть $\alpha = \alpha_1 + \alpha_2$. При каждом θ неравенства

$$x_{1-\alpha_2}(\theta) \leqslant \eta \leqslant x_{\alpha_1}(\theta) \tag{2}$$

выполняются с вероятностью $1 - \alpha$, близкой к единице. Обозначим функцию, обратную к $x_\gamma(\theta)$, т. е. решение уравнения

$$y = x_\gamma(\theta),$$

через

$$\theta = x_\gamma^{-1}(y).$$

Тогда неравенства (2) можно записать иначе:

$$x_{\alpha_1}^{-1}(\eta) \leqslant \theta \leqslant x_{1-\alpha_2}^{-1}(\eta). \tag{3}$$

Таким образом, неравенства (3) при любом θ выполняются с вероятностью $1-\alpha$. Обозначим $x_{\alpha_1}^{-1}(\eta)=\underline{\theta}(\eta)$, $x_{1-\alpha_2}^{-1}(\eta)=\bar{\theta}(\eta)$ и запишем (3) в следующем виде:

$$\mathsf{P}_\theta\{\underline{\theta}(\eta)\leqslant\theta\leqslant\bar{\theta}(\eta)\}=1-\alpha. \tag{4}$$

Интервал $\underline{\theta}(\eta)\leqslant\theta\leqslant\bar{\theta}(\eta)$ называется *доверительным интервалом* для параметра θ, а вероятность $1-\alpha$ — *доверительной вероятностью*. Следует различать смысл неравенств (2) и (3). В неравенстве (2) при любом θ случайная величина η попадает в указанный интервал с вероятностью $1-\alpha$. В неравенстве (3) параметр θ неслучайный, а концы интервала случайны, поэтому правильнее будет говорить, что при любом θ доверительный интервал (со случайными концами) покрывает параметр θ с доверительной вероятностью $1-\alpha$. Доверительный интервал (4), кроме доверительной вероятности $1-\alpha$, имеет еще одну характеристику — среднюю длину $\mathsf{M}[\bar{\theta}(\eta)-\underline{\theta}(\eta)]$. Мы должны стараться среди всех доверительных интервалов с доверительной вероятностью $1-\alpha$ выбрать тот, который имеет наименьшую длину. Если статистика $\eta=y(x_1, \ldots, x_n)$ уже выбрана, то мы можем варьировать разложение α на сумму $\alpha_1+\alpha_2$.

В дальнейшем мы встретимся со следующими двумя случаями.

Случай 1. Функция распределения $F(x;\theta)$ имеет вид $F(x-\theta)$. В этом случае $F(x-\theta)$ убывает с ростом θ. Легко видеть, что при этом $x_\gamma(\theta)=\theta+x_\gamma(0)$ и $x_\gamma^{-1}(y)=y-x_\gamma(0)$, поэтому доверительный интервал (3) имеет вид

$$\eta-x_{\alpha_1}(0)\leqslant\theta\leqslant\eta-x_{1-\alpha_2}(0). \tag{5}$$

Случай 2. Параметр θ положителен, и $F(x;\theta)=F\left(\frac{x}{\theta}\right)$, $F(0)=0$. В этом случае $F\left(\frac{x}{\theta}\right)$ при $x>0$ убывает с ростом θ, и $x_\gamma(\theta)=\theta x_\gamma(1)$, $x_\gamma^{-1}(y)=\frac{y}{x_\gamma(1)}$. Доверительный интервал (3) в этом случае имеет вид

$$\frac{\eta}{x_{\alpha_1}(1)}\leqslant\theta\leqslant\frac{\eta}{x_{1-\alpha_2}(1)}.$$

§ 64. Доверительные интервалы для параметров нормального распределения

Пусть независимая выборка (1) взята из нормального распределения с параметрами (a, σ).

а) Доверительный интервал для a при известном σ. Возьмем за статистику η среднее арифметическое $\bar{x}$. Это разумно, так как $\bar{x}$ есть достаточная статистика относительно a и является эффективной оценкой a. Как известно, $\bar{x}$ имеет нормальное распределение с параметрами $\left(a, \frac{\sigma}{\sqrt{n}}\right)$. Обозначим, как и раньше, через u_γ квантиль нормального распределения, т. е.

$$1-\Phi(u_\gamma)=\gamma.$$

Пусть $\alpha=\alpha_1+\alpha_2$. Так как $u_{1-\gamma}=-u_\gamma$, то неравенства

$$a-u_{\alpha_2}\frac{\sigma}{\sqrt{n}}\leqslant\bar{x}\leqslant a+u_{\alpha_1}\frac{\sigma}{\sqrt{n}} \tag{6}$$

выполняются с вероятностью $1-\alpha$. Разрешая неравенства (6) относительно a, имеем доверительный интервал для a

$$\bar{x}-u_{\alpha_1}\frac{\sigma}{\sqrt{n}}=\underline{a}\leqslant a\leqslant\bar{a}=\bar{x}+u_{\alpha_2}\frac{\sigma}{\sqrt{n}}, \tag{7}$$

