ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Н.С. АРКАШОВ, А.П. КОВАЛЕВСКИЙ

Теория вероятностей и случайные процессы

Утверждено Редакционно-издательским советом университета в качестве учебного пособия для студентов нематематических специальностей высших учебных заведений

Новосибирск
2014

УДК 519.21 (075.8)
А 822

Рецензенты: *В. И. Лотов,* д-р физ.-мат. наук,
проф. НГУ,
А. Г. Пинус, д-р физ.-мат. наук, проф.,
К. А. Джафаров, канд. физ.-мат. наук, доцент

Работа подготовлена на кафедре высшей математики
для студентов II курса

Аркашов Н. С.
А 822 Теория вероятностей и случайные процессы: Учеб. пособие / Н. С. Аркашов, А. П. Ковалевский. — Новосибирск: Изд-во НГТУ, 2014. — 238 с.

ISBN 978-5-7782-2382-0

Настоящее учебное пособие подготовлено для студентов II курса очного и заочного отделений всех направлений и специальностей, изучающих такие разделы высшей математики, как теория вероятностей и математическая статистика, в объеме семестрового курса.

Пособие содержит типовой расчет. В приложениях даны таблицы вероятностных распределений.

Все замечания по содержанию пособия просим передавать на кафедру высшей математики. Они будут с благодарностью приняты и учтены в следующих изданиях.

УДК 519.21 (075.8)

ISBN 978-5-7782-2382-0

Оглавление

Глава 1

Случайный эксперимент, события

§ 1.1. События, операции над событиями

Теория вероятностей изучает математические модели случайных экспериментов. Под случайным подразумевается такой эксперимент, исходы которого неоднозначно определяются начальными условиями. Простейшим примером такого эксперимента является подбрасывание монеты. В этом эксперименте возможны лишь два исхода: выпадение «герба» или «решетки», — при этом точно предсказать результат до проведения эксперимента невозможно.

Со всяким случайным экспериментом можно связать множество $\Omega = \{\omega\}$ всех его взаимоисключающих исходов. Это множество называют **пространством элементарных исходов**, а его элементы — **элементарными исходами**. Результатом проведения случайного эксперимента является некоторый элементарный исход $\omega \in \Omega$. **Событиями** называют подмножества пространства элементарных исходов Ω. Выражение «произошло событие А» означает $\omega \in A$, где ω — элементарный исход, явившийся результатом эксперимента. Любое «событие» в обычном понимании, то есть любой исход случайного эксперимента, может быть представлено некоторым подмножеством $A \subseteq \Omega$ при подходящем выборе пространства элементарных исходов. В дальнейшем не будем различать «события» в обычном понимании и события — подмножества Ω.

Событие называется **достоверным**, если в результате эксперимента оно непременно происходит; событие называется **невозможным**, если в результате эксперимента оно не может произойти; событие называется **случайным**, если в результате эксперимента оно может произойти, а может не произойти. Так как для любого элементарного исхода ω соотношение $\omega \in \Omega$ имеет место всегда, то все пространство Ω соответствует достоверному событию, пустое множество $\emptyset$ — невозможному событию, собственные подмножества $A \subset \Omega$ представляют случайные события.

Пусть A и B какие-нибудь события (подмножества Ω).

Объединением, или **суммой**, этих событий называется объединение $A \cup B$ множеств A и B. **Пересечением** или **произведением** событий называется их теоретико-множественное пересечение . Аналогично, **разностью** событий A и B называется разность $A \backslash B$ соответствующих множеств. **Противоположным** к событию A называется дополнение $\overline{A} = \Omega \setminus A$ множества A.

Появление события $A \cup B$ в результате эксперимента означает, что элементарный исход $\omega \in A \cup B$, а это имеет место, если $\omega \in A$ или $\omega \in B$. Поэтому можно сказать, что *объединение (сумма) событий происходит тогда и только тогда, когда происходит хотя бы одно из этих событий.*

Аналогично, *пересечение (произведение) событий происходит тогда и только тогда, когда эти события происходят совместно.*

Разность $A \setminus B$ событий происходит тогда и только тогда, когда происходит A, но не происходит B.

Противоположное событие $\overline{A}$ происходит тогда и только тогда, когда само A не происходит.

Говорят, что **событие A влечет событие B**, или **A содержится в B**, если A является подмножеством B: $A \subseteq B$. События называются **равными**, или **эквивалентными**, если они совпадают как множества: $A = B$. Очевидно, что $A = B$ *тогда и только тогда, когда каждое из этих событий влечет другое.*

Объединение и пересечение более чем двух событий определяются аналогично, т.е. как объединение и пересечение соответствующих подмножеств. Например,

$$\bigcup_{k=1}^{n} A_k \qquad \bigcap_{n=1}^{\infty} B_n$$

означают соответственно объединение конечного и пересечение бесконечного (счетного) множества событий.

События называют **непересекающимися**, или **несовместными**, если их пересечение есть невозможное событие. События из некоторой совокупности $A_1, A_2, ..., A_n, ...$ называют **попарно несовместными**, если любые два из них несовместны, т.е. $A_i \cap A_j = \emptyset$ при $i \neq j$.

Поскольку введенные операции над событиями тождественны соответствующим операциям над множествами, то они подчиняются всем аксиомам операций над множествами:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A \quad \text{(коммутативность)};$$

$A\cup(B\cup C)=(A\cup B)\cup C),\quad A\cap(B\cap C)=(A\cap B)\cap C)$ (ассоциативность);

$A\cap(B\cup C)=(A\cap B)\cup(A\cap C),\ A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$ (дистрибутивность);

$$\overline{\bigcup_n A_n}=\bigcap_n \overline{A_n},\quad \overline{\bigcap_n A_n}=\bigcup_n \overline{A_n}\quad \text{(двойственность)};$$

$$\overline{(\overline{A})}=A\quad \text{(отрицание отрицания)}.$$

Заметим в заключение, что для обозначения введенных операций объединения (суммы), пересечения (произведения), разности событий используются также традиционные алгебраические символы: «$+$», «$\cdot$», «$-$»:

$$A\cup B=A+B,\quad A\cap B=AB,\quad A\setminus B=A-B,\quad \bigcup_{k=1}^{n}A_k=\sum_{k=1}^{n}A_k.$$

§ 1.2. Решение типовых примеров

Пример 1.1. *Случайный эксперимент состоит в однократном подбрасывании игральной кости – правильного кубика с нанесенными на гранях числами от 1 до 6. Обозначим через* ω_k *исход, состоящий в появлении на верхней грани числа* k. *Тогда в качестве пространства элементарных исходов можно взять множество* $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$.

Пример 1.2. *Рассмотрим случайный эксперимент из предыдущего примера. Введем следующие «события»:* $A=\{$ *выпадение четного числа*$\}$, $B=\{$*выпадение числа, меньшего 3*$\}$, $C=\{$*выпадение дробного числа*$\}$. *При выборе пространства элементарных исходов из примера 1.1 указанные «события» очевидным образом представляются множествами:* $A=\{\omega_2,\omega_4,\omega_6\},\ B=\{\omega_1,\omega_2\},\ C=\emptyset$.

Пример 1.3. *Случайный эксперимент – двукратное подбрасывание игральной кости. Построить подходящее пространство элементарных исходов* Ω *для описания следующих событий:* $A=\{$*оба раза выпало число очков, кратное трем*$\}$, $B=\{$*сумма выпавших чисел не больше* $12\}$, $C=\{$*выпали одинаковые числа*$\}$, $D=\{$*произведение выпавших чисел делится на* $14\}$. *Найти подмножества* Ω, *образующие эти события, указать достоверные, невозможные и случайные события.*

Решение. Поскольку результатом эксперимента является пара чисел, выпавших на верхней грани при первом и втором подбрасывании игральной кости, то в качестве пространства элементарных исходов естественно выбрать множество всех упорядоченных наборов из двух чисел, каждое из которых может принимать любое из натуральных значений от 1 до 6, т.е. $\Omega = \{(i,j) : \ i = 1, \ldots, 6, \ j = 1, \ldots, 6\}$. Тогда указанные события совпадают со следующими подмножествами Ω :

$$A = \{(3,3), (3,6), (6,3), (6,6)\}; \ B = \Omega;$$

$$C = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}; \ D = \emptyset.$$

Событие B достоверно, так как оно происходит при любом элементарном исходе $(i,j) \in \Omega$, событие D невозможно (если бы оно могло произойти, то какое-нибудь из двух выпавших чисел должно делиться на 7), события A и C случайны. $\bigtriangledown$ [1]

§ 1.3. Задачи для самостоятельного решения

1.1 Что означают события $A \cup A$ и $A \cap A$?
1.2 Когда возможно равенство $A \cap B = A$?
1.3 События: A — хотя бы один из трех проверяемых приборов бракованный, B — все приборы доброкачественные. Что означают события: а) $A \cup B$; б) $A \cap B$?
1.4 Среди студентов, собравшихся на лекцию по теории вероятностей, наудачу выбирают одного. Пусть событие A заключается в том, что выбранный студент окажется юношей; событие B в том, что он не курит, а событие C - в том, что он живет в общежитии. Описать событие $AB\overline{C}$. Когда справедливы:

а) равенство $AB\overline{C} = A$; в) равенство $\overline{A} = B$;
б) включение $\overline{C} \subset B$; г) равенство $\overline{B} = B$?

1.5 Рабочий изготовил n деталей. Пусть событие A_i состоит в том, что i-я изготовленная им деталь имеет дефект. Записать событие, заключающееся в том, что:
а) ни одна из деталей не имеет дефектов;
б) хотя бы одна деталь имеет дефект;
в) ровно одна деталь имеет дефект;
г) не более двух деталей имеют дефекты;
д) по крайней мере две детали не имеют дефектов;

[1]Конец решения или доказательства.

е) точно две детали дефектны.

1.6 Преподаватель проводит занятия с группой из трех студентов. Событие А — первый студент потребует внимание преподавателя в течение часа; В — второй студент потребует внимание преподавателя в течение часа; С — третий студент потребует внимание преподавателя в течение часа. Что означают события: а) ABC; б) $A+B+C$; в) $A\overline{B}\ \overline{C}+\overline{A}B\overline{C}+\overline{A}\ \overline{B}C$; г) $AB\overline{C}+A\overline{B}C+\overline{A}BC$; д) $\overline{A}\ \overline{B}\ \overline{C}$; е) $A+B+C-ABC$?

1.7 Событие A — хотя бы одно из четырех имеющихся изделий бракованное, событие B — бракованных изделий среди них не менее двух. Что означают противоположные события $\overline{A}$ и $\overline{B}$?

1.8 Совместны ли события A и $\overline{A\cup B}$?

1.9 Доказать тождества:

а) $(\overline{A}\ +BC)(\overline{B}+AC)(\ \overline{C}+AB)=ABC+\ \overline{A}\ \overline{B}\ \overline{C}$.

в) $(A-B)+(A-C)=A-BC$.

1.10 Доказать, что следующие события достоверны: а) $(A+\overline{B})(\overline{A}+B)+(A+B)(\overline{A}+\overline{B})$;

б) $(A+B)(\overline{A}+B)+(A+\overline{B})(\overline{A}+\overline{B})$.

1.11 Доказать, что событие $(A+B)(\overline{A}+B)(A+\overline{B})(\overline{A}+\overline{B})$ невозможно.

1.12 Установить какие из следующих соотношений правильны:

а) $(A+B)\setminus C=A+(B\setminus C)$; д) $ABC=AB(B+C)$;

б) $(\overline{A+B})C=\overline{AB}C$; е) $(\overline{A+B})C=\overline{A}C+\overline{B}C$;

в) $\overline{A+B+C}=\overline{A}\ \overline{B}\ \overline{C}$; ж) $A\overline{B}C\subset A+B$;

г) $(\overline{A+B})C=C\setminus C(A+B)$; з) $(AB+BC+CA)\subset(A+B+C)$.

1.13 Из таблицы случайных чисел наудачу выбрано одно число. Событие A — выбранное число делится на 5; событие B — данное число оканчивается нулем. Что означают события $A\backslash B$ и $A\cap\overline{B}$?

Глава 2

Классическое вероятностное пространство

§ 2.1. Вероятностное пространство, классическое определение вероятности

Пусть пространство элементарных исходов конечно, то есть $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$. При этом вероятность любого события А вычисляется по формуле:

$$\mathbf{P}(A) = \frac{N(A)}{N}, \tag{2.1}$$

где $N(A)$ — число элементарных исходов, образующих событие А, $N = N(\Omega)$ — число всех элементарных исходов. Равенство (2.1) называют **классическим определением вероятности.** Из этого определения легко следует, что

M1. $\mathbf{P}(\Omega)$=1 и $\mathbf{P}(A) \geq 0$ для любого подмножества $A \subseteq \Omega$;

M2. для любых попарно несовместных множеств $A_1, A_2, ...$ выполняется равенство $\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$ (свойство σ–аддитивности).

Более того, о всякой функции заданной на множестве подмножеств из Ω и обладающей свойствами M1 и M2 мы в дальнейшем будем говорить как о вероятностной мере. В частности, из свойств M1 и M2 вытекают следующие очень полезные свойства вероятности.

1. $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$.
2. Если $A \subseteq B$, то $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A)$.
3. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$.

Докажем первое равенство. Поскольку события A и $\overline{A}$ не пересекаются, при этом $A \cup \overline{A} = \Omega$, то, используя свойства М1 и М2, получаем

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup \overline{A}) = \mathbf{P}(A) + \mathbf{P}(\overline{A}),$$

откуда собственно и получается требуемое равенство. Докажем второе равенство. Т.к. $A \subseteq B$, то $B = (B \setminus A) \bigcup A$. При этом события $B \setminus A$ и A не пересекаются, поэтому, применяя свойство М2, получаем $\mathbf{P}(B) = \mathbf{P}(B \setminus A) + \mathbf{P}(A)$, поэтому и $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A)$. Так же просто доказывается и третье равенство. Для этого достаточно заметить, что $A \cup B = (A \setminus AB) \cup AB \cup (B \setminus AB)$. Множества $A \setminus AB$, AB и $B \setminus AB$ попарно не пересекаются, следовательно, применяя М2 и второе соотношение, получаем

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

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

§ 2.2. Элементы комбинаторики

Пусть имеется конечное множество $S = \{a_1, a_2, \ldots, a_n\}$, которое будем называть *генеральной совокупностью*, а число его элементов n – *объемом* генеральной совокупности.

Определение 2.1. Выборкой объема *k из генеральной совокупности объема n называется упорядоченный набор из k элементов генеральной совокупности S: $(a_{j_1}, a_{j_2}, \ldots, a_{j_k})$. Выборка называется* **выборкой с возвращением***, если она может содержать повторяющиеся (одинаковые) элементы, или* **выборкой без возвращения***, если все ее элементы различны.*

Например, семизначный номер телефона представляет выборку объема 7 из генеральной совокупности объема 10 (количество цифр от 0 до 9). Поскольку цифры в номере могут повторяться, это выборка с возвращением. Названия трех карт, извлеченных по порядку из колоды, образуют выборку без возвращения объема 3 из генеральной совокупности всех карт колоды, например: (туз треф, туз бубей, дама пик).

Заметим, что названия «выборка с возвращением» и «выборка без возвращения» отражают процесс извлечения выборок из генеральной совокупности: извлекая по порядку каждый элемент выборки с возвращением, мы его отмечаем и возвращаем в генеральную совокупность. При извлечении выборки без возвращения, элементы в генеральную совокупность не возвращаются. Для выборки без возвращения распространено и другое название: *размещение из n элементов по k*. Выборки с возвращением называют также *размещениями с повторениями.*

Теорема 2.1. Число всех выборок с возвращением *(или число размещений с повторениями) из n элементов по k обозначается B_n^k и вычисляется формулой:*

$$B_n^k = n^k. \tag{2.2}$$

Число всех выборок без возвращения *(или число размещений) из n элементов по k обозначается A_n^k или $n^{[k]}$ и вычисляется по формуле:*

$$A_n^k = n^{[k]} = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}. \tag{2.3}$$

Доказательство. Извлечение выборки из генеральной совокупности можно представить как процесс заполнения следующей таблицы:

a_{j_1}	a_{j_2}	a_{j_3}	$\cdots$	$a_{j_{k-1}}$	a_{j_k}

При этом, если речь идет о выборке с возвращением, то первую клетку таблицы можно заполнить n способами, поскольку на месте элемента a_{j_1} может оказаться любой из n элементов генеральной совокупности; вторую клетку также можно заполнить n способами, поскольку элемент a_{j_2} извлекается из полной генеральной совокупности. Таким образом, число способов, которыми можно заполнить пару из двух первых клеток, равно $n \cdot n = n^2$. Рассуждая аналогично, можно сказать, что три клетки заполняются n^3 способами, а вся таблица из k клеток может быть заполнена числом способов, равным n^k. Это и есть число всех выборок с возвращением. Выведем теперь вторую формулу. Поскольку речь идет о выборке без возвращения, то первую клетку таблицы можно заполнить n способами; вторую соответственно $n-1$ способом, т.к. элемент, который мы поставили в первую клетку использоваться уже не может и т.д. и, наконец, k-ую клетку таблицы можно заполнить $n-k+1$ способами. Поскольку каждый способ заполнения 1-ой клетки свободно комбинируется со всеми способами заполнения 2-ой клетки и т.д. до k-ой клетки, то получаем $A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$. $\lhd$

Пример 2.1. *Случайный эксперимент состоит в произвольном размещении k различимых шаров по n ящикам $(k \leq n)$. Предполагается,* *что каждый шар случайно и равновозможно может оказаться в любом ящике. Найти вероятность того, что в результате эксперимента все шары окажутся в разных ящиках.*

Решение. Занумеруем все ящики и множество номеров $S = \{1, 2, ..., n\}$ будем считать генеральной совокупностью. Случайный эксперимент по размещению k шаров (будем считать, что они тоже занумерованы) можно представить как извлечение выборки объема k из генеральной совокупности объема n: эта выборка состоит из номеров ящиков, предназначенных для каждого из k шаров. Поскольку каждый шар может изначально оказаться в любом ящике, то номера ящиков для разных шаров могут повторяться, следовательно, выборка, представляющая каждый элементарный исход описываемого эксперимента, есть *выборка с возвращением* объема k из генеральной совокупности объема n. Общее число всех элементарных исходов равно $N = N(\Omega) = B_n^k = n^k$. Событие A=\{все шары окажутся в разных ящиках\} образуется такими элементарными исходами, при которых номера ящиков для различных шаров должны быть различны, следовательно, событие A есть множество всех выборок *без возвращения* из n элементов по k . Их число $N(A) = A_n^k = n^{[k]}$. Тогда вероятность события A равна

$$\mathbf{P}(A) = \frac{N(A)}{N} = \frac{A_n^k}{B_n^k} = \frac{n^{[k]}}{n^k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k}. \quad \triangleleft$$

Задача, рассмотренная в этом примере, является модельной, или эталонной, в том смысле, что на язык этой задачи можно перевести множество других задач с иным конкретным содержанием, и тогда ответ становится очевидным.

Пример 2.2. *Случайный эксперимент состоит в четырехкратном подбрасывании игральной кости. Найти вероятность того, что числа, появившиеся на верхней грани, во всех четырех бросаниях будут различны.*

Решение. Суть эксперимента не изменится, если вместо четырех бросаний одной кости подбрасывать поочередно четыре различные кости. Тогда условия и вопрос задачи становятся полностью идентичны рассмотренным в примере 2.1: кости соответствуют шарам, а числа на верхней грани – номерам ящиков. Для нашего случая число шаров $k = 4$, число ящиков $n = 6$, а вероятность интересующего события равна

$$\mathbf{P}(A) = \frac{6^{[4]}}{6^4} = \frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4} = \frac{5}{18}. \quad \triangleleft$$

Определение 2.2. Перестановкой из n элементов *называется выборка без возвращения, объем которой совпадает с объемом n генеральной совокупности.*

Термин «перестановка» отражает тот факт, что извлечение любой выборки без возвращения объема n из генеральной совокупности того же объема n равносильно выстраиванию всех элементов генеральной совокупности в определенном порядке, то есть некоторой перестановке, или упорядочению, генеральной совокупности.

Следствие 2.1. Число всех перестановок *из n элементов обозначается $\mathcal{P}_n$ и вычисляется формулой:*

$$\mathcal{P}_n = n! \tag{2.4}$$

Доказательство вытекает из равенств: $\mathcal{P}_n = A_n^n = n \cdot (n-1) \cdots 2 \cdot 1$.

Пример 2.3. *10 книг, среди которых имеется трехтомник А.С.Пушкина, случайным образом ставят на полку. Какова вероятность того, что три тома А.С.Пушкина окажутся стоящими рядом: а) в их естественном порядке; б) в любом порядке.*

Решение. Элементарный исход случайного эксперимента – расстановка 10 книг в определенном порядке – есть, очевидно, перестановка из 10 элементов. Число всех элементарных исходов равно числу перестановок $N = 10!$, и по условию все элементарные исходы равновозможны.
а) Обозначим интересующее нас событие: А={три тома А.С.Пушкина окажутся стоящими рядом в естественном порядке}.Чтобы подсчитать число перестановок, образующих событие А, заметим, что указанные три тома можно расположить на 10 местах в естественном порядке 8 способами, а именно: располагая первый том на любом месте с 1-го по 8-й, а второй и третий – рядом вслед за первым. Каждый из 8 способов расположения трехтомника можно сочетать с любым способом расстановки остальных семи томов на оставшихся 7 местах, что можно сделать 7! способами (число перестановок из 7 элементов). Таким образом, общее число перестановок из 10 книг, в которых три тома А.С.Пушкина стоят рядом в естественном порядке, равно $N(A) = 8 \cdot 7! = 8!$, а вероятность события А равна $\mathbf{P}(A) = \dfrac{N(A)}{N} = \dfrac{8!}{10!} = \dfrac{1}{9 \cdot 10} = \dfrac{1}{90}$.

б) Если нас интересует событие В={три тома А.С.Пушкина окажутся рядом в произвольном порядке}, то число перестановок, реализующих это событие равно $N(B) = N(A) \cdot 3!$. В самом деле, из каждой перестановки, реализующей событие А, можно получить 3! различных перестановок, реализующих событие В, переставляя три тома А.С.Пушкина (1,2,3) между собой 3! способами. Тогда вероятность события В равна:

$$\mathbf{P}(B) = \frac{N(B)}{N} = \frac{8! \cdot 3!}{10!} = \frac{3 \cdot 2 \cdot 1}{9 \cdot 10} = \frac{1}{15}. \quad \triangleleft$$

Определение 2.3. Сочетаниями из n элементов по k *называются выборки без возвращения объема k из генеральной совокупности объема n, различающиеся лишь по составу элементов.*

Поскольку для сочетания порядок расположения его элементов значения не имеет, то можно сказать, что сочетания из n элементов по k есть всевозможные *k-элементные подмножества* генеральной совокупности объема n.

Теорема 2.2. Число сочетаний *из n элементов по k обозначается* C_n^k *или* $\binom{n}{k}$ *и вычисляется формулой:*

$$C_n^k = \tbinom{n}{k} = \frac{A_n^k}{\mathcal{P}_k} = \frac{n^{[k]}}{k!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 2 \cdot 1} = \frac{n!}{k! \cdot (n-k)!}. \tag{2.5}$$

Доказательство. Обозначим через X искомое число всех сочетаний из n элементов по k. Из каждого сочетания, путем всевозможных перестановок его элементов, можно получить $\mathcal{P}_k = k!$ различных выборок без возвращения с данным составом элементов. Проделав это с каждым из X сочетаний, мы получим, очевидно, все выборки без возвращения объема k из генеральной совокупности объема n. Иначе говоря, справедливо соотношение $X \cdot \mathcal{P}_k = A_n^k$, из которого находим:

$$X = \frac{A_n^k}{\mathcal{P}_k} = \frac{n^{[k]}}{k!} \ . \quad \triangleleft$$

Выражения $C_n^k = \binom{n}{k}$, заданные равенствами (2.5), уже встречались читателю в курсе высшей математики, например, в известной формуле бинома Ньютона:

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}.$$

Поэтому выражения $C_n^k = \binom{n}{k}$ называются *биномиальными коэффициентами.* Ниже приводятся некоторые свойства биномиальных коэффициентов, которые бывают полезны при вычислениях.

Свойства биномиальных коэффициентов

1. $C_n^0 = C_n^n = 1$ (полагают по определению).
2. $C_n^k = C_n^{n-k}$.
3. $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$.
4. $\sum\limits_{k=0}^{n} C_n^k = C_n^0 + C_n^1 + \dots + C_n^n = 2^n$.

Свойства 2, 3 можно доказать, пользуясь равенствами (2.5), а свойство 4 получается из формулы бинома Ньютона, если в ней положить $a = b = 1$. Если вспомнить, что коэффициент C_n^k вычисляет число всех k-элементных подмножеств множества, состоящего из n элементов, то из свойства 4 вытекает, что *число всех подмножеств n-элементного множества равно* 2^n.

Пример 2.4. *(Урновая схема) В урне n шаров, неразличимых наощупь, из них n_1 – черных, $(n-n_1)$ – белых. Наугад извлекаются k шаров (без возвращения). Найти вероятность того, что среди них окажется k_1 черных и $(k-k_1)$ белых.*

Решение. Элементарный исход описанного эксперимента – выборка без возвращения и без интереса к порядку шаров в выборке, т.е. — сочетание из n по k. Равновозможность элементарных исходов обеспечивается тем, что шары в урне неразличимы. Следовательно, применимо классическое определение вероятности. Общее число элементарных исходов $N = N(\Omega) = C_n^k$. Введем событие $A_{k_1} =$ {среди выбранных шаров окажется k_1 черных и $(k-k_1)$ белых}. Тогда $N(A_{k_1}) = C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}$. В этой формуле первый сомножитель $C_{n_1}^{k_1}$ – это число способов выбрать k_1 черных шаров из n_1, имеющихся в урне, а второй $C_{n-n_1}^{k-k_1}$ — число способов выбрать $(k-k_1)$ белых шаров из имеющихся в урне $(n-n_1)$. Поскольку каждый способ выбора черных шаров можно сочетать с каждым способом выбора белых, общее число способов выбора комбинации «k_1 черных и $(k-k_1)$ белых» равно указанному произведению. Отсюда находим

$$\mathbf{P}(A_{k_1}) = \frac{N(A_{k_1})}{N} = \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}. \qquad \lhd \tag{2.6}$$

Задача, разобранная в этом примере, также является эталонной. Ниже приводится пример аналогичной задачи, сформулированной в других терминах, но с ответом, совпадающим с (2.6)

§ 2.3. Решение типовых примеров

Пример 2.5. *В партии из n изделий k бракованных. Определить вероятность того, что среди выбранных наудачу для проверки m изделий ровно l окажутся бракованными.*

Решение. Число возможных способов взять m изделий из n равно C_n^m. Благоприятствующими являются случаи, когда из общего числа бракованных изделий взято l (это можно сделать C_k^l способами), а остальные $m-l$ изделий небракованные, т.е. они взяты из общего числа $n-k$ (количество способов равно C_{n-k}^{m-l}). Поэтому число благоприятствующих случаев равно $C_k^l C_{n-k}^{m-l}$. Тогда искомая вероятность будет равна

$$p_l = \frac{C_k^l C_{n-k}^{m-l}}{C_n^m}.$$

Заметим, что набор вероятностей p_l, $l = 0, ..., k$, называют гипергеометрическим распределением. $\bigtriangledown$

Пример 2.6. *Бросаются две игральные кости. Какова вероятность, что сумма выпавших очков делится на 6?*

Решение. Пространство элементарных исходов — множество $\Omega = \{(i,j) : i = 1, ..., 6,\ j = 1, ..., 6\}$ всех упорядоченных пар чисел (i,j), где i и j принимают независимо друг от друга целые значения от 1 до 6. Стало быть, число всех таких пар $N = N(\Omega) = 6^2 = 36$. Естественно предположить, что обе игральные кости идентичны и симметричны, поэтому все элементарные исходы равновозможны, а значит, применимо классическое определение вероятности (2.1). Обозначив через $\mathrm{A} = \{$сумма выпавших очков делится на 6$\}$ интересующее нас событие, легко перечислить все элементарные исходы (i,j), образующие A:

$$A = \{(1,5);\ (2,4);\ (3,3);\ (4,2);\ (5,1);\ (6,6)\}.$$

Таким образом, $N(A) = 6$, и следовательно,

$$\mathbf{P}(A) = \frac{N(A)}{N} = \frac{6}{36} = \frac{1}{6}. \qquad \bigtriangledown$$

Пример 2.7. *Колода игральных карт (52 листа, 4 масти по 13 карт в каждой) тщательно перетасована. Наудачу берут 6 карт (без возвращения). Описать пространство элементарных исходов, а также найти вероятность того, что среди этих карт:*

а) окажется король пик;

б) будет ровно 5 карт одной масти.

Решение. Элементарный исход случайного эксперимента — выборка без возвращения объема 6 из генеральной совокупности объема 52. Так как в описании интересующих нас событий порядок элементов выборки, то есть порядок расположения шести выбранных карт, роли не играет, а важен лишь состав выбранных карт, то можно считать, что пространство элементарных исходов Ω составляют всевозможные сочетания из 52 по 6. Поскольку колода была тщательно перетасована, то можно считать все элементарные исходы равновозможными. Их общее число $N = N(\Omega) = C_{52}^6$.

а) Обозначив A={среди выбранных карт окажется король пик}, находим

$$\mathbf{P}(A) = \frac{C_1^1 \cdot C_{51}^5}{C_{52}^6} = \frac{1 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47 \cdot 6!}{5! \cdot 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47} = \frac{3}{26}.$$

б) Обозначая B={среди выбранных карт окажется 5 одной масти}, представим это событие в виде суммы попарно несовместных событий: $B = B_1 + B_2 + B_3 + B_4$, — где B_1={среди выбранных карт окажется ровно 5 пик}, B_2={среди выбранных карт окажется ровно 5 треф}, и т.д. Тогда

$$N(B) = N(B_1) + N(B_2) + N(B_3) + N(B_4) = 4N(B_1)$$

, и

$$\mathbf{P}(B) = \frac{N(B)}{N} = 4\frac{N(B_1)}{N} = 4\mathbf{P}(B_1).$$

Для нахождения $\mathbf{P}(B_1)$ мы используем тот факт, что пять пиковых карт из колоды можно выбрать C_{13}^5 способами, в то время как одну непиковую карту можно выбрать C_{39}^1 способами. Все способы выбора пяти пиковых карт свободно комбинируются со всеми способами выбора одной непиковой, следовательно, число всех способов выбора шести карт из колоды, среди которых ровно пять пиковых, равно $C_{13}^5 \cdot C_{39}^1$.

$$\mathbf{P}(B_1) = \frac{C_{13}^5 \cdot C_{39}^1}{C_{52}^6} = \frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 39 \cdot 6!}{5! \cdot 1! \cdot 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47},$$

$$\mathbf{P}(B) = 4\frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 39 \cdot 6!}{5! \cdot 1! \cdot 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47} \approx 0,0099.$$

§ 2.4. Задачи для самостоятельного решения

2.1 В корзине пять красных и четыре зеленых яблока. Некто берет наугад три яблока. Найти вероятность: а) того, что среди вынутых трех яблок будет ровно два зеленых; б) не более двух красных; в) по крайней мере одно зеленое; г) не менее трех красных.

2.2 Куб, все грани которого окрашены, распилен на 1000 кубиков одного размера. Определить вероятность того, что наудачу извлеченный кубик будет иметь ровно две окрашенные грани.

2.3 В лотерее разыгрывается 1000 билетов. Среди них два выигрыша по 50 руб., 5 по 20 руб., десять - по 10 руб. и 25 - по 5 руб. Некто покупает один билет. Найти вероятность: а) выигрыша не менее 20 руб.; б) какого-либо выигрыша.

2.4 На шести карточках написаны буквы В, Д, З, О, У, Х. После перетасовки вынимают наугад одну карточку за другой и:

а) раскладывают их в том порядке, в каком они были вынуты;

б) каждая из букв на вынутой карточке записывается, а сама карточка возвращается в колоду.

Найти вероятность того, что получится слово "воздух".

2.5 Ребенок играет с 10 буквами разрезной азбуки: А, А, А, Е, И, К, М, М, Т, Т. Какова вероятность того, что при случайном расположении букв в ряд он получит слово МАТЕМАТИКА?

2.6 В лифт 9-этажного дома на первом этаже вошли 5 человек. Известно, что каждый из них с равной вероятностью может выйти на любом из этажей, начиная со второго. Найти вероятность того, что:

а) все пятеро выйдут на пятом этаже;

б) все пятеро выйдут одновременно (на одном и том же этаже);

в) все пятеро выйдут на разных этажах;

г) на первых трех этажах не выйдет ни один человек;

д) все пассажиры выйдут на первых шести этажах;

е) на одном этаже выйдут три пассажира, а на другом два?

2.7 Какова вероятность того, что в четырехзначном номере случайно выбраного в большом городе автомобиля:

а) все цифры разные;

б) две пары одинаковых цифр;

в) только две одинаковые цифры;

г) только три одинаковые цифры;

д) все цифры одинаковые;

е) сумма двух первых цифр равна сумме двух последних цифр?

2.8 В ящике 10 красных и 6 синих пуговиц. Какова вероятность того, что две наудачу вынутые пуговицы будут одноцветными?

2.9 Из колоды карт (52 листа) наудачу вынимаются три карты. Найти вероятность того, что:

а) среди них окажется ровно один туз;

б) среди них окажется хотя бы один туз;

в) это будут тройка, семерка и туз (в любом порядке).

2.10 Участник лотереи «спортлото» из 49 наименований видов спорта называет шесть. Выигрыш определяется тем, сколько наименований он угадал из шести других наименований, которые определяются в момент розыгрыша лотереи с помощью специального устройства, реализующего случайный выбор. С какой вероятностью участник угадает все шесть наименований? Пять наименований и т.д.?

2.11 Из колоды карт (52 листа) наудачу извлекаются три карты. Найти вероятность того, что это будут тройка, семерка, туз.

2.12 Из 28 костей домино случайно выбирают две. Найти вероятность того, что из них можно составить «цепочку» согласно правилам игры.

2.13 Имеется пять отрезков, длины которых равны соответственно 1, 3, 5, 7 и 9 единицам. Определить вероятность того, что с помощью взятых наудачу трех отрезков из данных пяти можно построить треугольник.

2.14 Из десяти билетов выигрышными являются два. Определить вероятность того, что среди взятых наудачу пяти билетов: а) ровно один выигрышный; б) ровно два выигрышных; в) хотя бы один выигрышный.

2.15 На полке в случайном порядке расставлено n книг, среди которых находится двухтомник Д. Лондона. Предполагая, что различные расположения книг равновероятны, найти вероятность того, что оба тома расположены рядом.

2.16 Бросается 6 игральных костей. Найти вероятности следующих событий: $A =${выпадут 3 единицы, две тройки и одна шестерка}, $B =${выпадут разные цифры}, $C =${выпадут три одинаковые цифры}.

2.17 52 карты раздаются четырем игрокам (каждому по 13 карт). Найти вероятности следующих событий: $A =${каждый игрок получит туза}, $B =${один из игроков получит все 13 карт одной масти}, $C =${все тузы попадут к одному из игроков}.

2.18 Один школьник, желая подшутить над своими товарищами, собрал в гардеробе все пальто, а потом развесил их в случайном порядке. Какова вероятность p_n того, что хотя бы одно пальто попало на прежнее место, если всего в гардеробе n крючков и на них висело n пальто? Найти предел p_n при $n \to \infty$.

2.19 В купейный вагон (9 купе по 4 места) семи пассажирам продано семь билетов. Найти вероятность того,что оказались занятыми:

а) ровно два купе; б) ровно три купе

2.20 8 студентов, дополнительно зачисленных в университет, случайным образом распределяются по четырем группам. Найти вероятность того, что:

а) в каждую группу попадут по 2 студента;

б) все окажутся в одной группе;
в) по 4 студента попадут в две группы.

Глава 3

Геометрическая вероятность

В геометрической вероятностной модели пространство элементарных исходов $\Omega \subset R^d$ есть некоторое подмножество d-мерного евклидова пространства R^d, имеющее конечный ненулевой объем $\mu(\Omega)$ (в частности, при $d = 1$ это подмножество действительной прямой, имеющее конечную длину; при $d = 2$ это подмножество плоскости, имеющее конечную площадь; при $d = 3$ это подмножество трехмерного пространства, имеющее конечный объем).