являющийся частным случаем (5). Доверительная вероятность (7) равна $1-\alpha$, а его длина $\left(u_{\alpha_2}+u_{\alpha_1}\right)\frac{\sigma}{\sqrt{n}}$. Эта длина будет наименьшей, если взять $\alpha_1=\sigma_2=\alpha/2$. Пусть, например, $\alpha_2>\alpha_1$; тогда $u_{\alpha_1}>u_{\alpha_2}$. Пусть $\Delta>0$ таково, что $\alpha_2-\Delta>\alpha_1+\Delta$; тогда $u_{\alpha_1}>u_{\alpha_1+\Delta}>u_{\alpha_2-\Delta}>$ $>u_{\alpha_2}$. Из неравенств

$$\Delta=1-\Phi\left(u_{\alpha_2}\right)-\left(1-\Phi\left(u_{\alpha_2-\Delta}\right)\right)=$$

$$=\frac{1}{\sqrt{2\pi}}\int\limits_{u_{\alpha_2}}^{u_{\alpha_2-\Delta}}e^{-\frac{x^2}{2}}\,dx\geqslant\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2_{\alpha_1-\Delta}}{2}}\left(u_{\alpha_2-\Delta}-u_{\alpha_2}\right),$$

$$\Delta=1-\Phi\left(u_{\alpha_1-\Delta}\right)-\left(1-\Phi\left(u_{\alpha_1}\right)\right)=$$

$$=\frac{1}{\sqrt{2\pi}}\int\limits_{u_{\alpha_1+\Delta}}^{u_{\alpha_1}}e^{-\frac{x^2}{2}}\,dx\geqslant\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2_{\alpha_1+\Delta}}{2}}\left(u_{\alpha_1}-u_{\alpha_1+\Delta}\right)$$

имеем

$$u_{\alpha_2-\Delta}-u_{\alpha_2}\leqslant \Delta\sqrt{2\pi}\,e^{\frac{u^2_{\alpha_2-\Delta}}{2}}<\Delta\sqrt{2\pi}\,e^{\frac{u^2_{\alpha_1+\Delta}}{2}}\leqslant u_{\alpha_1}-u_{\alpha_1+\Delta}$$

или

$$u_{\alpha_2-\Delta}+u_{\alpha_1+\Delta}<u_{\alpha_1}+u_{\alpha_2},$$

т. е. при $\alpha_1\neq\alpha_2$ доверительный интервал не имеет наименьшей длины.

б) Доверительный интервал для a при неизвестном σ. Прежде всего докажем важное свойство выборочного среднего

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

и дисперсии

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

для выборки (1) из нормального распределения.

Теорема 1. *Статистики $\bar{x}$ и s^2 для выборки* (1) *из нормального распределения независимы. Случайная величина $s^2(n-1)/\sigma^2$ имеет χ^2-распределение с $(n-1)$-й степенью свободы.*

Доказательство. Случайные величины $x_i'=\frac{x_i-a}{\sigma}$ независимы и имеют нормальное распределение с параметрами (0, 1). Обозначим $\bar{x}'=\frac{1}{n}\sum_{i=1}^{n}x_i'$, $s'^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i'-\bar{x}')^2$. Тогда $\bar{x}=a+\sigma\bar{x}'$, $s^2=\sigma^2 s'^2$. Докажем, что $\bar{x}'$ и s'^2 независимы и что $s'^2(n-1)$ имеет χ^2-распределение с $(n-1)$-й степенью свободы. Случайный вектор $(x_1', \ldots, x_n')$ имеет сферическое нормальное распределение с плотностью

$$\frac{1}{(2\pi)^{n/2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}x_i'^2\right]. \tag{8}$$

Пусть $y = Cx'$ — ортогональное преобразование, заданное соотношениями

$$y_1 = \sqrt{n}\,\bar{x}' = \frac{x_1'}{\sqrt{n}} + \ldots + \frac{x_n'}{\sqrt{n}}, \tag{9}$$

$$y_k = \sum_{i=1}^{n} c_{ki} x_i', \qquad k = 2, \ldots, n.$$

Всегда можно подобрать коэффициенты c_{ki} так, чтобы равенства (9) задавали ортогональное преобразование. Тогда $y_1, y_2, \ldots, y_n$ также будут иметь сферическое нормальное распределение с плотностью (8). Так как $y_1 = \sqrt{n}\,\bar{x}'$ и $\sum_{i=1}^{n} x_i'^2 = \sum_{i=1}^{n} y_i^2$ (из-за ортогональности преобразования C), то

$$(n-1)\,s'^2 = \sum_{i=1}^{n} (x_i' - \bar{x}')^2 = \sum_{i=1}^{n} x_i'^2 - n\bar{x}^2 =$$
$$= \sum_{i=1}^{n} y_i^2 - y_1^2 = \sum_{i=2}^{n} y_i^2,$$

поэтому $(n-1)\,s'^2$ имеет распределение χ^2 с $(n-1)$-й степенью свободы. Теорема доказана.

Следствием только что доказанной теоремы является

Теорема 2. *Пусть* (1) — *выборка из нормального распределения. Статистика*

$$\tau = \frac{(\bar{x} - a)\sqrt{n}}{s}, \tag{10}$$

называемая *отношением Стьюдента, имеет распределение Стьюдента с* $(n-1)$*-й степенью свободы.*

Доказательство. $\dfrac{\bar{x} - a}{\sigma}\sqrt{n}$ имеет нормальное распределение с параметрами $(0, 1)$, а s/σ не зависит от $\bar{x}$ и равно $\sqrt{\chi_{n-1}^2/(n-1)}$, где χ_{n-1}^2 имеет χ^2-распределение с $(n-1)$-й степенью свободы. Поэтому отношение (10) имеет распределение Стьюдента с $(n-1)$-й степенью свободы. Теорема доказана.

Для построения доверительного интервала для a при неизвестном σ воспользуемся отношением Стьюдента (10). Пусть $S_n(t)$ — функция распределения Стьюдента с n степенями свободы. Обозначим $t_\gamma(n)$ квантиль распределения $S_n(t)$, т. е. корень уравнения

$$S_n(t) = 1 - \gamma.$$

Так как распределение Стьюдента симметрично, то $t_{1-\gamma}(n) = -t_\gamma(n)$ и при построении доверительного интервала надо брать $\alpha_1 = \alpha_2 = \alpha/2$. Неравенство

$$-\frac{s}{\sqrt{n}} t_{\alpha/2}(n-1) \leqslant \bar{x} - a \leqslant \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1)$$

выполняется с вероятностью $1-\alpha$. Это дает нам доверительный интервал

$$\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1) \leqslant a \leqslant \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1).$$

в) Доверительный интервал для σ при известном a. Статистика

$$\frac{\eta}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{x_i - a}{\sigma}\right)^2$$

является достаточной для параметра σ и имеет χ^2-распределение с n степенями свободы. Обозначим через $K_n(x)$ функцию распределения η/σ^2 и через $k_\gamma(n)$ квантиль $K_n(x)$, т. е. корень уравнения

$$K_n(x) = 1 - \gamma.$$

Пусть $\alpha = \alpha_1 + \alpha_2$. Тогда неравенства

$$k_{1-\alpha_1}(n) \leqslant \frac{\eta}{\sigma^2} \leqslant k_{\alpha_2}(n)$$

выполняются с вероятностью $1-\alpha$. Это дает доверительный интервал

$$\sqrt{\frac{\eta}{k_{\alpha_2}(n)}} \leqslant \sigma \leqslant \sqrt{\frac{\eta}{k_{1-\alpha_1}(n)}}. \tag{11}$$

Можно доказать, что этот интервал будет иметь наименьшую длину среди всех доверительных интервалов этого вида с доверительной вероятностью $1-\alpha$, если α_1 и α_2 выбраны так, что плотность $k_n(x) = K_n'(x)$ удовлетворяет равенству

$$k_n(k_{1-\alpha_1}(n)) = k_n(k_{\alpha_2}(n)).$$

г) Доверительный интервал для σ при неизвестном a. В этом случае за основную статистику η возьмем эмпирическую дисперсию. По теореме 1 $\frac{s^2(n-1)}{\sigma^2}$ имеет χ^2-распределение с $(n-1)$-й степенью свободы. Это приводит к доверительному интервалу, аналогичному (11):

$$s\sqrt{\frac{n-1}{k_{\alpha_2}(n-1)}} \leqslant \sigma \leqslant s\sqrt{\frac{n-1}{k_{1-\alpha_1}(n-1)}}$$

с доверительной вероятностью $1-(\alpha_1+\alpha_2)$.

§ 65. Доверительные интервалы для вероятности успеха в схеме Бернулли

Той же самой процедуры построения доверительных интервалов можно придерживаться и в том случае, когда основное распределение дискретно. Продемонстрируем это на схеме Бернулли. Пусть μ — число успехов при n испытаниях в схеме Бернулли. Функция распределения

$$F(m;\,p) = \mathsf{P}\{\mu \leqslant m\} = \sum_{k=0}^{m} C_n^k p^k (1-p)^{n-k},$$

рассматриваемая в целых точках $m = 0, 1, 2, \ldots, n$, убывает с ростом p, так как

$$\frac{d}{dp}F(m;\,p) =$$

$$= \sum_{k=0}^{m} C_n^k \cdot kp^{k-1}(1-p)^{n-k} - \sum_{k=0}^{m} C_n^k (n-k)\, p^k (1-p)^{n-k-1} =$$

$$= -nC_{n-1}^m p^m (1-p)^{n-m-1} < 0.$$

Обозначим $m_\gamma(p)$ такое наименьшее целое число, для которого

$$1 - F(m_\gamma(p);\,p) \geqslant 1-\gamma.$$

Тогда $m_\gamma(p)-1$ есть такое наибольшее число, что

$$F(m_\gamma(p)-1;\,p) < \gamma.$$

Пусть $\alpha_1+\alpha_2 = \alpha$. Тогда с вероятностью $\geqslant 1-\alpha$

$$m_{1-\alpha_1}(p) \leqslant \mu \leqslant m_{\alpha_2}(p).$$

Обозначим решение уравнения $y = m_\gamma(p)$ относительно p через $m_\gamma^{-1}(y)$. Тогда неравенства

$$\underline{p} = m_{\alpha_2}^{-1}(\mu) \leqslant p \leqslant m_{1-\alpha_1}^{-1}(\mu) = \bar{p}, \tag{12}$$

задающие доверительный интервал, выполняются с вероятностью $\geqslant 1-\alpha$. Число $1-\alpha$ называется в этом случае *коэффициентом доверия*. Для нахождения границ доверительных интервалов (12) можно пользоваться таблицами биномиального распределения, но такие таблицы громоздки и не очень распространены. Поэтому часто используют предельную теорему Муавра — Лапласа об асимптотической нормальности

$$\eta = \frac{\mu - np}{\sqrt{np(1-p)}}$$

для построения приближенных доверительных интервалов. Неравенство $|\eta| \leqslant u_{\alpha/2}$ при больших n выполняется с вероятностью $\approx 1-\alpha$. Это дает

$$\frac{\mu}{n} - u_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leqslant p \leqslant \frac{\mu}{n} + u_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, \tag{13}$$

однако неравенства (13) не задают доверительного интервала, так как в них имеется справа и слева зависимость от p. Поскольку $\frac{\mu}{n} \xrightarrow{P} p$, то из (13) получают доверительный интервал

$$\frac{\mu}{n} - u_{\alpha/2}\sqrt{\frac{\mu}{n^2}\left(1-\frac{\mu}{n}\right)} \leqslant p \leqslant \frac{\mu}{n} + u_{\alpha/2}\sqrt{\frac{\mu}{n^2}\left(1-\frac{\mu}{n}\right)}.$$

Другой подход к построению приближенного доверительного интервала для p основан на следующей теореме.