Предположим, что вероятность попадания в любое подмножество $A \subset \Omega$ пропорциональна мере этого подмножества $\mu(A)$ (т.е. длине, площади или объему) и не зависит от вида и расположения множества A. Тогда вероятность события A (вероятность попадания в множество A) определяется формулой:

$$\mathbf{P}(A) = \frac{\mu(A)}{\mu(\Omega)}. \tag{3.1}$$

§ 3.1. Решение типовых примеров

Пример 3.1. *На плоскости начерчены параллельные прямые, находящиеся друг от друга на расстоянии $2a$. На плоскость наудачу брошена монета радиуса $r < a$. Какова вероятность того, что монета пересечет одну из прямых?*

Решение. Интересующее нас событие однозначно описывается положением монеты относительно ближайшей из прямых. Поэтому в качестве элементарного исхода данного эксперимента можно взять значение величины x — расстояние от центра монеты до ближайшей прямой. Тогда пространство элементарных исходов Ω представляется множеством

$\Omega = \{x : 0 \le x \le a\}$, то есть отрезком $[0, a]$ на числовой оси. Событие A ={монета пересечет одну из прямых} происходит тогда и только тогда, когда $x \le r$. Следовательно, оно представляется множеством $A = \{x : 0 \le x \le r\}$, то есть отрезком $[0, r] \subset \Omega$. Вероятность события A находим по формуле (3.1):

$$\mathbf{P}(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{\mu([0, r]}{\mu([0, a])} = \frac{r}{a}. \qquad \nabla$$

Пример 3.2. *(Задача о встрече). Два лица A и B условились встретится в определенном месте между двумя и тремя часами дня. Лицо A ждет другого в течение 10 минут, после чего уходит; лицо B ждет другого в течение 15 минут. Предполагая, что каждый из них может прийти в любое время в течение указанного часа, найти вероятности следующих событий: C ={встреча состоится}, D ={встреча состоится, но во второй половине часа}.*

Решение. Не ограничивая общности, можно считать, что встреча назначена на промежуток времени между 0 и 1. Обозначим через (x, y) моменты прихода A и B соответственно. Пространством элементарных исходов является множество $\Omega = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1\}$, то есть единичный квадрат на плоскости. Событие C можно представить в виде $C = C_1 \cup C_2$, где C_1 ={встреча состоится, причем первым придет A}, C_2 ={встреча состоится, причем первым придет B}. Оно соответствует множеству $C = C_1 \cup C_2$, где $C_1 = \{(x, y) \in \Omega : x \le y \le x + \frac{1}{6}\}$, $C_2 = \{(x, y) \in \Omega : y \le x \le y + \frac{1}{4}\} = \{(x, y) \in \Omega : x - \frac{1}{4} \le y \le x\}$. Изображаем эти множества на рис. 3.1. При этом множество C_1 есть множество точек квадрата Ω, заключенных между прямыми $y = x$ и $y = x + \frac{1}{6}$; множество C_2 — это часть Ω, заключенная между прямыми $y = x$ и $y = x - \frac{1}{4}$; объединение их C заштриховано параллельно диагонали квадрата. Равновозможность прихода обоих лиц в течение часа позволяет использовать геометрическое определение вероятности (3.1), согласно которому

$$\mathbf{P}(C) = \frac{\mu(C)}{\mu(\Omega)} = \frac{1 - \frac{1}{2}(\frac{5}{6})^2 - \frac{1}{2}(\frac{3}{4})^2}{1} = \frac{107}{288}.$$

Событие D можно представить в виде пересечения $D = C \cap E$, где E ={хотя бы одно из двух лиц придет во второй половине часа}. Соответствующее множество $E = \{(x, y) \in \Omega : y > \frac{1}{2} \;\; x > \frac{1}{2}\}$ выделено на рис.3.1 горизонтальной штриховкой, а $D = C \cap E$ - множество с двойной штриховкой. Находя площадь этого множества (площадь единичного квадрата

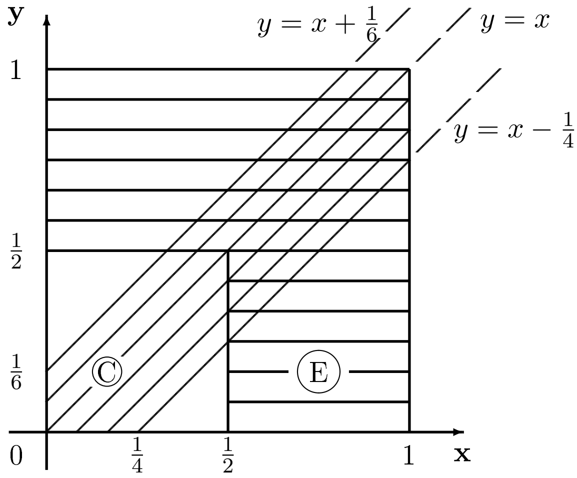


Рис. 3.1:

минус площадь квадрата со стороной $1/2$ минус площади двух трапеций) и применяя снова геометрическое определение вероятности, получим:

$$\mathbf{P}(D) = \frac{\mu(D)}{\mu(\Omega)} = \frac{1 - (\frac{1}{2})^2 - \frac{1}{2}(\frac{1}{3} + \frac{5}{6})\frac{1}{2} - \frac{1}{2}(\frac{1}{4} + \frac{3}{4})\frac{1}{2}}{1} = \frac{5}{24}. \qquad \triangledown$$

§ 3.2. Задачи для самостоятельного решения

3.1 В квадрат с вершинами (0,0), (0,1), (1,0), (1,1) наудачу брошена точка. Пусть (X, Y) будут ее координаты. Найти:

а) $\mathbf{P}\{|2X - Y| < 1/4\}$; г) $\mathbf{P}\{XY < 3/4\}$;
б) $\mathbf{P}\{\min(X, Y) < 3/4\}$; д) $\mathbf{P}\{\max(X, Y) < 1/5\}$;
в) $\mathbf{P}\{(X + Y)/2 < 1/3\}$; е) $\mathbf{P}\{X + 4Y < 1/2\}$.

3.2 На бесконечную шахматную доску со стороной квадрата a бросается наудачу монета диаметра $2r < a$. Найти вероятность того, что:
а) монета попадет целиком внутрь квадрата;
б) монета пересечет не более одной стороны квадрата.

3.3 В интервале времени $[0, T]$ в случайный момент u появляется сигнал длительности Δ. Приемник включается в случайный момент $v \in [0, T]$ на время t. Предположив, что точка (u, v) равномерно распределена на квадрате $[0, T] \times [0, T]$, найти вероятность обнаружения сигнала.

3.4 В квадрат с вершинами (0,0), (0,1), (1,0), (1,1) наудачу брошена точка. Пусть (X, Y) будут ее координаты. Доказать, что для любых $0 \le x, y \le 1$ выполнено:
$\mathbf{P}\{X < x, Y < y\} = \mathbf{P}\{X < x\}\mathbf{P}\{Y < y\} = xy.$
Для $0 < z < 1$ найти:

а) $\mathbf{P}\{\lvert X - Y\rvert < z\}$;	г) $\mathbf{P}\{XY < z\}$;
б) $\mathbf{P}\{\min(X, Y) < z\}$;	д) $\mathbf{P}\{\max(X, Y) < z\}$;
в) $\mathbf{P}\{(X + Y)/2 < z\}$;	е) $\mathbf{P}\{X + 2Y < z\}$.

3.5 Стержень длины l разломан в двух наудачу выбранных точках. С какой вероятностью из полученных отрезков можно составить: а) треугольник; б) остроугольный треугольник?

3.6 На окружности наудачу выбраны три точки A, B, C. Найти вероятность того, что треугольник ABC будет:

а)остроугольным;	г)правильным;
б)тупоугольным;	д)равнобедренным.
в)прямоугольным;	

3.7 На перекрестке установлен автоматический светофор, в котором одну минуту горит зеленый свет и полминуты красный, затем снова одну минуту — зеленый и полминуты — красный и т.д. В случайный момент времени к перекрестку подъезжает легковой автомобиль. Какова вероятность того,что он проедет перекресток без остановки?

3.8 Точка взята наудачу внутри круга радиусом R. Найдите вероятность того, что эта точка окажется от центра на расстоянии, меньшем r.

3.9 Наудачу выбирают два числа из промежутка $[0,\ 1]$. Какова вероятность того, что их сумма больше или равна 1, а их разность меньше либо равна 0?

3.10 Двое студентов условились встретиться в определенном месте и в определенный день между 12 и 13 часами. Пришедший первым ждет второго в течение 10 мин, после чего уходит. Найти вероятность встречи студентов, если приход каждого из них независим и равновозможен в любой промежуток времени.

Глава 4

Условные вероятности

§ 4.1. Определения и примеры

Для любых событий A, B, где $\mathbf{P}(B) > 0$, **условной вероятностью события** A **при условии** B называется число

$$\mathbf{P}(A/B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}. \tag{4.1}$$

Условная вероятность возникает в ситуации, когда имеется частичная информация о результатах случайного эксперимента («произошло событие B»), и в этих условиях требуется найти вероятность события A.

§ 4.2. Решение типовых примеров

Пример 4.1. *Брошены три игральные кости. Чему равна вероятность того, что на одной из них выпала единица, если на всех трех костях выпали разные числа?*

Решение. Первый способ. Обозначим события: $A = \{$на одной из костей выпала единица$\}$, $B = \{$на всех костях выпали разные числа$\}$. Пространство элементарных исходов Ω есть множество всех упорядоченных наборов чисел (i, j, k), выпавших соответственно на 1-й, 2-й и 3-й костях. Поскольку каждое из этих чисел может принимать любое из значений от 1 до 6, то Ω есть множество всех выборок с возвращением объема 3 из шести. Общее число всех элементарных исходов равно $N(\Omega) = B_6^3 = 6^3$. Событие B образовано всеми выборками без возвращения, их число $N(B) = A_6^3 = 6^{[3]}$, вероятность события B равна

$\mathbf{P}(B) = \frac{6^{[3]}}{6^3} = \frac{6\cdot5\cdot4}{6^3} = \frac{5}{9}$. Событие AB образовано такими выборками без возвращения (i, j, k), у которых на одной из трех позиций стоит «1», а на остальных - различные числа от 2 до 6. Число таких выборок равно $N(AB) = 3\cdot5\cdot4$, а вероятность $\mathbf{P}(AB) = \frac{3\cdot5\cdot4}{6^3} = \frac{5}{18}$. Тогда условную вероятность находим по формуле (4.1):

$$\mathbf{P}(A/B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \frac{\frac{5}{18}}{\frac{5}{9}} = \frac{1}{2}.$$

Второй способ. Поскольку известно, что на всех костях выпали разные числа, то есть произошло событие B, то пространством элементарных исходов можно считать множество B. В этом случае появление события A равносильно появлению пересечения AB. Тогда условная вероятность события A вычисляется по классическому определению:

$$\mathbf{P}(A/B) = \frac{N(AB)}{N(B)} = \frac{3\cdot5\cdot4}{6^{[3]}} = \frac{3\cdot5\cdot4}{6\cdot5\cdot4} = \frac{1}{2}. \qquad \triangledown$$

§ 4.3. Задачи для самостоятельного решения

4.1 Брошены две игральные кости. Чему равна вероятность того, что сумма выпавших на них очков равна 8, если известно, что эта сумма есть четное число?

4.2 Брошены две игральные кости. Какова вероятность того, что на первой кости выпало 4 очка, если известно, что на второй кости выпало больше очков, чем на первой?

4.3 Известно, что при бросании 10 игральных костей выпала по крайней мере одна единица. Какова при этом вероятность того, что выпали две или более единицы?

4.4 В ящике лежат 12 красных, 8 зеленых и 10 синих шаров. Наудачу вынимают два шара. Какова вероятность того, что вынуты шары разного цвета, если известно, что среди них нет синего?

4.5 Игральная кость бросается до тех пор, пока не выпадет единица. Известно, что для этого потребовалось четное число бросаний. Найти вероятность того, что единица выпадет при втором бросании.

4.6 Пусть $\mathbf{P}(A_1/A_2) = \mathbf{P}(A_1)$. Показать справедливость следующих равенств:

а) $\mathbf{P}(A_1/\overline{A}_2) = \mathbf{P}(A_1)$; б) $\mathbf{P}(\overline{A}_1/A_2) = \mathbf{P}(\overline{A}_1)$.

4.7 Двое играют в игру, поочередно вынимая шар из урны, содержащей 3 белых и 5 черных шаров. Выигравшим считается тот, кто первым вынет белый шар. Чему равна вероятность выигрыша для начавшего игру, если после каждого неудачного эксперимента шар возвращается в урну?

4.8 Письмо находится в столе с вероятностью p, причем с равной вероятностью оно может быть в любом из восьми ящиков стола. Мы пересмотрели 7 ящиков и письма не обнаружили. Какова вероятность, что письмо:

а) находится в восьмом ящике;

б) отсутствует в столе?

4.9 Из полной колоды карт (52 листа) вынимают сразу две карты. Одну из них смотрят — она оказалась дамой, после этого вынутые карты перемешивают и одну из них берут наугад. Найти вероятность, что она окажется тузом.

4.10 Четыре шара последовательно размещаются в четырех ящиках. Какова вероятность того, что один из ящиков будет содержать ровно три шара, если известно, что первые два шара оказались в разных ячейках?

4.11 Два игрока играют в безобидную игру (у обоих шансы победить одинаковы, ничьих нет), и они договорились, что тот, кто первым выиграет шесть партий, получит весь приз. Предположим, что на самом деле игра остановилась до того, как один из них выиграл приз, например, первый игрок выиграл пять партий, а второй — три. Как справедливо следует разделить приз?

4.12 Доказать, что если $C \subseteq B \subseteq A$, то $\mathbf{P}(C/A) = \mathbf{P}(C/B)\mathbf{P}(B/A)$.

Глава 5

Независимые события

Два события A, B называются **независимыми**, если для них выполняется соотношение:

$$\mathbf{P}(AB) = \mathbf{P}(A) \cdot \mathbf{P}(B). \tag{5.1}$$

Если $\mathbf{P}(B) > 0$, то независимость событий A, B равносильна выполнению равенства

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

которое означает, что вероятность события A не зависит от появления B.

События $A_1, A_2, ..., A_n$ называются **независимыми в совокупности**, если для любого конечного набора из этих событий выполняется соотношение:

$$\mathbf{P}(A_{i_1} \cdot A_{i_2} \cdots A_{i_k}) = \mathbf{P}(A_{i_1}) \cdot \mathbf{P}(A_{i_2}) \cdots \mathbf{P}(A_{i_k}). \tag{5.2}$$

События $A_1, A_2, ..., A_n$ называются **попарно независимыми** , если любые два из них независимы.

§ 5.1. Решение типовых примеров

Пример 5.1. *Точка $\xi = (\xi_1, \xi_2)$ выбрана наудачу в квадрате $[0,1]^2$. Пусть $A = \{\xi_1 \le 1/2\}, B = \{\xi_2 \le 1/2\}$ $C = \{(\xi_1 - 1/2)(\xi_2 - 1/2) > 0\}$. Выяснить, независимы ли эти события:*
а) в совокупности; *б) попарно.*

Решение. Пространство элементарных исходов - единичный квадрат $\Omega = [0,1]^2$. Событие A — левая его половина, событие B — нижняя, событие C представляет объединение двух квадратов

$$AB = \{\xi_1 \le 1/2,\ \xi_2 \le 1/2\} \quad \overline{A}\,\overline{B} = \{\xi_1 \ge 1/2,\ \xi_2 \ge 1/2\},$$

заштрихованных на рис.5.1. Вероятности этих событий найдем с помощью

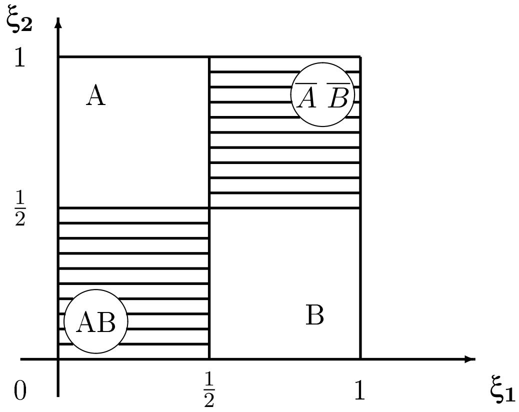


Рис. 5.1:

геометрического определения:

$$\mathbf{P}(A) = \frac{1}{2}, \quad \mathbf{P}(B) = \frac{1}{2}, \quad \mathbf{P}(C) = \frac{1}{2}.$$

Вероятности попарных пересечений равны:

$$\mathbf{P}(AB) = \frac{1}{4} = \mathbf{P}(A) \cdot \mathbf{P}(B), \ \mathbf{P}(AC) = \frac{1}{4} = \mathbf{P}(A) \cdot \mathbf{P}(C),$$

$$\mathbf{P}(BC) = \frac{1}{4} = \mathbf{P}(B) \cdot \mathbf{P}(C).$$

Таким образом, данные события попарно независимы. Так как $ABC = AC = BC = AB$, то

$$\mathbf{P}(ABC) = \frac{1}{4} \neq \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(C),$$

а значит, события A, B, C не являются независимыми в совокупности. $\bigtriangledown$

§ 5.2. Задачи для самостоятельного решения

5.1 Доказать, что если события A и B несовместны, $\mathbf{P}(A) > 0$ и $\mathbf{P}(B) > 0$, то события A и B зависимы.

5.2 Доказать, что если события A и B независимы, то также независимы события: а) A и $\overline{B}$; б) $\overline{A}$ и B; в) $\overline{A}$ и $\overline{B}$.

5.3 Рассмотрим все возможные перестановки четырех букв a, b, c и d. Определить, зависимы или нет события $A = \{a \text{ предшествует } b\}$ и $B = \{c \text{ предшествует } d\}$.

5.4 Брошены две игральные кости. Рассмотрим три события: A — на первой кости выпало нечетное число, B — на второй кости выпало нечетное число, C — сумма чисел на обеих костях нечетна. Проверить, зависимы или нет события A, B, C:

а) в совокупности; б) попарно.

5.5 Пусть событие A таково, что оно не зависит от самого себя. Доказать, что тогда $\mathbf{P}(A) = 0$ или $\mathbf{P}(A) = 1$.

5.6 Брошены последовательно три монеты. Определить, зависимы или нет события $A = \{\text{выпал герб на первой монете}\}$ и $B = \{\text{выпала хотя бы одна решка}\}$.

5.7 На рис. 5.2 приведена схема соединения элементов, образующих электрическую цепь. Предполагается, что отказы элементов являются независимыми в совокупности событиями. Считается известной надежность p_k k-го элемента, то есть вероятность его безотказной работы (соответственно $q_k = 1 - p_k$ — вероятность его отказа). Вычислить надежность p всей цепи.

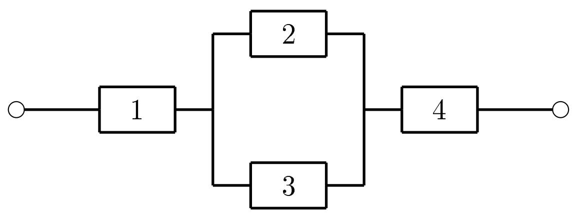


Рис. 5.2:

Рис. 5.2.

5.8 Бросаются три правильные монеты. Пусть событие A состоит в том, что выпало по крайней мере две решки, а событие B — в том, что хотя бы на одной монете выпал герб. Описать событие A. Определить, зависимы или нет события A и B.

5.9 Бросаются две игральные кости. Пусть событие A состоит в том, что выпавшая сумма очков нечетна, а событие B — в том, что хотя бы на одной из костей выпала единица. Описать событие AB. Определить, зависимы или нет события A и B.

5.10 Пусть $0 < \mathbf{P}(A) < \mathbf{P}(B) < 1$. Следует ли отсюда независимость событий A и B?

5.11 Чтобы найти нужную книгу, студент решил обойти 3 библиотеки. Для каждой библиотеки одинаково вероятно, есть в фондах эта книга или нет, и если книга есть в фондах, то с вероятностью 0,5 она не занята другим читателем. Какова вероятность того, что студент найдет книгу, если известно, что библиотеки комплектуются независимо одна от другой?

Глава 6

Независимые испытания

§ 6.1. Формулы Бернулли

Рассмотрим случайный эксперимент, состоящий из последовательности (конечной или бесконечной) отдельных однотипных экспериментов $I_1, I_2, ..., I_n, ...$, которые будем называть **испытаниями**. Будем предполагать, что в каждом испытании возможны лишь два исхода: появление некоторого события A либо противоположного события $\overline{A}$, при этом появление A будем называть «успехом», а появление $\overline{A}$ — «неудачей». Обозначим A_i ={появление A (или «успех») в i-ом испытании}. Будем предполагать, что вероятность «успеха» в каждом испытании одна и та же, т.е. $\mathbf{P}(A_i) = p, \quad \mathbf{P}(\overline{A_i}) = 1 - p = q$ при всех i.

Обозначим через ν_n число «успехов» в n независимых испытаниях. Очевидно, ν_n может принимать любые целые значения от 0 до n. Введем события $B_k = \{\nu_n = k\}, \quad k = 0, 1, \ldots, n$. Вероятности событий B_k вычисляются по следующим **формулам Бернулли**:

$$\mathbf{P}(B_k) = \mathbf{P}(\nu_n = k) = C_n^k p^k q^{n-k}, \; k = 0, 1, ..., n. \tag{6.1}$$

Полиномиальная схема.

Пусть имеется n независимых испытаний, но в каждом испытании может наступить один из k возможных исходов: $R_1, R_2, \ldots, R_k$, где для каждого из n испытаний вероятности исходов одни и те же: $\mathbf{P}(R_j) = p_j, \quad j = 1, 2, ..., k$, так что $p_1 + p_2 + \ldots + p_k = 1$. Соответствующее вероятностное пространство, в отличие от схемы Бернулли, называется **полиномиальной схемой.** Обозначим ν_n^j число исходов R_j в n независимых испытаниях. Следующая теорема обобщает формулы Бернулли на случай полиномиальной схемы.

Теорема 6.1. *Для любых натуральных* $n_1, n_2, ..., n_k$ *таких, что* $n = n_1 + n_2 + ... + n_k$,

$$\mathbf{P}(\nu_n^1 = n_1, \nu_n^2 = n_2, ..., \nu_n^k = n_k) = \frac{n!}{n_1! \cdot n_2! \cdots n_k!}\ p_1^{n_1} \cdot p_2^{n_2} \cdots p_k^{n_k}. \qquad (6.2)$$

§ 6.2. Решение типовых примеров

Пример 6.1. *При передаче сообщения вероятность искажения одного знака равна 1/10. Какова вероятность того, что сообщение из 10 знаков: а) не будет искажено, б) содержит ровно 3 искажения, в) содержит не более трех искажений?*

Решение. Передачу $n = 10$ знаков можно рассматривать как n независимых испытаний, искажение знака будем считать успехом, вероятность успеха $\mathbf{P}(\text{успеха}) = p = 1/10$. Искомые вероятности находим по формулам Бернулли:
а) $\mathbf{P}(\nu_n = 0) = C_n^0 p^0 q^n = q^n = (\frac{9}{10})^{10}$;
б) $\mathbf{P}(\nu_n = 3) = C_n^3 p^3 q^{n-3} = C_{10}^3 (\frac{1}{10})^3 (\frac{9}{10})^7$;
в) $\mathbf{P}(\nu_n \leq 3) = \mathbf{P}\{\bigcup\limits_{k=0}^{3} (\nu_n = k)\} = \sum\limits_{k=0}^{3} C_n^k p^k q^{n-k} = (\frac{9}{10})^{10} + C_{10}^1 \frac{1}{10} (\frac{9}{10})^9 + $ $+ C_{10}^2 (\frac{1}{10})^2 (\frac{9}{10})^8 + C_{10}^3 (\frac{1}{10})^3 (\frac{9}{10})^7$. $\bigtriangledown$

Пример 6.2. *В одном учебном заведении обучаются 730 студентов. День рождения наудачу выбранного студента приходится на определенный день года с вероятностью* $1/365$ *для каждого из 365 дней. Найти вероятность того, что найдутся три студента, имеющие один и тот же день рождения.*

Решение. Найдем вероятность дополнительного события, то есть найдем вероятность, того что ровно два студента родились в каждый из месяцев года. Для этого воспользуемся полиномиальной схемой. В нашем случае, очевидно, $n = 730$, причем вероятность рождения наудачу выбранного студента в определенный день одинакова и равна $p_i = 1/365$, $i = 1...365$. Под событием A_i, $i = 1...365$ понимается рождение выбранного студента в i-ом месяце. Тогда нам собственно нужно найти вероятность, что в 730 испытаниях произойдет ровно 2 исхода типа A_1, ровно два исхода типа A_2,..., ровно два исхода типа A_{365} или, что то же самое, $\nu_{730}^j = 2$, $j = 1...365$. В таком случае вероятность дополнительного события равна:

$$\frac{730!}{(2!)^{365}}(1/365)^{730}.$$

Искомая же вероятность равна $1 - \frac{730!}{(2!)^{365}}(1/365)^{730}$ $\triangledown$.

§ 6.3. Задачи для самостоятельного решения

6.1 Испытание заключается в бросании трех игральных костей. Найти вероятность того, что в пяти независимых испытаниях ровно два раза выпадет по три единицы.

6.2 Наблюдениями установлено, что в некоторой местности в сентябре в среднем бывает 12 дождливых дней. Какова вероятность того, что из восьми случайно взятых в этом месяце дней три дня окажутся дождливыми?

6.3 Пусть вероятность попадания в цель равна $1/5$. Производится 10 независимых выстрелов. Какова вероятность попадания в цель по меньшей мере дважды?

6.4 Вероятность получения удачного результата при проведении сложного химического опыта равна $2/3$. Найти наивероятнейшее число удачных опытов, если их общее число равно 7.

6.5 Вероятность попадания в цель при каждом выстреле из орудия равна $4/5$. Сколько нужно произвести выстрелов, чтобы наивероятнейшее число попаданий было равно 20?

6.6 Каждую секунду с вероятностью p независимо от других моментов времени по дороге проезжает автомашина. Пешеходу для перехода дороги необходимо 3 с. Какова вероятность того, что подошедший к дороге пешеход будет ожидать возможности перехода: а) 3 с; б) 4 с; в) 5с?

6.7 Двое по очереди бросают монету. Выигрывает тот, кто первым получит «герб». Найти вероятность событий:
а) игра закончится до 4 бросания;
б) выиграет начавший игру (первый игрок);
в) выиграет второй игрок.

6.8 В схеме Бернулли p — вероятность исхода 1 и $q = 1 - p$ — вероятность исхода 0. Найти вероятность того, что цепочка 00 (два нуля подряд) появится раньше цепочки 10. В частности, вычислить эту вероятность при $p = 1/2$.

6.9 Технический контроль проверяет изделия, каждое из которых независимо от других изделий может с вероятностью p оказаться дефектным.

а) Какова вероятность того, что из 10 проверенных изделий только одно оказалось дефектным?
б) Найти вероятность того, что первым дефектным оказалось k-е проверенное изделие.
в) Найти вероятность того, что последующие 10 изделий окажутся годными при условии, что предыдущие $l = 5$ изделий были также годными. Зависит ли эта вероятность от l?
г) Найти вероятность того, что число обнаруженных при проверке годных изделий между двумя последовательными дефектными равно k.

6.10 В некоторой группе людей дальтоники составляют 1%. Найти вероятность того, что среди 100 человек:
а) нет дальтоников; б) дальтоников два или больше.

6.11 Отрезок $[0, 10]$ точками 1, 2, 3, 4, 7 разделен на четыре отрезка длины 1 и на два отрезка длины 3. Пусть $A_1, ..., A_8$ — независимые случайные точки, имеющие равномерное распределение на отрезке $[0, 10]$. Какова вероятность того, что из этих точек в два каких-либо отрезка длиной 1 попадет по 2 точки, а в каждый из оставшихся отрезков — по одной точке?

6.12 На отрезок AB длины a брошены наудачу, независимо одна от другой, шесть точек. Найти вероятность того, что:
а) две точки будут находится от точки A на расстоянии, меньшем b, а четыре — на расстоянии большем b;
б) две точки будут находится от точки A на расстоянии, меньшем c, одна — на расстоянии, не меньшем c и не большем b, а три точки — на расстоянии большем b.

6.13 Какова вероятность получить каждую грань дважды при бросании двенадцати игральных костей?

Глава 7

Формулы полной вероятности и Байеса

§ 7.1. Полная группа событий

Предположим, что для A можно найти конечную или бесконечную последовательность событий $\{H_k\}$, которые удовлетворяют следующим условиям:
C1. События $\{H_k\}$ попарно несовместны, т.е. $H_i \cap H_j = \emptyset$ при $i \neq j$.
C2. Событие A происходит, лишь когда происходит одно из событий H_k, т.е.

$$A \subseteq \bigcup_k H_k.$$

C3. $\mathbf{P}(H_k) > 0$ при всех k.

Набор событий, удовлетворяющий этим условиям, называют **полной группой событий (гипотез).** Условия C1–C2 равносильны тому, что **вместе с событием A выполняется одно из событий $\{H_k\}$, причем только одно.** Если событие A и последовательность событий $\{H_k\}$ удовлетворяют условиям C1–C3, то вероятность $\mathbf{P}(A)$ может быть вычислена по следующей **формуле полной вероятности:**

$$\mathbf{P}(A) = \sum_k \mathbf{P}(A/H_k)\mathbf{P}(H_k). \tag{7.1}$$

В тех же условиях, если произошло событие A положительной вероятности, то с учетом этого события новые, т. е. условные вероятности событий H_j вычисляются по *формуле Байеса*

$$\mathbf{P}(H_j/A) = \frac{\mathbf{P}(A/H_j)\mathbf{P}(H_j)}{\sum_k \mathbf{P}(A/H_k)\mathbf{P}(H_k)}.$$

Эти вероятности называются *апостериорными* (т. е. вычисляемыми после осуществления события A).

§ 7.2. Решение типовых примеров

Пример 7.1. *Из урны, содержащей 3 белых и 2 черных шара, вынимаются наудачу два шара и перекладываются во вторую урну, содержащую 4 белых и 4 черных шара.* а) *Найти вероятность вынуть после этого из второй урны белый шар.* б) *Пусть из второй урны был вынут белый шар. Какова вероятность того, что при перекладывании из первой урны во вторую были положены два белых шара?*

Решение. Пусть A ={из второй урны вынут белый шар}. В качестве полной группы событий, связанных с этим событием, можно рассмотреть все возможные исходы первой фазы эксперимента, т.е. перекладывания двух шаров из первой урны во вторую. Обозначим, например, $H_{\text{ЧЧ}}$ ={из первой урны вынуты два черных шара} , $H_{\text{ЧБ}}$ ={из первой урны вынуты один черный и один белый шары}, $H_{\text{ББ}}$ ={из первой урны вынуты два белых шара}. Очевидно, что эти события попарно несовместны, и всегда происходит одно из них, следовательно, они образуют полную группу событий. Для вычисления вероятности события А применим формулу полной вероятности:

$$\mathbf{P}(A) = \mathbf{P}(A/H_{\text{ЧЧ}})\mathbf{P}(H_{\text{ЧЧ}}) + \mathbf{P}(A/H_{\text{ЧБ}})\mathbf{P}(H_{\text{ЧБ}}) + \mathbf{P}(A/H_{\text{ББ}})\mathbf{P}(H_{\text{ББ}}). \quad (7.2)$$

Найдем вероятности гипотез:

$$\mathbf{P}(H_{\text{ЧЧ}}) = \frac{C_3^0 C_2^2}{C_5^2} = \frac{1}{10}, \quad \mathbf{P}(H_{\text{ЧБ}}) = \frac{C_3^1 C_2^1}{C_5^2} = \frac{6}{10},$$

$$\mathbf{P}(H_{\text{ББ}}) = \frac{C_3^1 C_2^0}{C_5^2} = \frac{3}{10}.$$

Условные вероятности находим, исходя из известного состава второй урны при осуществлении той или иной гипотезы:

$$\mathbf{P}(A/H_{\text{ЧЧ}}) = \frac{4}{10} = \frac{2}{5}, \quad \mathbf{P}(A/H_{\text{ЧБ}}) = \frac{5}{10} = \frac{1}{2}, \quad \mathbf{P}(A/H_{\text{ББ}}) = \frac{6}{10} = \frac{3}{5}.$$

Подставляя в (7.2) найденные вероятности, получим:

$$\mathbf{P}(A) = \frac{1}{10} \cdot \frac{2}{5} + \frac{6}{10} \cdot \frac{1}{2} + \frac{3}{10} \cdot \frac{3}{5} = \frac{13}{25}.$$

Ответим на второй вопрос задачи. Искомую вероятность $\mathbf{P}(H_{\text{ББ}}/A)$ находим по формуле Байеса:

$$\mathbf{P}(H_{\text{ББ}}/A) = \frac{\mathbf{P}(A/H_{\text{ББ}})\mathbf{P}(H_{\text{ББ}})}{\mathbf{P}(A/H_{\text{ЧЧ}})\mathbf{P}(H_{\text{ЧЧ}}) + \mathbf{P}(A/H_{\text{ЧБ}})\mathbf{P}(H_{\text{ЧБ}}) + \mathbf{P}(A/H_{\text{ББ}})\mathbf{P}(H_{\text{ББ}})}$$

$$= \frac{\mathbf{P}(A/H_{\text{ББ}})\mathbf{P}(H_{\text{ББ}})}{\mathbf{P}(A)} = \frac{\frac{3}{10}\cdot\frac{3}{5}}{\frac{13}{25}} = \frac{9}{26}. \qquad \triangledown$$

Пример 7.2. *По каналу связи может быть передана одна из трех последовательностей букв: $AAAA$, $BBBB$, $CCCC$, причем априорные вероятности каждой из последовательностей есть соответственно 3/10, 2/5, и 3/10. Известно, что под действием шумов вероятность правильного приема каждой из переданных букв равна 3/5, а вероятности перевода каждой буквы в любую другую одинаковы и равны 1/5. Предполагается, что буквы искажаются независимо друг от друга. Найти вероятность того, что была передана последовательность $AAAA$, если на приемном устройстве получена последовательность $ABCA$.*

Решение. Обозначим D={на приемном устройстве получена последовательность ABCA}, а в качестве гипотез выберем события: H_A ={передана последовательность AAAA }, H_B ={передана последовательность BBBB}, H_C ={передана последовательность CCCC}. Из условия задачи очевидно, что события H_A, H_B, H_C образуют полную группу событий, а их вероятности равны

$$\mathbf{P}(H_A) = \frac{3}{10},\ \mathbf{P}(H_B) = \frac{2}{5},\ \mathbf{P}(H_C) = \frac{3}{10}.$$

Требуется найти условную вероятность $\mathbf{P}(H_A/D)$, которую вычислим по формуле Байеса:

$$\mathbf{P}(H_A/D) = \frac{\mathbf{P}(D/H_A)\mathbf{P}(H_A)}{\mathbf{P}(D/H_A)\mathbf{P}(H_A) + \mathbf{P}(D/H_B)\mathbf{P}(H_B) + \mathbf{P}(D/H_C)\mathbf{P}(H_C)}. \tag{7.3}$$

Найдем условные вероятности, входящие в правую часть (7.3). Имеем: $\mathbf{P}(D/H_A) = \mathbf{P}(ABCA/H_A)$, последняя вероятность есть вероятность приема последовательности букв «ABCA» вместо переданной последовательности «AAAA», то есть вероятность правильного приема первой и четвертой букв A и перевода второй буквы A в B и третьей — в C. Так как правильный прием или искажения каждой из передаваемых букв являются независимыми событиями, то вероятность указанного пересечения событий равна произведению вероятностей:

$$\mathbf{P}(D/H_A) = \mathbf{P}(ABCA/H_A) = \mathbf{P}(A/H_A)\cdot\mathbf{P}(B/H_A)\cdot\mathbf{P}(C/H_A)\cdot\mathbf{P}(A/H_A) =$$

$$= \frac{3}{5}\cdot\frac{1}{5}\cdot\frac{1}{5}\cdot\frac{3}{5} = \frac{9}{625}.$$

Аналогично находим условные вероятности:

$$\mathbf{P}(D/H_B) = \mathbf{P}(ABCA/H_B) = \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} = \frac{3}{625},$$

$$\mathbf{P}(D/H_C) = \mathbf{P}(ABCA/H_C) = \frac{3}{625}.$$

Подставляя в (7.3) все найденные вероятности, получаем

$$\mathbf{P}(H_A/D) = \frac{\frac{9}{625} \cdot \frac{3}{10}}{\frac{9}{625} \cdot \frac{3}{10} + \frac{3}{625} \cdot \frac{2}{5} + \frac{3}{625} \cdot \frac{3}{10}} = \frac{9}{16}. \qquad \triangledown$$

Пример 7.3. *Три стрелка производят по одному выстрелу в одну и ту же мишень. Вероятности попадания в мишень при одном выстреле для каждого из стрелков соответственно равны* p_1, p_2, p_3. *Какова вероятность того, что второй стрелок промахнулся, если после выстрелов в мишени оказалось две пробоины?*

Решение. Обозначим $A =${после выстрелов в мишени оказалось две пробоины}, $H_i =${из трех стрелков промахнулся i-й, а два других попали}, $i = 1, 2, 3$. Требуется найти условную вероятность $\mathbf{P}(H_2/A)$, которую мы найдем по формуле Байеса. Для этого найдем вероятности гипотез:

$$\mathbf{P}(H_1) = \mathbf{P}(\overline{B_1} B_2 B_3) = \mathbf{P}(\overline{B_1}) \cdot \mathbf{P}(B_1) \cdot \mathbf{P}(B_3) = (1 - p_1) p_2 p_3.$$

Здесь использованы обозначения $B_i =${попадание i-го стрелка}, $\overline{B_i} =$ $=${промах i-го стрелка} и предположение, что результаты стрельбы отдельных стрелков являются независимыми событиями. Аналогично находим:

$$\mathbf{P}(H_2) = \mathbf{P}(B_1 \overline{B_2} B_3) = p_1(1 - p_2)p_3, \quad \mathbf{P}(H_3) = \mathbf{P}(B_1 B_2 \overline{B_3}) = p_1 p_2 (1 - p_3).$$

Заметим, что все условные вероятности равны 1: $\mathbf{P}(A/H_1) = \mathbf{P}(A/H_2) = \mathbf{P}(A/H_3) = 1$. Тогда подставляя все эти вероятности в формулу Байеса, получаем

$$\mathbf{P}(H_2/A) = \frac{p_1(1 - p_2)p_3}{(1 - p_1)p_2 p_3 + p_1(1 - p_2)p_3 + p_1 p_2 (1 - p_3)}. \qquad \triangledown$$

§ 7.3. Задачи для самостоятельного решения

7.1 В первой урне находятся 1 белый и 9 черных шаров, во второй — 1 черный и 5 белых. Из каждой урны удалили по одному шару, выбранному наугад, а оставшиеся шары ссыпали в третью урну.

а) Найти вероятность того, что шар, вынутый наугад из третьей урны, окажется белым.

б) Если шар, вынутый из третьей урны, оказался белым, то какова вероятность того, что из первых двух урн были удалены черные шары?

7.2 В первой урне находятся 1 белый шар и 4 красных, во второй — 1 белый и 7 красных. В первую урну добавили два шара, выбранных наугад из второй урны.

а) Найти вероятность того, что шар, вынутый наугад из пополненной первой урны, окажется белым.

б) Поставить вопрос, ответ на который можно найти с помощью формулы Байеса (и ответить на него).

7.3 Имеется 5 урн следующего состава: в первой и второй урнах по 2 белых и 3 черных шара в каждой; в третьей и четвертой урнах — по 1 белому и 4 черных шара; в пятой урне — 4 белых и 1 черный шар. Из одной наудачу выбранной урны взят шар. Он оказался белым. Чему равна при этом вероятность того, что он вынут из пятой урны?

7.4 В трех урнах содержатся белые и черные шары: в первой урне — 2 белых и 3 черных шара, во второй — 2 белых и 2 черных шара, в третьей — 3 белых и 1 черный шар. Из первой урны вынут наудачу шар и переложен во вторую. Далее из второй урны вынут наудачу шар и переложен в третью. Наконец из третьей урны шар переложен в первую.

а) Какой состав шаров в первой урне наиболее вероятен?

б) С какой вероятностью состав шаров во всех урнах не изменится?

7.5 Из урны, в которой было $m \geq 3$ белых и n черных, потеряли один шар неизвестного цвета. Для того чтобы определить состав шаров в урне, из нее наудачу были вынуты два шара. Найти вероятность того, что был потерян белый шар, если известно, что вынутые шары оказались белыми.

7.6 В урне лежат 12 шаров, из них 8 черных и 4 белых. Три игрока A, B и C поочередно вынимают шары. Выигрывает тот, кто первым вынет белый шар. Оценить шансы на успех каждого игрока.

7.7 При некоторых условиях стрельбы стрелок А поражает мишень с вероятностью $p_1 = 3/5$, стрелок В — с вероятностью $p_2 = 1/2$, стрелок С — с вероятностью $p_3 = 2/5$. Стрелки дали залп по мишени, и две пули попали в цель. Что вероятнее: попал С в мишень или нет?

7.8 Изделия некоторого производства удовлетворяют стандарту с вероятностью 0,96. Предлагается упрощенная система испытаний, дающая положительный результат с вероятностью 0,98 для изделий, удовлетворяющих стандарту, а для изделий, которые не удовлетворяют стандарту, с вероятностью 0,05. Какова вероятность того, что:

а) изделие будет забраковано;

б) изделие, выдержавшее испытание, удовлетворяет стандарту?

7.9 Некоторое изделие выпускается двумя заводами, причем объем продукции второго завода в k раз превосходит объем продукции первого. Доля брака у первого завода p_1, у второго — p_2. Изделия, выпущенные заводами за одинаковый промежуток времени, перемешали и пустили в продажу. Какова вероятность того, что вы приобрели изделие со второго завода, если оно оказалось бракованным?

7.10 В семи урнах содержится по 3 белых и 2 черных шара, а в трех урнах по 7 белых и 3 черных шара. Какова вероятность, что из урны, взятой наудачу, будет извлечен белый шар? Найти вероятность, что шар извлечен из урны с 7 белыми и 3 черными шарами, если он оказался белым.

Глава 8

Распределения случайных величин

§ 8.1. Случайная величина и функция распределения

Случайной величиной (с.в.) называется всякая функция $X: \ \Omega \to R$, сопоставляющая каждому элементарному исходу $\omega \in \Omega$ действительное число $X(\omega)$.

Функцией распределения (ф.р.) случайной величины X называется функция $F_X : R \to R$, задаваемая соотношением:

$$F_X(t) = \mathbf{P}(X < t) = \mathbf{P}\{\omega : X(\omega) < t\}, \quad t \in R. \tag{8.1}$$

Функция распределения обладает свойствами:

F1. $0 \le F_X(t) \le 1 \quad \forall\, t \in R.$

F2. $F_X \nearrow$ (неубывающая функция).

F3. $F_X(-\infty) = \lim\limits_{t\to-\infty} F_X(t) = 0, \qquad F_X(+\infty) = \lim\limits_{t\to+\infty} F_X(t) = 1$.

F4. $F_X(t)$ непрерывна слева и имеет конечный предел справа: для всех $t \in R$

$$\lim_{y\to t-0} F_X(y) = F_X(t); \quad \lim_{y\to t+0} F_X(y) = F_X(t+0).$$

F5. Для любых $a < b$ вероятность попадания с.в. X в интервал $[a,b)$ равна

$$\mathbf{P}\{a \le X < b\} = F_X(b) - F_X(a) .$$

F6. Для любого $a \in R$ вероятность попадания с.в. X в точку a равна скачку ф.р. в точке a:

$$\mathbf{P}\{X = a\} = F_X(a+0) - F_X(a).$$

§ 8.2. Дискретное и абсолютно непрерывное распределения

Множество G называется **счетным**, если оно бесконечно, но все элементы его можно занумеровать числами натурального ряда: $G = \{x_1, x_2, ..., x_n, ...\}$. Множество называют **не более чем счетным**, если оно конечно или счетно. Говорят, что с.в. X имеет **дискретное распределение**, если с вероятностью 1 она может принимать не более чем счетное множество значений. Иначе говоря, с.в. X имеет дискретное распределение, если существует не более чем счетное множество $G = \{x_1, x_2, ..., x_n, ...\}$, такое, что

$$p_i = \mathbf{P}(X = x_i) > 0, \; i = 1, 2, ... \qquad \sum_i p_i = 1. \tag{8.2}$$

Упрощая терминологию, будем говорить в этом случае, что **случайная величина** X **дискретна**. Если с.в. X дискретна, то для нее можно указать две последовательности: последовательность $x_1, x_2, ...$ возможных значений X и последовательность $p_1, p_2, ...$ вероятностей этих значений. Таблица, составленная из этих последовательностей, называется **рядом распределения** или **таблицей распределения** X:

X	x_1	x_2	...	x_i	...
$\mathbf{P}(X = x_i)$	p_1	p_2	...	p_i	...

(8.3)

Если с.в. X имеет дискретное распределение (8.3), то вероятность ее попадания в множество $A \subseteq R$ вычисляется по формуле:

$$\mathbf{P}(X \in A) = \sum_{x_i \in A} p_i. \tag{8.4}$$

В частности, функция распределения случайной величины X выражается через ряд распределения формулой:

$$F_X(t) = \sum_{x_i < t} p_i. \tag{8.5}$$

Если предположить, что все значения случайной величины X упорядочены по возрастанию, то есть

$$x_1 < x_2 < \ldots < x_n < \ldots,$$

то функция распределения будет иметь график, изображенный на рис. 8.1.

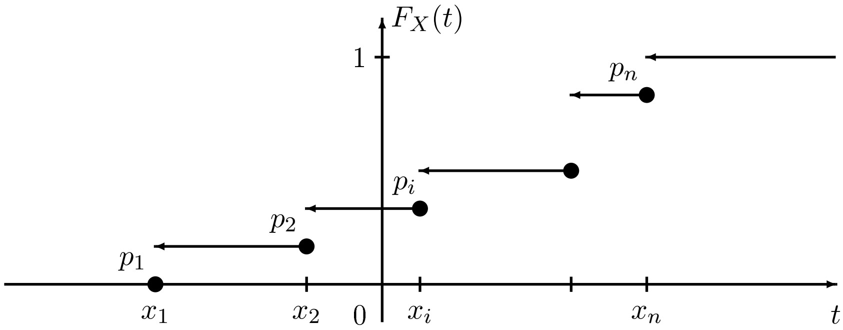


Рис. 8.1: График функции распределения $F_X(t)$.

.

Функция распределения дискретной случайной величины X является **ступенчатой**, то есть она сохраняет постоянные значения в интервалах (x_{i-1}, x_i), а в точках x_i имеет разрыв первого рода (скачок), равный вероятности $p_i = \mathbf{P}(X = x_i)$.

Говорят, что с.в. X имеет **абсолютно непрерывное распределение**, если ее функция распределения $F_X(t)$ может быть представлена в виде:

$$F_X(t) = \int_{-\infty}^{t} f_X(u)du, \tag{8.6}$$

где $f_X(u)$ — неотрицательная, интегрируемая на всей числовой оси функция, которая называется **плотностью распределения** случайной величины X. Если распределение с.в. X абсолютно непрерывно, то ее плотность распределения обладает свойствами:
f1. Для всех $t \in \mathbf{R}$ плотность распределения неотрицательна:

$$f_X(t) \geq 0.$$

f2. Интеграл от плотности распределения по всей числовой прямой равен единице (**условие нормировки**):

$$\int_{-\infty}^{\infty} f_X(t)dt = 1.$$

f3. Вероятность того, что случайная величина принимает значения из множества, равна интегралу от плотности распределения по этому множеству:

$$\mathbf{P}(X \in A) = \int\limits_A f_X(t)dt$$

для любого множества $A \subseteq \mathbf{R}$, для которого интеграл в правой части определен.

f4. Функция распределения $F_X(t)$ непрерывна на всей числовой оси, а в точках непрерывности плотности $f_X(t)$ ф.р. $F_X(t)$ дифференцируема, причем

$$F'_X(t) = f_X(t).$$

f5. Вероятность попадания случайной величины в точку равна нулю: для любого $a \in \mathbf{R}$ выполнено

$$\mathbf{P}(X = a) = 0.$$

f6. Для любых $a < b$

$$\mathbf{P}(a \leq X < b) = \mathbf{P}(a \leq X \leq b) = \mathbf{P}(a < X \leq b) = \mathbf{P}(a < X < b) = \int_a^b f_X(t)dt.$$

Вероятности попадания в промежуток, фигурирующие в последнем свойстве, вычисляются как площадь криволинейной трапеции, заштрихованной на рис. 8.2:

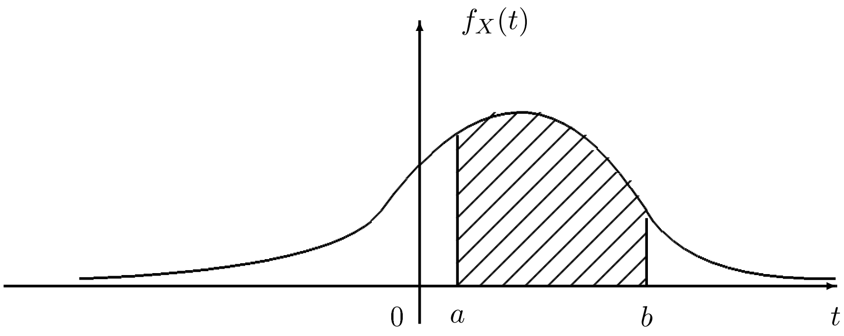


Рис. 8.2: Плотность распределения $f_X(t)$.

Отметим, что свойства f1 и f2 являются характеристическими свойствами плотности распределения, то есть любая функция, удовлетворяющая этим свойствам, является плотностью распределения некоторой случайной величины.

§ 8.3. Примеры распределений случайных величин

1. Распределение Бернулли B_p. Это распределение с. в. X, принимающей лишь два значения 1 и 0 с вероятностями p и $q = 1 - p$ соответственно. Как мы видели, такое распределение имеет случайная величина, равная числу «успехов» при одном испытании Бернулли. Очевидно, распределение X дискретно, и ряд распределения задается равенством:

$$\mathbf{P}(X = k) = \begin{cases} p, & \text{если } k = 1; \\ 0, & \text{если } k = 0. \end{cases}$$

Тот факт, что с. в. X имеет распределение Бернулли, будем обозначать следующим образом: $X \Subset B_p$, где p — параметр, $0 < p < 1$.

2. Биномиальное распределение $B_{n,p}$. Говорят, что с. в. Y имеет биномиальное распределение с параметрами (n, p), где $n = 1, 2, \ldots$, $0 < p < 1$ (обозначаем $Y \Subset B_{n,p}$), если она может принимать значения $0, 1, \ldots, n$ с вероятностями, вычисляемыми по формулам Бернулли:

$$p_k = \mathbf{P}(Y = k) = C_n^k p^k q^{n-k}, \; k = 0, 1, \ldots, n.$$

Как мы знаем, биномиальное распределение имеет случайная величина Y, равная числу «успехов» в n независимых испытаниях Бернулли. Очевидно, распределение Бернулли есть частный случай биномиального при $n = 1: \; B_p = B_{1,p}$.

3. Распределение Пуассона Π_λ. Говорят, что с. в. Z имеет распределение Пуассона с параметром $\lambda > 0$ (обозначаем $Z \Subset \Pi_\lambda$), если она может принимать целые неотрицательные значения с вероятностями, вычисляемыми по формулам:

$$p_k = \mathbf{P}(Z = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots \, .$$

Распределение Пуассона является предельным для биномиального $B_{n,p}$, когда n велико, а p мало (см. более точно теоремы Пуассона 10.6, 10.7). Приближенно можно считать, что закону распределения Пуассона удовлетворяют такие случайные величины, как число вызовов, поступивших на телефонную станцию за определенный промежуток времени, или число отказавших элементов сложной аппаратуры, состоящей из большого числа таких элементов, за некоторый промежуток времени.

4. Геометрическое распределение G_p. Говорят, что с. в. X имеет геометрическое распределение с параметром p, где $0 < p < 1$ (обозначаем $X \Subset G_p$), если она может принимать положительные целые значения с вероятностями, вычисляемыми по формулам:

$$p_k = \mathbf{P}(X = k) = pq^{k-1}, \quad k = 1, 2, \ \dots\ , \quad q = 1 - p.$$

Название этого распределения объясняется тем, что вероятности p_k ряда распределения образуют бесконечно убывающую геометрическую прогрессию. Геометрическое распределение имеет с. в. X, равная числу проведенных испытаний Бернулли до появления первого «успеха».

Замечание 8.1. *Все вышеприведенные примеры представляют дискретные распределения. Читателю предлагается проверить, что для них выполняются обязательные для всякого дискретного распределения условия p1 – p2.*

5. Равномерное распределение $U_{[a;\ b]}$. Говорят, что с. в. X имеет равномерное распределение на отрезке $[a;\ b]$ (обозначаем $X \Subset U_{[a;\ b]}$), если распределение X абсолютно непрерывно, а плотность распределения постоянна на отрезке $[a, b]$ и равна нулю вне этого отрезка:

$$f_X(t) = \begin{cases} C, & \text{если } t \in [a, b], \\ 0, & \text{если } t \notin [a, b]. \end{cases}$$

Константу C можно найти, воспользовавшись обязательным для всякой плотности условием нормировки *f2*:

$$1 = \int\limits_{-\infty}^{\infty} f_X(t)dt == \int\limits_{-\infty}^{a} 0dt + \int\limits_{a}^{b} Cdt + \int\limits_{b}^{\infty} 0dt = 0 + C(b-a) + 0 = C(b-a).$$

Отсюда $C = \dfrac{1}{b-a}$, и плотность равномерного распределения $U_{[a;\ b]}$ принимает следующий окончательный вид:

$$f_X(t) = \begin{cases} \dfrac{1}{b-a}, & \text{если } t \in [a, b], \\ 0, & \text{если } t \notin [a, b]. \end{cases} \tag{8.7}$$

Нетрудно построить графики функции и плотности равномерного распределения $U_{[a;\ b]}$, они представлены на рис. 8.3.

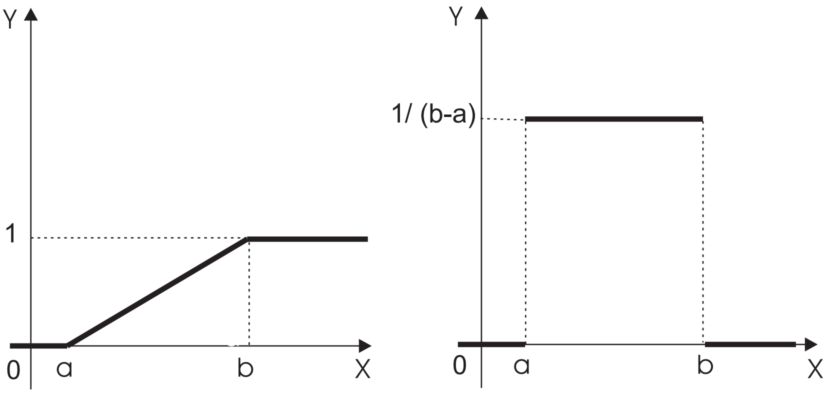


Рис. 8.3: Функция и плотность распределения равномерного закона.

Упражнение 8.1. *Найдите аналитическое выражение для ф.р. равномерного распределения* $U_{[a;\ b]}$.

Полезно отметить еще одно свойство равномерного распределения, часто используемое при вычислении вероятностей.

Если с. в. X *имеет равномерное распределение на отрезке* $[a, b]$, *то вероятность ее попадания в любой промежуток, целиком содержащийся в отрезке* $[a, b]$, *вычисляется согласно геометрическому определению вероятности. Например, если* $[c, d] \subset [a, b]$, *то*

$$\mathbf{P}(X \in [c, d]) = fracd - cb - a. \tag{8.8}$$

Доказательство. Поскольку вероятность попадания в промежуток вычисляется интегралом от плотности распределения по этому промежутку, то

$$\mathbf{P}(X \in [c, d]) = \int\limits_c^d f_X(t)dt = \int\limits_c^d \frac{1}{b-a}dt = \frac{1}{b-a}\int\limits_c^d dt = \frac{d-c}{b-a}.$$

5. Экспоненциальное (показательное) распределение E_α. Говорят, что с. в. X имеет показательное распределение с параметром $\alpha > 0$ (обозначаем $X \Subset E_\alpha$), если распределение X абсолютно непрерывно с плотностью распределения

$$f_X(t) = \begin{cases} Ce^{-\alpha t}, & \text{если } t \geq 0, \\ 0, & \text{если } t < 0. \end{cases}$$

Константу C можно найти, как и в предыдущем примере, воспользовавшись условием нормировки *f2*:

$$1 = \int\limits_{-\infty}^{\infty} f_X(t)dt = \int\limits_{-\infty}^{0} f_X(t)dt + \int\limits_{0}^{\infty} f_X(t)dt =$$

$$= 0 + C\int\limits_{0}^{\infty} e^{-\alpha t}dt = \frac{C}{-\alpha}e^{-\alpha t}\,|_0^{\infty} = \frac{C}{\alpha}.$$

Отсюда $C = \alpha$, и плотность экспоненциального распределения E_α имеет вид

$$f_X(t) = \begin{cases} \alpha e^{-\alpha t}, & \text{если } t \geq 0, \\ 0, & \text{если } t < 0. \end{cases} \tag{8.9}$$

Закону экспоненциального распределения подчиняются такие случайные величины, как время безотказной работы радиоаппаратуры или время между поступлением двух последовательных вызовов на станцию скорой медицинской помощи.

Упражнение 8.2. *Найти аналитическое выражение для ф.р. показательного распределения E_α и построить графики функции и плотности распределения.*

Определение 8.1. Говорят, что случайная величина X имеет *нормальное распределение*, или *распределение Гаусса*, с параметрами (a, σ^2), если распределение X абсолютно непрерывно с плотностью

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-a)^2}{2\sigma^2}}, \quad -\infty < t < \infty. \tag{8.10}$$

Нормальное, или гауссовское, распределение занимает особое место в теории вероятностей в силу своей распространенности в приложениях и теоретической важности для многих фундаментальных результатов теории вероятностей.

Относительно параметров нормального распределения предполагается, что a — любое действительное число, называемое *средним*, а $\sigma > 0$ называется *стандартным, или средним квадратическим, отклонением.*

Тот факт, что с. в. X имеет нормальное распределение с параметрами (a, σ^2), будем обозначать: $X \Subset N_{a,\sigma^2}$. Нормальное распределение $N_{0,1}$

с параметрами $a = 0, \sigma^2 = 1$ называют *стандартным нормальным распределением*, или *стандартным нормальным законом.*

Отметим, что если $X \in N_{a,\sigma^2}$, то случайная величина X может быть представлена в виде

$$X = a + \sigma Z, \tag{8.11}$$

где случайная величина Z имеет стандартное нормальное распределение.

Смысл параметров (a, σ^2) нормального закона можно понять из графика плотности, представленного на рис. 8.4.

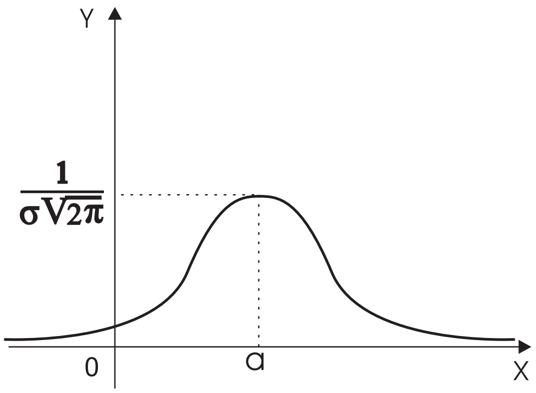


Рис. 8.4: Плотность распределения нормального закона.

Найдем теперь ф. р. нормально распределенной случайной величины. Будем обозначать через $\Phi(t)$ ф. р. стандартного нормального закона. Тогда

$$\Phi(t) = \int\limits_{-\infty}^{t} \varphi(u)du = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{t} e^{-\frac{u^2}{2}} \sigma du. \tag{8.12}$$

Функция $\Phi(t)$, определенная этим равенством, называется *функцией Лапласа*, ее значения протабулированы, и соответствующие таблицы имеются во всех руководствах по теории вероятностей или математической статистике. В настоящем учебнике она помещена в приложении а вконце книги. Для практического использования таблицы полезны некоторые свойства функции $\Phi(t)$. Они содержатся в следующей лемме.

Лемма 8.1. *Для функции распределения* $\Phi(t)$ *выполняются соотношения: для всех* $t > 0$

$$\Phi(-t) = 1 - \Phi(t), \quad \Phi(0) = \frac{1}{2}. \tag{8.13}$$

Доказательство. Рассмотрим график плотности $y = \varphi(x)$ (рис. 8.5).

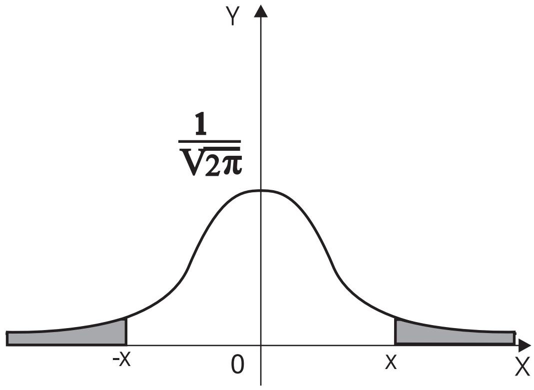


Рис. 8.5: График плотности распределения стандартного нормального закона.

Обозначим через Z с. в., имеющую стандартное нормальное распределение. Так как в силу (8.12)

$$\Phi(-t) = \mathbf{P}(Z < -t) = \int\limits_{-\infty}^{-t} \varphi(u)du,$$

$$1 - \Phi(t) = 1 - \mathbf{P}(Z < u) = \mathbf{P}(Z \geq u) = \int\limits_{t}^{\infty} \varphi(u)du,$$

то значения $\Phi(-t)$ и $1 - \Phi(t)$ равны площадям, заштрихованным на рис. 3.10, которые совпадают ввиду симметрии плотности $y = \varphi(x)$. Тем самым доказано первое из соотношений (8.13). Второе следует из равенств (**??**) при $a = 0$:

$$\Phi(0) = \mathbf{P}(Z < 0) = \mathbf{P}(Z \leq 0) = \frac{1}{2}.$$

Замечание 8.2. *Доказанная лемма позволяет пользоваться таблицами функции Лапласа* $\Phi(t)$ *только при положительных* t.

Следующая лемма выражает ф. р. произвольного нормального закона N_{a,σ^2} через функцию Лапласа.

Лемма 8.2. *Функция распределения с. в. $X \in N_{a,\sigma^2}$ вычисляется по формуле*
$$\Phi_{a,\sigma^2}(t) = \Phi\left(\frac{t-a}{\sigma}\right). \tag{8.14}$$

Доказательство. Вспомним, что согласно (8.11) справедливо представление $X = a+\sigma Z$, где Z имеет стандартное нормальное распределение. Выразим ф. р. $\Phi_{a,\sigma}(t)$ по определению функции распределения:
$$\Phi_{a,\sigma^2}(t) = \mathbf{P}(X < t) = \mathbf{P}(a+\sigma Z < t) = \mathbf{P}\left(Z < \frac{t-a}{\sigma}\right) = \Phi\left(\frac{t-a}{\sigma}\right).$$

§ 8.4. Генерирование случайных чисел

Рассмотим генерирование последовательности независимых одинаково распределенных случайных величин, имеющих распределение F. Для простоты будем предполагать, что F — строго возрастающая функция. Отметим, что здесь мы рассмотрим только один способ генерирования случайных величин.

Оказывается, случайная величина $F^{-1}(\omega)$, где ω имеет равномерное распределение на отрезке $[0,1]$, распределена по закону F. Это можно установить, воспользовавшись монотонностью F. Действительно,
$$\mathbf{P}(F^{-1}(\omega) < t) = \mathbf{P}(\omega < F(t)) = F(t),$$
в силу того, что ω равномерно распределена на отрезке $[0;\ 1]$. Поэтому для того, чтобы сгенерировать последовательность $X_1, X_2, \ldots$ независимых случайных величин с распределением F, достаточно сгенерировать последовательность $\omega_1, \omega_2, \ldots$ независимых равномерно распределенных на $[0,1]$ случайных величин и положить $X_i = F^{-1}(\omega_i)$, $i = 1, 2, \ldots$. Для генерирования последовательности равномерно распределенных на отрезке $[0,1]$ случайных величин можно использовать, например, пакет программ MS Excel. Для этого в пункте меню «Вставка» нужно открыть пункт «Функция» и в категории «математические» найти встроенную функцию

СЛЧИС()

— эта функция будет выдавать при каждом ее запуске числа, равномерно распределенные на отрезке $[0,1]$. В следующем примере мы для конкретного F сгенерируем последовательность независимых случайных величин.

Пример 8.1. *Сгенерировать последовательность экспоненциально распределенных случайных величин.*

Решение. Говорят, что случайная величина X имеет **экспоненциальное распределение с параметром** α (обозначение: $X \Subset \mathbf{E}_\alpha$), если ее распределение имеет вид

$$F_X(t) = \begin{cases} 1 - \exp(-\alpha t), & \text{если } t > 0, \\ 0, & \text{если } t \leq 0. \end{cases} \tag{8.15}$$

Тогда $F^{-1}(t) = -\frac{1}{\alpha}\ln(1-t)$. Далее генерируем последовательность равномерно распределенных на отрезке $[0,1]$ случайных величин: ω_1, $\omega_2, \ldots$. В итоге последовательность

$$F^{-1}(\omega_i) = -\frac{1}{\alpha}\ln(1-\omega_i),\ i = 1, 2, \ldots$$

и является искомой.

$\bigtriangledown$

§ 8.5. Решение типовых примеров

Пример 8.2. *Точка* $\omega = (\omega_1, \omega_2)$ *выбирается наудачу в треугольнике с вершинами в точках* (0,0), (2,1) *и* (2,0). *Найти функцию распределения случайной величины* X*, если:*
а) $X(\omega) = \omega_1$*;* *б)* $X(\omega) = \omega_2$.

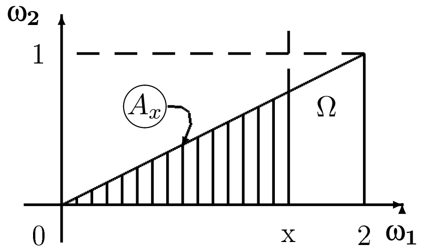
Решение. а) Согласно определению (8.1), функция распределения $F_X(x)$ есть вероятность того, что случайная величина примет значение, меньшее x.

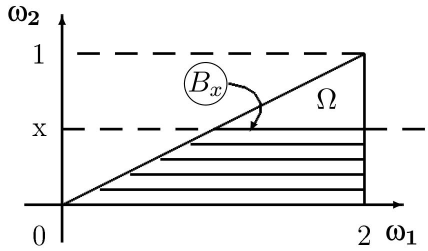
Так как точка $\omega = (\omega_1, \omega_2)$ выбирается в пределах треугольника Ω (см. рис.**??**), то с.в. $X(\omega) = \omega_1$ может принимать свои значения лишь в пределах отрезка $[0, 2]$.

Поэтому при $x \leq 0$ имеем $F_X(x) = \mathbf{P}(X < x) = 0$, а при $x > 2$ получаем $F_X(x) = \mathbf{P}(X < x) = 1$.

Пусть теперь $0 < x \leq 2$. Тогда событие $\{X < x\} = \{\omega_1 < x\}$ означает, что точка $\omega = (\omega_1, \omega_2)$ окажется левее прямой $\omega_1 = x$, то есть окажется в области A_x, заштрихованной на рис. 8.6.

Рис. 8.6: Вычисление функций распределения в примере 8.2.





Вероятность попадания в эту область находим согласно геометрическому определению

$$F_X(x) = \mathbf{P}(X < x) = \mathbf{P}(\omega \in A_x) = \frac{\mu(A_x)}{\mu(\Omega)} = \frac{x^2}{4}.$$

Таким образом, искомая функция распределения с.в. X равна

$$F_X(x) = \begin{cases} 0, & \text{если } x \le 0, \\ \frac{x^2}{4}, & \text{если } 0 < x \le 2, \\ 1, & \text{если } x > 2. \end{cases}$$

б) Аналогично предыдущему, находим:

$$F_X(x) = \mathbf{P}(X < x) = \mathbf{P}(\omega_2 < x) = 0$$

при $x \le 0$;

$$F_X(x) = \mathbf{P}(X < x) = \mathbf{P}(\omega_2 < x) = 1$$

при $x > 1$. Пусть $0 < x \le 1$. Тогда имеем (см. рис. **??**):

$$F_X(x) = \mathbf{P}(X < x) = \mathbf{P}(\omega_2 < x) = \frac{\mu(B_x)}{\mu(\Omega)} = \frac{1-(1-x)^2}{1} = 2x - x^2.$$

Полное выражение для ф.р. случайной величины X имеет вид:

$$F_X(x) = \begin{cases} 0, & \text{если } x \le 0, \\ 2x - x^2, & \text{если } 0 < x \le 1, \\ 1, & \text{если } x > 1. \end{cases} \qquad \triangledown$$

Пример 8.3. *Монету бросают $n = 3$ раза. Найти ряд распределения и функцию распределения: а) числа выпадений герба; б) разности чисел выпадения герба и решетки.*

Решение. а) Обозначая X — число выпадений герба, заметим, что с.в. X может принимать лишь конечное множество значений: 0, 1, 2, 3. Следовательно, распределение X дискретно, а ряд распределения определяется формулами Бернулли (6.1):

$$p_k = \mathbf{P}(X = k) = C_n^k p^k q^{n-k}, \; k = 0, 1, ..., n.$$

Подставляя в эти формулы значения $p = q = \frac{1}{2}$, $n = 3$, находим:

$$p_0 = C_3^0 \frac{1}{2^3} = \frac{1}{8}, \;\; p_1 = C_3^1 \frac{1}{2^3} = \frac{3}{8}, \;\; p_2 = \frac{3}{8}, \;\; p_3 = \frac{1}{8}.$$

Найденный ряд распределения можно представить таблицей:

i	0	1	2	3
$\mathbf{P}(X = i)$	1/8	3/8	3/8	1/8

Функцию распределения проще всего представить графически подобно рис. 8.7.

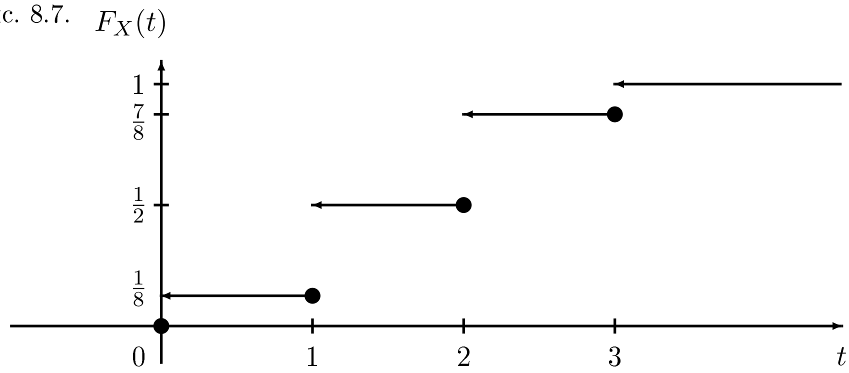


Рис. 8.7: График $F_X(t)$ в примере 8.3.

Аналитически функцию распределения $F_X(t)$ можно задать формулой:

$$F_X(t) = \begin{cases} 0, & \text{если } t \le 0, \\ 1/8, & \text{если } 0 < t \le 1, \\ 1/2, & \text{если } 1 < t \le 2, \\ 7/8, & \text{если } 2 < t \le 3, \\ 1, & \text{если } t > 3. \end{cases}$$

б) Обозначая Y — разность чисел выпадения герба и решетки, заметим, что $Y = X - (3 - X) = 2X - 3$. Следовательно, Y может принимать конечное множество значений: $-3, -1, 1, 3$. Поэтому распределение случайной величины Y дискретно, а ее ряд распределения определяется из соотношений:

$$p_i = \mathbf{P}(X = i) = \mathbf{P}(2X - 3 = 2i - 3) = \mathbf{P}(Y = 2i - 3), \quad i = 0, 1, 2, 3.$$

Подставляя значения p_i из таблицы распределения X, получим ряд распределения для Y:

k	-3	-1	1	3
$\mathbf{P}(Y = k)$	$1/8$	$3/8$	$3/8$	$1/8$

Функцию распределения $F_Y(t)$ строим так же, как в пункте а), с помощью ряда распределения Y. $\triangledown$

Пример 8.4. *Говорят, что случайная величина X имеет* **равномерное распределение на отрезке [a; b]** *(обозначение: $X \Subset U_{[a;\ b]}$), если ее распределение абсолютно непрерывно с плотностью*

$$f_X(t) = \begin{cases} C = const, & \text{если } t \in [a;\ b], \\ 0, & \text{если } t \notin [a;\ b]. \end{cases} \tag{8.16}$$

Считая, что $X \Subset U_{[0;\ 1]}$, найти плотности распределения случайных величин: $Y = 2X + 1$; $Z = X^2$.

Решение. Из вида плотности распределения с.в. $X \Subset U_{[a;\ b]}$ вытекает, что $\mathbf{P}(a \le X \le b) = 1$. Так как $Y = 2X + 1$, то все значения Y с вероятностью 1 принадлежат отрезку $[2a + 1;\ 2b + 1]$. Подставляя $a = 0$, $b = 1$, находим: $\mathbf{P}(1 \le Y \le 3) = 1$. Следовательно,

$$F_Y(t) = \mathbf{P}(Y < t) = \begin{cases} 0, & \text{если } t \le 1, \\ \mathbf{P}(2X + 1 < t), & \text{если } 1 < t \le 3, \\ 1, & \text{если } t > 3. \end{cases} \tag{8.17}$$

Вычислим вероятность в средней строке последней формулы, пользуясь функцией распределения случайной величины $X \Subset U[0;\ 1]$:

$$F_X(t) = \begin{cases} 0, & \text{если } t \le 0, \\ t, & \text{если } 0 < t \le 1, \\ 1, & \text{если } t > 1. \end{cases} \tag{8.18}$$

$$\mathbf{P}(2X+1<t)=\mathbf{P}\left(X<\frac{t-1}{2}\right)=F_X\left(\frac{t-1}{2}\right)=\frac{t-1}{2},$$

так как $1<t\leq 3$, а значит, и $0<\frac{t-1}{2}\leq 1$. Подставляя результат этого вычисления в (8.17), получаем окончательно:

$$F_Y(t)=\begin{cases}0, & \text{если } t\leq 1,\\ \frac{t-1}{2}, & \text{если } 1<t\leq 3,\\ 1, & \text{если } t>3.\end{cases}$$

Плотность распределения с.в. Y находим в соответствии со свойством f4 дифференцированием ф.р. $F_Y(t)$:

$$f_Y(t)=F'_Y(t)=\begin{cases}0, & \text{если } t\notin[1;\ 3],\\ \frac{1}{2}, & \text{если } 1<t<3.\end{cases}$$

Функцию распределения с.в. $Z=X^2$ находим аналогично предыдущему, причем с самого начала заметим, что $\mathbf{P}(Z\in[0;\ 1])=1$, а следовательно,

$$F_Z(t)=\mathbf{P}(Z<t)=\begin{cases}0, & \text{если } t\leq 0,\\ \mathbf{P}(Z<t), & \text{если } 0<t\leq 1,\\ 1, & \text{если } t>1.\end{cases}\tag{8.19}$$

Пусть $0<t\leq 1$. Тогда

$$F_Z(t)=\mathbf{P}(Z<t)=\mathbf{P}(X^2<t)=\mathbf{P}(|X|<\sqrt{t})=\mathbf{P}(X<\sqrt{t}),$$

так как $\mathbf{P}(X\in[0;\ 1])=1$. Вероятность $\mathbf{P}(X<\sqrt{t})$ вычисляем, пользуясь известным выражением (8.18) для ф.р. $X\Subset U[0;\ 1]$ и учитывая, что $0<\sqrt{t}\leq 1$:

$$\mathbf{P}(X<\sqrt{t})=\sqrt{t}.$$

Подставляя результаты проделанных вычислений в (8.19), получаем

$$F_Z(t)=\mathbf{P}(Z<t)=\begin{cases}0, & \text{если } t\leq 0,\\ \sqrt{t}, & \text{если } 0<t\leq 1,\\ 1, & \text{если } t>1.\end{cases}$$

Наконец, плотность $f_Z(t)$ находим дифференцированием функции распределения $F_Z(t)$:

$$f_Z(t)=F'_Z(t)=\begin{cases}0, & \text{если } t\notin[0,1],\\ \frac{1}{2\sqrt{t}}, & \text{если } 0<t<1.\end{cases}\qquad\triangledown$$

§ 8.6. Задачи для самостоятельного решения

8.1 Вероятность попадания баскетбольного мяча в кольцо при бросании начинающим спортсменом равна $1/4$. Мяч бросают до первого попадания, но дают не более 2 попыток. Найти ряд распределения числа промахов. Построить график функции распределения.