Теорема 3. *Пусть последовательность ξ_n такова, что $\mathsf{M}\xi_n = a$, $\mathsf{D}\xi_n = \sigma_n^2 \to 0$ и распределение ξ_n асимптотически нормально с параметрами (a, σ_n). Предположим, что функция $g(x)$ ограничена, $|g(x)| \leqslant K$, и имеет непрерывные производные $g'(x)$, $g''(x)$ в окрестности точки $x = a$, $g'(a) \neq 0$. Тогда случайная величина $\eta_n = g(\xi_n)$ асимптотически нормальна с параметрами $(g(a), |g'(a)|\sigma_n)$.*

Доказательство. Пусть при $|x-a|<\varepsilon$ имеет место разложение Тейлора

$$g(x)=g(a)+g'(a)(x-a)+r(x)(x-a)^2, \tag{14}$$

где $|r(x)|\leqslant K_1$ при некотором $K_1<\infty$. Обозначим $A_n=\{\omega: |\xi_n(\omega)-a|<\sigma_n^{1/4}\}$. Для $n\geqslant n_0$ таких, что $\sigma_n\leqslant\varepsilon^4$, мы можем, пользуясь разложением (14), написать

$$\eta_n=g(\xi_n)I_{A_n}+g(\xi_n)I_{\bar{A}_n}=$$
$$=I_{A_n}\left(g(a)+g'(a)(\xi_n-a)+r(\xi_n)(\xi_n-a)^2\right)+I_{\bar{A}_n}g(\xi_n).$$

Тогда $\eta_n=\eta_n'+\delta_n$, где

$$\eta_n'=g(a)+g'(a)(\xi_n-a),$$

$$\delta_n=-g(a)I_{\bar{A}_n}-g'(a)(\xi_n-a)I_{\bar{A}_n}+$$
$$+I_{A_n}r(\xi_n)(\xi_n-a)^2+g(\xi_n)I_{\bar{A}_n}. \tag{15}$$

Так как правая часть равенства

$$\frac{\eta_n'-g(a)}{|g'(a)|\sigma_n}=\frac{g'(a)}{|g'(a)|}\cdot\frac{\xi_n-a}{\sigma_n}$$

имеет в пределе нормальное распределение с параметрами (0, 1), то η_n' асимптотически нормально с параметрами $(g(a), |g'(a)|\sigma_n)$. Докажем, что $\frac{\delta_n}{\sigma_n}\xrightarrow{\mathsf{P}}0$. Для этого обратимся к (15) и докажем соответствующую сходимость по вероятности к нулю для каждого слагаемого. По неравенству Чебышева (32) из § 17

$$\mathsf{P}(\bar{A}_n)\leqslant\frac{\sigma_n^2}{\sqrt{\sigma_n}}=\sigma_n^{3/2}.$$

По неравенству Чебышева (31) из § 17 для любого ε

$$\mathsf{P}\left\{|g(a)|I_{\bar{A}_n}>\varepsilon\sigma_n\right\}\leqslant\frac{K\mathsf{P}(\bar{A}_n)}{\varepsilon\sigma_n}\leqslant\frac{K}{\varepsilon}\sqrt{\sigma_n}\to 0.$$

Аналогично доказывается $\mathsf{P}\left\{|g(\xi_n)|I_{\bar{A}_n}>\varepsilon\sigma_n\right\}\to 0$. Далее по неравенству Чебышева

$$\mathsf{P}\left\{|g'(a)|\cdot|\xi_n-a|I_{\bar{A}_n}>\varepsilon\sigma_n\right\}\leqslant\frac{|g'(a)|\mathsf{M}|\xi_n-a)|I_{\bar{A}_n}}{\varepsilon\sigma_n}$$

и по неравенству Коши — Буняковского

$$\mathsf{M}\,|\xi_n - a|\, I_{\bar{A}_n} \leqslant \sqrt{\mathsf{M}\,(\xi_n - a)^2\, \mathsf{M} I_{\bar{A}_n}} \leqslant \sigma_n \cdot \sigma_n^{3/4},$$

поэтому

$$\mathsf{P}\left\{|g'(a)| \cdot |\xi_n - a|\, I_{\bar{A}_n} > \varepsilon\sigma_n\right\} \leqslant \frac{|g'(a)|\,\sigma_n^{3/4}}{\varepsilon} \to 0.$$

И, наконец,

$$\mathsf{P}\left\{\left|I_{A_n} r(\xi_n)(\xi_n - a)^2\right| > \varepsilon\sigma_n\right\} \leqslant \frac{K_1 \mathsf{M}\,(\xi_n - a)^2}{\varepsilon\sigma_n} = \frac{K_1\sigma_n}{\varepsilon} \to 0.$$

Итак, мы доказали, что $\frac{\delta_n}{\sigma_n} \xrightarrow{\mathsf{P}} 0$. Так как в сумме справа

$$\frac{\eta_n - g(a)}{|g'(a)|\,\sigma_n} = \frac{\eta'_n - g(a)}{|g'(a)|\,\sigma_n} + \frac{\delta_n}{|g'(a)|\,\sigma_n}$$

распределение первого слагаемого сходится к нормальному с параметрами $(0, 1)$, а второе по вероятности стремится к нулю, то распределение $\frac{\eta_n - g(a)}{|g'(a)|\,\sigma_n}$ сходится к нормальному с параметрами $(0, 1)$. Теорема доказана.