8.2 По мишени одновременно стреляют 2 стрелка, вероятности попаданий которых равны соответственно 0,3 и 0,6. Найти ряд распределения числа попаданий в мишень. Построить график функции распределения.

8.3 Вероятность попадания в мишень равна 0,6 при каждом выстреле. Стрельба ведется одиночными выстрелами до первого попадания, пока не будет израсходован боезапас. Найти ряд распределения числа произведенных выстрелов, если боезапас составляет 2 единицы. Построить график функции распределения.

8.4 Игральную кость бросают $n = 6$ раз. Найти ряд распределения и функцию распределения: а) числа выпадений шестерки; б) разности чисел выпадения шестерки и тройки.

8.5 Вероятность отказа сервера при каждом из независимых подключений с помощью модема равна 0,3. Попытки подключения производятся до установления связи. Найти ряд распределения числа произведенных попыток подключения, если число попыток ограничено двумя. Построить график функции распределения. **8.6** Игральную кость бросают до первого появления шестерки. Найти ряд распределения и функцию распределения числа проведенных бросаний.

8.7 Монету бросают до появления двух гербов. Найти ряд распределения и функцию распределения числа проведенных бросаний.

8.8 Дискретная случайная величина X имеет ряд распределения

x_i	-1	0	1
$\mathbf{P}(X = x_i)$	1/3	1/3	1/3

Построить ряды распределения следующих случайных величин:

а) $2X + 5$; г) 2^X;

б) $X^2 + 1$; д) $\min(X, 1)$;

в) $|X|$; е) $1/(3 - X)$.

8.9 Решить задачу **8.8**, если дискретная случайная величина X имеет ряд распределения

x_i	-2	-1	0	1	2
$\mathbf{P}(X = x_i)$	1/10	1/5	3/10	3/10	1/10

8.10 Скорость пешехода на дистанции в 1 км является случайной величиной, равномерно распределенной на отрезке от 2 км/ч до 6 км/ч. Найти вероятность того, что время, затраченное на преодоление дистанции, превысит 24 минуты.

8.11 Закон Рэлея с плотностью распределения

$$f(x) = \begin{cases} Axe^{-x^2/2} & \text{при } x \geq 0; \\ 0 & \text{при } x < 0 \end{cases}$$

в ряде случаев описывает распределение срока службы электронной аппаратуры. Найти коэффициент A, функцию распределения и вероятность того, что случайная величина, распределенная по закону Рэлея, примет значение, большее 1.

8.12 Случайная величина X имеет плотность распределения

$$f(x) = \begin{cases} A\cos 4x & \text{при } x \in [0; \pi/8]; \\ 0 & \text{при } x \notin [0; \pi/8]. \end{cases}$$

Найти коэффициент A и функцию распределения. Построить графики плотности распределения и функции распределения.

8.13 Непрерывная функция распределения случайной величины X задана выражением

$$F(t) = \begin{cases} 0 & \text{при } t \leq 0; \\ At^3 & \text{при } 0 < t \leq 2; \\ 1 & \text{при } t > 0. \end{cases}$$

Найти коэффициент A, плотность распределения, а также $\mathbf{P}(0 < X < 1)$. Построить графики функции распределения и плотности распределения.

8.14 На окружность радиуса R с центром в начале координат наудачу брошена точка. Найти функцию и плотность распределения:

а) абсциссы точки попадания;

б) длины хорды, соединяющей точку попадания с точкой $(-R, 0)$.

8.15 Равнобедренный треугольник образован единичным вектором в направлении оси абцисс и единичным вектором в случайном направлении в R^2. Найти функцию распределения длины третьей стороны.

8.16 Из точки $(0, a)$ проведена прямая под углом φ к оси ординат. Найти функцию распределения точки пересечения этой прямой с осью абсцисс, если угол φ равномерно распределен в промежутке:

а) $(0, \pi/2)$; б) $(-\pi/2, \pi/2)$.

8.17 На отрезок оси ординат между точками $(0, 0)$ и $(0, R)$ наудачу брошена точка. Через точку попадания проведена хорда окружности $x^2 + y^2 = R^2$, перпендикулярная оси ординат. Найти распределение длины этой хорды.

8.18 Говорят, что случайная величина X имеет показательное распределение с параметром $\alpha > 0$ $(X \Subset E(\alpha))$, если ее распределение абсолютно непрерывно с плотностью:

$$f_X(t) = \begin{cases} ce^{-\alpha t}, & \text{если } t > 0, \\ 0, & \text{если } t \leq 0. \end{cases}$$

Найти значение входящей в определение $f_X(t)$ постоянной c и функцию распределения $F_X(t)$. Построить графики плотности распределения и функции распределения.

8.19 Можно ли подобрать постоянную c так, чтобы функция ct^{-4} была плотностью распределения на множестве:

а) $[1;\infty)$; в) [-2;-1];

б) $(0;\infty)$; г) [-3;0).

8.20 Пусть случайная величина X распределена равномерно на отрезке [0,1] $(X \Subset U[0;1])$. Найти плотности распределения следующих случайных величин:

а) $-\ln X$; в) $-\ln(1-X)$;

б) $X - 1/X$; г) e^{X-1}.

8.21 Пусть случайная величина X имеет показательное распределение с параметром α, то есть абсолютно непрерывное с плотностью

$$f_X(t) = \begin{cases} \alpha e^{-\alpha t}, & \text{если } t > 0, \\ 0, & \text{если } t \leq 0. \end{cases}$$

Найти плотности распределения следующих случайных величин:

а) $\sqrt{X}$; г) $\ln(\alpha X)$;

б) X^2; д) $e^{-\alpha X}$;

в) $2X$; е) $\min(X, X^2)$.

8.21 Пусть случайная величина X имеет плотность распределения p. Найти плотности распределения следующих величин:

а) $aX + b$, $a, b \in R$, $a \neq 0$; г) X^3;

б) X^{-1}; д) e^X;

в) X^2; е) $|X - 1|$.

Глава 9

Математическое ожидание

§ 9.1. Определение и свойства

Пусть $X = X(\omega)$ — случайная величина, заданная на пространстве элементарных исходов Ω.

Математическим ожиданием с.в. X называется число, обозначаемое $\mathbf{E}X$ или $\mathbf{M}X$, которое определяется следующим образом.
Если X имеет дискретное распределение: $p_n = \mathbf{P}(X = x_n)$, $n = 1, 2, \ldots$, то

$$\mathbf{E}X = \sum_{n=1}^{\infty} x_n p_n. \tag{9.1}$$

Если X имеет абсолютно непрерывное распределение с плотностью $f_X(x)$, то

$$\mathbf{E}X = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{9.2}$$

Говорят, что **математическое ожидание существует (конечно),** если ряд в правой части (9.1) или интеграл в (9.2) **абсолютно сходится.** Математическое ожидание (м.о.) $\mathbf{E}X$ можно рассматривать как **среднее значение** случайной величины. Математическое ожидание обладает следующими свойствами:
E1. М.о. постоянной равно этой постоянной: $\mathbf{E}C = C$.
E2. М.о. линейно, то есть для любых постоянных C_1, C_2 и любых с.в. X_1, X_2 выполнено равенство

$$\mathbf{E}(C_1 X_1 + C_2 X_2) = C_1 \mathbf{E}X_1 + C_2 \mathbf{E}X_2$$

при условии, что математические ожидания в правой части существуют. В частности, **постоянный множитель можно выносить из-под знака**

математического ожидания:

$$\mathbf{E}(C \cdot X) = C \cdot \mathbf{E}X, \tag{9.3}$$

а **математическое ожидание суммы равно сумме математических ожиданий**:

$$\mathbf{E}(X_1 + X_2 + \ldots + X_n) = \mathbf{E}X_1 + \mathbf{E}X_2 + \ldots + \mathbf{E}X_n. \tag{9.4}$$

Е3. Если с.в. $X \geq 0$ с вероятностью 1, то $\mathbf{E}X \geq 0$. При этом $\mathbf{E}X = 0$ равносильно $\mathbf{P}\{X = 0\} = 1$.
Е4. Если с.в. X_1, X_2 независимы, то математическое ожидание их произведения равно произведению математических ожиданий:

$$\mathbf{E}(X_1 \cdot X_2) = \mathbf{E}X_1 \cdot \mathbf{E}X_2 \tag{9.5}$$

при условии существования м.о. в правой части. Это свойство распространяется на любое число с.в.:

$$\mathbf{E}(X_1 \cdot X_2 \cdots X_n) = \mathbf{E}X_1 \cdot \mathbf{E}X_2 \cdots \mathbf{E}X_n, \tag{9.6}$$

если с.в. X_1, X_2, ..., X_n независимы.

Случайные величины, удовлетворяющие условию (9.5), называются **некоррелированными.** Таким образом, свойство Е4 утверждает, что если с.в. независимы, то они некоррелированы.

В дальнейшем нам понадобится формула для вычисления $\mathbf{E}g(X)$, где g — некоторая функция. Если случайная величина X имеет абсолютно непрерывное распределение, то

$$\mathbf{E}g(X) = \int_{-\infty}^{\infty} g(t) f_X(t)\, dt. \tag{9.7}$$

В случае дискретного распределения формула для вычисления $\mathbf{E}g(X)$ принимает вид:

$$\mathbf{E}g(X) = \sum_{n=1}^{\infty} g(x_n) p_n. \tag{9.8}$$

В заключение рассмотрим примеры вычисления математических ожиданий.

Пример 9.1. *Найти м.о. случайной величины X, имеющей распределение Бернулли B_p.*

Решение. Это распределение с. в. X, принимающей лишь два значения 1 и 0 с вероятностями p и $q = 1-p$ соответственно. Напомним, что распределение Бернулли имеет случайная величина, равная числу «успехов» при одном испытании Бернулли. Распределение X дискретно, и ряд распределения задается равенством

$$\mathbf{P}(X = k) = \begin{cases} p, & \text{если } k = 1 \\ 0, & \text{если } k = 0. \end{cases}$$

Математическое ожидание

$$\mathbf{E}X = \sum_{k=1}^{\infty} x_k p_k = 1 \cdot p + 0 \cdot q = p.$$

Пример 9.2. *Найти м.о. случайной величины Y, имеющей биномиальное распределение $B_{n,\ p}$.*

Решение. Напомним, что с. в. Y имеет биномиальное распределение с параметрами (n, p), где $n = 1, 2, \ldots;\ 0 < p < 1$ ($Y \Subset B_{n,p}$), если она может принимать значения $0, 1, \ldots, n$ с вероятностями, вычисляемыми по формулам Бернулли:

$$p_k = \mathbf{P}(Y = k) = C_n^k p^k q^{n-k},\ k = 0, 1, ..., n.$$

Такое распределение имеет с. в. Y, равная числу «успехов» в n независимых испытаниях Бернулли. Поскольку с. в. Y есть число «успехов» в n независимых испытаниях Бернулли, то ее можно представить в виде

$$Y = X_1 + X_2 + ... + X_n, \tag{9.9}$$

где с. в. X_k равна числу «успехов» в одном k-м испытании Бернулли $(k = 1, 2, \ldots, n)$, т. е. она принимает значения 1 или 0 в зависимости от того, был ли «успех» в k-м испытании. Ясно, что все X_k имеют распределение Бернулли B_p. Тогда, применяя следствие **??** и результат предыдущего примера, находим более простой ответ:

$$\mathbf{E}Y = \mathbf{E}(X_1 + X_2 + ... + X_n) = \mathbf{E}X_1 + \mathbf{E}X_2 + ... + \mathbf{E}X_n = np.$$

Пример 9.3. *Найти м.о. случайной величины Y, имеющей распределение Пуассона Π_λ.*

Решение. Напомним, что с. в. Y имеет распределение Пуассона с параметром $\lambda > 0$, ($Y \Subset \Pi_\lambda$), если она может принимать целые неотрицательные значения с вероятностями, вычисляемыми по формулам:

$$p_k = \mathbf{P}(Y = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, \dots .$$

Для вычисления м.о. применим его определение (**??**):

$$\mathbf{E}Y = \sum_{k=1}^{\infty} x_k p_k = \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} =$$

$$= e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda.$$

В последних вычислениях была использована формула Маклорена для показательной функции:

$$e^{\lambda} = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^{k-1}}{(k-1)!} + \dots = \sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}.$$

Упражнение 9.1. *Покажите, что если с. в. Y имеет геометрическое распределение G_p ($Y \Subset G_p$), то ее математическое ожидание равно* $\mathbf{E}Y = \dfrac{1}{p}$.

Пример 9.4. *Найти м.о. случайной величины X, имеющей равномерное распределение $U_{[a;\ b]}$.*

Решение. С. в. $\xi \Subset U_{[a;\ b]}$ имеет абсолютно непрерывное распределение с плотностью

$$f_X(t) = \begin{cases} \dfrac{1}{b-a}, & \text{если } t \in [a, b], \\ 0, & \text{если } t \notin [a, b]. \end{cases}$$

М.о. находим по определению:

$$\mathbf{E}X = \int_{-\infty}^{\infty} t f_X(t)dt = \int_a^b \frac{t}{b-a}\,dt = \frac{1}{b-a}\cdot\frac{t^2}{2}\bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

Пример 9.5. *Найти м.о. случайной величины X, имеющей нормальное распределение N_{a,σ^2}.*

Решение. Напомним, что с. в. $X \Subset N_{a,\sigma^2}$ имеет абсолютно непрерывное распределение с плотностью

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}}, \quad -\infty < t < \infty.$$

М.о. находим по определению:

$$\mathbf{E}X = \int\limits_{-\infty}^{\infty} t f_X(t) dt = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

Вычислим последний интеграл, прибегнув к замене переменной:

$$\frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x e^{-\frac{(t-a)^2}{2\sigma^2}} = \left| \begin{array}{ll} y = \frac{t-a}{\sigma}, & t = \sigma y + a \\ dt = \sigma dy, & -\infty < y < \infty \end{array} \right| =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} (\sigma y + a) e^{-\frac{y^2}{2}} \sigma dy =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} y e^{-\frac{y^2}{2}} dy + \frac{a}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = 0 + a \cdot 1 = a.$$

Упражнение 9.2. *Докажите, что если с. в. X имеет экспоненциальное распределение E_α ($X \Subset E_\alpha$), то ее математическое ожидание равно* $\mathbf{E}X = \dfrac{1}{\alpha}$.

§ 9.2. Моменты и дисперсия

Пусть $k > 0$. **Моментом порядка** k случайной величины X, или **начальным моментом порядка** k, называется число

$$\alpha_k = \mathbf{E}X^k,$$

если соответствующее математическое ожидание существует. Аналогично определяются: **абсолютный момент порядка** k

$$\beta_k = \mathbf{E}|X|^k,$$

центральный момент порядка k

$$\mu_k = \mathbf{E}(X - \mathbf{E}X)^k,$$

абсолютный центральный момент порядка k

$$\nu_k = \mathbf{E}|X - \mathbf{E}X|^k.$$

Вычисляются моменты с помощью формул

$$\mathbf{E}X^k = \sum_{n=1}^{\infty} x_n^k p_n, \quad \mathbf{E}X^k = \int\limits_{-\infty}^{\infty} x^k f_X(x)dx$$

для дискретного и абсолютно непрерывного распределений соответственно.

Говорят, что **момент случайной величины не существует**, если соответствующий ряд или интеграл не является абсолютно сходящимся.

Теорема о существовании моментов утверждает, что из существования момента порядка $k > 0$ случайной величины x следует существование момента любого порядка $l > 0$, меньшего, чем k. В частности, из существования второго момента следует существование первого момента, то есть математического ожидания.

Дисперсией случайной величины X называют центральный момент второго порядка и обозначают ее $\mathbf{D}X$ либо $\mathbf{Var}X$:

$$\mathbf{D}X = \mathbf{Var}X = \mu_2 = \mathbf{E}(X - \mathbf{E}X)^2.$$

При вычислении дисперсии удобно пользоваться следующим ее представлением:

$$\mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2, \tag{9.10}$$

легко вытекающим из определения. В случае абсолютно непрерывного распределения дисперсия вычисляется также по такой формуле

$$\mathbf{D}X = \int\limits_{-\infty}^{\infty} (t - \mathbf{E}X)^2 f_X(t)dt.$$

И аналогично в случае дискретного распределения

$$\mathbf{D}X = \sum_{n=1}^{\infty} (x_n - \mathbf{E}X)^2 p_n.$$

В то время как математическое ожидание представляет среднее значение случайной величины, дисперсия характеризует средний квадрат отклонения случайной величины от ее математического ожидания. Дисперсия имеет следующие свойства.

D1. Дисперсия неотрицательна: $\mathbf{D}X \geq 0$, — и обращается в нуль тогда и только тогда, когда с.в. неслучайна, т. е. $\mathbf{P}(X = C = const) = 1$.

D2. Постоянный множитель выносится из-под знака дисперсии с квадратом:

$$\mathbf{D}(CX) = C^2\mathbf{D}X.$$

D3. Прибавление к с.в. константы не изменяет дисперсии:

$$\mathbf{D}(X + C) = \mathbf{D}X.$$

D4. Если случайные величины $X_1, X_2, ..., X_n$ попарно некоррелированы, тем более, **если они независимы, то дисперсия их суммы равна сумме дисперсий:**

$$\mathbf{D}(X_1 + X_2 + ... + X_n) = \mathbf{D}X_1 + \mathbf{D}X_2 + ... + \mathbf{D}X_n. \tag{9.11}$$

Стандартным (среднеквадратическим) отклонением σ_X называется корень из дисперсии:

$$\sigma_X = \sqrt{\mathbf{D}X}.$$

Стандартное отклонение имеет ту же размерность, что и исходная случайная величина.

Рассмотрим некоторые примеры вычисления дисперсий.

Пример 9.6. *Найти дисперсию случайной величины* X*, имеющей распределение Бернулли* B_p.

Решение. Напомним, что это распределение с. в. X, принимающей лишь два значения 1 и 0 с вероятностями p и $q = 1-p$ соответственно. В примере 9.1 было найдено м.о. $\mathbf{E}X = p$. Тогда, используя формулу (**??**), находим: $\mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \mathbf{E}X^2 - p^2$. Заметим, что с. в. X^2 имеет то же распределение, что и X, а именно: она принимает те же два значения 1 и 0 с вероятностями p и $q = 1 - p$ соответственно. Следовательно, ее м.о. совпадает с м.о. X, то есть $\mathbf{E}X^2 = p$. Учитывая эти соображения, находим окончательно:

$$\mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = p - p^2 = p(1 - p) = pq.$$

Пример 9.7. *Найти дисперсию случайной величины Y, имеющей биномиальное распределение $B_{n,p}$.*

Решение. Найдем $\mathbf{D}Y$, используя свойства дисперсии. Как и при решении примера 9.2, случайную величину Y — число «успехов» в n независимых испытаниях Бернулли — можно представить в виде (9.9):

$$Y = X_1 + X_2 + ... + X_n,$$

где с. в. X_k равна числу «успехов» в одном k-м испытании Бернулли $(k = 1, 2, \ldots, n)$. При этом, поскольку с. в. X_k связана с k-м испытанием Бернулли, а испытания независимы, то с. в. X_1, X_2, ... , X_n независимы. Тогда дисперсия суммы равна сумме дисперсий:

$$\mathbf{D}Y = \mathbf{D}(X_1 + X_2 + ... + X_n) = \mathbf{D}X_1 + \mathbf{D}X_2 + ... + \mathbf{D}X_n = npq,$$

так как все X_k имеют распределение Бернулли B_p и, по предыдущему примеру, $\mathbf{D}X_k = pq$.

Упражнение 9.3. *Показать, что если с. в. Y имеет распределение Пуассона Π_λ, то*

$$\mathbf{D}Y = \mathbf{E}Y = \lambda.$$

Упражнение 9.4. *Найти дисперсию с. в. Y, имеющей геометрическое распределение G_p.*

Упражнение 9.5. *Показать, что если с. в. X имеет равномерное распределение $U_{[a;\ b]}$, то*

$$\mathbf{D}X = \frac{(b-a)^2}{12}.$$

Пример 9.8. *Найти дисперсию случайной величины X, имеющей нормальное распределение N_{a,σ^2}.*

Решение. Найдем дисперсию стандартного нормального закона. Вспомним, что, согласно примеру 9.5, для случайной величины Z, имеющей стандартное нормальное распределение, математическое ожидание равно нулю, и потому

$$\mathbf{D}Z = \mathbf{E}Z^2 - (\mathbf{E}Z)^2 = \mathbf{E}Z^2 == \frac{1}{sqrt2\pi}\int_{-infty}^{\infty} t^2 e^{-t^2/2} dt.$$

Этот интеграл вычислим по частям:

$$\mathbf{D}Z = \left| \begin{array}{ll} u = t, & du = dt \\ dv = te^{-\frac{t^2}{2}}dt, & v = \int te^{-\frac{t^2}{2}}dt = -e^{-\frac{t^2}{2}} \end{array} \right| =$$

$$= \frac{1}{\sqrt{2\pi}}(-te^{-\frac{t^2}{2}})\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt = 0 + 1 = 1.$$

В последних вычислениях были использованы соотношения:

$$\lim_{t\to\pm\infty} te^{-\frac{t^2}{2}} = \lim_{t\to\pm\infty} \frac{t}{e^{\frac{t^2}{2}}} = 0, \quad \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt = 1,$$

первое из которых получается из правила Лопиталя, а второе — из свойства плотности нормального распределения.

Итак, для стандартного нормального распределения $\mathbf{D}Z = 1$.

Случайную величину с произвольным нормальным распределением представим в виде $X = a + \sigma Z$, где Z имеет стандартное нормальное распределение. По свойствам дисперсии,

$$\mathbf{D}X = \mathbf{D}(a + \sigma Z) = \mathbf{D}(\sigma Z) = \sigma^2\mathbf{D}Z = \sigma^2 \cdot 1 = \sigma^2.$$

Таким образом, *параметры нормального распределения* N_{a,σ^2} *представляют соответственно математическое ожидание и дисперсию случайной величины*:

$$a = \mathbf{E}X, \quad \sigma^2 = \mathbf{D}X.$$

§ 9.3. Числовые характеристики случайных векторов

Ковариацией случайных величин X_1, X_2 называется число:

$$\mathbf{cov}(X_1, X_2) = \mathbf{E}(X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2). \tag{9.12}$$

Для вычисления ковариации можно использовать следующую формулу:

$$\mathbf{cov}(X_1, X_2) = \mathbf{E}(X_1X_2) - \mathbf{E}X_1 \cdot \mathbf{E}X_2. \tag{9.13}$$

Ковариация для дискретных случайных величин обычно считается по следующей формуле:

$$\mathbf{cov}(X_1, X_2) = \sum_{k=1}^{\infty}\sum_{n=1}^{\infty} x_n y_k \mathbf{P}(X_1 = x_n, X_2 = y_k) - \mathbf{E}X_1\mathbf{E}X_2.$$

Заметим, что **если с.в.** X_1, X_2 **независимы, то ковариация их равна нулю**: $\mathbf{cov}(X_1, X_2) = 0$. В общем же случае (т.е. когда с.в. не обязательно независимы), справедлива следующая формула, связывающая ковариацию с дисперсиями с.в. $X_1, X_2, X_1 + X_2$:

$$\mathbf{D}(X_1 + X_2) = \mathbf{D}X_1 + 2\mathbf{cov}(X_1, X_2) + \mathbf{D}X_2. \tag{9.14}$$

Последнее соотношение допускает обобщение на случай n случайных слагаемых:

$$\mathbf{D}(X_1+X_2+...+X_n) = \mathbf{D}X_1+\mathbf{D}X_2+...+\mathbf{D}X_n+2\sum_{1\le i<j\le n}\mathbf{cov}(X_i, X_j). \tag{9.15}$$

Коэффициентом корреляции случайных величин X_1, X_2 называется величина $\rho = \rho(X_1, X_2)$, равная отношению:

$$\rho(X_1, X_2) = \frac{\mathbf{cov}(X_1, X_2)}{\sqrt{\mathbf{D}X_1}\sqrt{\mathbf{D}X_2}} = \frac{\mathbf{E}(X_1X_2) - \mathbf{E}X_1 \cdot \mathbf{E}X_2}{\sqrt{\mathbf{D}X_1}\sqrt{\mathbf{D}X_2}}. \tag{9.16}$$

Коэффициент корреляции $\rho = \rho(X_1, X_2)$ широко используется как **мера зависимости между с.в.** X_1, X_2, ввиду следующих своих свойств:
ρ1. $|\rho| \le 1$;
ρ2. Если с.в. X_1, X_2 независимы, то $\rho = \rho(X_1, X_2) = 0$;
ρ3. $|\rho| = 1$ тогда и только тогда, когда X_1, X_2 связаны линейной зависимостью: $X_1 = aX_2 + b$, при этом

$$(\rho = 1) \ \Leftrightarrow \ (a > 0); \qquad (\rho = -1) \ \Leftrightarrow \ (a < 0).$$

§ 9.4. Решение типовых примеров

Пример 9.9. *Стрельба по цели ведется до первого попадания, но дается не более двух попыток. Каково математическое ожидание и дисперсия числа выстрелов, которые будут сделаны, если вероятность попадания в каждом выстреле равна $0,2$?*

Решение. Число сделанных выстрелов — это случайная величина X, которая может принимать лишь два значения 1 или 2, с вероятностями $0,2$ (это означает, что сразу произошло попадание и стрельба закончилась) или $0,8$ соответственно. Причем $\mathbf{P}(X = 2) = 0,8 \cdot 0,2 + 0,8 \cdot 0,8 = 0,8$, т. е. две попытки могут произойти в двух случаях: либо в первый раз не попали, но во второй раз попали, либо и в первый, и во второй раз не попали. Тогда математическое ожидание легко находится по формуле (9.1):

$$\mathbf{E}X = x_1p_1 + x_2p_2 = 1 \cdot 0,2 + 2 \cdot 0,8 = 1,8.$$

Дисперсия считается по формуле:

$$\mathbf{D}X = x_1^2 p_1 + x_2^2 p_2 - (\mathbf{E}X)^2 = 1^2 \cdot 0,2 + 2^2 \cdot 0,8 - 1,8^2 = 0,16. \qquad \bigtriangledown$$

Пример 9.10. *Двумерный случайный вектор (X, Y) имеет распределение, заданное таблицей:*

Y \ X	*0*	*1*
-1	*0,10*	*0,15*
0	*0,15*	*0,25*
1	*0,20*	*0,15*

а) Найти математические ожидания, дисперсии и коэффициент корреляции с. в. X, Y.
б) Найти коэффициент корреляции $\rho(X+Y, X-2Y)$.

Решение. а) Чтобы вычислить математические ожидания и дисперсии, найдем частные распределения с.в. X, Y, записывая их в дополнительные строку и столбец данной таблицы совместного распределения:

Y \ X	0	1	$\mathbf{P}(Y=k)$
-1	0,10	0,15	0,25
0	0,15	0,25	0,40
1	0,20	0,15	0,35
$\mathbf{P}(X=n)$	0,45	0,55	

Используя найденные ряды распределения с.в. X, Y, вычисляем их м.о. и дисперсии:

$$\mathbf{E}X = 0 \cdot 0,45 + 1 \cdot 0,55 = 0,55; \mathbf{E}Y = -1 \cdot 0,25 + 0 \cdot 0,40 + 1 \cdot 0,35 = 0,10;$$

$$\mathbf{E}X^2 = 0^2 \cdot 0,45 + 1^2 \cdot 0,55 = 0,55; \mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = 0,55 - 0,55^2 = 0,2475;$$

$$\mathbf{E}Y^2 = (-1)^2 \cdot 0,25 + 0 + 1^2 \cdot 0,35 = 0,6; \mathbf{D}Y = \mathbf{E}Y^2 - (\mathbf{E}Y)^2 = 0,6 - 0,1^2 = 0,59.$$

Чтобы найти коэффициент корреляции, найдем сначала ковариацию, используя соотношение (9.13) и таблицу распределения X, Y:

$$\mathbf{E}(XY) = \sum_{n,k} (n \cdot k) \mathbf{P}(X=n, Y=k) = 0 \cdot (-1) \cdot 0,10 + 0 \cdot 0 \cdot 0,15 + 0 \cdot 1 \cdot 0,20 +$$

$$+1 \cdot (-1) \cdot 0,15 + 1 \cdot 0 \cdot 0,25 + 1 \cdot 1 \cdot 0,15 = 0;$$

$$\mathbf{cov}(X,Y)=\mathbf{E}(XY)-\mathbf{E}X\cdot\mathbf{E}Y=0-0,55\cdot 0,1=-0,055;$$

$$\rho(X,Y)=\frac{\mathbf{cov}(X,Y)}{\sqrt{\mathbf{D}X}\sqrt{\mathbf{D}Y}}=\frac{-0,055}{\sqrt{0,2475}\sqrt{0,59}}=-0,144.$$

Величина коэффициента корреляции близка к нулю, поэтому можно считать, что с.в. X,Y слабо зависимы.

б) Используя определение ковариации, получаем

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

$$=\mathbf{D}X-\mathbf{cov}(X,Y)-2\mathbf{D}Y=0,2475-0,055-2\cdot 0,59=-0,9875.$$

Дисперсии с.в. $X+Y$, $X-2Y$ найдем, пользуясь соотношением (9.14):

$$\mathbf{D}(X+Y)=\mathbf{D}X+\mathbf{D}Y+2\mathbf{cov}(X,Y)=0,2475-2\cdot 0,055+0,59=0,7275;$$

$$\mathbf{D}(X-2Y)=\mathbf{D}X+4\mathbf{D}Y-4\mathbf{cov}(X,Y)=0,2475-4\cdot 0,055+4\cdot 0,59=2,3875.$$

Остается вычислить коэффициент корреляции:

$$\rho(X+Y,X-2Y)=\frac{\mathbf{cov}(X+Y,X-2Y)}{\sqrt{\mathbf{D}(X+Y)\mathbf{D}(X-2Y)}}=\frac{-0,9875}{\sqrt{0,7275\cdot 2,3875}}=-0,749. \quad \triangledown$$

Пример 9.11. *Из цифр 0, 1, 2, 3, 4 выбирают наудачу две цифры без возвращения. Найти коэффициент корреляции между меньшей и большей из выбранных цифр.*

Решение. Число способов выбрать 2 цифры из 5 без возвращения и без учета порядка равняется $C_5^2=\frac{5!}{2!3!}=10$. Поэтому вероятность каждой допустимой комбинации двух цифр равна $1/10=0,1$.

Обозначим через X меньшую из двух цифр, через Y — большую. Получим следующую таблицу двумерного распределения:

X \ Y	1	2	3	4
0	0,1	0,1	0,1	0,1
1	0	0,1	0,1	0,1
2	0	0	0,1	0,1
3	0	0	0	0,1

Составим таблицы одномерных распределений:

i	0	1	2	3
$\mathbf{P}(X=i)$	0,4	0,3	0,2	0,1

j	1	2	3	4
$\mathbf{P}(X=i)$	0,1	0,2	0,3	0,4

Для сокращения вычислений полезно отметить, что случайная величина Y распределена так же, как $4-X$. Найдем числовые характеристики случайной величины X. Математическое ожидание и второй момент равняются

$$\mathbf{E}X = \sum_{i=0}^{3} x_i p_i = 0 \cdot 0,4 + 1 \cdot 0,3 + 2 \cdot 0,2 + 3 \cdot 0,1 = 0,7;$$

$$\mathbf{E}X^2 = \sum_{i=0}^{3} (x_i)^2 p_i = 0^2 \cdot 0,4 + 1^2 \cdot 0,3 + 2^2 \cdot 0,2 + 3^2 \cdot 0,1 = 0 + 0,3 + 0,8 + 0,9 = 2.$$

Найдем дисперсию:

$$\mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = 2 - 0,7^2 = 2 - 0,49 = 1,51.$$

Для нахождения числовых характеристик случайной величины Y вспомним, что она распределена так же, как $4-X$ (проверьте, что числовые характеристики для Y можно вычислить и непосредственно, и результаты совпадают с теми, что получены ниже):

$$\mathbf{E}Y = \mathbf{E}(4-X) = 4 - \mathbf{E}X = 4 - 0,7 = 3,3;$$

$$\mathbf{D}Y = \mathbf{D}(4-X) = \mathbf{D}(-X) = (-1)^2 \mathbf{D}X = \mathbf{D}X = 1,51.$$

Вычислим математическое ожидание произведения случайных величин XY по таблице совместного распределения, пользуясь формулой

$$\mathbf{E}XY = \sum_{i=0}^{3} \sum_{j=1}^{4} i \cdot j \cdot \mathbf{P}(X=i,\ Y=j).$$

Выписанная двойная сумма состоит из 16 слагаемых, каждое из которых получается произведением значений случайных величин и вероятности в каждой из 16 клеток таблицы. Выпишем те 10 слагаемых, для которых вероятности не равны нулю:

$$\mathbf{E}XY = 0 \cdot 1 \cdot 0,1 + 0 \cdot 2 \cdot 0,1 + 0 \cdot 3 \cdot 0,1 + 0 \cdot 4 \cdot 0,1 +$$

$$+ 1 \cdot 2 \cdot 0,1 + 1 \cdot 3 \cdot 0,1 + 1 \cdot 4 \cdot 0,1 + 2 \cdot 3 \cdot 0,1 + 2 \cdot 4 \cdot 0,1 + 3 \cdot 4 \cdot 0,1 =$$
$$= 0 + 0 + 0 + 0 + 0,2 + 0,3 + 0,4 + 0,6 + 0,8 + 1,2 = 3,5.$$

Следовательно, ковариация случайных величин X и Y равна

$$\mathbf{cov}(X,\ Y) = \mathbf{E}XY - \mathbf{E}X \cdot \mathbf{E}Y = 3,5 - 0,7 \cdot 3,3 = 3,5 - 2,31 = 1,19,$$

и коэффициент корреляции

$$\rho(X,Y) = \frac{\mathbf{cov}(X,\ Y)}{\sqrt{\mathbf{D}X}\sqrt{\mathbf{D}Y}} = \frac{1,19}{\sqrt{1,51^2}} = \frac{1,19}{1,51} \approx 0,79.$$

§ 9.5. Задачи для самостоятельного решения

9.1 Дискретная случайная величина X имеет ряд распределения

x_i	-1	0	1	2
$\mathbf{P}(X = x_i)$	$1/5$	$1/10$	$3/10$	$2/5$

.

Найти математические ожидания и дисперсии случайных величин:

а) X; в) X^2;

б) $|X|$; г) 2^X.

9.2 Пусть случайная величина X принимает значения -2, -1, 0, 1 и 2 с вероятностью $1/5$ каждое. Найти математические ожидания и дисперсии случайных величин:

а) X; в) $|X|$;

б) $-X$; г) X^2.

9.3 Имеется шесть ключей, из которых только один подходит к замку. Составить ряд распределения числа попыток при открывании замка, если испробованный ключ в последующих опробованиях не участвует. Чему равны математическое ожидание и дисперсия этой случайной величины?

9.4 Из урны, содержащей пять белых и три черных шара, последовательно вынимают шары, причем операция извлечения продолжается до появления белого шара. Составить ряд распределения числа извлеченных черных шаров и вычислить математическое ожидание и дисперсию, если известно, что: а) вынутые шары в урну не возвращаются; б) вынутые шары возвращаются в урну.

Найти математическое ожидание и дисперсию соответствующих случайных величин в задачах 8.1 — 8.7.

9.5 Игрок бросает в автомат жетон стоимостью 10 рублей. В случае выигрыша игрок получает 50 рублей. Вероятность выигрыша составляет 0,16. Найти математическое ожидание и дисперсию выигрыша игрока в такой игре.

9.6 Диаметр круга измерен приближенно. Считая, что его величина равномерно распределена на отрезке $[a, b]$, найти среднее значение и дисперсию площади круга.

9.7 Случайные величины X Y независимы, причем X имеет нормальное распределение с параметрами 2 и $1/2$, а Y - равномерное распределение на отрезке $[0, 4]$. Найти:

а) $\mathbf{E}(X+Y)$; г) $\mathbf{E}(X-Y^2)$;

б) $\mathbf{E}XY$; д) $\mathbf{D}(X+Y)$;

в) $\mathbf{E}X^2$; е) $\mathbf{D}(X-Y)$.

9.8 Бросается n игральных костей. Найти математическое ожидание и дисперсию суммы очков на всех костях.

9.9 Двумерное распределение пары целочисленных случайных величин X и Y задается с помощью таблицы,

	$X=-2$	$X=0$	$X=1$
$Y=-1$	$1/6$	$1/6$	$1/6$
$Y=2$	$1/6$	$1/6$	$1/6$

где в пересечении столбца $X = i$ и строки $Y = j$ находится вероятность: $\mathbf{P}\{X = i,\ Y = j\}$. Найти:

а) $\mathbf{E}X,\ \mathbf{D}X$; в) $\mathbf{Cov}(X, Y)$;

б) $\mathbf{E}Y,\ \mathbf{D}Y$; г) $\mathbf{E}(X-2Y),\ \mathbf{D}(X-2Y)$.

9.10 Двумерный случайный вектор (X, Y) имеет распределение, заданное таблицей:

X / Y	-1	0	1
-2	$1/8$	$1/12$	$7/24$
0	$1/12$	$1/12$	$1/16$
1	$3/24$	$1/12$	$1/16$

Найти математические ожидания, дисперсии и коэффициент корреляции с. в. X, Y, а также математическое ожидание и дисперсию с.в. $X-3Y+1$.
9.11 Найти коэффициент корреляции $\rho(X,\ X+Y)$, где X и Y независимы, одинаково распределены и имеют конечный второй момент, применить полученную формулу для случая, когда X и Y имеют стандартное нормальное распределение.
9.12 Пусть с.в. X имеет равномерное на отрезке $[-1,1]$ распределение. Найти коэффициент корреляции $\rho(X,\ X^2)$.
9.13 Случайная величина Z имеет равномерное распределение на отрезке $[0,\ 1]$. Найти коэффициент корреляции случайных величин $Y_1,\ Y_2$, если:

а) $Y_1 = aZ,\ \ Y_2 = bZ\ (a, b > 0)$;

б) $Y_1 = aZ,\ \ Y_2 = bZ\ (a < 0 < b)$;

в) $Y_1 = Z,\ \ Y_2 = Z^2$.

9.14 Найти коэффициент корреляции между числом единиц и числом шестерок при трех бросаниях правильной игральной кости.

Глава 10

Предельные теоремы

Еще во введении, говоря о закономерностях случайных явлений, мы уточнили, что эти закономерности проявляются в результате проведения большого числа случайных экспериментов. Пример такой закономерности — устойчивость частоты события — можно наблюдать, проведя достаточно большое число реальных экспериментов, или воспользовавшись статистическими данными наблюдений того или иного случайного явления (демографические данные, метеорологические наблюдения и т.д.). Наиболее яркие результаты теории вероятностей, присущие именно этой науке, связаны с открытием фактов, наблюдаемых *только* при проведении большого числа случайных экспериментов. Такого рода результаты называют *предельными теоремами* теории вероятностей. Наиболее важными и известными предельными теоремами являются *закон больших чисел* (ЗБЧ) и *центральная предельная теорема* (ЦПТ).

§ 10.1. Закон больших чисел

В простейшем случае закон больших чисел заключается в следующем: *среднее арифметическое большого числа независимых, одинаково распределенных случайных величин ведет себя как величина неслучайная, равная математическому ожиданию.* Это означает, что среднее арифметическое $\dfrac{X_1 + X_2 + \ldots + X_n}{n}$ при $n \to \infty$ ведет себя весьма устойчиво, в то время как отдельные слагаемые $X_1,\ X_2,\ \ldots, X_n$ могут испытывать значительные случайные отклонения. Иначе говоря, при $n \to \infty$ имеет место сходимость:

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \quad \to \quad a = \mathbf{E}X_1,$$

где смысл сходимости и дополнительные условия уточняются ниже в точных формулировках. Здесь же отметим, что ЗБЧ проявляется во многих реальных «случайных экспериментах»: среднее количество осадков, выпадающих в данной местности за год, вычисляемое по результатам многолетних наблюдений, оказывается величиной весьма стабильной; результат измерения физических величин вычисляется обыкновенно как среднее арифметическое достаточно большого числа реальных измерений, чтобы уменьшить влияние случайных ошибок, возникающих при отдельных измерениях, и др.

Пусть заданы последовательность случайных величин $Y_1,\ Y_2,\ \ldots$ и случайная величина Y.

Определение 10.1. Будем говорить, что последовательность случайных величин $Y_1,\ Y_2,\ \ldots$ *сходится с вероятностью 1* (или *сходится почти наверное*) к случайной величине Y, если $Y_n(\omega) \to Y(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$ за исключением, быть может, ω из множества нулевой вероятности.

Будем использовать обозначение для сходимости почти наверное:

$$Y_n \to Y \quad \text{п. н.}$$

Из определения следует, что $Y_n \to Y$ п. н. тогда и только тогда, когда $Y_n - Y \to 0$ п. н., что в свою очередь равносильно сходимости $|Y_n - Y| \to 0$ п. н.

Отметим, что если плотность распределения случайной величины X_1 положительна почти всюду на (конечном или бесконечном) интервале $(a;\ b)$, и $(X_1,\ \ldots,\ X_n)$ — выборка, то $\min\{X_i\} \to a$ п. н., и $\max\{X_i\} \to b$ п. н.

Важным свойством сходимости почти наверное является свойство сходимости функций от случайных величин.

Теорема 10.1. *Пусть $Y_n \to Y$ п. н., $g(x)$ — непрерывная функция. Тогда $g(Y_n) \to g(Y)$ п. н.*

Доказательство. По определению непрерывной функции, событие $Y_n \to Y$ влечет событие $g(Y_n) \to g(Y)$. Следовательно,

$$\mathbf{P}(g(Y_n) \to g(Y)) \geq \mathbf{P}(Y_n \to Y) = 1.$$

То же свойство имеет место и для функции произвольного числа переменных. Докажем его для случая двух переменных.

Теорема 10.2. *Пусть $Y_n \to Y$ п. н., $Z_n \to Z$ п. н., $g(x, y)$ — непрерывная функция двух переменных. Тогда $g(Y_n, Z_n) \to g(Y, Z)$ п. н.*

Доказательство. По определению непрерывной функции, пересечение событий $Y_n \to Y$, $Z_n \to Z$ влечет событие $g(Y_n, Z_n) \to g(Y, Z)$. Следовательно, $\mathbf{P}(g(Y_n, Z_n) \to g(Y, Z)) \geq \mathbf{P}(Y_n \to Y, Z_n \to Z)$. Так как $\mathbf{P}(Y_n \to Y) = 1$, $\mathbf{P}(Z_n \to Z) = 1$, то объединение этих множеств также имеет вероятность 1. Отсюда по формуле вероятности объединения получаем, что вероятность пересечения также равна 1. Следовательно, $\mathbf{P}(g(Y_n, Z_n) \to g(Y, Z)) = 1$.

Следующая важная теорема носит название *закона больших чисел Колмогорова.*

Теорема 10.3. *Пусть X_1, X_2, ... — независимые одинаково распределенные случайные величины с конечным математическим ожиданием $\mathbf{E}X_1$. Обозначим $S_n = X_1 + \ldots + X_n$. Тогда $S_n/n \to \mathbf{E}X_1$ п. н.*

Доказательство этой теоремы можно найти, например, в [1].

§ 10.2. Центральная предельная теорема

Рассмотрим несколько результатов, которые объединяются под названием центральной предельной теоремы и занимают особое место в теории вероятностей. Упрощенно говоря, центральная предельная теорема (ЦПТ) формулируется так: *сумма большого числа малых независимых случайных величин приближенно имеет нормальное распределение.* Благодаря ЦПТ нормальное распределение имеет, пожалуй, наибольшее распространение в различных прикладных областях. Так, например, общепринято считать результат любого физического измерения нормально распределенной случайной величиной, поскольку в результат измерения неизбежно входит некоторая ошибка, которая состоит из большого числа мелких независимых ошибок, вызываемых различными факторами: неточность инструмента, меняющиеся параметры среды, условия измерения и т.д.

Сходимость центрированных и нормированных сумм случайных величин в центральной предельной теореме имеет место в некотором специальном смысле, более слабом, чем сходимость почти наверное.

Дадим определение сходимости по распределению. Последовательность случайных величин Y_1, Y_2, ... называется *сходящейся по распределению* к случайной величине Y с непрерывной функцией распределения

$F_Y(t)$, если последовательность функций распределения сходится к предельной функции распределения в каждой точке: для любого действительного t выполнено

$$F_{Y_n}(t) \to F_Y(t)$$

при $n \to \infty$.

Рассмотрим последовательность независимых одинаково распределенных случайных величин $\{X_n\}_{n=1}^{\infty}$ и обозначим через $S_n = \sum\limits_{k=1}^{n} X_k$ последовательность частичных сумм, составленных из первых n случайных величин X_k. Относительно распределений случайных величин X_k будем предполагать, что они имеют два конечных момента: $\mathbf{E}X_k = a$, $\mathbf{D}X_k = \sigma^2 > 0$, $k = 1, 2, \ldots$. Тогда и суммы S_n имеют конечные математическое ожидание и дисперсию:

$$\mathbf{E}S_n = na, \quad \mathbf{D}S_n = n\sigma^2.$$

Перейдем к последовательности центрированных и нормированных сумм:

$$\widetilde{S_n} = \frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} = \frac{\sum\limits_{k=1}^{n} X_k - na}{\sigma\sqrt{n}}.$$

Очевидно, $\mathbf{E}\widetilde{S_n} = 0$, $\mathbf{D}\widetilde{S_n} = 1$. Обозначим также через $\widetilde{F_n}(x) = \mathbf{P}(\widetilde{S_n} < x)$ функции распределения центрированных и нормированных сумм $\widetilde{S_n}$, а через $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du$ — функцию распределения стандартного нормального закона $N(0,1)$.

Теорема 10.4. (*ЦПТ для одинаково распределенных слагаемых.*) *Пусть $\{X_n\}_{n=1}^{\infty}$ – последовательность независимых и одинаково распределенных случайных величин, имеющих конечные математическое ожидание $\mathbf{E}X_n = a$ и дисперсию $\mathbf{D}X_n = \sigma^2 > 0$. Тогда для этой последовательности центральная предельная теорема выполняется в следующем виде: для всех $x \in \mathbf{R}$ выполнено*

$$\widetilde{F_n}(x) = \mathbf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} < x\right) \;\to\; \Phi(x), \quad n \to \infty, \tag{10.1}$$

или для всех $x_1, x_2 \in \mathbf{R}$ $(x_1 < x_2)$ *выполнено*

$$\lim_{n\to\infty} \mathbf{P}\left(x_1 \le \frac{S_n - na}{\sigma\sqrt{n}} < x_2\right) = \Phi(x_2) - \Phi(x_1), \quad n \to \infty, \tag{10.2}$$

то есть центрированные и нормированные суммы случайных величин сходятся по распределению к случайной величине, имеющей стандартное нормальное распределение.

Важным следствием теоремы 10.4 является следующая

Теорема 10.5. *(Муавра — Лапласа). Пусть* ν *— число «успехов» в* n *независимых испытаниях схемы Бернулли, и* p *— вероятность «успеха» в каждом испытании. Тогда справедливы следующие равносильные предельные соотношения:*

$$\mathbf{P}\left(\frac{\nu - np}{\sqrt{npq}} < x\right) \to \Phi(x), \quad n \to \infty, \tag{10.3}$$

$$\mathbf{P}\left(x_1 \le \frac{\nu - np}{\sqrt{npq}} < x_2\right) \to \Phi(x_2) - \Phi(x_1), \quad n \to \infty, \tag{10.4}$$

где

$$q = 1 - p\,; \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du.$$

Доказательство. Поскольку случайная величина ν есть число «успехов» в n независимых испытаниях Бернулли, то ее можно представить в виде суммы независимых случайных величин:

$$\nu = S_n = X_1 + X_2 + ... + X_n,$$

где случайная величина X_k равна числу «успехов» в одном k-м испытании Бернулли ($k = 1, 2, \ldots, n$), т. е. она принимает значения 1 или 0 в зависимости от того, был ли «успех» в k-м испытании. Ясно, что все X_k имеют одно и то же распределение Бернулли B_p. Если вспомнить, что их математическое ожидание и дисперсия равны $\mathbf{E}X_k = a = p$ и $\mathbf{D}X_k = \sigma^2 = pq$ соответственно, то все условия теоремы 10.4 выполнены, а доказываемые соотношения (10.3)–(10.4) получаются подстановкой новых обозначений в соотношения (10.1)–(10.2) из теоремы 10.4.

§ 10.3. Теорема Пуассона

Теорема Пуассона дает другое приближение в схеме Бернулли. Напомним, что *распределением Пуассона* с параметром $\lambda > 0$ называется дискретное распределение вида

$$\mathbf{P}\{X = k\} = \pi_k = e^{-\lambda}\frac{\lambda^k}{k!}; \quad k = 0, 1, \ldots \,.$$

Условия сближения биномиального распределения с распределением Пуассона, упрощенно говоря, состоят в том, что число n испытаний велико, а вероятность p «успеха» в каждом испытании мала. Более точная формулировка содержится в следующей теореме.

Теорема 10.6. *(Пуассон). Пусть n — число испытаний в схеме Бернулли — неограниченно возрастает, а вероятность «успеха» — одна и та же для всех испытаний — зависит от n : $p = p(n)$; причем $p = p(n) \to 0$ таким образом, что $np \to \lambda > 0$ при $n \to \infty$. Тогда для любого фиксированного целого $k \geq 0$ выполняется предельное соотношение:*

$$P_{n,p}(k) \to \pi_k, \quad n \to \infty, \tag{10.5}$$

где

$$P_{n,p}(k) = C_n^k p^k q^{n-k}; \quad \pi_k = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Доказательство. Будем использовать следующее стандартное для математического анализа обозначение эквивалентных переменных:

$$\alpha_n \sim \beta_n \quad \Leftrightarrow \quad \lim_{n\to\infty}\frac{\alpha_n}{\beta_n} = 1 \quad \Leftrightarrow \alpha_n = \beta_n + o(\beta_n).$$

Из условий доказываемой теоремы следует, что $p \sim \dfrac{\lambda}{n}$, $q = 1-p \to 1$ при $n \to \infty$. Докажем соотношение (10.5) сначала для $k = 0$:

$$\lim_{n\to\infty} P_{n,p}(0) = \lim_{n\to\infty} C_n^0 p^0 q^n = \lim_{n\to\infty} q^n = \lim_{n\to\infty}(1-p)^n = \lim_{n\to\infty}[(1-p)^{\frac{1}{p}}]^{pn} =$$

$$= \lim_{n\to\infty}[(1-p)^{\frac{1}{p}}]^{(\frac{\lambda}{n}+o(\frac{\lambda}{n}))\cdot n} = e^{-\lambda}.$$

Последнее равенство следует из того, что основание степени $(1 - p)^{1/p} \to e^{-1}$, а показатель $(\frac{\lambda}{n} + o(\frac{\lambda}{n}))\cdot n = \lambda + o(\frac{\lambda}{n})\cdot n \to \lambda$. Таким образом,

$$P_{n,p}(0) \to e^{-\lambda} = \pi_0, \quad n \to \infty. \tag{10.6}$$

Найдем теперь предел отношения

$$\frac{P_{n,p}(k)}{P_{n,p}(k-1)} = \frac{p\cdot(n-k+1)}{k\cdot q}.$$

При этом, поскольку $n \to \infty$, а $k \geq 1$ фиксировано, можно считать, что $k \leq n$. Будем также использовать тот факт, что предел не меняется, если любой сомножитель в «допредельном» выражении заменить на эквивалентный.

$$\lim_{n\to\infty}\frac{P_{n,p}(k)}{P_{n,p}(k-1)}=\lim_{n\to\infty}\frac{p\cdot(n-k+1)}{k\cdot q}=\lim_{n\to\infty}\frac{\frac{\lambda}{n}\cdot(n-k+1)}{k\cdot 1}=\frac{\lambda}{k}.$$

Таким образом,

$$\frac{P_{n,p}(k)}{P_{n,p}(k-1)}\to\frac{\lambda}{k},\quad n\to\infty. \tag{10.7}$$

Наконец, для любого целого $k\geq 1$ представим вероятность $P_{n,p}(k)$ в виде

$$P_{n,p}(k)=P_{n,p}(0)\cdot\frac{P_{n,p}(1)}{P_{n,p}(0)}\cdot\frac{P_{n,p}(2)}{P_{n,p}(1)}\cdots\frac{P_{n,p}(k)}{P_{n,p}(k-1)}.$$

Заменяя в правой части этого представления каждый сомножитель его пределом из (10.6) и (10.7), найдем предел произведения

$$\lim_{n\to\infty}P_{n,p}(k)=\lim_{n\to\infty}P_{n,p}(0)\cdot\lim_{n\to\infty}\frac{P_{n,p}(1)}{P_{n,p}(0)}\cdots\lim_{n\to\infty}\frac{P_{n,p}(k)}{P_{n,p}(k-1)}=$$

$$=e^{-\lambda}\cdot\frac{\lambda}{1}\cdot\frac{\lambda}{2}\cdot\,\cdots\,\cdot\frac{\lambda}{k}=e^{-\lambda}\cdot\frac{\lambda^k}{k!}=\pi_k.$$

Пример 10.1. *Известно, что левши составляют в среднем 1% всего населения. Оценить вероятность того, что по меньшей мере трое левшей окажется среди 200 человек.*

Решение. Естественно использовать схему Бернулли с параметрами $n=200$; $p=0,01$. Обозначим через ν число левшей, оказавшихся в данной группе (число «успехов» в n независимых испытаниях). Тогда точное значение искомой вероятности выражается с помощью формул Бернулли:

$$\mathbf{P}(\nu\geq 3)=\mathbf{P}\left(\bigcup_{k=3}^{200}(\nu=k)\right)=\sum_{k=3}^{200}\mathbf{P}(\nu=k)=\sum_{k=3}^{200}C_{200}^k p^k q^{200-k}.$$

При подстановке значений $p=0,01$; $q=0,99$ получим выражение, весьма трудоемкое для вычислений. Впрочем, перейдя к противоположному событию, мы значительно уменьшим число слагаемых в сумме. Кроме того, заменив вероятности биномиального распределения $C_n^k p^k q^{n-k}=P_{n,p}(k)$ вероятностями $e^{-\lambda}\cdot\frac{\lambda^k}{k!}=\pi_k$ распределения Пуассона, мы получим более простое приближенное значение искомой вероятности:

$$\mathbf{P}(\nu\geq 3)=1-\mathbf{P}(\nu<3)=1-\sum_{k=0}^{2}\mathbf{P}(\nu=k)\approx 1-\pi_0-\pi_1-\pi_2.$$

Вычислим значения вероятностей $\pi_k = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ при $\lambda = np = 2$. Получим

$$\mathbf{P}(\nu \geq 3) \approx 1 - 0,68 = 0,32.$$

Замечание 10.1. *В теореме 10.6 устанавливается лишь сам факт приближенного равенства $P_{n,p}(k) \approx \pi_k$ при больших значениях n. Однако для практического использования этого приближения необходимо знать оценку допускаемой погрешности. Если, применяя указанное приближенное равенство, мы допускаем ошибку, сравнимую с результатом вычисления, то использование приближенного равенства теряет смысл. Мы увидим позже, что в предыдущем примере ошибка приближения Пуассона не превышает 0,01.*

В формулируемой ниже теореме устанавливается не только возможность замены биномиального распределения распределением Пуассона, но и дается оценка погрешности, возникающей при такой замене. Доказательство этой теоремы выходит за рамки элементарного курса теории вероятностей, поэтому мы ограничимся только ее формулировкой.

Теорема 10.7. *Пусть ν — число «успехов» в n независимых испытаниях схемы Бернулли, а p — вероятность «успеха» в каждом испытании. Тогда для любого числового множества $A \subseteq R$ справедливо неравенство*

$$|\mathbf{P}(\nu \in A) - \sum_{k \in A} \pi_k| = |\sum_{k \in A} P_{n,p}(k) - \sum_{k \in A} \pi_k| \leq \min(p;\ np^2), \qquad (10.8)$$

где

$$\lambda = np\,; \quad P_{n,p}\,(k) = C_n^k p^k q^{n-k}\,; \quad \pi_k = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Возвращаясь к примеру 10.2, полученный при его решении ответ представим в виде

$$\mathbf{P}(\nu \geq 3) = 0,32 + \Delta,$$

где погрешность Δ оценивается с помощью теоремы 10.7:

$$|\Delta| \leq \min(p,\ np^2) = \min(0,01,\ 0,02) = 0,01.$$

Хотелось бы также иметь возможность оценивать ошибку, возникающую при использовании нормального приближения биномиального распределения. В некоторой степени эту возможность предоставляет следующий результат, который мы приводим из [13] без доказательства.

Теорема 10.8. *(Берри — Ессен). Пусть* $X_1, X_2, ..., X_n$ *— независимые случайные величины, имеющие одинаковое распределение. Пусть также* $\mathbf{E}X_1 = 0$, $\mathbf{E}X_1^2 = o^2 > 0$, $\mathbf{E}|X_1|^3 < \infty$, $\rho = \dfrac{\mathbf{E}|X_1|^3}{\sigma^3}$. *Тогда выполняется следующее* неравенство Берри — Ессена*:*

$$\sup_x \left| \mathbf{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} X_k < x \right) - \Phi(x) \right| \leq A\frac{\rho}{\sqrt{n}}, \tag{10.9}$$

где $\Phi(x)$ *— функция распределения стандартного нормального закона.*

Следствие 10.1. *Ошибка, возникающая при использовании теоремы Муавра — Лапласа, оценивается следующим неравенством:*

$$\Delta = \sup_x \left| \mathbf{P}\left(\frac{\nu - np}{\sqrt{npq}} < x \right) - \Phi(x) \right| \leq A\frac{p^2 + q^2}{\sqrt{npq}}. \tag{10.10}$$

Известно, что в качестве константы A *можно взять 0,4. Итак, при применении предельных теорем к схеме Бернулли будем действовать следующим образом: вычислим максимально возможные погрешности в теоремах Муавра—Лапласа и Пуассона и сравним их. Если*

$$0,4\frac{p^2 + q^2}{\sqrt{np(1-p)}} \geq \min(np^2,\ p),$$

то будем использовать теорему Пуассона. В противном случае будем использовать теорему Муавра—Лапласа.

Если в условиях примера 10.2 положить $p = 0,5$, то для погрешности Δ можно гарантировать лишь следующую границу:

$$|\Delta| \leq \min(p,\ np^2) = \min(0,5,\ 50) = 0,5.$$

Таким образом, использование приближения Пуассона при $p = 0,5$ недопустимо.

Пример 10.2. *Проектируется телефонная станция на 300 номеров. Предполагается, что каждый из пользователей, независимо от других, пользуется телефонной связью в среднем одну минуту в час. Каким должно быть минимальное число каналов связи, чтобы с вероятностью, не меньшей 90 %, любой вызов не получил бы отказа.*

Решение. Будем рассматривать действия каждого из $n = 300$ абонентов в течение некоторого короткого промежутка времени как независимые испытания, а «успехом» будем считать то, что абонент воспользовался телефонной связью. Вероятность «успеха» по условию равна $p = 1/60$. Обозначим через ν число абонентов, воспользовавшихся телефонной связью за рассматриваемый промежуток времени, т. е. ν — число «успехов» в n независимых испытаниях схемы Бернулли. Тогда вопрос задачи сводится к нахождению минимального натурального числа n_0, удовлетворяющего неравенству:

$$\mathbf{P}(\nu \leq n_0) \geq 0,9. \tag{10.11}$$

Чтобы найти такое n_0, выразим левую часть этого неравенства через n_0 и исходные данные задачи, используя приближение Пуассона:

$$\mathbf{P}(\nu \leq n_0) = \sum_{k=0}^{n_0} P_{n,p}(k) = \sum_{k=0}^{n_0} \pi_k + \Delta. \tag{10.12}$$

В нашем случае $\lambda = np = 300 \cdot \dfrac{1}{60} = 5$, а погрешность используемого приближения Пуассона оценивается неравенством: $|\Delta| \leq \min(2p,\ np^2) = \min(1/30,\ 1/12) = 0,033$. Для выполнения неравенства (10.11) необходимо потребовать, чтобы сумма в правой части (10.12) удовлетворяла неравенству

$$\sum_{k=0}^{n_0} \pi_k \geq 0,933,$$

страхуясь от возможной ошибки Δ. Вычисляя вероятности по формуле Пуассона, при $\lambda = 5$ видим, что сумма

$$\sum_{k=0}^{n_0} \pi_k \ = \ \sum_{k=0}^{n_0} e^{-\lambda} \frac{\lambda^k}{k!}$$

впервые превышает значение 0,933 при $n_0 = 9$. Стало быть, искомое минимальное число каналов связи равно $n_0 = 9$.

§ 10.4. Решение типовых примеров

Пример 10.3. *Театр, вмещающий 1000 человек, имеет два разных входа. Около каждого входа имеется свой гардероб. Сколько мест должно быть в каждом гардеробе для того, чтобы в среднем в 99 случаях из 100 все зрители могли раздеться в гардеробе того входа, через который они* *вошли. Предполагается, что зрители приходят: а) парами независимо одна пара от другой; б) поодиночке независимо друг от друга.*

Доказательство. Приведем решение в условиях б), оставив пункт а) читателю в качестве упражнения. Предположим, что оба входа равноправны в том смысле, что всякий зритель выбирает любой из них с вероятностью 0,5. Выберем какой-нибудь один вход, для которого подсчитаем требуемое число мест в гардеробе. Будем считать «успехом» то, что очередной пришедший зритель выбрал данный вход. Тогда, если ν есть число «успехов» в $n = 1000$ независимых испытаниях, то ν означает число зрителей, пришедших в театр через данный вход. Вопрос задачи сводится к тому, чтобы найти наименьшее натуральное число K, удовлетворяющее неравенству:

$$\mathbf{P}\,(\nu \le K) \ge 0,99 \iff \mathbf{P}\,(\nu > K) \le 0,01.$$

Поскольку с.в. ν принимает лишь целочисленные значения, то $\mathbf{P}\,(\nu \le K) = \mathbf{P}\,(\nu < K+1)$. Применим к последней вероятности теорему Муавра — Лапласа:

$$\mathbf{P}\,(\nu < K+1) = \mathbf{P}\left(\frac{\nu - np}{\sqrt{npq}} < \frac{K+1-np}{\sqrt{npq}}\right) \to \Phi(x),$$

где

$$x = \frac{K+1-np}{\sqrt{npq}} = \frac{K+1-1000\cdot 0,5}{\sqrt{1000\cdot 0,5\cdot 0,5}} = \frac{K-499}{5\sqrt{10}}.$$

По таблицам функции нормального распределения $\Phi(x)$ находим такое x, для которого $\Phi(x) = 0,99$, и по этому значению x найдем соответствующее K:

$$x = 2,33 \quad \Leftrightarrow \quad \frac{K-499}{5\sqrt{10}} = 2,33 \quad \Leftrightarrow \quad K = 499 + 5\sqrt{10}\cdot 2,33 \approx 535,84.$$

Поскольку K целое число, то следует взять $K = 536$.

Замечание 10.2. *В приведенном решении мы воспользовались тем, что с.в.* ν*, согласно теореме Муавра — Лапласа, имеет приближенно нормальное распределение. Однако биномиальное распределение* $B(n,p)$*, которое имеет с.в.* ν*, может быть приближено распределением Пуассона* $\Pi(\lambda)$*, где* $\lambda = np$*, причем возможная ошибка при этом приближении не превосходит величины* $\Delta = \min(np^2,\ p)$*, которая в условиях нашего примера равна*

$$\Delta = \min(1000\cdot 0,5^2,\ 0,5) = \min(250;\ 0,5) = 0,5.$$

Ясно, что вычислять вероятность с такой ошибкой не имеет смысла, поэтому приближение Пуассона в данном случае неприменимо.

Возвращаясь к примеру 10.3, вычислим оценку возможной ошибки при использовании нормального приближения (теоремы Муавра — Лапласа):

$$\Delta \leq 0,4\frac{p^2+q^2}{\sqrt{npq}} = 0,4\frac{0,5^2+0,5^2}{\sqrt{1000\cdot 0,5^2}} = \frac{0,4}{10\sqrt{10}} \approx 0,013.$$

Как видим, возможная ошибка приемлема для использования данного приближения. Однако, если уж быть совсем скрупулезными, следует заметить, что полученный при решении примера ответ $K = 536$ обеспечивает оговоренное в примере условие (чтобы все зрители могли раздеться в гардеробе того входа, через который они вошли) не с вероятностью 0,99, а лишь с вероятностью $0,99 - \Delta = 0,977$.

Пример 10.4. *Урожайность куста картофеля равна 0 кг с вероятностью 0,1; 1 кг с вероятностью 0,2; 1,5 кг с вероятностью 0,2; 2 кг с вероятностью 0,3; 2,5 кг с вероятностью 0,2. На участке посажено 900 кустов.*
а) В каких пределах с вероятностью 0,95 будет находиться урожай?
б) Какое наименьшее количество кустов надо посадить, чтобы с вероятностью, не меньшей 0,975 , урожай был не менее тонны?

Решение. Обозначим X_k — урожайность k-го куста. Тогда $S_n = X_1 + X_2 + ... + X_n$ — урожай, полученный с n кустов. Используя данное в условии задачи распределение с.в. X_k , найдем ее моменты:

$$\mathbf{E}X_k = 0\cdot 0,1 + 1\cdot 0,2 + 1,5\cdot 0,2 + 2\cdot 0,3 + 2,5\cdot 0,2 = 1,6;$$

$$\mathbf{D}X_k = \mathbf{E}X_k^2 - (\mathbf{E}X_k)^2 = 1\cdot 0,2 + 1,5^2\cdot 0,2 + 2^2\cdot 0,3 + 2,5^2\cdot 0,2 - 1,6^2 = 0,54.$$

а) Нужно найти $x_1 < x_2$, такие,что $\mathbf{P}(x_1 \leq S_n \leq x_2) = 0,95$ при $n = 900$. Такая задача, однако, не решается однозначно (два неизвестных при одном уравнении). Поэтому, в качестве границ, между которыми окажется значение S_n, обычно выбирают границы промежутка, симметричного относительно математического ожидания $\mathbf{E}S_n$. Таким образом, нужно найти такое $l > 0$, для которого

$$\mathbf{P}(\mathbf{E}S_n - l \leq S_n \leq \mathbf{E}S_n + l) = 0,95.$$

Преобразуем левую часть этого уравнения к виду, удобному для применения центральной предельной теоремы.

$$\mathbf{P}(\mathbf{E}S_n - l \leq S_n \leq \mathbf{E}S_n + l) = \mathbf{P}\left(-l \leq S_n - na \leq l\right) =$$

$$= \mathbf{P}\left(-\frac{l}{\sigma\sqrt{n}} \le \frac{S_n - na}{\sigma\sqrt{n}} \le \frac{l}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{l}{\sigma\sqrt{n}}\right) - \Phi\left(-\frac{l}{\sigma\sqrt{n}}\right), \qquad (10.13)$$

где $a = \mathbf{E}\xi_k = 1,6$; $\sigma = \sqrt{\mathbf{D}\xi_k} = \sqrt{0,54}$. В силу (10.13), будем искать такое $x = \frac{l}{\sigma\sqrt{n}}$, для которого (см. (10.13))

$$\Phi(x) - \Phi(-x) = 2\Phi(x) - 1 = 0,95 \iff \Phi(x) = 0,975,$$

а затем найдем l. Для этого воспользуемся таблицей значений функции распределения стандартного нормального закона $\Phi(x)$ (см. приложение в конце книги), откуда найдем $x = 1,96$, при котором $\Phi(x) = 0,975$. Тогда

$$l = x\sigma\sqrt{n} = 1,96 \cdot \sqrt{0,54 \cdot 900} \approx 43,2.$$

Учитывая значение $\mathbf{E}S_n = na = 900 \cdot 1,6 = 1440$, из (10.13) окончательно получаем:

$$\mathbf{P}(1440 - 43 \le S_n \le 1440 + 43) = \mathbf{P}(1397 \le S_n \le 1483) \approx 0,95.$$

б) В данном случае неизвестно n, и его нужно найти из условия:

$$\mathbf{P}(S_n \ge 1000) \ge 0,975.$$

Эквивалентными преобразованиями приведем левую часть этого неравенства к виду, удобному для применения центральной предельной теоремы:

$$\mathbf{P}(S_n \ge 1000) = \mathbf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} \ge \frac{1000 - na}{\sigma\sqrt{n}}\right) =$$

$$= 1 - \mathbf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} < \frac{1000 - na}{\sigma\sqrt{n}}\right) \approx 1 - \Phi(x), \quad x = \frac{1000 - na}{\sigma\sqrt{n}}.$$

Будем искать x, при котором $1 - \Phi(x) \ge 0,975$ или (см. (**??**)) $\Phi(-x) \ge 0,975$. Обратившись к таблице значений $\Phi(x)$, находим, что

$$\Phi(-x) \ge 0,975 \iff -x \ge 1,96 \iff -\frac{1000 - na}{\sigma\sqrt{n}} \ge 1,96 \iff$$

$$\iff na - 1000 \ge 1,96\sigma\sqrt{n} \iff na - 1,96\sigma\sqrt{n} - 1000 \ge 0.$$

Учитывая, что $a > 0$, решаем последнее квадратное неравенство относительно $\sqrt{n} > 0$:

$$\sqrt{n} \ge \frac{1,96\,\sigma + \sqrt{1,96^2\sigma^2 + 4000a}}{2a} \iff n \ge \left(\frac{1,96\,\sigma + \sqrt{1,96^2\sigma^2 + 4000a}}{2a}\right)^2.$$

Подставляя в последнее неравенство значения $a = 1,6$ и $\sigma = \sqrt{0,54}$, находим окончательно $n \ge 648$. $\triangledown$

§ 10.5. Задачи для самостоятельного решения

10.1 Игрок в каждой игре (независимо от результатов других игр) выигрывает 80 рублей с вероятностью 0,1, проигрывает 20 рублей с вероятностью 0,9. Найти, к какой величине сходится средний выигрыш за n игр при $n \to \infty$.

10.2 Пусть $X_1,\ X_2,\ \ldots$ — случайные числа, то есть независимые случайные величины, имеющие равномерное распределение на отрезке от 0 до 1. Найти пределы п. н. следующих выражений при $n \to \infty$:

$$\text{а)}\ \frac{{X_1}^2 + \ldots + {X_n}^2}{n}; \qquad \text{в)}\ \frac{1}{n}\left(\frac{1}{1+X_1} + \ldots + \frac{1}{1+X_n}\right);$$

$$\text{б)}\ \frac{\sqrt{{X_1}^2 + \ldots + {X_n}^2}}{n}; \qquad \text{г)}\ \operatorname{arctg}\left(\frac{2}{n}\left(X_1 + \ldots + X_n\right)\right).$$

10.3 Какова вероятность того, что в 100 партиях одинаковых по силе противников один из них выиграет более 70 раз? Ничьих нет.

10.4 Каждая буква текста может оказаться опечаткой с вероятностью 10^{-4}. Какова вероятность того, что в тексте из 40 000 букв окажется более 2 опечаток?

10.5 Для лица, дожившего до 20-летнего возраста, вероятность смерти на 21-м году жизни равна 0,006. Застрахована группа в 10000 человек 20-летнего возраста, причем каждый застрахованный внес 120 рублей страховых взносов за год. В случае смерти застрахованного страховое учреждение выплачивает наследникам 10000 рублей. Каковы вероятности, что:
а) к концу года страховое учреждение окажется в убытке;
б) его доход превысит 600000 рублей; 400000 рублей?

Какой минимальный страховой взнос следует учредить, чтобы в тех же условиях с вероятностью 0.95 доход был не менее 4000000 рублей?

10.6 Студент получает на экзамене 5 с вероятностью 0,2; 4 с вероятностью 0,4; 3 с вероятностью 0,3 и 2 с вероятностью 0,1. За время обучения он сдает 40 экзаменов. Найти пределы, в которых с вероятностью 0,95 лежит средний балл студента.

10.7 Известно, что вероятность рождения мальчика приблизительно равна 0.515. Какова вероятность того, что среди 10 тыс. новорожденных окажется мальчиков не больше, чем девочек?

10.8 Известно, что вероятность выпуска сверла повышенной хрупкости (брак) равна 0.02. Сверла укладываются в коробки по 100 шт. Чему равна вероятность того, что в коробке не окажется бракованных сверл? Какое наименьшее количество сверл нужно класть в коробку для того, чтобы с вероятностью, не меньшей 0.9, в ней было не менее 100 исправных?

10.9 Вероятность выхода из строя за время T одного конденсатора равна 0.05. Определить вероятность того, что за время T из 100 конденсаторов выйдут из строя а) не менее 5 конденсаторов; б) менее 13 конденсаторов.

10.10 Вероятность угадывания 6 номеров в спортлото (6 из 49) равна $7,2 \cdot 10^{-8}$. При подсчете оказались заполненными 5 млн. карточек. Какова вероятность, что никто не угадал все 6 номеров? Какое наименьшее количество карточек нужно заполнить, чтобы с вероятностью не менее 0,9 хотя бы один угадал 6 номеров?

10.11 1000 раз бросается игральная кость. Найти пределы, в которых с вероятностью, большей 0.95, будет лежать сумма выпавших очков.

10.12 Некоторая машина состоит из 10 тыс. деталей. Каждая деталь независимо от других деталей может оказаться неисправной с вероятностью p_i, причем для $n_1 = 1000$ деталей $p_1 = 0.0003$, для $n_2 = 2000$ деталей $p_2 = 0.0002$, и для $n_3 = 7000$ деталей $p_3 = 0.0001$. Машина не работает, если в ней неисправны хотя бы две детали. Найти вероятность того, что машина не будет работать.