Применяя эту теорему к случайной величине

$$\eta_n = 2 \arcsin \sqrt{\frac{\mu}{n}},$$

получаем асимптотическую нормальность η_n с параметрами $\left(2 \arcsin \sqrt{p},\ \sqrt{\frac{1}{n}}\right)$, так как случайная величина μ_n/n асимптотически нормальна с параметрами $\left(p,\ \sqrt{\frac{p(1-p)}{n}}\right)$, а функция $2 \arcsin \sqrt{x}$ удовлетворяет условиям теоремы 3 и

$$(2 \arcsin \sqrt{x})' = \frac{1}{\sqrt{x(1-x)}}.$$

Выбирая квантиль нормального распределения $u_{\alpha/2}$, мы можем построить доверительный интервал для p, исходя из того, что неравенство

$$\left|2 \arcsin \sqrt{\frac{\mu}{n}} - 2 \arcsin \sqrt{p}\right| \leqslant \frac{u_{\alpha/2}}{\sqrt{n}}$$

выполняется при больших n приблизительно с вероятностью $1-\alpha$. Отсюда получаем неравенства

$$2\arcsin\sqrt{\frac{\mu}{n}}-\frac{u_{\alpha/2}}{\sqrt{n}}\leqslant 2\arcsin\sqrt{p}\leqslant \leqslant 2\arcsin\sqrt{\frac{\mu}{n}}+\frac{u_{\alpha/2}}{\sqrt{n}}$$

и приближенный доверительный интервал

$$\sin^2\left(\arcsin\sqrt{\frac{\mu}{n}}-\frac{u_{\alpha/2}}{2\sqrt{n}}\right)\leqslant \leqslant p\leqslant \sin^2\left(\arcsin\sqrt{\frac{\mu}{n}}+\frac{u_{\alpha/2}}{2\sqrt{n}}\right).$$

Задачи

1. По независимым выборкам $x_1, \ldots, x_{n_1}$ и $y_1, \ldots, y_{n_2}$ из двух нормальных распределений с параметрами (a_1, σ_1) и (a_2, σ_2) соответственно построить доверительный интервал с доверительной вероятностью $1-\alpha$ для разности a_1-a_2, если σ_1 и σ_2 известны.

2. Построить доверительный интервал для той же разности a_1-a_2, что и в задаче 1, если $\sigma_1=\sigma_2=\sigma$, где σ неизвестно.

3. С помощью теоремы 3 для параметра a пуассоновского закона построить приближенный доверительный интервал с доверительной вероятностью $1-\alpha$ по независимой выборке $x_1, x_2, \ldots, x_n$.

4. По независимой выборке $x_1, \ldots, x_n$ из равномерного распределения $b(0, \theta)$ построить доверительный интервал для параметра θ с доверительной вероятностью $1-\alpha$.

ОТВЕТЫ К ЗАДАЧАМ

Глава 1

6. $\dfrac{C_6^3 C_{43}^3}{C_{49}^6} = 0{,}017650\ \ldots$ **7.** а) 1/3; б) 1/5. **8.** 0,025.

9. 1/3. **10.** $\displaystyle\sum_{k=0}^{5} \frac{C_{10}^k C_{10}^{10-k}}{C_{20}^{10}} = \frac{1}{2} + \frac{\left(C_{10}^5\right)^2}{2C_{20}^{10}} = 0{,}67186.$

11. а) $\dfrac{10 \cdot 9 \cdot 8 \cdot 7}{10^4} = 0{,}504;$ б) $\dfrac{C_4^2 \cdot 10 \cdot 9 \cdot 8}{10^4} = 0{,}432;$

в) $\dfrac{C_4^2 C_{10}^2 + 4 \cdot 10 \cdot 9}{10^4} = 0{,}063;$ г) $\dfrac{10}{10^4} = 0{,}001.$

12. $p_1 = \dfrac{(a-2r)^2}{a^2},$ $p_2 = \dfrac{4r}{a} - \dfrac{\pi r^2}{a^2} - \dfrac{4r^2}{a^2},$ $p_3 = 0,$

$= \dfrac{\pi r^2}{a^2}.$

13. $\dfrac{1}{8}(1 - 2r/a)^2.$ **14.** $1 - \pi/4.$

Глава 2

1. 0,7. **2.** $\dfrac{M(N-M)(M-1)}{N(N-1)(N-2)}.$ **4.** $\dfrac{4l}{(a+b)\pi}.$ **5.** $\dfrac{4r}{a} - \dfrac{4r^2}{a^2}.$

6. $1/n.$ **7.** $2/n.$ **8.** $a/(a+b-2r).$

9. а) $1 - 3q^3 + 2q^4;$ б) $1 - 4q^3 + 3q^4.$ **10.** 1/3.

11. $6\pi(1-\pi/4)^4 \cdot 4^{-5} = 0{,}000039.$

Глава 3

1. $\mathsf{P}\{\xi = k\} = 1/28$ при $k = 0, 1, 11, 12,$
$\mathsf{P}\{\xi = k\} = 2/28$ при $k = 2, 3, 9, 10;$
$\mathsf{P}\{\xi = k\} = 3/28$ при $k = 4, 5, 7, 8;$
$\mathsf{P}\{\xi = 6\} = 4/28.$

2. $\mathsf{P}\{\eta = 0\} = \mathsf{P}\{\eta = \pm\sqrt{3}/2\} = 1/3.$

3. а) $\mathsf{M}\xi = np,$ $\mathsf{D}\xi = npq;$

б) $\mathsf{M}\xi = n\frac{M}{N}$, $\mathsf{D}\xi = n\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$; в) $\mathsf{M}\xi = \frac{N+1}{2}$, $\mathsf{D}\xi = \frac{N^2-1}{12}$.

4. $\sum\limits_{k=0}^{n}\frac{(-1)^k}{k!}$. 6. $\mathsf{M}\xi_i = np_i$, $\mathsf{D}\xi_i = np_i(1-p_i)$, $\operatorname{Cov}(\xi_i, \xi_j) = -np_ip_j$, $i \neq j$. 8. $\sqrt{0,1} = 0,316227\ldots$ 9. 49/4.

11. $\mathsf{P}\{450 \leqslant \mu \leqslant 550\} \geqslant 0,9$.

Глава 4

1. $t = 71$. 2. 16 587. 3. 1340. 4. 0,00029.
5. $(1-1/12)^{30} = 0,07351$; $e^{-30/12} = 0,08208$.