10.13 Игральная кость подбрасывается до тех пор, пока общая сумма очков не превысит 700. Оценить вероятность того, что для этого потребуется более 210 бросаний.

10.14 Математическое ожидание количества выпадающих осадков в течение года в данной местности составляет 60 см. В предположении, что количество осадков распределено по нормальному закону, оценить вероятность того, что количество осадков в предстоящем году будет не больше 120 см, если известно стандартное отклонение — 20см.

Оценить вероятность того, что выпадет не менее 180 см осадков в условиях, описанных выше.

Глава 11

Выборка. Оценивание параметров

§ 11.1. Выборка и вариационный ряд

В математической статистике рассматривается ситуация, когда распределение наблюдаемой в случайном эксперименте величины X неизвестно (хотя бы частично), зато исследователь располагает результатами эксперимента (статистическими данными), по которым он должен сделать выводы о неизвестном распределении случайной величины X. К этому стоит добавить, что задачей математической статистики является использовать результаты эксперимента по возможности оптимальным образом.

Основным объектом исследования в математической статистике является **выборка** $\vec{X} = (X_1, X_2, ..., X_n)$, то есть набор значений случайной величины X, полученных в результате n независимых воспроизведений эксперимента. Иначе говоря, выборка представляет собой случайный вектор, координаты которого — **элементы выборки** $X_1, X_2, ..., X_n$ — независимые случайные величины, имеющие общее распределение с функцией распределения $F(t)$. Будем говорить в этом случае, что имеется **случайная выборка** $\vec{X}$ из распределения F, и обозначать сокращенно: $\vec{X} \Subset F$. Число n называется **объемом выборки**. Конкретный набор числовых значений случайных величин $X_1, X_2, ..., X_n$, полученный в результате эксперимента, будем называть **реализацией** выборки и обозначать $\vec{x} = (x_1, x_2, ..., x_n)$.

Если элементы выборки $X_1, \ldots, X_n$ упорядочить по возрастанию, то получится новый набор случайных величин, называемый **вариационным рядом**:

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n-1)} \leq X_{(n)}.$$

Случайная величина $X_{(k)}$, $k = 1, \ldots, n$ называется k**-м членом вариационного ряда**, или k**-й порядковой статистикой**. В частности, $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

§ 11.2. Эмпирическая функция распределения, гистограмма

Эмпирической функцией распределения $F_n^*(t)$ называется частота элементов выброрки, меньших заданного t. Эмпирическая функция распределения, соответствующая выборке $\vec{X} = (X_1, X_2, ..., X_n)$, может быть построена по этой выборке с помощью любой из следующих формул:

$$F_n^*(t) = \frac{\{\text{количество } X_i : X_i < t\}}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}(X_i < t), \qquad (11.1)$$

где функция

$$\mathbf{I}(X_i < t) = \begin{cases} 1, & \text{если } X_i < t, \\ 0 & \text{иначе} \end{cases}$$

— индикатор события $\{X_i < t\}$.

Заметим, что эмпирическая функция распределения, соответствующая случайной выборке $\vec{X}$, сама является случайной, поскольку определяется через элементы выборки X_1, $X_2, \ldots, X_n$, являющиеся случайными величинами. В то же время любая реализация $\vec{x} = (x_1, x_2, \ldots, x_n)$ выборки $\vec{X}$ порождает соответствующую реализацию эмпирической функции распределения (по той же формуле (11.1)), которая является обычной (а не случайной) функцией распределения.

С помощью вариационного ряда (или его реализации) эмпирическая функция распределения может быть построена графически.

Эмпирическая функция распределения $F_n^*(t)$ является выборочным аналогом неизвестной теоретической функции распределения $F(t)$, ее называют также **оценкой** для $F(t)$. Выборочным аналогом для теоретической плотности распределения $f(t)$ является **гистограмма**, или **эмпирическая плотность распределения**, которая строится по выборке $\vec{X} = (X_1, X_2, ..., X_n)$ следующим образом.

Пусть $h > 0$ — произвольное число. Разобьем область значений изучаемой случайной величины (например, всю числовую ось) на промежутки $\Delta_k = [z_{k-1}, z_k)$ длины h и построим ступенчатую функцию $f_n^*(t)$, которая

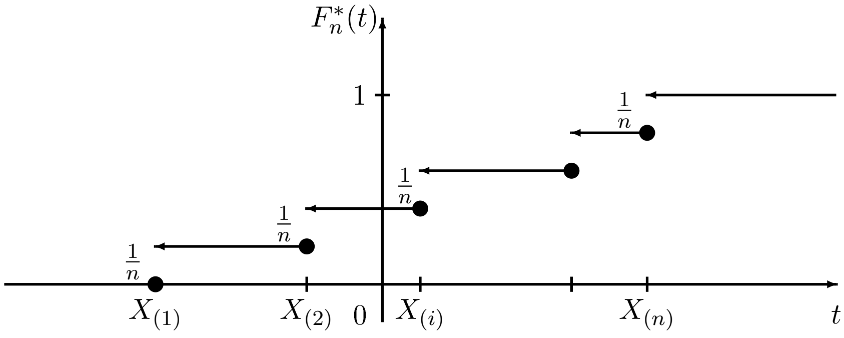


Рис. 11.1: Эмпирическая функция распределения $F_n^*(t)$.

на каждом промежутке Δ_k принимает постоянное значение, вычисляемое по любой из формул:

$$f_n^*(t) = \frac{\{\text{количество } X_i : X_i \in \Delta_k\}}{nh} = \frac{1}{nh}\sum_{i=1}^{n} \mathbf{I}(X_i \in \Delta_k) = \frac{\nu_k}{nh},\ t \in \Delta_k, \tag{11.2}$$

где ν_k - число элементов выборки, попавших в промежуток Δ_k. Так построенная функция называется гистограммой с шагом h и имеет график, изображенный на рис. 11.2. Заметим, что площадь каждого прямоугольни-

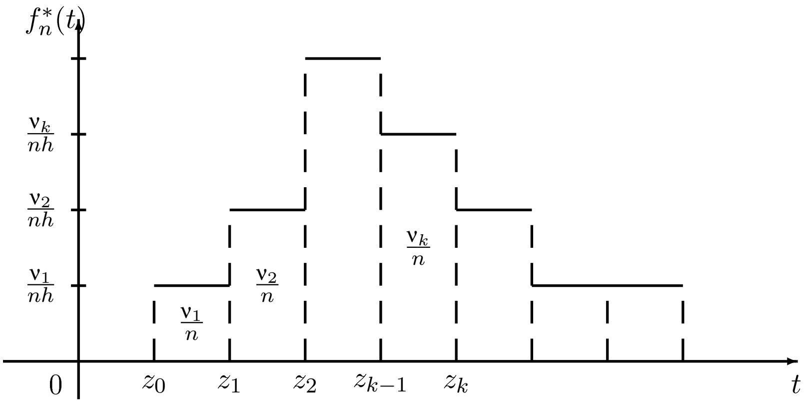


Рис. 11.2: Гистограмма $f_n^*(t)$.

ка гистограммы равна $\frac{\nu_k}{n}$, то есть частоте попадания в соответствующий интервал Δ_k.

Иногда шаг гистограммы h выбирают следующим образом. Сначала расчитывают число интервалов K по *формуле Стеджеса*

$$K = [\log_2 n] + 1. \tag{11.3}$$

Здесь n — объем выборки, $[\cdot]$ — целая часть числа. Потом длина интервала расчитывается по формуле

$$h = \frac{X_{(n)} - X_{(1)}}{K}.$$

При построении гистограммы последний промежуток выбирается замкнутым: $\Delta_K = [z_{K-1};\ z_K]$. Величину $X_{(n)} - X_{(1)} = \max\{X_i\} - \min\{X_i\}$ называют размахом выборки.

В некоторых случаях более точной оценкой для плотности, то есть оценкой, более точно аппроксимирующей неизвестную плотность распределения, является **полигон частот**. Это кусочно-линейная ломаная, которая строится из гистограммы путем последовательного соединения отрезками прямых середин верхних оснований прямоугольников, составляющих гистограмму (см. рис. 11.3).

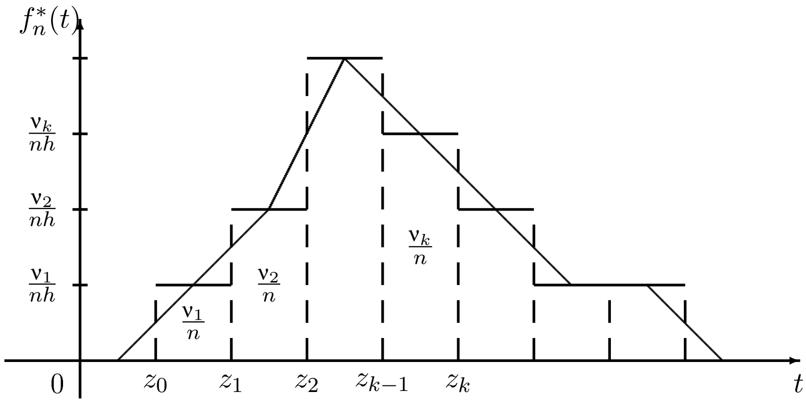


Рис. 11.3: Гистограмма и полигон частот.

§ 11.3. Выборочные моменты

По выборке $\vec{X} = (X_1, X_2, ..., X_n)$ можно построить эмпирические (выборочные) аналоги числовых характеристик распределения. Наиболее употребительными являются выборочное математическое ожидание, или **выборочное среднее**, $\overline{X}$, и **выборочная дисперсия** S^2:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{11.4}$$

Подобно выборочным среднему и дисперсии определяются выборочные моменты порядка k

$$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k,$$

которые являются эмпирическими аналогами моментов $\alpha_k = \mathbf{E}X_i^k$.

Пример 11.1. *Предполагая известными соответствующие теоретические моменты, доказать, что* $\mathbf{E}\overline{X^k} = \alpha_k$.

Приведенное соотношение означает, что математические ожидания эмпирических моментов совпадают с соответствующими теоретическими моментами. Это свойство называется **несмещенностью**: говорят, что эмпирические моменты являются **несмещенными оценками** для соответствующих теоретических.

Решение. Из свойств математического ожидания получаем:

$$\mathbf{E}\overline{X^k} = \mathbf{E}\frac{1}{n}\sum_{i=1}^{n} X_i^k = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}X_i^k = \frac{1}{n}\sum_{i=1}^{n}\alpha_k = \frac{1}{n}\cdot n\alpha_k = \alpha_k.$$

В то же время центральные эмпирические моменты являются смещенными оценками для своих теоретических аналогов.

Отметим, что выборочная дисперсия вычисляется аналогично дисперсии.

Следствие 11.1. $S^2 = \overline{X^2} - (\overline{X})^2$.

Доказательство.

Раскроем скобки в определении S^2:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 - 2\frac{\overline{X}}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\sum_{i=1}^{n}(\overline{X})^2 = \overline{X^2} - 2(\overline{X})^2 + (\overline{X})^2 = \overline{X^2} - (\overline{X})^2.$$

Вычислим математическое ожидание статистики S^2:

$$\mathbf{E}S^2 = \mathbf{E}\overline{X^2} - \mathbf{E}\overline{X}^2 = \mathbf{E}\overline{X^2} - (\mathbf{E}\overline{X})^2 - \mathbf{D}\overline{X} = \frac{n-1}{n}\mathbf{D}X_1.$$

Итак, эта оценка является асимптотически несмещенной.

Для того, чтобы получить несмещенную оценку дисперсии, делят S^2 на $\frac{n-1}{n}$.

Несмещенная выборочная дисперсия — это статистика

$$S_0{}^2 = \frac{n}{n-1} S^2.$$

Для нее выполнено свойство

$$\mathbf{E} S_0{}^2 = \mathbf{D} X_1.$$

Отметим, что корень из несмещенной выборочной дисперсии S_0 не является несмещенной оценкой для стандартного отклонения σ_X, так как $\mathbf{E}\sqrt{Y} \neq \sqrt{\mathbf{E}Y}$.

§ 11.4. Статистики и оценки

Задача оценивания параметров возникает в ситуации, когда распределение F не является полностью неизвестным, а известен его математический вид $F = F(t, \theta)$, содержащий неизвестный параметр θ (или несколько, тогда θ - многомерный параметр). Задача состоит в том, чтобы по выборке $\vec{X}$ вычислить приближенное значение $\theta^*(\vec{X})$ для неизвестного параметра, причем сделать это в том или ином смысле оптимальным образом. Это задача **точечного оценивания**. Другой подход состоит в построении по выборке $\mathbf{X}$ интервала $(\theta_-(\vec{X});\ \theta_+(\vec{X}))$, который накрывает неизвестное значение параметра θ с заданной (высокой) вероятностью. Этот подход называется **интервальным оцениванием**, а $(\theta_-(\vec{X});\ \theta_+(\vec{X}))$ называется **доверительным интервалом**.

Пусть $\vec{X} \Subset F(t, \theta)$, причем параметр θ может принимать значения из множества Θ, которое называется **параметрическим множеством**. Будем называть **статистикой** любую случайную величину вида $T(\vec{X})$, которая является функцией **только от элементов выборки**. **Оценкой параметра** θ называется статистика $\tilde{\theta} = \tilde{\theta}(\vec{X})$, которая принимает значения из параметрического множества Θ.

Оценка $\tilde{\theta}$ называется **несмещенной** оценкой параметра θ, если для любого $\theta \in \Theta$ выполнено

$$\mathbf{E}\tilde{\theta} = \theta. \tag{11.5}$$

Договоримся указывать в обозначении статистики объем выборки, если это необходимо подчеркнуть: $\tilde{\theta} = \tilde{\theta}_n$.

Оценка $\tilde{\theta}_n$ называется **(сильно) состоятельной оценкой параметра** θ, если для любого $\theta \in \Theta$ при $n \to \infty$ имеет место сходимость

$$\tilde{\theta}_n \xrightarrow{\text{п. н.}} \theta, \tag{11.6}$$

то есть $\mathbf{P}\{\tilde{\theta}_n \to \theta\} = 1$.

К следующему примеру мы будем часто возвращаться в дальнейшем.

Пример 11.2. *(Задача о расписании автобусов). Придя на остановку, пассажир пытается оценить длительность интервалов между автобусами выбранного им маршрута. Он анкетирует других пассажиров, ожидающих этот автобус, и у каждого из n пассажиров выясняет время, проведенное им на остановке, получая таким образом выборку $X_1, \ldots, X_n$. Предполагается, что $X_1, \ldots, X_n$ образуют выборку из равномерного распределения $U_{[0;\ \theta]}$, где $\theta > 0$ — неизвестный параметр — интервал времени между автобусами.*

Первый пассажир предлагает для оценки параметра θ использовать выборочное среднее, т. е. получить оценку в виде $\widetilde{\theta}_1 = c_1\overline{X}$.

Второй пассажир предлагает использовать самое большое время ожидания, т. е. получить оценку в виде $\widetilde{\theta}_2 = c_2 X_{(n)}$.

Третий пассажир предлагает сложить самое большое и самое маленькое время ожидания: $\widetilde{\theta}_3 = X_{(n)} + X_{(1)}$.

Вычислить константы c_1, c_2, обеспечивающие несмещенность оценок $\widetilde{\theta}_1$, $\widetilde{\theta}_2$. Проверить несмещенность оценки $\widetilde{\theta}_3$. Исследовать сильную состоятельность всех полученных оценок.

Решение. Вычислим математическое ожидание статистики $\widetilde{\theta}_1$:

$$\mathbf{E}\widetilde{\theta}_1 = c_1\mathbf{E}\overline{X} = c_1\mathbf{E}X_1 = c_1\theta/2.$$

Условие несмещенности выполнено, если

$$\mathbf{E}\widetilde{\theta}_1 = c_1\theta/2 = \theta.$$

Отсюда с необходимостью $c_1 = 2$. Итак, мы получили первую несмещенную оценку: $\widetilde{\theta}_1 = 2\overline{X}$. Ее сильная состоятельность следует из усиленного закона больших чисел: так как $\overline{X} \to \mathbf{E}X_1 = \theta/2$ п. н., то $\widetilde{\theta}_1 = 2\overline{X} \to \theta$ п. н. в силу непрерывности функции $y(t) = 2t$.

Исследование оценки $\widetilde{\theta}_2$ значительно более трудоемко. Найдем распределение статистики $X_{(n)}$. Ее функция распределения равна

$$F_{X_{(n)}}(y) = \mathbf{P}\{X_{(n)} < y\} = \mathbf{P}\{\max(X_1, \ldots, X_n) < y\} =$$

$$= \mathbf{P}\{\bigcap_{i=1}^{n}(X_i < y)\} = \prod_{i=1}^{n}\mathbf{P}(X_i < y) = F^n(y),$$

где

$$F(y) = \begin{cases} 0, & \text{если } y \le 0 \\ \frac{y}{\theta}, & \text{если } 0 < y \le \theta, \\ 1, & \text{если } y > \theta. \end{cases}$$

— функция распределения закона $U_{[0;\ \theta]}$. Дифференцируя $F_{X_{(n)}}(y)$, найдем плотность распределения случайной величины $X_{(n)}$:

$$f_{X_{(n)}}(y) = nF^{n-1}(y)F'(y) = nF^{n-1}(y)f(y).$$

Подставляя в последнее равенство функцию распределения закона $U_{[0;\ \theta]}$ и его плотность

$$f(y) = \begin{cases} \frac{1}{\theta}, & \text{если } y \in (0, \theta), \\ 0, & \text{если } y \notin [0, \theta], \end{cases}$$

находим плотность распределения $X_{(n)}$:

$$f_{X_{(n)}}(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & \text{если } y \in (0, \theta), \\ 0, & \text{если } y \notin [0, \theta]. \end{cases} \tag{11.7}$$

Для проверки несмещенности найдем математическое ожидание оценки:

$$\mathbf{E}_\theta \widetilde{\theta} = \mathbf{E}_\theta X_{(n)} = \int_0^\theta y \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{n+1} \frac{y^{n+1}}{\theta^n}\bigg|_0^\theta = \frac{n}{n+1}\theta.$$

Отсюда следует, что оценка $\widetilde{\theta}_2 = c_2\widetilde{\theta} = c_2 X_{(n)}$ является несмещенной для параметра θ при условии $c_2 n\theta/(n+1) = \theta$. Отсюда находим $c_2 = (n+1)/n$. Мы получили вторую несмещенную оценку: $\widetilde{\theta}_2 = (n+1)X_{(n)}/n$.

Проверим сильную состоятельность $\widetilde{\theta}_2$. Согласно отмеченному свойству сходимости почти наверное, максимум из независимых одинаково распределенных случайных величин, имеющих на интервале положительную плотность, сходится п. н. к правому концу интервала:

$$X_{(n)} \to \theta \text{ п. н.}$$

Так как $(n+1)/n \to 1$ как числовая последовательность, то в силу непрерывности функции $g(x, y) = xy$ имеет место сходимость

$$\widetilde{\theta}_2 = (n+1)/n \cdot X_{(n)} \to 1 \cdot \theta = \theta \text{ п. н.}$$

Для доказательства несмещенности третьей оценки заметим, что минимум выборки $X_{(1)}$ распределен симметрично максимуму $X_{(n)}$ относительно середины отрезка $\theta/2$, т. е. для всех t выполнено

$$\mathbf{P}\{X_{(1)} < t\} = \mathbf{P}\{\theta - X_{(n)} < t\}.$$

Отсюда

$$\mathbf{E}X_{(1)} = \theta - \mathbf{E}X_{(n)} = \theta/(n+1),$$

$\mathbf{E}_\theta \widetilde{\theta}_3 = \theta$ — оценка несмещенная.

В силу той же симметрии получаем, что $X_{(1)} \to 0$ п. н., и в силу непрерывности функции $g(x,y) = x + y$ имеет место сходимость $\widetilde{\theta}_3 \to \theta$ п. н.

§ 11.5. Оценки методом моментов

Наиболее распространенными методами нахождения оценок являются *метод моментов* и *метод максимального правдоподобия.*

Метод моментов (одномерный случай)

Пусть $\theta \in \Theta$ - одномерный параметр, и $g : R \longrightarrow R$ некоторая числовая функция. Тогда по данной выборке $\vec{X} = (X_1, X_2, ..., X_n)$ можно построить выборку $g(\vec{X}) = (g(X_1), g(X_2), ..., g(X_n))$. Обозначим

$$\overline{g(X)} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

выборочное среднее этой выборки. С другой стороны, можно найти теоретическое среднее выборки $g(\vec{X})$:

$$m_g(\theta) = \mathbf{E}g(X_i).$$

Определение 11.1. *Оценкой метода моментов* (ОММ) называется такое значение $\theta^*_g = \theta^*_g(X)$, при котором теоретическое среднее выборки $\mathbf{g}(\mathbf{X})$ совпадает с выборочным средним:

$$m_g(\theta^*_g) = \overline{g(X)}, \qquad (11.8)$$

то есть ОММ является решением уравнения (11.8) относительно неизвестного θ^*_g.

Если при этом оказывается, что функция $m_g(\theta)$ непрерывна и строго монотонна, то для нее существует обратная m_g^{-1}, и ОММ имеет вид:

$$\theta^*_g(X) = m_g^{-1}(\overline{g(X)}).$$

В качестве функции g чаще всего выбирают степенные функции: $g(x) = x^k$, где $k = 1, 2, \ldots$. В этом случае теоретическое среднее выборки $g(\vec{X})$ совпадает с теоретическим моментом соответствующего порядка, например, если $g(X) = x$, то $m_g(\theta) = \mathbf{E}X_i = \alpha_1(\theta)$; если $g(X) = x^2$, то $m_g(\theta) = \mathbf{E}X_i^2 = \alpha_2(\theta)$, и т.д. При этом уравнение (11.8) для нахождения ОММ приобретает вид:

$$\alpha_k(\theta^*) = \overline{X^k}. \tag{11.9}$$

Оценка по методу моментов в этом случае называется *оценкой по k-тому моменту* и обозначается θ_k^*

Отметим, что если функция $m_g(\theta) = \mathbf{E}g(X_1)$ непрерывна и строго монотонна, то оценка по методу моментов $\theta_g^*(X) = m_g^{-1}(\overline{g(X)})$ сильно состоятельна.

Метод моментов (многомерный случай)

Пусть $\vec{X} \Subset \mathbf{F}_\theta$, где параметр $\theta \in \Theta$, подлежащий оцениванию, — многомерный. Рассмотрим для простоты двумерный случай, то есть $\theta = (\theta_1, \theta_2)$. Тогда для однозначного нахождения двух неизвестных θ_1, θ_2 одного уравнения (11.8) (или (11.9)) недостаточно. Оценкой метода моментов в этом случае называется решение (θ_1^*, θ_2^*) системы уравнений вида:

$$\begin{cases} m_{g_1}(\theta_1, \theta_2) = \overline{g_1(X)}, \\ m_{g_2}(\theta_1, \theta_2) = \overline{g_2(X)}. \end{cases} \tag{11.10}$$

В качестве функций g_1, g_2 можно выбрать, как и раньше, степенные функции $g_i(x) = x^k$, где $k = 1, 2, \ldots$. Тогда уравнения системы (11.10) получаются как результат приравнивания эмпирических моментов выборки $\vec{X}$ соответствующим теоретическим. Например, приравнивая первые два момента, получим систему:

$$\begin{cases} \alpha_1(\theta_1, \theta_2) = \overline{X}, \\ \alpha_2(\theta_1, \theta_2) = \overline{X^2}. \end{cases} \tag{11.11}$$

Как и раньше, вместо вторых моментов можно приравнивать дисперсии.

§ 11.6. Решение типовых примеров

Пример 11.3. *По данной реализации выборки* $\vec{x} = (3, 8, 6, 4, 6, 1, 5, 4, 9, 4)$ *построить реализацию вариационного ряда,*

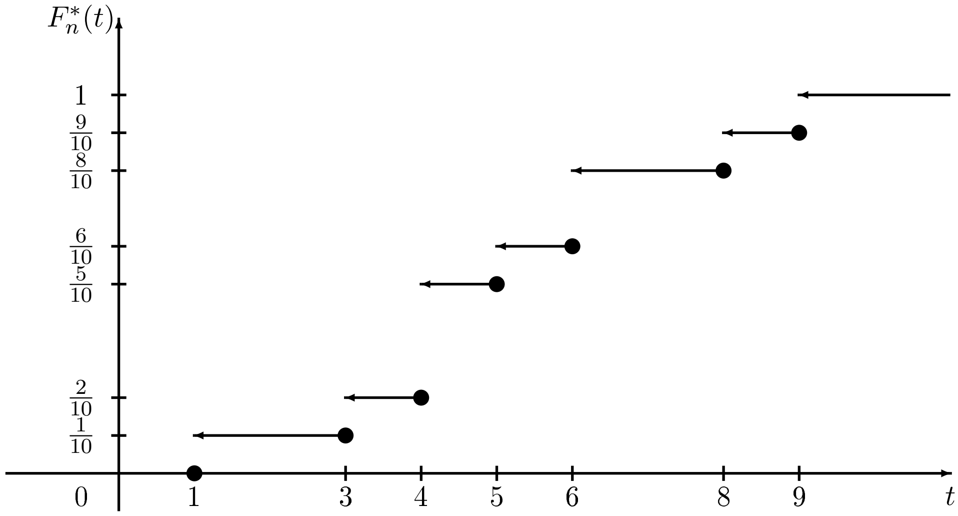


Рис. 11.4: Эмпирическая функция распределения $F_n^*(t)$.

графики реализаций эмпирической функции распределения и гистограммы. Число интервалов для построения гистограммы выбрать по формуле Стеджеса.

Решение. Реализацию вариационного ряда образуем из элементов данной реализации выборки, расположив их в порядке возрастания:

$$1,\ 3,\ 4,\ 4,\ 4,\ 5,\ 6,\ 6,\ 8,\ 9. \tag{11.12}$$

Объем выборки $n = 10$. График реализации эмпирической функции распределения строим с помощью полученной реализации вариационного ряда: это ступенчатая функция со скачками в точках вариационного ряда, принимающая значение 0 в промежутке $(-\infty, 1]$ и имеющая скачки в точках $x_{(i)}$, равные частоте элемента $x_{(i)}$. Например, скачок в точке $x_{(1)} = 1$ равен $\frac{1}{10}$, скачок в точке $x_{(2)} = 3$ равен $\frac{1}{10}$ и т.д. График реализации эмпирической функции распределения изображен на рис.11.4

Расчитаем число промежутков по формуле Стеджеса: $K = [\log_2 10] + 1 = 3 + 1 = 4$. Размах выборки равен $9 - 1 = 8$, шаг гистограммы $h = 8/4 = 2$. Разобьем отрезок $[1;\ 9]$ на промежутки длины $h = 2$:

$$\Delta_1 = [1;\ 3);\ \Delta_2 = [3;\ 5);\ \Delta_3 = [5;\ 7);\ \Delta_4 = [7;\ 9].$$

Число элементов выборки, попавших в интервал Δ_1, равно $\nu_1 = 1$. Аналогично находим:

$$\nu_2 = 4; \quad \nu_3 = 3; \quad \nu_4 = 2.$$

Вычисляя значения функции $f_n^*(t) = \frac{\nu_k}{nh}, t \in \Delta_k$ на каждом из интервалов Δ_k, строим гистограмму (рис.11.5):

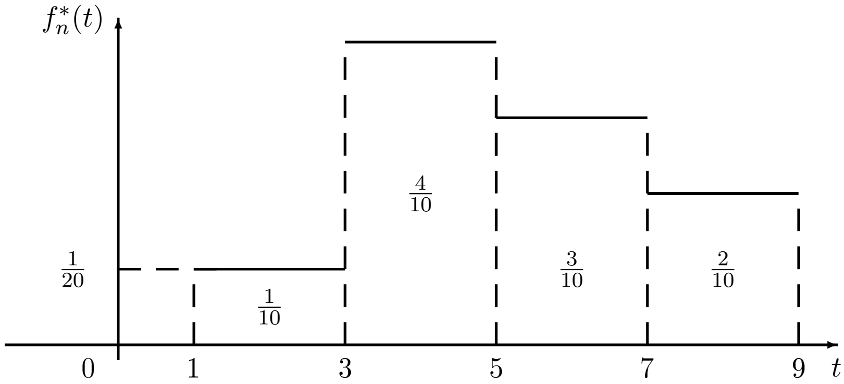


Рис. 11.5: Гистограмма $f_n^*(t)$.

Пример 11.4. *Пусть $\vec{X} \Subset \Pi_\lambda$, где $\lambda > 0$ — неизвестный параметр. Найти оценки параметра λ по а) первому и б) второму моментам.*

Решение. а) Так как для распределения Пуассона $\mathbf{E}X_i = \lambda$, то λ_1^* получается сразу из равенств (11.8) или (11.9): заменяя λ на λ_1^*, а $\mathbf{E}X_i$ на $\overline{X}$, получаем $\lambda_1^* = \overline{X}$.

б) В этом случае вычисляем второй момент распределения Пуассона:

$$\mathbf{E}X_i^2 = (\mathbf{E}X_i)^2 + \mathbf{D}X_i = \lambda + \lambda^2.$$

Приравнивая эту функцию второму выборочному моменту и заменяя λ на λ_2^*, получим уравнение:

$$\lambda_2^* + (\lambda_2^*)^2 = \overline{X^2},$$

из которого находим λ_2^* :

$$\lambda_2^* = -\frac{1}{2} \pm \sqrt{\frac{1}{4} + \overline{X^2}}.$$

Так как $\lambda_2^* > 0$, то из двух решений выбираем одно — положительное, и ОММ имеет вид:

$$\lambda_2^*(X) = -\frac{1}{2} + \sqrt{\frac{1}{4} + \overline{X^2}}.$$

Замечание 11.1. *Из двух найденных оценок* λ_1^* *представляется предпочтительней. Во-первых, она несмещенная, так как*

$$\mathbf{E}_\lambda \lambda_1^* = \mathbf{E}_\lambda \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}_\lambda X_i = \frac{1}{n} n\lambda = \lambda;$$

во-вторых, она состоятельная в силу усиленного закона больших чисел (УЗБЧ).

В то же время, оценка λ_2^* *менее удобна для исследования, хотя она является состоятельной (проверьте, используя УЗБЧ). Например, исследовать для нее свойство несмещенности — технически трудная задача.*

Пример 11.5. *Пусть* $\vec{X} \Subset U_{[\theta_1;\ \theta_2]}$, *где* $-\infty < \theta_1 < \theta_2 < \infty$ *— неизвестные параметры. Найти ОММ.*

Решение. Вычислим моменты первых двух порядков равномерного распределения

$$\alpha_1(\theta_1, \theta_2) = \mathbf{E}X_i = \frac{\theta_1 + \theta_2}{2},$$

$$\mathbf{D}X_i = \frac{(\theta_2 - \theta_1)^2}{12}.$$

Составим систему уравнений, приравнивая теоретические и эмпирические математическое ожидание и дисперсию:

$$\begin{cases} \mathbf{E}X_1 = \overline{X}, \\ \mathbf{D}X_1 = S^2, \end{cases} \iff \begin{cases} \frac{\theta_1+\theta_2}{2} = \overline{X}, \\ \frac{(\theta_2-\theta_1)^2}{12} = S^2, \end{cases} \iff \begin{cases} \theta_1 + \theta_2 = 2\overline{X}, \\ \theta_2 - \theta_1 = \sqrt{12}S. \end{cases}$$

Решая последнюю систему относительно неизвестных θ_1, θ_2 (вычитая и складывая уравнения системы), получим оценки ММ:

$$\theta_1^* = \overline{X} - \sqrt{3}S, \quad \theta_2^* = \overline{X} + \sqrt{3}S.$$

§ 11.7. Задачи для самостоятельного решения

11.1 По данной реализации выборки $\mathbf{x} = (0,0,1,1,0,0,0,0,0,1)$:

а) построить графики эмпирической функции распределения и гистограммы;

б) вычислить выборочные среднее и дисперсию.

11.2 Проводились опыты с бросанием одновременно 12 игральных костей. Наблюдаемую случайную величину X считали равной числу костей, на которых выпало не больше трех очков. Пусть ν_i - число опытов, в которых наблюдалось значение $X = i$; $i = 0, 1, ..., 12$. Данные для $n = 4096$ опытов приведены в следующей таблице:

i	0	1	2	3	4	5	6	7	8	9	10	11	12
ν_i	0	7	60	198	430	731	948	847	536	257	71	11	0

а) Построить гистограмму и сравнить ее с графиком функции $y = ce^{-\frac{x^2}{6}}$.

б) Вычислить выборочные среднее и дисперсию.

11.3 Измерен рост (в см) студентов одной учебной группы. Результаты измерений дали выборку (171; 186; 164; 190; 158; 181; 176; 180; 174; 157; 176; 169; 164; 186).

а) Построить реализацию гистограммы.

б) Вычислить реализации выборочного среднего, выборочной дисперсии и выборочного стандартного отклонения S. На одном графике с гистограммой построить график плотности нормального закона с параметрами $\overline{X}$, S^2.

11.4 Пассажир маршрутного такси измерил 8 раз время ожидания такси и получил следующие результаты (в минутах): 8; 4; 5; 4; 2; 15; 1; 6. У него есть две гипотезы относительно графика движения такси: либо график движения соблюдается, и время ожидания имеет равномерное распределение на отрезке $[0;\ \theta]$, либо график движения не соблюдается, и время ожидания имеет показательное распределение с параметром λ.

а) Вычислить реализации оценок параметров θ и λ, использовав оценки $\widetilde{\theta}_2 = (n+1)X_{(n)}/n$ и $\widetilde{\lambda}_2 = \frac{n-1}{n\overline{X}}$.

б) Построить на одном графике реализацию эмпирической функции распределения и теоретические функции распределения равномерного и показательного законов, в которые вместо неизвестных параметров подставлены реализации их оценок.

в) Построить на одном графике реализацию гистограммы и теоретические плотности распределения равномерного и показательного законов, в которые вместо неизвестных параметров подставлены реализации их оценок.

г) На основании проведенного исследования сделать вывод о том, какая из гипотез выглядит более соответствующей экспериментальным данным.

11.5 Пусть $\vec{X} \Subset N(a, \sigma^2)$. Вычислить $\mathbf{E}\overline{X}$, $\mathbf{D}\overline{X}$. Какое распределение имеет случайная величина $\overline{X}$?

11.6 Дана выборка $\mathbf{X} \Subset \Pi(\lambda)$, $\lambda > 0$ — неизвестный параметр. Проверить, что статистики

$$T_1 = \overline{X}, \quad T_2 = \frac{1}{n}\sum_{i=1}^{n} I(X_i = k), \quad T_3 = \frac{X_1 + X_n}{2}$$

являются несмещенными оценками соответственно для λ, $\frac{\lambda^k}{k!}e^{-\lambda}$ и λ. Являются ли эти оценки состоятельными?

11.7 По выборке $(X_1, \ldots, X_n)$ из бернуллиевского распределения B_p с неизвестным параметром $p \in (0;\ 1)$ построить оценки параметра p:

а) по первому моменту;

б) по второму моменту;

в) по произвольному k-му моменту.

Можно ли отдать предпочтение какой-либо из построенных оценок? Исследовать их состоятельность и несмещенность.

11.8 По выборке $(X_1, \ldots, X_n)$ из биномиального распределения $B_{m,p}$ построить оценки методом моментов:

а) параметра p по первому и по второму моменту при известном $m > 0$;

б) параметров p и m.

Исследовать состоятельность построенных оценок.

11.9 При каких значениях параметра $\theta > 0$ распределения Парето с плотностью

$$f_\theta(t) = \begin{cases} \frac{\theta}{t^{\theta+1}}, & t \geq 1, \\ 0, & t < 1 \end{cases}$$

существует оценка параметра по первому моменту? Можно ли построить состоятельную оценку методом моментов в случае, когда оценки по первому моменту не существует?

11.10 По выборке $(X_1, \ldots, X_n)$ из распределения Лапласа с плотностью $f_\lambda(t) = \frac{\lambda}{2}e^{-\lambda|t|}$, $t \in \mathbf{R}$, построить оценку параметра $\lambda > 0$ методом моментов.

11.11 Пусть дана выборка из нормального распределения с параметрами α и σ^2. Используя метод моментов, построить оценки

а) неизвестного математического ожидания α;

б) неизвестной дисперсии σ^2, если α известно;

в) неизвестной дисперсии σ^2, если α неизвестно.

Исследовать полученные оценки на несмещенность и состоятельность.

11.12 Используя метод моментов, оценить параметр θ равномерного распределения на отрезке

а) $[-\theta;\ \theta]$, $\theta > 0$; б) $[\theta;\ \theta+1]$.

Исследовать полученные оценки на несмещенность и состоятельность.

11.13 С помощью метода моментов найти оценки параметров a и b равномерного распределения на отрезке $[a, b]$. Будут ли они состоятельными?

11.14 Используя метод моментов, построить бесконечную последовательность различных оценок параметра θ равномерного распределения на отрезке $[0;\ \theta]$. Будут ли полученные оценки состоятельными?

11.15 С помощью метода моментов построить оценку параметра $\theta > 0$, если распределение выборки имеет плотность

а) $\theta t^{\theta-1}$ при $t \in [0;\ 1]$; б) $2t/\theta^2$ при $t \in [0;\ \theta]$.

Исследовать полученные оценки на состоятельность.

11.16 Дана выборка из распределения с плотностью

$$f_\theta(t) = \begin{cases} 3t^2\theta^{-3}, & t \in [0;\ 1], \\ 0, & t \notin [0;\ 1]. \end{cases}$$

Найти оценку параметра $\theta > 0$ методом моментов, исследовать ее на несмещенность и состоятельность.

11.17 Методом моментов найти оценку параметра $\alpha > 0$ по выборке из показательного распределения с плотностью $f_\alpha(t) = \alpha e^{-\alpha t}$, $t > 0$. Будет ли оценка несмещенной и состоятельной?

11.18 По выборке $(X_1, \ldots, X_n)$ методом моментов найти две различные оценки параметра $p \in (0, 1)$, если известно, что

$P\{X_1 = 1\} = p/2,\ P\{X_1 = 2\} = p/2,\ P\{X_1 = 3\} = 1 - p.$

Будут ли полученные оценки несмещенными и состоятельными?

Глава 12

Оценки максимального правдоподобия. Сравнение оценок

§ 12.1. Метод максимального правдоподобия

Пусть $\vec{X} \Subset F(t, \theta)$, $\theta \in \Theta$. Предположим, что теоретическое распределение либо абсолютно непрерывно с плотностью $f(t, \theta) = f_{X_i}(t)$, либо дискретно, при этом для ряда распределения будем использовать то же обозначение: $f(t, \theta) = \mathbf{P}(X_i = t)$. **Функцией правдоподобия, соответствующей выборке** $\vec{X}$, называется функция

$$\Pi(\theta) = \Pi(\vec{X},\ \theta) = \prod_{i=1}^{n} f(X_i,\ \theta). \tag{12.1}$$

Оценкой максимального правдоподобия (ОМП) называется такое значение параметра $\theta = \hat{\theta}(\vec{X})$, при котором функция правдоподобия принимает наибольшее значение, то есть

$$\Pi(\vec{X},\ \hat{\theta}) = \max_{\theta \in \Theta} \Pi(\vec{X},\ \theta). \tag{12.2}$$

Если функция правдоподобия дифференцируема при всех $\theta \in \Theta$, то значение $\theta = \hat{\theta}$ должно быть решением уравнения

$$\Pi'(\theta) = 0, \tag{12.3}$$

которое называется уравнением правдоподобия, или эквивалентного уравнения

$$\frac{d}{d\theta}\ln\Pi(\theta)=0 \iff \sum_{i=1}^{n}\frac{d}{d\theta}\ln f(X_i,\theta)=0. \tag{12.4}$$

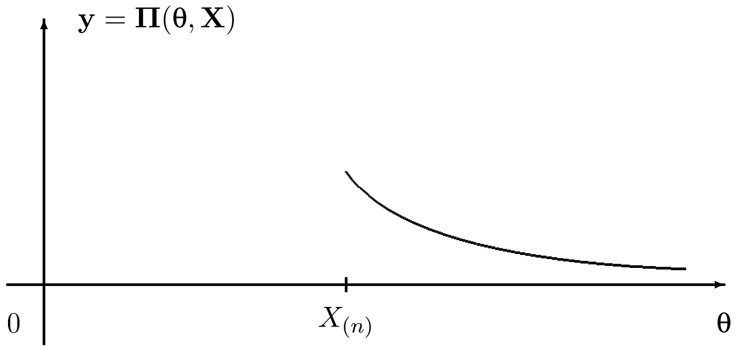


Рис. 12.1: Функция правдоподобия для $\vec{X}\Subset U_{[0,\theta]}$.

Рассмотрим теперь случай многомерного параметра, предположив опять для простоты, что $\theta=(\theta_1,\theta_2)$ - двумерный параметр. Тогда для нахождения ОМП нужно найти точку $\hat\theta=(\hat\theta_1,\hat\theta_2)$ наибольшего значения функции двух переменных $\Pi(\theta_1,\theta_2)$. В частности, если функция правдоподобия дифференцируема, то для решения этой задачи, вместо уравнения правдоподобия (12.3) или (12.4), нужно найти решение следующей системы уравнений:

$$\begin{cases}\frac{\partial\Pi(\theta_1,\theta_2)}{\partial\theta_1}=0,\\ \\ \frac{\partial\Pi(\theta_1,\theta_2)}{\partial\theta_2}=0.\end{cases} \tag{12.5}$$

§ 12.2. Сравнение оценок: среднеквадратический подход

Пусть $\vec{X}\Subset F(t,\theta)$, $\theta\in\Theta$, и $\widetilde\theta=\widetilde\theta(\vec{X})$ — какая-нибудь оценка параметра θ. Так как оценка является случайной величиной, то даже свойство несмещенности не гарантирует близость ее конкретной реализации $\widetilde\theta(\vec{x})$ к оцениваемому параметру. Если оценка является состоятельной, то такая близость гарантируется с заданной вероятностью, но только при достаточно больших объемах выборки n. При фиксированном объеме выборки наиболее распространенной «мерой близости» оценки к оцениваемому параметру является **квадратическая характеристика**, или среднее значение квадрата отклонения $\mathbf{E}(\widetilde{\theta}-\theta)^2$.

Из двух оценок $\widetilde{\theta}_1$ считается **лучше**, чем $\widetilde{\theta}_2$, если при всех $\theta \in \Theta$ выполняется неравенство

$$\mathbf{E}(\widetilde{\theta}_1-\theta)^2 \leq \mathbf{E}(\widetilde{\theta}_2-\theta)^2, \tag{12.6}$$

а хотя бы для одного θ неравенство в (12.6) оказывается строгим.

Заметим, что квадратическая характеристика оценки не меньше ее дисперсии, и равенство достигается для несмещенных оценок:

$$\mathbf{E}(\widetilde{\theta}-\theta)^2 = \Big(\mathbf{E}(\widetilde{\theta}-\theta)\Big)^2 + \mathbf{D}(\widetilde{\theta}-\theta) = \Big(\mathbf{E}\widetilde{\theta}-\theta\Big)^2 + \mathbf{D}\widetilde{\theta} \geq \mathbf{D}\widetilde{\theta}.$$

Если $\widetilde{\theta}$ — несмещенная оценка параметра θ, то есть $\mathbf{E}\widetilde{\theta}=\theta$, то для нее

$$\mathbf{E}(\widetilde{\theta}-\theta)^2 = \Big(\mathbf{E}\widetilde{\theta}-\theta\Big)^2 + \mathbf{D}\widetilde{\theta} = \mathbf{D}\widetilde{\theta}.$$

Отметим, что при среднеквадратическом подходе к сравнению оценок нельзя найти наилучшую в классе всех оценок (в частности, существуют несравнимые оценки).

Теорема 12.1. *В невырожденной статистической задаче (то есть в ситуации, когда по выборке нельзя однозначно определить неизвестный параметр) не существует наилучшей в классе всех оценок.*

Доказательство. Предположим, что существует наилучшая в классе всех оценок параметра θ оценка $\check{\theta}$. Рассмотрим следующую дурацкую оценку $\tilde{\theta}_d=\theta_0$. Эта оценка равняется константе θ_0 для любых выборочных значений, то есть никак не использует информацию, представленную выборкой. Однако если θ_0 — возможное значение параметра θ, то нельзя исключить ситуации, когда $\theta=\theta_0$ (дурацкая оценка может оказаться наиболее верной, если правильно угадывает значение неизвестного параметра). В этом случае квадратичная характеристика этой оценки равна 0:

$$\mathbf{E}(\tilde{\theta}_d-\theta)^2 = \mathbf{E}(\theta_0-\theta_0)^2 = 0.$$

Но если оценка $\check{\theta}$ не хуже, чем $\tilde{\theta}_d$, то ее квадратичная характеристика всегда не больше, чем квадратичная характеристика оценки $\tilde{\theta}_d$ для любого θ. В частности, при $\theta=\theta_0$ получаем:

$$\mathbf{E}(\check{\theta}-\theta_0)^2 \leq \mathbf{E}(\theta_0-\theta_0)^2 = 0.$$

Отсюда $\mathbf{E}(\check{\theta} - \theta_0)^2 = 0$, то есть $\check{\theta} = \theta_0$ с вероятностью 1. Это противоречит невырожденности статистической задачи — предполагалось, что по выборке нельзя однозначно определить неизвестный параметр. Итак, не может существовать наилучшей в классе всех оценок. Теорема доказана.

Для того, чтобы избежать необходимости сравнивать получаемые оценки с вырожденными оценками (рассмотренными в доказательстве теоремы), нужно ограничить класс рассматриваемых оценок. Как правило, сравнивают только несмещенные оценки. Среди несмещенных оценок наилучшая оценка параметра для заданного параметрического семейства может существовать. Ее называют *эффективной* оценкой. Эффективная оценка имеет наименьшую дисперсию из всех несмещенных оценок.

К сожалению, такое определение эффективной оценки непригодно для практического использования, так как для проверки оптимальности одной оценки требуется сравнивать дисперсии всех оценок всех несмещенных оценок. Поэтому желательно иметь критерий, позволяющий проверять оптимальность оценки на основании характеристик распределения только этой оценки. Один из таких критериев основан на неравенстве Рао-Крамера, которое сформулировано ниже.

Для формулировки точного результата введем дополнительное условие. Пусть функция распределения $F(t, \theta)$ рассматриваемой модели имеет плотность или ряд распределения, которые мы, как и прежде, обозначаем одинаково: $f(t, \theta)$). Будем предполагать, что функция $f(t, \theta)$ удовлетворяет некоторым аналитическим условиям, которые будем называть **условиями регулярности** (условия (R)), и суть которых заключается в возможности менять порядок дифференцирования по θ и интегрирования по $\vec{x}$ функции правдоподобия $\Pi(\vec{x}, \theta)$, соответствующей $f(t, \theta)$. Точная формулировка этих условий довольно сложна, приведем в качестве примера условие из [1], достаточное для (R): функция $\sqrt{f(t, \theta)}$ дифференцируема по $\theta \in \Theta$, и функция $\mathbf{i}(\theta)$, называемая информацией по Фишеру и определяемая равенством

$$\mathbf{i}(\theta) = \mathbf{E}\left(\frac{\partial \ln f(X_i, \theta)}{\partial \theta}\right)^2, \tag{12.7}$$

существует, строго положительна и непрерывна по θ для всех $\theta \in \Theta$.

Примером модели, для которой не выполнены условия регулярности, является модель $\vec{X} \Subset U_{[0;\ \theta]}$, $\theta > 0$.

Теорема 12.2. (Неравенство Рао-Крамера.)
Пусть $\vec{X} \Subset F(t, \theta)$, $\theta \in \Theta$, *выполнены условия регулярности. Тогда для*

любой несмещенной оценки $\tilde{\theta}$ *параметра* θ *выполняется неравенство:*

$$\mathbf{D}\tilde{\theta} \geq \frac{1}{n\mathbf{i}(\theta)}. \tag{12.8}$$

Если для некоторой несмещенной оценки $\check{\theta}$ окажется, что ее дисперсия совпадает с правой частью неравенства Рао-Крамера (говорят, что для этой оценки в неравенстве Рао-Крамера достигается равенство), то оценка $\check{\theta}$ является эффективной, так как для любой другой несмещенной оценки $\tilde{\theta}$ неравенство (12.8) продолжает выполняться и, следовательно, $\mathbf{D}\tilde{\theta} \geq \mathbf{D}\check{\theta}$ для всех $\theta \in \Theta$.

§ 12.3. Решение типовых примеров

Пример 12.1. *В условиях примера 11.4 найти ОМП неизвестного параметра* λ.

Решение. Для распределения Пуассона Π_λ ряд распределения имеет вид:

$$f(t,\lambda) = \mathbf{P}(X_i = t) = e^{-\lambda}\frac{\lambda^t}{t!}.$$

Искомая оценка должна быть решением уравнения правдоподобия (12.3) или (12.4). Для решения этого уравнения вычислим последовательно:

$$f(X_i,\lambda) = e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}, \quad \ln f(X_i,\lambda) = -\lambda + X_i \ln\lambda - \ln(X_i!),$$

$$\frac{d}{d\lambda}\ln f(X_i,\lambda) = -1 + \frac{1}{\lambda}X_i,$$

$$\sum_{i=1}^{n}\frac{d}{d\lambda}\ln f(X_i,\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n}X_i = -n + \frac{1}{\lambda}n\overline{X}.$$

Тогда уравнение (12.4) и его решение имеют вид:

$$-n + \frac{1}{\lambda}n\overline{X} = 0 \iff \lambda = \overline{X}.$$

Заметим, что вторая производная логарифмической функции правдоподобия

$$\frac{d^2}{d\lambda^2}\ln\Pi(\lambda) = -\frac{1}{\lambda^2}n\overline{X} < 0$$

при всех λ, так как при нашем предположении $\vec{X} \Subset \Pi_\lambda$ все элементы выборки $X_1, X_2, ..., X_n$, а значит, и выборочное среднее $\overline{X}$, с вероятностью единица неотрицательны. Значит, найденное решение $\lambda = \overline{X}$ уравнения правдоподобия является единственной точкой максимума функций $\Pi(\lambda)$ и $\ln \Pi(\lambda)$, а следовательно, статистика $\hat{\lambda} = \overline{X}$ является ОМП параметра λ.

Пример 12.2. *Пусть* $\vec{X} \Subset U_{[0,\theta]}$, *где* $\theta > 0$. *Найти ОМП для параметра* θ.

Решение. Найдем функцию правдоподобия, соответствующую выборке $\vec{X}$ из равномерного распределения $U_{[0,\theta]}$. Плотность распределения закона $U_{[0,\theta]}$ при $t = X_i$ равна:

$$f(X_i, \theta) = \begin{cases} \frac{1}{\theta}, & \text{если } X_i \in [0, \theta], \\ 0, & \text{если } X_i \notin [0, \theta] \end{cases} \tag{12.9}$$

Тогда функция правдоподобия вычисляется следующим образом:

$$\Pi(\theta) = \Pi(\vec{X},\ \theta) = \prod_{i=1}^{n} f(X_i, \theta) = \begin{cases} \frac{1}{\theta^n}, & \text{если } X_i \in [0, \theta] \text{ для всех } i = 1, 2, ..., n; \\ 0, & \text{иначе.} \end{cases}$$

Это соотношение можно переписать в следующих равносильных формах:

$$\Pi(\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{если } \max\limits_{1 \le i \le n} X_i \le \theta; \\ 0, & \text{иначе.} \end{cases} \iff \Pi(\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{если } \theta > X_{(n)}; \\ 0, & \text{иначе.} \end{cases}$$

Последнее задание функции $\Pi(\theta) = \Pi(\vec{X},\ \theta)$ позволяет легко изобразить ее график (см. рис. 12.1). Из графика видно, что своего наибольшего значения функция $\Pi(\theta)$ достигает при $\theta = X_{(n)}$. Следовательно, оценка максимального правдоподобия имеет вид: $\hat{\theta}(\vec{X}) = X_{(n)}$.

Пример 12.3. *В условиях примера 11.5 найти ОМП неизвестного параметра* $\theta = (\theta_1, \theta_2)$.

Решение. Найдем функцию правдоподобия, соответствующую выборке $\vec{X}$ из равномерного распределения $U_{[\theta_1,\theta_2]}$. Так как плотность распределения закона $U_{[\theta_1,\theta_2]}$ при $t = X_i$ равна

$$f(X_i, \theta) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \text{если } X_i \in (\theta_1, \theta_2), \\ 0, & \text{если } X_i \notin [\theta_1, \theta_2], \end{cases}$$

то функция правдоподобия представляется в виде

$$\Pi(\theta_1, \theta_2) = \prod_{i=1}^{n} p(X_i, \theta) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)^n}, & \text{если все} X_i \in [\theta_1, \theta_2]; \\ 0, & \text{иначе.} \end{cases}$$

Или по-другому:

$$\Pi(\theta_1, \theta_2) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)^n}, & \text{если } \theta_2 \geq X_{(n)},\ \theta_1 \leq X_{(1)}; \\ 0, & \text{иначе.} \end{cases} \tag{12.10}$$

Из последнего равенства видно, что функция правдоподобия отлична от нуля (более того, строго положительна) лишь при значениях (θ_1, θ_2), удовлетворяющих неравенствам:

$$\theta_1 \leq X_{(1)} \leq X_{(n)} \leq \theta_2.$$

Значит, своего наибольшего значения функция $\Pi(\theta_1, \theta_2)$ достигает лишь при таких (θ_1, θ_2). Однако при таких значениях (θ_1, θ_2) разность $(\theta_2 - \theta_1)$ принимает свое наименьшее значение $(X_{(n)} - X_{(1)})$ при $\theta_2 = X_{(n)}$, $\theta_1 = X_{(1)}$. А значит, в силу (12.10), функция $\Pi(\theta_1, \theta_2)$ принимает свое наибольшее значение при тех же значениях (θ_1, θ_2), то есть искомая ОМП имеет вид: $\hat{\theta_2} = X_{(n)}$, $\hat{\theta_1} = X_{(1)}$.

Пример 12.4. *Пусть* $\vec{X} \Subset U_{[0;\,\theta]}$, $\theta > 0$. *Сравнить с помощью среднеквадратического подхода оценки параметра* θ : $\theta_1^* = 2\overline{X}$ *и* $\hat{\theta} = X_{(n)}$.

Решение. Проверим сначала свойство несмещенности обеих оценок. Вычисляем математические ожидания, используя результаты предыдущего параграфа (см. решение примера 11.2):

$$\mathbf{E}\theta_1^* = \mathbf{E}(2\overline{X}) = 2\mathbf{E}\overline{X} = 2\frac{\theta}{2} = \theta, \quad \mathbf{E}\hat{\theta} = \mathbf{E}X_{(n)} = \frac{n}{n+1}\theta.$$

Видим, что из двух оценок $\theta_1^* = 2\overline{X}$ является несмещенной, а $\hat{\theta} = X_{(n)}$ — смещенной. Чтобы выяснить, какая из оценок лучше, вычислим для каждой квадратичную характеристику. Для несмещенной оценки она совпадает с дисперсией:

$$\mathbf{E}(\theta_1^* - \theta)^2 = \mathbf{D}\theta_1^* = \mathbf{D}(2\overline{X}) = 4\mathbf{D}\frac{1}{n}\sum_{i=1}^{n} X_i = 4\frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}_\theta X_i =$$

$$= 4\frac{1}{n^2}n\mathbf{D}_\theta X_1 = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}. \tag{12.11}$$

При вычислении квадратичной характеристики оценки $\hat{\theta}_n = X_{(n)}$ мы будем использовать плотность ее распределения, найденную при решении примера 11.2.

$$\mathbf{E}(X_{(n)} - \theta)^2 = \mathbf{E}X_{(n)}^2 - 2\theta\mathbf{E}X_{(n)} + \theta^2 = \int\limits_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} dy - 2\theta\frac{n\theta}{n+1} + \theta^2 =$$

$$= \frac{n}{n+2}\theta^2 - \frac{2n}{n+1}\theta^2 + \theta^2 = \frac{2\theta^2}{(n+1)(n+2)}. \qquad (12.12)$$

Сравнивая квадратичные характеристики, вычисленные в (12.11) и (12.12), видим, что

$$\frac{\theta^2}{3n} \geq \frac{2\theta^2}{(n+1)(n+2)}$$

для всех $\theta > 0$ и для всех $n \geq 1$.

Следовательно, ОМП $\hat{\theta} = X_{(n)}$ лучше в среднеквадратичном, чем ОММ $\theta_1^* = 2\overline{X}$.

Пример 12.5. *Исследовать с помощью неравенства Рао-Крамера оптимальность оценки* $\overline{X}$ *в моделях:*

а) $\vec{X} \Subset E_{\frac{1}{\theta}}$, $\theta > 0$;

б) $\vec{X} \Subset \Pi_\lambda$, $\lambda > 0$.

Решение. а) Вычислим дисперсию оценки, применяя свойства дисперсии и учитывая, что $\mathbf{D}X_i = \theta^2$:

$$\mathbf{D}\theta^* = \mathbf{D}\overline{X} = \mathbf{D}\frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n^2}\sum_{i=1}^n \mathbf{D}X_i = \frac{1}{n^2} n\mathbf{D}X_1 = \frac{\theta^2}{n}. \qquad (12.13)$$

Найдем правую часть неравенства Рао-Крамера для рассматриваемой модели, для этого вычислим последовательно:

$$f(X_i, \theta) = \frac{1}{\theta}e^{-\frac{X_i}{\theta}}; \quad \ln f(X_i, \theta) = -\ln\theta - \frac{X_i}{\theta}; \quad \frac{\partial \ln f(X_i, \theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{X_i}{\theta^2};$$

$$\mathbf{i}(\theta) = \mathbf{E}\left(\frac{\partial \ln f(X_i, \theta)}{\partial\theta}\right)^2 = \mathbf{E}\left(-\frac{1}{\theta} + \frac{X_i}{\theta^2}\right)^2 = \mathbf{E}\left(\frac{X_i - \theta}{\theta^2}\right)^2 =$$

$$= \frac{1}{\theta^4}\mathbf{E}(X_i - \theta)^2 = \frac{\mathbf{D}X_i}{\theta^4} = \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2};$$

$$\frac{1}{n\mathbf{i}(\theta)} = \frac{\theta^2}{n}. \tag{12.14}$$

Сравнивая (12.13) и (12.14), видим, что для оценки $\theta^* = \overline{X}$ в неравенстве Рао-Крамера достигается равенство, следовательно, она эффективна.

б) Прежде всего вспомним, что для распределения Пуассона Π_λ математическое ожидание и дисперсия равны $\mathbf{E}X_i = \lambda$, $\mathbf{D}X_i = \lambda$. Тогда дисперсия нашей оценки равна:

$$\mathbf{D}\lambda^* = \mathbf{D}\overline{X} = \mathbf{D}\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}X_i = \frac{1}{n^2}n\mathbf{D}X_1 = \frac{\lambda}{n}. \tag{12.15}$$

Аналогично пункту а), вычисляем правую часть неравенства Рао-Крамера:

$$f(X_i,\lambda) = e^{-\lambda}\frac{\lambda^{X_i}}{X_i!};\ \ln f(X_i,\lambda) = -\lambda + X_i\ln\lambda - \ln(X_i!);\ \frac{\partial \ln f(X_i,\lambda)}{\partial\lambda} = -1 + \frac{X_i}{\lambda};$$

$$\mathbf{i}(\lambda) = \mathbf{E}\left(\frac{\partial \ln f(X_i,\lambda)}{\partial\lambda}\right)^2 = \mathbf{E}\left(\frac{X_i}{\lambda} - 1\right)^2 = \mathbf{E}\left(\frac{X_i - \lambda}{\lambda}\right)^2 =$$

$$= \frac{1}{\lambda^2}\mathbf{E}(X_i - \lambda)^2 = \frac{1}{\lambda^2}\mathbf{D}X_i = \frac{1}{\lambda};$$

$$\frac{1}{n\mathbf{i}(\lambda)} = \frac{\lambda}{n}. \tag{12.16}$$

Сравнивая (12.15) и (12.16), видим, что для оценки $\lambda^* = \overline{X}$ в неравенстве Рао-Крамера достигается равенство, следовательно, она эффективна.

§ 12.4. Задачи для самостоятельного решения

12.1 По выборке $(X_1,\ \ldots,\ X_n)$ из бернуллиевского распределения B_p с неизвестным параметром $p \in (0;\ 1)$ построить оценку параметра p методом максимального правдоподобия. (Указание: показать, что вероятность попадания в точку t для элементов выборки равна $f(t,\ p) = p^t(1-p)^{1-t}$, где t может принимать только два значения: 0 и 1). Исследовать состоятельность и несмещенность полученной оценки.
12.2 По выборке $(X_1,\ \ldots,\ X_n)$ из биномиального распределения $B_{m,p}$ построить оценку максимального правдоподобия параметра p при известном $m > 0$. Исследовать состоятельность и несмещенность оценки.
12.3 По выборке из показательного распределения E_α построить оценку

максимального правдоподобия параметра $\alpha > 0$. Исследовать состоятельность оценки.

12.4 Построить оценку максимального правдоподобия по выборке из распределения Парето с плотностью

$$f_\theta(t) = \begin{cases} \frac{\theta}{t^{\theta+1}}, & t \geq 1, \\ 0, & t < 1 \end{cases}.$$

Доказать состоятельность полученной оценки.

12.5 По выборке $(X_1, \ \ldots, \ X_n)$ из распределения Лапласа с плотностью $f_\lambda(t) = \frac{\lambda}{2}e^{-\lambda|t|}, \quad t \in \mathbf{R}$, построить оценку параметра $\lambda > 0$ методом максимального правдоподобия.

12.6 Пусть дана выборка из нормального распределения с параметрами α и σ^2. Используя метод максимального правдоподобия, построить оценки

а) неизвестного математического ожидания α;

б) неизвестной дисперсии σ^2, если α известно;

в) неизвестной дисперсии σ^2, если α неизвестно.

Исследовать полученные оценки на несмещенность и состоятельность.

12.7 Используя метод максимального правдоподобия, оценить параметр θ равномерного распределения на отрезке

а) $[-\theta; \ \theta]$, $\theta > 0$; б) $[\theta; \ \theta + 1]$.

Исследовать полученные оценки на несмещенность и состоятельность.

12.8 С помощью метода максимального правдоподобия построить оценку параметра $\theta > 0$, если распределение выборки имеет плотность

а) $\theta t^{\theta-1}$ при $t \in [0; \ 1]$; б) $2t/\theta^2$ при $t \in [0; \ \theta]$.

Исследовать полученные оценки на состоятельность.

12.9 Дана выборка из распределения с плотностью

$$f_\theta(t) = \begin{cases} 3t^2\theta^{-3}, & t \in [0; \ 1], \\ 0, & t \notin [0; \ 1]. \end{cases}$$

Найти оценку параметра $\theta > 0$ методом максимального првдоподобия, исследовать ее на несмещенность и состоятельность.

12.10 По выборке $(X_1, \ldots, X_n)$ методом максимального правдоподобия найти оценку параметра $p \in (0, 1)$, если известно, что

$P\{X_1 = 1\} = p/2, \ P\{X_1 = 2\} = p/2, \ P\{X_1 = 3\} = 1 - p.$

Будет ли полученная оценка несмещенной и состоятельной?

12.11 По реализации $\vec{x} = (4; \ 5; \ 2)$ выборки из равномерного распределения на отрезке $[0; \ \theta]$ найти реализации оценок параметра θ, предложенных в примере 11.2.

12.12 По реализации $\vec{x} = (0; \ 2; \ 0; \ 3)$ выборки из распределения Пуассона с параметром $\lambda > 0$ найти реализации оценок параметра λ по первому и второму моментам, и оценки максимального правдоподобия.

12.13 По реализации $\vec{x} = (-2;\ 3;\ 4;\ -2;\ 1)$ выборки из равномерного распределения на отрезке $[\theta_1;\ \theta_2]$ найти реализации оценок параметров, предложенных в примере 11.5, и оценки максимального правдоподобия из примера 12.4.

12.14 Дана выборка $\vec{X} \Subset \mathrm{U}_{[0,\theta]}$, $\theta > 0$ — неизвестный параметр. Сравнить, какая из оценок для параметра θ лучше в среднеквадратичном: $\theta_1^* = 2\overline{X}$, $\theta_2^* = \frac{n+1}{n} X_{(n)}$.

12.15 Пусть $\vec{X} \Subset F(t, \theta)$, где $\theta = \mathbf{E}_\theta X_1$, $\mathbf{D} X_1 < \infty$. Показать, что оценка $\theta_1^* = \overline{X}$ является наилучшей в среднеквадратичном среди всех несмещенных оценок вида

$$\theta^* = C_1 X_1 + C_2 X_2 + \cdots + C_n X_n, \quad C_1 + C_2 + \cdots + C_n = 1.$$

12.16 Исследовать с помощью неравенства Рао-Крамера оптимальность ОМП для неизвестного параметра в моделях

а) $\vec{X} \Subset \mathrm{B}_p$, $0 < p < 1$;

б) $\vec{X} \Subset \mathrm{B}_{m,p}$, $0 < p < 1$, m — известно.

в) $\vec{X} \Subset \mathrm{N}_{\theta,1}$, $-\infty < \theta < \infty$.

г) $\vec{X} \Subset \mathrm{N}_{0,\theta}$, $0 < \theta < \infty$.

д) $\vec{X} \Subset \mathrm{G}_{1/\theta}$, $\theta > 1$.

12.17 Дана выборка из распределения с плотностью

$$f_\theta(t) = \begin{cases} e^{\theta - t}, & t \geq \theta, \\ 0, & t < \theta. \end{cases}$$

Найти оценку для θ а) методом моментов; б) методом максимального правдоподобия. Будут ли полученные оценки несмещенными и состоятельными?

12.18 Вычислить смещения оценок в задаче **12.17** и получить исправленные несмещенные оценки.

Глава 13

Статистическая обработка в пакете Excel

Пакет программ Microsoft Excel для ОС Windows не является специализированным пакетом статистического анализа, но широко распространен и снабжен набором функций, достаточным для решения большинства статистических задач.

§ 13.1. Пример статистической обработки

Рассмотрим процедуры статистического анализа на примере искусственно сгенерированной выборки.

Пример 13.1. *Сгенерировать реализацию выборки объема* $n = 30$ *по формуле* $x_i = 1 - 100 \ln u_i$, *где* u_i *— случайные числа — образуют реализацию выборки из равномерного распределения на* $[0;\ 1]$. *Построить реализацию вариационного ряда и гистограммы, выбрав число промежутков по формуле Стеджеса. Выдвинуть две двухпараметрических гипотезы о распределении выборочных значений. Оценить параметры распределений методом моментов (по первому и второму моментам) и методом максимального правдоподобия. На основании полученных реализаций оценок построить реализации оценок функций распределения. Сделать вывод о наиболее адекватной модели.*

Решение. Получим реализацию выборки в столбике `A` электронной таблицы. Для этого в ячейку `A1` введем формулу

=1-LN(СЛЧИС())*100

(здесь `СЛЧИС()` — *математическая* функция, реализующая независимые случайные числа, равномерно распределенные на отрезке от 0 до 1).

Скопируем содержимое ячейки в ячейки `A2-A30`. Скопируем *значения* столбика `A` в тот же столбик (для этого щелкнем правой кнопкой мыши по букве `A` и в выпадающем меню выберем *специальная вставка* $\Rightarrow$ *значения.*

Копирование значений фиксирует реализацию выборки, сохраняя значения от последующего пересчета.

Вычислим количество промежутков по формуле Стеджеса: в ячейку `B1` введем формулу

=ЦЕЛОЕ(LOG(30;2))+1

В ячейках `B2, B3, B4, B5` найдем последовательно наибольшее и наименьшее значения, размах реализации выборки и длину промежутка:

ячейка	формула
B2	=МАКС(A:A)
B3	=МИН(A:A)
B4	=B2-B3
B5	=B4/B1

Последовательно прибавляя длину промежутка к минимальному значению, хранящемуся в ячейке `A1`, получаем в столбике `C` правые границы промежутков: 61,5; 117; 173; 229; 284 (округленно). Отметим, что здесь надо специально позаботиться о том, чтобы все элементы попали левее самой правой границы промежутка, для этого прибавим к самой правой границе 1, получив 285 вместо 284.

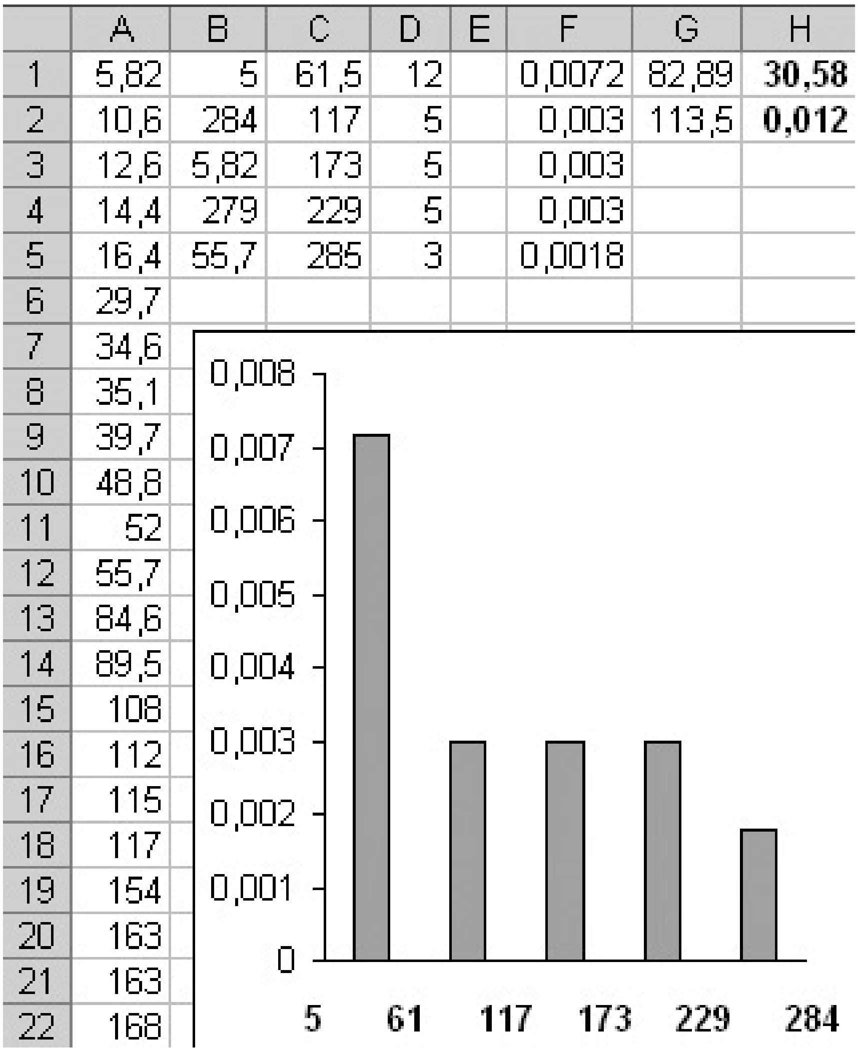
Подсчитаем количества элементов, попавших в каждый из промежутков. Для этого воспользуемся функцией `ЧАСТОТА`. Введем в ячейку `D1` формулу

=ЧАСТОТА(A1:A30;C1:C5)

Затем выделим ячейки `D1:D5`, нажмем клавишу F2 и введем формулу как формулу массива, нажав клавиши `CTRL+SHIFT+ENTER`.

В столбике `F` получим значения гистограммы, разделив значения столбика `D` на $n = 30$ и на длину промежутка, хранящуюся в ячейке `B5`. Построим гистограмму по столбику `F` с помощью функции *диаграмма* (см. рис. 13.1).

По виду гистограммы нам предстоит решить, какие гипотезы о распределении выборки следует выдвинуть. Вспомним, как выглядят графики плотности распределения изученных нами двухпараметрических семейств распределений: равномерного, сдвинутого показательного, Парето, нормального.



	A	B	C	D	E	F	G	H
1	5,82	5	61,5	12		0,0072	82,89	**30,58**
2	10,6	284	117	5		0,003	113,5	**0,012**
3	12,6	5,82	173	5		0,003		
4	14,4	279	229	5		0,003		
5	16,4	55,7	285	3		0,0018		
6	29,7							
7	34,6							
8	35,1							
9	39,7							
10	48,8							
11	52							
12	55,7							
13	84,6							
14	89,5							
15	108							
16	112							
17	115							
18	117							
19	154							
20	163							
21	163							
22	168							

Рис. 13.1: Таблица Excel и гистограмма выборочных данных.