Глава 5

1. $p_{i,\,i+1} = \frac{5-i}{5}$, $p_{i,\,i-1} = \frac{i}{5}$; $p_{ij} = 0$ для остальных i, j; $p_{ii}(2) = \frac{5+10i-2i^2}{25}$, $p_{i,\,i-2}(2) = \frac{i(i-1)}{25}$, $p_{i,\,i+2}(2) = \frac{(5-i)(4-i)}{25}$, остальные $p_{ij}(2) = 0$; $p_j = C_5^j/2^5$, $j = 0, 1, 2, 3, 4, 5$.

2. Переходные вероятности $p_{i_1i_2i_3,\, i_1i_3i_2} = p_{i_1i_2i_3,\, i_3i_1i_2} = 0,5$, остальные $p_{i_1i_2i_3,\, j_1j_2j_3} = 0$, все предельные вероятности $p_{i_1i_2i_3} = 1/6$.

3. $p_{00}(t) = p_{33}(t) = 1$, $p_{10}(t) = q\cdot\frac{1-(pq)^{\left[\frac{t+1}{2}\right]}}{1-pq}$, $p_{20}(t) = q^2\frac{1-(pq)^{\left[\frac{t}{2}\right]}}{1-pq}$, $p_{13}(t) = p^2\frac{1-(pq)^{\left[\frac{t}{2}\right]}}{1-pq}$, $p_{23}(t) = p\,\frac{1-(pq)^{\left[\frac{t+1}{2}\right]}}{1-pq}$; при нечетных t: $p_{11}(t) = p_{22}(t) = 0$, $p_{12}(t) = p\cdot(pq)^{\frac{t-1}{2}}$, $p_{21}(t) = q\cdot(pq)^{\frac{t-1}{2}}$; при четных t: $p_{12}(t) = p_{21}(t) = 0$, $p_{11}(t) = p_{22}(t) = (pq)^{t/2}$; $p^*_{20} = \frac{q^2}{1-pq}$, $p^*_{23} = \frac{p}{1-pq}$, $p^*_{10} = \frac{q}{1-pq}$, $p^*_{13} = \frac{p^2}{1-pq}$, $p^*_{00} = p^*_{33} = 1$, остальные $p^*_{ij} = 0$.

4. $p_{\alpha\beta,\,\beta0}(1) = q$, $p_{\alpha\beta,\,\beta1}(1) = p$, $p_{\alpha\beta,\,\gamma\delta}(1) = 0$ при $\beta \neq \gamma$; при $t > 1$ $p_{\alpha\beta,\,00}(t) = q^2$, $p_{\alpha\beta,\,01}(t) = p_{\alpha\beta,\,10}(t) = pq$, $p_{\alpha\beta,\,11}(t) = p^2$.
5. Нет.

Глава 6

1. $$F_\xi(z) = \begin{cases} 0 & \text{при } z < 0, \\ \dfrac{z + z\ln ab - z\ln z}{ab} & \text{при } 0 \leqslant z \leqslant ab, \\ 1 & \text{при } ab < z; \end{cases}$$

$$p_\xi(z) = \begin{cases} \dfrac{1}{ab} \ln \dfrac{ab}{z} & \text{при} \quad 0 \leqslant z \leqslant ab, \\ 0 & \text{в остальных случаях.} \end{cases}$$

2. $F(x) = (2lx - x^2)/l^2, \quad 0 \leqslant x \leqslant l;$
$p(x) = 2(l - x)/l^2, \quad 0 \leqslant x \leqslant l.$

3. а) $F_1(x) \cdot \ldots \cdot F_n(x)$; б) $1 - \prod_{i=1}^{n} (1 - F_i(x))$.

4. $p_{\xi_{(k)}}(x) = \dfrac{n!}{(k-1)!\,(n-k)!} F^{k-1}(x)\,(1 - F(x))^{n-k}\, p(x),$

$$p_{\xi_{(k)},\, \xi_{(l)}}(x, y) = \frac{n!}{(k-1)!\,(l-k-1)!\,(n-l)!} F^{k-1}(x)\,(F(y) - F(x))^{l-k-1} \times$$
$$\times (1 - F(y))^{n-l}\, p(x)\, p(y), \quad x < y; \quad p_{\xi_{(k)}\xi_{(l)}}(x, y) = 0 \quad \text{при} \quad x > y.$$

7. При $\lambda_1 \neq \lambda_2$: $p_{\xi_1+\xi_2}(x) = \lambda_1\lambda_2 \dfrac{e^{-\lambda_1 x} - e^{-\lambda_2 x}}{\lambda_2 - \lambda_1}, \quad x \geqslant 0$; при $\lambda_1 = \lambda_2 = \lambda$: $p_{\xi_1+\xi_2}(x) = \lambda^2 x e^{-\lambda x}, \ x \geqslant 0.$

8. $p_n(x) = \dfrac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, \quad x \geqslant 0.$ 9. $F_\eta(x) = x, \quad 0 \leqslant x \leqslant 1.$

Глава 7

1. $\mathsf{M}\xi^{2n+1} = 0$, $\mathsf{M}\xi^{2n} = \dfrac{(2n)!}{2^n n!} \sigma^{2n}$. 2. $e^{\sigma^2/2}$.

3. $\displaystyle\sum_{k=0}^{[n/2]} C_n^{2k} a^{n-2k} \frac{(2k)!}{2^k k!} \sigma^{2k}.$

4. $\mathsf{D}\eta_n = \displaystyle\prod_{i=1}^{n} \left(\sigma_i^2 + a_i^2\right) - \prod_{i=1}^{n} a_i^2.$

5. $\mathsf{M}\eta_n = 1/n$.

6. $\mathsf{M}\xi^\beta = \dfrac{\Gamma(\alpha+\beta)}{\lambda^\beta \Gamma(\alpha)}$ при $\alpha + \beta > 0$.

7. $\mathsf{M}\xi = \mathsf{M}\eta = 0, \quad \mathsf{D}\xi = \mathsf{D}\eta = R^2/4.$

Глава 8

1. $\varphi(s) = \dfrac{s^N - 1}{N(s-1)}, \quad \varphi'(s) = \dfrac{(N-1)s^N - Ns^{N-1} + 1}{N(s-1)^2}, \quad \varphi''(s) =$
$$= \frac{(N-1)(N-2)s^N - 2N(N-2)s^{N-1} + N(N-1)s^{N-2} - 2}{N(s-1)^3}; \quad \lim_{s\uparrow 1} \varphi(s) = 1,$$
$$\varphi'(1) = \lim_{s\uparrow 1} \varphi'(s) = \frac{N-1}{2}, \quad \varphi''(1) = \lim_{s\uparrow 1} \varphi''(s) = \frac{(N-1)(N-2)}{3},$$
$$\mathsf{M}\xi = \varphi'(1) = \frac{N-1}{2}, \quad \mathsf{D}\xi = \frac{N^2-1}{12}.$$

2. $\mathsf{P}\{\xi = n\} = (2n)!/2^{2n}(n!)^2(2n-1), \ n = 1, 2, \ldots, \ \mathsf{M}\xi = \infty.$

3. $(1-\varphi(s))/(1-s)$. 4. $\varphi_t(s)=1-A^t(1-s)\Big/\left[\frac{B(A^t-1)}{2A(A-1)}(1-s)+1\right]$ при $A\neq 1$; $\varphi_t(s)=1-(1-s)\Big/\left(\frac{Bt}{2}(1-s)+1\right)$ при $A=1$; $q=1$, при $A\leqslant 1$; $q=1-2A(A-1)/B$ при $A>1$.

Глава 9

1. $(1+t^2)^{-1}$. 2. $\frac{1}{2}\left[(1+it)^{-\beta}+(1-it)^{-\alpha}\right]$.

Глава 10

1. $2\Phi_0(\sqrt{12})=0{,}99947$. 3. $0\leqslant\beta<1$, $2\alpha>\beta-1$.

4. $\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x/\sqrt{2\lambda}} e^{-\frac{u^2}{2}}\,du$. 5. $0{,}99952$.

Глава 11

1. $\frac{2}{t_1t_2}\left[\frac{t_1e^{it_1}-t_2e^{it_1}}{t_1-t_2}-1\right]$. 2. $f_{\xi_1\xi_2}(t_1, t_2)=e^{it_2}f(t_1-t_2)$.

3. $\mathsf{P}\{|\xi|\leqslant 1, |\eta|\leqslant 1\}=0{,}46594$, $\mathsf{P}\{\xi^2+\eta^2\leqslant 1\}=0{,}39347$.

4. $$\mathsf{P}\{\eta\leqslant x\}=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\int\limits_{-\infty}^{x/\sqrt{2}}\left(1+\frac{u^2}{n}\right)^{-\frac{n+1}{2}}du.$$

Глава 12

5. Справедливо.

Глава 13

1. $\mathrm{Cov}(\bar{x}, \bar{y})=-\frac{1}{N(N-1)}\sum\limits_{i=1}^{N}(X_i-\overline{X})^2$, $\rho(\bar{x}, \bar{y})=$
$$=-\sqrt{\frac{n_1n_2}{(N-n_1)(N-n_2)}}.$$

2. $\frac{n-1}{n(N-1)}\sum\limits_{i=1}^{N}(X_i-\overline{X})^2$. 3. $\frac{n-1}{n}\sigma^2$.

4. $\mathsf{M}x_{(k)}=\frac{ka}{n+1}$, $\mathsf{D}x_k=\frac{k(n-k+1)}{(n+1)^2(n+2)}a^2$.

Глава 14

1. Если $\min\limits_{1\leqslant i\leqslant n} x_i<1$, то принимать H_0; если $\max\limits_{1\leqslant i\leqslant n} x_i>2$, то принимать H_1; в остальных случаях с вероятностью 1/2 принимать H_0. В этом случае $\max(\alpha, \beta)=2^{-n-1}$.

2. Критическое множество оптимального критерия: $x_1 + x_2 + \ldots$ $\ldots + x_n \leqslant \lambda_0 x_{1-\alpha}$, где $\frac{1}{\Gamma(n)} \int\limits_{x_{1-\alpha}}^{\infty} x^{n-1} e^{-x}\, dx = 1 - \alpha$. Мощность:

$$W(\lambda_1) = \frac{1}{\Gamma(n)} \int\limits_0^{\frac{\lambda_0}{\lambda_1} x_{1-\alpha}} x^{n-1} e^{-x}\, dx.$$

3. Критическое множество: $\bar{x} - \bar{y} \geqslant u_\alpha \sigma$, где $\bar{x} = \frac{1}{n_1} \sum\limits_{i=1}^{n_1} x_i$, $\bar{y} = \frac{1}{n_2} \sum\limits_{i=1}^{n_2} y_i$, $\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$, u_α — α-квантиль нормального распределения.

4. $\chi^2 = \sum\limits_{i=0}^{9} \frac{(\nu_i - 100)^2}{100} = 5{,}86 < \chi^2_{0,05}(9) = 16{,}9$, поэтому данные таблицы не противоречат гипотезе случайности.

Глава 15

1. $\hat{a} = \mu$, $\hat{\sigma}^2 = \frac{\mu(n-\mu)}{n-1}$. 2. $\hat{a} = nI_{\{\xi=0\}}$, $\hat{\sigma}^2 = nI_{\{\xi=1\}}$.
3. $\varphi_m(\xi) = (1-m)^\xi$. 4. а), б) $x_1 + x_2 + \ldots + x_n$; в) $\max\limits_{1 \leqslant i \leqslant n} x_i$.
5. $\hat{a} = \bar{x}$. 6. $\hat{\theta} = \max\limits_{1 \leqslant i \leqslant n} x_i$.

Глава 16

1. $\bar{x} - \bar{y} - u_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leqslant a_1 - a_2 \leqslant \bar{x} - \bar{y} + u_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, где $u_{\alpha/2}$ — $\alpha/2$-квантиль нормального распределения, $\bar{x} = \frac{1}{n_1} \sum\limits_{i=1}^{n_1} x_i$, $\bar{y} = \frac{1}{n_2} \sum\limits_{i=1}^{n_2} y_i$.

2. $\bar{x} - \bar{y} - t_{\alpha/2}(n_1 + n_2 - 2) \sqrt{\frac{s_1^2 \cdot (n_1 - 1) + s_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}} \leqslant a_1 - a_2 \leqslant$ $\leqslant \bar{x} - \bar{y} + t_{\alpha/2}(n_1 + n_2 - 2) \sqrt{\frac{s_1^2 \cdot (n_1 - 1) + s_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}}$, где $\bar{x}$, $\bar{y}$

те же, что в задаче 1, $s_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2$, $s_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_i-\bar{y})^2$, $t_{\alpha/2}(n_1+n_2-2)$ — $\alpha/2$-квантиль распределения Стьюдента с n_1+n_2-2 степенями свободы.

3. $\left(\sqrt{\bar{x}}-\frac{u_{\alpha/2}}{2\sqrt{n}}\right)^2\leqslant a\leqslant\left(\sqrt{\bar{x}}+\frac{u_{\alpha/2}}{2\sqrt{n}}\right)^2$, где $\bar{x}=(x_1+\ldots+x_n)/n$, $u_{\alpha/2}$ — $\alpha/2$-квантиль нормального распределения.

4. $\dfrac{\max\limits_{1\leqslant i\leqslant n} x_i}{\sqrt[n]{1-\alpha}}\leqslant\theta\leqslant\max\limits_{1\leqslant i\leqslant n} x_i$.

ТАБЛИЦЫ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Значения функции $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2}\, du$

x	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	398	438	478	517	557	596	636	675	714	753
0,2	793	832	871	910	948	987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	406	443	480	517
0,4	554	591	628	664	700	736	772	808	844	879
0,5	915	950	985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	357	389	422	454	486	517	549
0,7	580	611	642	673	703	734	764	794	823	852
0,8	881	910	939	967	995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	289	315	340	365	389
1,0	413	437	461	485	508	581	554	577	599	621
1,1	643	665	686	708	729	749	770	790	810	830
1,2	849	869	888	907	925	944	962	980	997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	177
1,4	192	207	222	236	251	265	279	292	306	319
1,5	332	345	357	370	382	394	406	418	429	441
1,6	452	463	474	484	495	505	515	525	535	545
1,7	554	564	573	582	591	599	608	616	625	633

Продолжение

x	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
1,8	641	649	656	664	671	678	686	693	699	706
1,9	713	719	726	732	738	744	750	756	761	767
2,0	772	778	783	788	793	798	803	808	812	817
2,1	821	826	830	834	838	842	846	850	854	857
2,2	861	864	868	871	875	878	881	884	887	890
2,3	893	896	898	901	904	906	909	911	913	916
2,4	918	920	922	925	927	929	931	932	934	936
2,5	938	940	941	943	945	946	948	949	951	952
2,6	953	955	956	957	959	960	961	962	963	964
2,7	965	966	967	968	969	970	971	972	973	974
2,8	974	975	976	977	977	978	979	979	980	981
2,9	981	982	982	983	984	984	985	985	985	986
3,0	987	987	987	988	988	989	989	989	990	990

Квантили нормального распределения, определяемые равенством $\alpha = \frac{1}{\sqrt{2\pi}} \int_{u_\alpha}^{\infty} e^{-\frac{x^2}{2}}\, dx$

α	0,001	0,005	0,010	0,015	0,020	0,025	0,030	0,035	0,040	0,045	0,050
u_α	3,0902	2,5758	2,3263	2,1701	2,0537	1,9600	1,8808	1,8119	1,7507	1,6954	1,6449

ЛИТЕРАТУРА

Боровков А. А. Теория вероятностей. — М.: Наука, 1976.

Гихман И. И., Скороход А. В., Ядренко М. Н. Теория вероятностей и математическая статистика. — Киев: Вища школа, 1979.

Гнеденко Б. В. Курс теории вероятностей. — М.: Наука, 1969.

Колмогоров А. Н. Основные понятия теории вероятностей. — М.: Наука, 1974.

Крамер Г. Математические методы статистики. — М.: Мир, 1975.

Леман Э. Проверка статистических гипотез. — М.: Наука, 1964.

Севастьянов Б. А., Чистяков В. П., Зубков А. М. Сборник задач по теории вероятностей. — М.: Наука, 1980.

Феллер В. Введение в теорию вероятностей и ее приложения. — М.: Мир, т. 1, 1964; т. 2, 1967.

Ширяев А. Н. Вероятность. — М.: Наука, 1980.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Борис Александрович Севастьянов

КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Редактор В. В. *Абгарян*

Технический редактор *Л. В. Лихачева*

Корректор *Н. Д. Дорохова*

ИБ № 11780

Сдано в набор 02.02.82. Подписано к печати 11.10.82. Формат 84×108¹/₃₂. Бумага тип. № 1. Литературная гарнитура. Высокая печать. Условн. печ. л. 13,44. Уч-изд. л. 13,44. Тираж 50 000 экз. Заказ № 75. Цена 60 коп.

Издательство «Наука»
Главная редакция физико-математической литературы
117071, Москва, В-71, Ленинский проспект, 15.

Ленинградская типография № 2 головное предприятие ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном Комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.