Заметим, что только сдвинутое показательное распределение и распределение Парето имеют плотности, похожие на полученную гистограмму (рис. 13.2). На рисунке слева изображен график плотности сдвинутого показательного распределения, справа — распределения Парето. Напомним, что формулы для плотностей распределений имеют следующий вид.

$$f_{\alpha,\theta}(t) = \begin{cases} \alpha e^{-\alpha(t-\theta)}, & \text{если } t \geq \theta; \\ 0 & \text{иначе;} \end{cases} \qquad f_{\gamma,h}(t) = \begin{cases} \gamma h^{\gamma} t^{-(\gamma+1)}, & \text{если } t \geq h; \\ 0 & \text{иначе.} \end{cases}$$

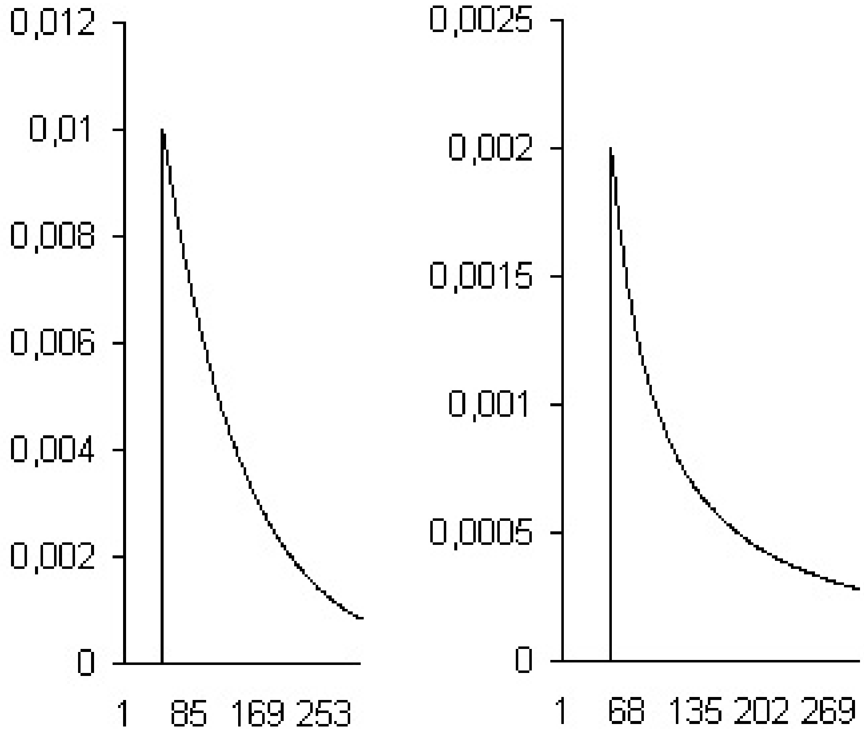


Рис. 13.2: Плотности сдвинутого показательного распределения и распределения Парето

У сдвинутого показательного распределения параметр α положительный, а параметр θ — любое действительное число. У распределения Парето оба параметра γ и h положительны. Соответствующие функции распределения имеют вид

$$F_{\alpha,\theta}(t) = \begin{cases} 1 - e^{-\alpha(t-\theta)}, & \text{если } t \geq \theta; \\ 0 & \text{иначе;} \end{cases} \qquad F_{\gamma,h}(t) = \begin{cases} 1 - h^{\gamma}t^{-\gamma}, & \text{если } t \geq h; \\ 0 & \text{иначе.} \end{cases}$$

Построим оценки параметров по первому и второму моментам. Для сдвинутого показательного распределения элементы выборки X_i равны $X_i = \theta + Y_i$, где Y_i образуют выборку из показательного распределения с параметром α, а θ — параметр сдвига. Как известно, $\mathbf{E}Y_i = 1/\alpha$, $\mathbf{D}Y_i = 1/\alpha^2$. Пользуясь свойствами математического ожидания и дисперсии, получаем систему уравнений

$$\begin{cases} \mathbf{E}X_i = \theta + 1/\alpha, \\ \mathbf{D}X_i = 1/\alpha^2. \end{cases}$$

Выразим параметры:

$$\begin{cases} \alpha = (\mathbf{D}X_i)^{-1/2}, \\ \theta = \mathbf{E}X_i - (\mathbf{D}X_i)^{1/2}. \end{cases}$$

Заменим математическое ожидание и дисперсию на выборочное среднее $\overline{X}$ и выборочную дисперсию S^2, а параметры α и θ на их оценки α^* и θ^*. Получим оценки параметров:

$$\begin{cases} \alpha^* = S^{-1}, \\ \theta^* = \overline{X} - S. \end{cases}$$

Найдем реализации этих оценок. Выборочное стандартное отклонение S — это функция `СТАНДОТКЛОНП`, а выборочное среднее — функция `СРЗНАЧ`. Вычислим их значения в ячейках `G1` и `G2`, введя туда функции `=СТАНДОТКЛОНП(A:A)` и `=СРЗНАЧ(A:A)`. В ячейках `H1` и `H2` получим реализации оценок θ^* и α^*. Для того, чтобы понять, насколько хороши оценки методом моментов, построим графики реализаций параметрической оценки функции распределения $F(t, \alpha^*, \theta^*)$ и эмпирической функции распределения $F_n^*(t)$. Получим формулу для интервала дискретизации dt переменной t, исходя из того, чтобы dt было целой степенью числа 10, и множество выборочных значений делилось не менее чем на 100 интервалов. Обозначив через $R = X_{(n)} - X_{(1)}$ размах выборки, получаем:

$$R/100 \geq dt, \quad dt = 10^k, \quad dt \leq 10^{\lg R - 2}.$$

Выбирая в качестве dt наибольшее из таких чисел, приходим к формуле

$$dt = 10^{[\lg R] - 2},$$

где $[\cdot]$ — целая часть числа.

Поскольку в нашем примере размах выборки равен 279, получаем $[\lg 279] = 2$, и $dt = 1$. Найдем значения оценки функции распределения по формуле

=ЕСЛИ(СТРОКА()<H$1;0;1-EXP(-H$2*(СТРОКА()-H$1)))

и скопируем эту формулу в ячейки `I1:I285`.

Получим значения эмпирической функции распределения в тех же точках. Для этого создадим вспомогательный столбик `M`, содержащий границы промежутков дискретизации, скопировав функцию `=СТРОКА()` в ячейки `M1:M285`. Потом подсчитаем, сколько элементов выборки попало в каждый из промежутков. Для этого воспользуемся функцией `ЧАСТОТА`. Введем в ячейку `P1` формулу

=ЧАСТОТА(A1:A30;M1:M285)

Затем выделим ячейки `P1:P285`, нажмем клавишу F2 и введем формулу как формулу массива, нажав клавиши `CTRL+SHIFT+ENTER`. Теперь получим значения эмпирической функции распределения в столбике `J`, введя в первую ячейку формулу

=СУММ(P$1:P1)/30

и скопировав ее в остальные ячейки. Здесь $30 = n$ — объем выборки.

Построим диаграмму по столбикам `I` и `J` (рис. 13.3).

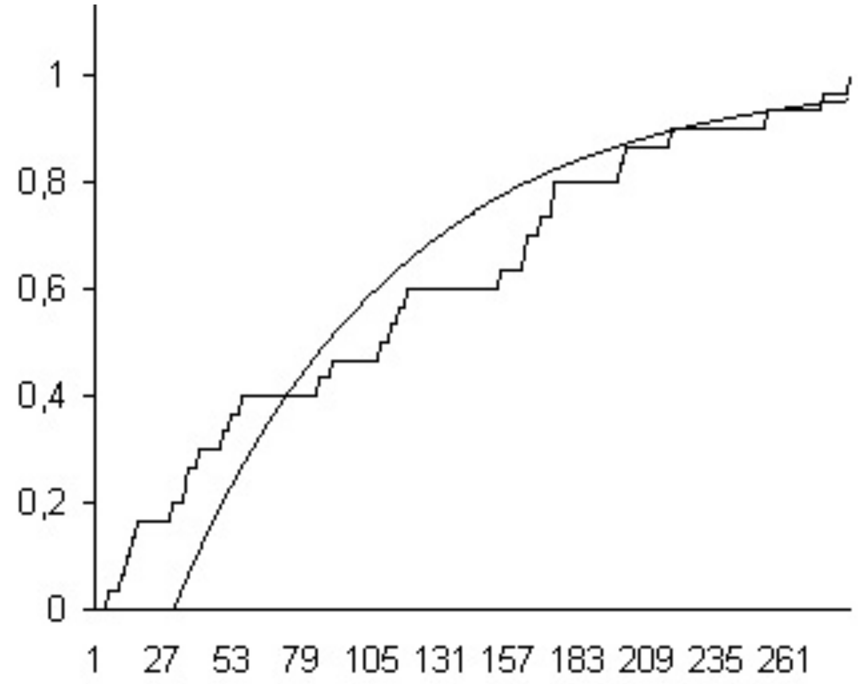


Рис. 13.3: Оценка функции сдвинутого показательного распределения методом моментов.

Теперь получим оценки максимального правдоподобия для параметров α и θ сдвинутого показательного распределения. Заметим, что плотность распределения
$$f_{\alpha,\theta}(t) = \begin{cases} \alpha e^{-\alpha(t-\theta)}, & \text{если } t \geq \theta; \\ 0 & \text{иначе;} \end{cases}$$
непрерывна по параметру $\alpha > 0$ и разрывна по параметру θ. Сначала найдем оценку параметра θ непосредственно отысканием точки максимума функции правдопобобия. Функция правдоподобия равна

$$\Pi(\vec{X}, \alpha, \theta) = \begin{cases} \prod_{i=1}^{n}(\alpha e^{-\alpha(X_i-\theta)}), & \text{если все } X_i \geq \theta; \\ 0 & \text{иначе;} \end{cases}$$

или
$$\Pi(\vec{X}, \alpha, \theta) = \begin{cases} \alpha^n e^{-\alpha(\sum_{i=1}^{n} X_i - n\theta)}, & \text{если } \theta \leq \min\{X_i\}; \\ 0 & \text{иначе.} \end{cases}$$

Зависимость функции правдоподобия от параметра θ изображена на рис 13.4. Ее максимум достигается в точке $\widehat{\theta} = \min\{X_i\}$, которая является оценкой максимального правдоподобия параметра θ.

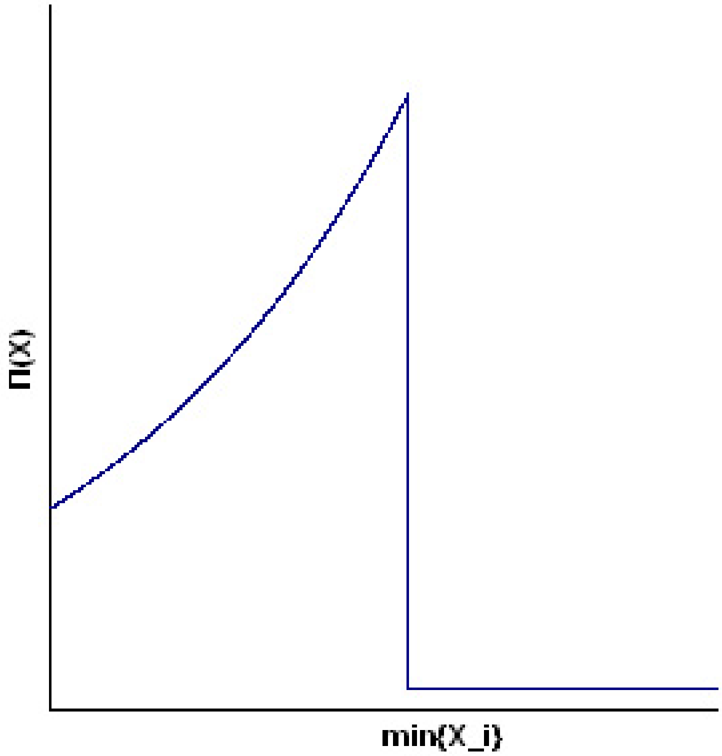


Рис. 13.4: Зависимость функции правдоподобия от параметра θ.

Найдем оценку максимального правдоподобия параметра α. Для этого последовательно вычислим

$$\ln f(t, \alpha, \theta) = \ln \alpha - \alpha(t - \theta)$$

при $t \geq \theta$;

$$\frac{\partial}{\partial \alpha} \ln f(t, \alpha, \theta) = \frac{1}{\alpha} - (t - \theta)$$

при $t \geq \theta$;

$$\frac{\partial}{\partial\alpha}\ln\Pi(\vec{X},\alpha,\theta)=\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f(X_i,\alpha,\theta)=\sum_{i=1}^{n}\left(\frac{1}{\alpha}-(X_i-\theta)\right),$$

если все $X_i \geq \theta$. Приравнивая производную логарифма функции правдоподобия к нулю, получаем уравнение для определения оценки параметра α:

$$\sum_{i=1}^{n}\left(\frac{1}{\alpha}-(X_i-\theta)\right)=0,$$

решением которого является

$$\alpha=\frac{n}{\sum_{i=1}^{n}X_i-n\theta}=\frac{1}{\overline{X}-\theta}.$$

Поскольку параметр θ неизвестен, заменим его на оценку максимального правдоподобия $\widehat{\theta}=\min\{X_i\}$ и получим

$$\widehat{\alpha}=\frac{1}{\overline{X}-\min\{X_i\}}.$$

Условие $X_i \geq \widehat{\theta}$ оказывается выполненным автоматически.

Найдем реализации оценок максимального правдоподобия и построим графики реализаций параметрической оценки функции распределения $F(t,\widehat{\alpha},\widehat{\theta})$ (как для оценок методом моментов) и эмпирической функции распределения $F_n^*(t)$. График приведен на рис. 13.5.

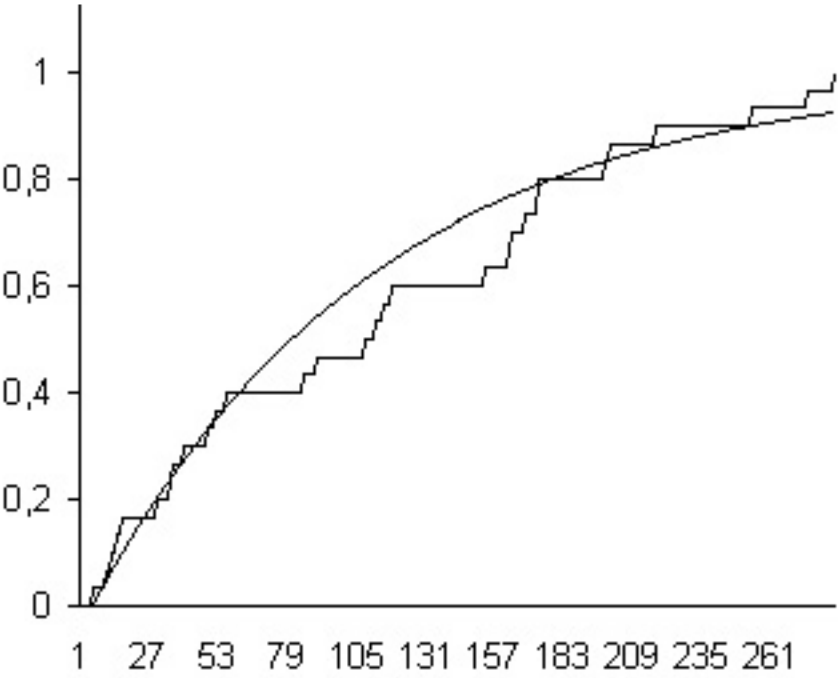


Рис. 13.5: Оценка функции сдвинутого показательного распределения методом максимального правдоподобия.

Сравнивая результат с рис. 13.3, видим, что использование оценок максимального правдоподобия позволяет более точно приблизить эмпирическую функцию распределения.

Построим оценки параметров распределения Парето по первому и второму моментам. Вспомним, что плотность распределения задается формулой

$$f_{\gamma,h}(t) = \begin{cases} \gamma h^{\gamma} t^{-(\gamma+1)}, & \text{если } t \geq h; \\ 0 & \text{иначе.} \end{cases}$$

Вычислим $\mathbf{E}X_i$ и $\mathbf{E}X_i^2$:

$$\mathbf{E}X_i = \int_{-\infty}^{\infty} t f_{\gamma,h}(t)dt = \int_h^{\infty} t\gamma h^{\gamma} t^{-(\gamma+1)}dt =$$

$$= \gamma h^{\gamma} \int_h^{\infty} t^{-\gamma} dt = \gamma h^{\gamma} \frac{t^{-\gamma+1}}{-\gamma+1}\Big|_h^{\infty} = \frac{\gamma h^{\gamma} h^{-\gamma+1}}{\gamma-1} = \frac{\gamma h}{\gamma-1},$$

если $\gamma > 1$ (в противном случае математическое ожидание не существует). Аналогично

$$\mathbf{E}X_i^2 = \int_{-\infty}^{\infty} t^2 f_{\gamma,h}(t)dt = \int_h^{\infty} t^2 \gamma h^\gamma t^{-(\gamma+1)}dt =$$

$$= \gamma h^\gamma \int_h^{\infty} t^{-\gamma+1}dt = \gamma h^\gamma \frac{t^{-\gamma+2}}{-\gamma+2}|_h^{\infty} = \frac{\gamma h^\gamma h^{-\gamma+2}}{\gamma-2} = \frac{\gamma h^2}{\gamma-2},$$

если $\gamma > 2$ (в противном случае второй момент не существует).

Получаем систему уравнений

$$\begin{cases} \mathbf{E}X_i = \frac{\gamma h^2}{\gamma-2}, \\ \mathbf{E}X_i^2 = \frac{\gamma h^2}{\gamma-2}. \end{cases}$$

Выразим параметры:

$$h = \frac{\gamma-1}{\gamma}\mathbf{E}X_i,$$

$$\frac{\gamma(\gamma-1)^2}{\gamma^2(\gamma-2)}(\mathbf{E}X_i)^2 = \mathbf{E}X_i^2,$$

$$(\gamma-1)^2(\mathbf{E}X_i)^2 = \gamma(\gamma-2)\mathbf{E}X_i^2.$$

Получаем квадратное уравнение:

$$\mathbf{D}X_i\gamma^2 - 2\mathbf{D}X_i\gamma - (\mathbf{E}X_i)^2 = 0.$$

Решая его и выбирая положительный корень, получаем:

$$\gamma = 1 + \sqrt{1 + \frac{(\mathbf{E}X_i)^2}{\mathbf{D}X_i}}.$$

Заменим математическое ожидание и дисперсию на выборочное среднее $\overline{X}$ и выборочную дисперсию S^2, а параметры h и γ на их оценки h^* и γ^*. Получим оценки параметров:

$$\begin{cases} \gamma^* = 1 + \sqrt{1 + \frac{(\overline{X})^2}{S^2}}, \\ h = \frac{\gamma^*-1}{\gamma^*}\overline{X}. \end{cases}$$

Отметим, что оценка параметра γ всегда не меньше числа 2, что соответствует требованию к параметру, обеспечивающему конечность второго момента. Найдем реализации оценок по выборке и построим графики реализаций параметрической оценки функции распределения $F(t, h^*, \gamma^*)$ по формуле

$$=\text{ЕСЛИ}(\text{СТРОКА}() < R\$1; 0; 1 - (\text{СТРОКА}()/R\$1)\hat{}(-R\$2))$$

и эмпирической функции распределения $F_n^*(t)$. График приведен на рис. 13.6.

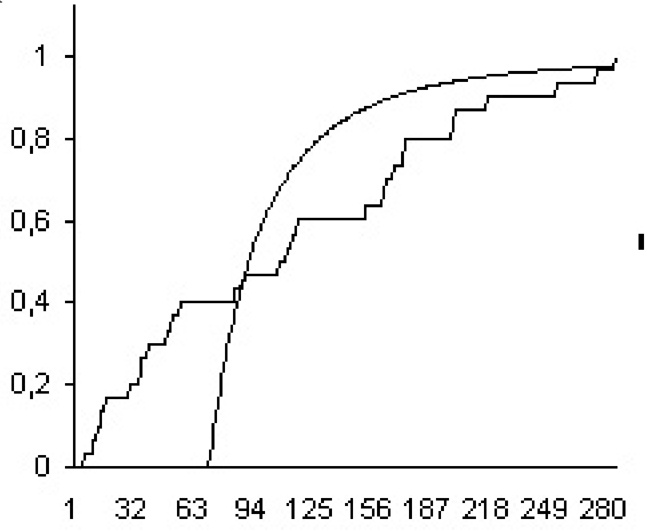


Рис. 13.6: Оценка функции распределения Парето методом моментов.

Из рисунка видно, что приближение в этом случае оказывается очень неудачным.

Получим оценки параметров распределения Парето методом максимального правдоподобия.

Сначала найдем оценку параметра h непосредственно отысканием точки максимума функции правдопобобия. Для этого запишем функцию правдоподобия

$$\Pi(\vec{X}, h, \gamma) = \begin{cases} \prod_{i=1}^{n}(\gamma h^{\gamma} X_i^{-(\gamma+1)}), & \text{если все } X_i \geq \theta; \\ 0 & \text{иначе.} \end{cases}$$

Зависимость функции правдоподобия от параметра h имеет тот же характер, что и в случае сдвинутого показательного распределения зависимость от параметра θ. Она изображена схематично на рис. 13.4. Ее максимум достигается в точке $\widehat{h} = \min\{X_i\}$, которая является оценкой максимального правдоподобия параметра h.

Найдем оценку максимального правдоподобия параметра γ. Для этого последовательно вычислим

$$\ln f(t, h, \gamma) = \ln \gamma + \gamma \ln h - (\gamma + 1) \ln t$$

при $t \geq h$;

$$\frac{\partial}{\partial \gamma} \ln f(t, h, \gamma) = \frac{1}{\gamma} + \ln h - \ln t$$

при $t \geq h$. Приравнивая производную логарифма функции правдоподобия к нулю, получаем уравнение для определения оценки параметра γ:

$$\sum_{i=1}^{n} \left(\frac{1}{\gamma} + \ln h - \ln X_i \right) = 0,$$

решением которого является

$$\gamma = \frac{1}{\overline{\ln X} - \ln h}.$$

Поскольку параметр h неизвестен, заменим его на оценку максимального правдоподобия $\widehat{h} = \min\{X_i\}$ (так же мы поступали при нахождении оценок для сдвинутого показательного распределения) и получим

$$\widehat{\gamma} = \frac{1}{\overline{\ln X} - \ln(\min\{X_i\})}.$$

Условие $X_i \geq \widehat{h}$ оказывается выполненным автоматически.

Найдем реализации оценок максимального правдоподобия. Отметим, что для нахождения выборочного усреднения логарифма $\overline{\ln X}$ нужно предварительно в отдельном столбике вычислить логарифмы всех выборочных значений, скопировав функцию `=LN(A1)`, и затем вычислить среднее из 30 значений логарифмов.

Построим графики реализаций параметрической оценки функции распределения $F(t, \widehat{h}, \widehat{\gamma})$ (как для оценок методом моментов) и эмпирической функции распределения $F_n^*(t)$.

График приведен на рис. 13.7.

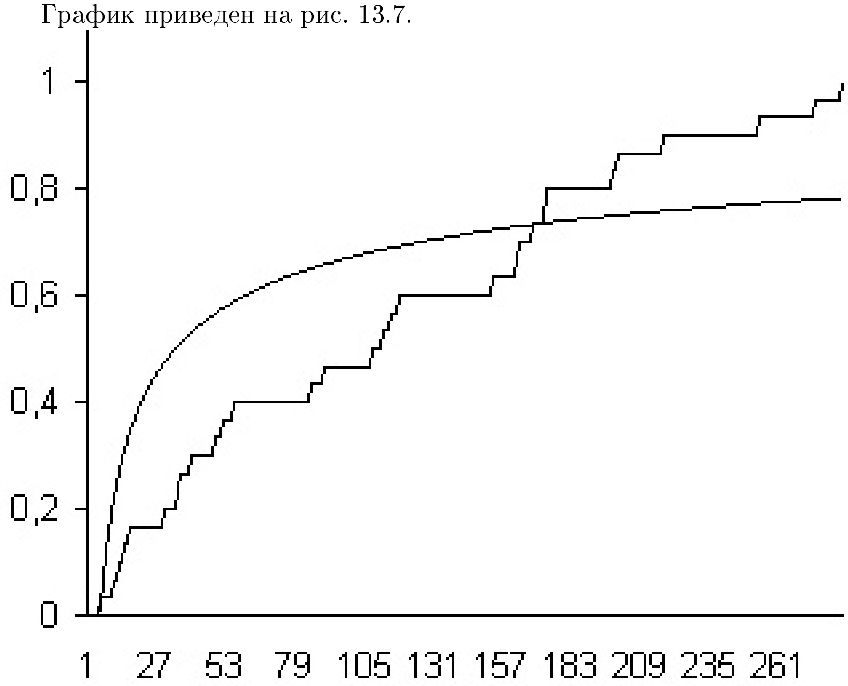


Рис. 13.7

Анализируя график, видим, что для распределения Парето и оценки максимального правдоподобия не дают хорошего приближения эмпирической функции распределения. На основании проведенного исследования можно заключить, что более адекватной моделью является модель сдвинутого показательного распределения, и лучший метод оценивания ее параметров — метод максимального правдоподобия. Оценки максимального правдоподобия здесь получаются смещенными, однако мы не будем обсуждать, как можно уменьшить смещение.

§ 13.2. Задачи для самостоятельного решения

Задача 13.1

В тексте задачи через № обозначен номер студента по списку группы.

1. Для выборки $X_1, \ \ldots, \ X_n$ из равномерного распределения на $[0; \ \theta]$ получить оценки параметра θ методом моментов на основании первого, второго, №+2-го момента. Вычислить $E(X_1 + №)e^{X_1/№}$ и на этом основании получить оценку параметра θ через усреднение соответствующей функции по выборке.
2. Для той же выборки найти оценку максимального правдоподобия параметра θ, вычислить ее математическое ожидание и исправить ее, получив несмещенную оценку.
3. Генерировать реализацию выборки объема $n = 100+№$ из равномерного распределения на $[0; \ \theta]$, приняв $\theta = №$.
4. Вычислить реализации всех полученных оценок. Подсчитать абсолютные погрешности оценивания и ранжировать оценки по абсолютной погрешности.

Задача 13.2

Случайная величина имеет логарифмически нормальное распределение, если ее логарифм распределен по нормальному закону.

1. Для выборки $X_1, \ \ldots, \ X_n$ из логарифмически нормального распределения с параметрами $a, \ \sigma$ получить оценки параметров методом моментов и методом максимального правдоподобия.
2. Генерировать реализацию выборки объема $n = 100$ по формуле $U_1U_2U_3/U_4$, где $U_1, \ldots, U_4$ — случайные числа, равномерно распределенные на $[0; \ 1]$. Построить гистограмму, выбрав число промежутков группирования по формуле Стеджеса.
3. По реализации выборки вычислить реализации всех полученных оценок.
4. * Найти теоретические значения параметров. Подсчитать абсолютные погрешности оценивания и ранжировать оценки по абсолютной погрешности.

Глава 14

Интервальное оценивание

§ 14.1. Определение доверительного интервала

Пусть имеется выборка объема n из распределения, известного с точностью до параметра: $\vec{X} \Subset F(t, \theta)$, $\theta \in \Theta$. *Доверительным интервалом с уровнем доверия* γ (γ-доверительным интервалом) для неизвестного параметра θ называют случайный интервал $(\theta_-;\ \theta_+) \subset \Theta$, построенный по выборке, который накрывает неизвестное значение параметра с вероятностью, равной γ, или по крайней мере стремящейся к γ с ростом объема выборки, то есть

$$\mathbf{P}\{\theta \in (\theta_-;\ \theta_+)\} \to \gamma \tag{14.1}$$

при $n \to \infty$.

В случае, когда вместо сходимости выполняется точное равенство, доверительный интервал называется *точным*.

θ_-, θ_+ — это оценки параметра θ, называемые *нижней и верхней доверительными границами*. Число $\gamma \in (0;\ 1)$ — уровень доверия, или доверительная вероятность, — выбирается заранее и отражает, как сказано в [7], «степень готовности мириться с возможностью ошибки»: чем менее мы готовы мириться с возможной ошибкой, тем большее (более близкое к единице) значение γ должны устанавливать.

§ 14.2. Распределения, связанные с нормальным

При построении доверительных интервалов для параметров нормального распределения мы будем использовать два специальных распределения, связанных с нормальным: распределение хи-квадрат и распределение Стьюдента. Название «распределение Стьюдента» связано с именем английского статистика К. Госсета, который подписывал свои работы псевдонимом «Стьюдент».

Случайная величина Z_n имеет *распределение хи-квадрат с n степенями свободы*, если

$$Z_n = {X_1}^2 + \ldots + {X_n}^2,$$

где $X_1, \ \ldots, \ X_n$ — независимые случайные величины со стандартным нормальным распределением. Отметим, что «число степеней свободы» — это просто традиционное название для параметра n распределения хи-квадрат. Параметр n — положительное целое число. В частности, при $n = 1$ получаем квадрат одной случайной величины со стандартным нормальным распределением: $Z_1 = X^2$, где $X \Subset N_{0,\,1}$.

Будем использовать следующее обозначение: $Z_n \Subset \chi^2_n$.

Отметим следующие свойства распределения хи-квадрат.

Следствие 14.1. *Пусть $Z_n \Subset \chi^2_n$. Тогда*
1) $\mathbf{E}Z_n = n$;
2) $Z_n/n \to 1$ почти наверное при $n \to \infty$.

Доказательство.
Во-первых,

$$\mathbf{E}Z_1 = \mathbf{E}X^2 = \mathbf{D}X + (\mathbf{E}X)^2,$$

где X имеет стандартное нормальное распределение, и потому $\mathbf{E}X = 0$, $\mathbf{D}X = 1$. Следовательно,

$$\mathbf{E}Z_1 = 1 + 0^2 = 1.$$

1) По определению распределения хи-квадрат,

$$\mathbf{E}Z_n = \mathbf{E}({X_1}^2 + \ldots + {X_n}^2) = \mathbf{E}{X_1}^2 + \ldots + \mathbf{E}{X_n}^2 = n\mathbf{E}{X_1}^2 = n \cdot 1 = n.$$

2) Так как Z_n — сумма независимых одинаково распределенных случайных величин, то справедлив закон больших чисел Колмогорова:

$$Z_n/n = \frac{{X_1}^2 + \ldots + {X_n}^2}{n} \to \mathbf{E}{X_1}^2 = 1$$

почти наверное при $n \to \infty$. Доказательство завершено.

Случайная величина Y_n имеет *распределение Стьюдента с n степенями свободы*, если

$$Y_n = \frac{X}{\sqrt{Z_n/n}},$$

где случайные величины X и Z_n независимы, причем X имеет стандартное нормальное распределение, а Z_n имеет распределение хи-квадрат с n степенями свободы. Здесь, как и у распределения хи-квадрат, n — это просто положительный целый параметр.

Будем использовать следующее обозначение: $Y_n \Subset T_n$.

Отметим следующие свойства распределения Стьюдента.

Следствие 14.2. *Пусть $Y_n \Subset T_n$. Тогда*

1) для любого t выполнено $\mathbf{P}\{Y_n < -t\} = \mathbf{P}\{Y_n > t\}$, то есть распределение Стьюдента симметрично;

2) $Y_n \to X$ почти наверное при $n \to \infty$, где X имеет стандартное нормальное распределение.

Доказательство.

1) Симметрия следует из симметрии стандартного нормального распределения:

$$\mathbf{P}\{Y_n < -t\} = \mathbf{P}\{X < -t\sqrt{Z_n/n}\} = \mathbf{P}\{X > t\sqrt{Z_n/n}\} = \mathbf{P}\{Y_n > t\}.$$

2) Сходимость следует из свойства сходимости почти наверное, непрерывности функции $x/\sqrt{y}$ и из свойства распределения хи-квадрат. Доказательство завершено.

§ 14.3. Точные доверительные интервалы

Наиболее распространенной ситуацией, когда возможно построение точных доверительных интервалов, является случай нормального распределения: $\vec{X} \Subset \mathrm{N}_{a,\sigma^2}$, — когда хотя бы один из его параметров неизвестен. В этом случае известно совместное распределение наиболее употребительных оценок $\overline{X}$ и S^2 параметров a и σ^2, с помощью которого и строятся соответствующие доверительные интервалы. Основные результаты содержатся в следующей теореме, которую примем без доказательства.

Теорема 14.1. (Теорема Фишера) *Пусть* $\vec{X} \Subset \mathrm{N}_{a,\sigma^2}$. *Тогда верны следующие 4 факта:*

1) $\dfrac{\sqrt{n}(\overline{X}-a)}{\sigma} \Subset N_{0,1}$.

2) $\dfrac{\sum_{i=1}^{n}(X_i-a)^2}{\sigma^2} \Subset \chi^2_n$.

3) $\dfrac{nS^2}{\sigma^2} \Subset \chi^2_{n-1}$.

4) $\dfrac{\sqrt{n-1}\left(\overline{X}-a\right)}{S} \Subset T_{n-1}$.

Отметим, что первое утверждение теоремы следует сразу же из свойств нормального распределения, второе — из определения распределения хи-квадрат с учетом того факта, что $\frac{X_i-a}{\sigma}$ имеет стандартное нормальное распределение. Третье утверждение — нетривиальный факт, прокомментировать который можно следующим образом. Вспомним, что

$$nS^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2.$$

Если выборка состоит из одного элемента X_1, то есть $n=1$, то

$$\frac{nS^2}{\sigma^2} = \frac{(X_1-X_1)^2}{\sigma^2} = 0,$$

— случайная величина имеет распределение, вырожденное в точке ноль, которое можно по определению считать распределением хи-квадрат с нулевым числом степеней свободы.

Если выборка состоит из двух элементов $(X_1;\ X_2)$, то есть $n=2$, то

$$\frac{nS^2}{\sigma^2} = \frac{(X_1-(X_1-X_2)/2)^2+(X_2-(X_1-X_2)/2)^2}{\sigma^2} = \frac{(X_1-X_2)^2}{2\sigma^2}.$$

Отметим, что $\mathbf{E}(X_1-X_2)=a-a=0$. Так как X_1 и X_2 независимы, то $\mathbf{D}(X_1-X_2)=\mathbf{D}X_1+\mathbf{D}X_2=2\sigma^2$. Согласно свойствам нормального распределения, $X_1-X_2 \Subset \mathrm{N}_{0,2\sigma^2}$, то есть $X_1-X_2=\sqrt{2}\sigma X$, где X имеет стандартное нормальное распределение. Поэтому

$$\frac{nS^2}{\sigma^2} = \frac{2\sigma^2X^2}{2\sigma^2} = X^2 \Subset \chi^2_1.$$

Доказательство для $n \geq 2$ оказывается существенно более сложным, и приводить его здесь мы не будем.

Четвертое утверждение следует из третьего и определения распределения Стьюдента.

§ 14.4. Асимптотические доверительные интервалы

Если распределение не является нормальным, точный доверительный интервал, как правило, не удается построить. Поэтому строят асимптотический доверительный интервал, применяя центральную предельную теорему (теорема 10.4), которая утверждает, что для всех $t_1, t_2 \in \mathbf{R}$ $(t_1 < t_2)$ выполнено (10.2)

$$\lim_{n\to\infty} \mathbf{P}\left(t_1 \leq \frac{n\overline{X} - na}{\sigma\sqrt{n}} < t_2\right) = \Phi(t_2) - \Phi(t_1), \quad n \to \infty$$

то есть центрированные и нормированные суммы случайных величин $n\overline{X} = X_1 + \ldots + X_n$ сходятся по распределению к случайной величине, имеющей стандартное нормальное распределение.

Здесь $a = \mathbf{E}X_1$, $\sigma^2 = \mathbf{D}X_1$ — математическое ожидание и дисперсия элементов выборки.

Если выбрать $t_2 = -t_1 = A$ и принять доверительный уровень равным γ, то

$$\lim_{n\to\infty} \mathbf{P}\left(-A \leq \frac{n\overline{X} - na}{\sigma\sqrt{n}} < A\right) = \Phi(A) - \Phi(-A) = 2\Phi(A) - 1 = \gamma, \quad n \to \infty$$

откуда получаем

$$\Phi(A) = (\gamma + 1)/2. \tag{14.2}$$

По заданному γ можно найти A с помощью таблиц нормального распределения или программных приложений. Отметим без доказательства следующее свойство сходимости по распределению: если Y_n сходится по распределению к Y, а Z_n сходится почти наверное к 1, то их произведение $Y_n Z_n$ сходится по распределению к Y. Выберем

$$Y_n = \frac{n\overline{X} - na}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\overline{X} - a)}{\sigma}, \qquad Z_n = \frac{\sigma}{S}.$$

Вспомним, что $S = \sqrt{\overline{X^2} - (\overline{X})^2} \to \sigma$ почти наверное, и по свойству сходимости почти наверное $Z_n \to 1$ п. н. Следовательно,

$$Y_n Z_n = \frac{\sqrt{n}(\overline{X} - a)}{\sigma} \cdot \frac{\sigma}{S} = \frac{\sqrt{n}(\overline{X} - a)}{S}$$

сходится по распределению к стандартной нормальной случайной величине, то есть

$$\lim_{n\to\infty} \mathbf{P}\left(-A \le \frac{\sqrt{n}(\overline{X} - a)}{S} < A\right) = \Phi(A) - \Phi(-A) = 2\Phi(A) - 1 = \gamma, \quad n \to \infty, \tag{14.3}$$

где константа A выбирается по формуле 14.2.

Чтобы для неизвестного параметра θ найти доверительный интервал асимптотического уровня γ, нужно для исследуемого однопараметрического семейства распределений найти зависимость $\mathbf{E}X_1 = a = a(\theta)$ и подставить полученное выражение в 14.3, то есть решить относительно параметра θ двойное неравенство

$$-A \le \frac{\sqrt{n}(\overline{X} - a(\theta))}{S} < A \tag{14.4}$$

(для этого нужно, чтобы функция $a(\theta)$ была непрерывной и строго монотонной). Получившиеся границы дверительного интервала будем обозначать через θ_- и θ_+.

§ 14.5. Решение типовых примеров

Пример 14.1. *Пусть* $\vec{X} \Subset \mathrm{N}_{a,\sigma^2}$, $a \in R$. *Построить доверительный интервал* $(a_-;\ a_+)$ *для параметра* a, *считая* σ^2 *известным. Вычислить реализацию доверительного интервала с уровнем доверия* $\gamma = 0,95$, *располагая данными:* $n = 10$, $\overline{X} = 2,7$, $\sigma^2 = 4$.

Решение. Для построения доверительного интервала используем оценку $\overline{X}$, распределение которой известно. Для заданной доверительной вероятности γ найдем такое $A > 0$, что

$$\gamma = \mathbf{P}\left\{\left|\sqrt{n}\frac{\overline{X} - a}{\sigma}\right| < A\right\} = \mathbf{P}\left\{-\frac{\sigma A}{\sqrt{n}} < \overline{X} - a < \frac{\sigma A}{\sqrt{n}}\right\}. \tag{14.5}$$

Таким образом, нужно искать $\varepsilon_1 = -\frac{\sigma A}{\sqrt{n}}$, $\varepsilon_2 = \frac{\sigma A}{\sqrt{n}}$ такие, что выполняется равенство:

$$\mathbf{P}\left\{\varepsilon_1 < \overline{X} - a < \varepsilon_2\right\} = \gamma.$$

Для этого вернемся к (14.5). В силу теоремы 14.1, случайная величина, стоящая под знаком модуля, имеет стандартное нормальное распределение, поэтому вероятность в правой части можно выразить через функцию распределения $\Phi(t)$ закона $N_{0,1}$, и тогда уравнение (14.5) приобретает вид:

$$2\Phi(A) - 1 = \gamma \iff \Phi(A) = \frac{1+\gamma}{2}, \tag{14.6}$$

где $\Phi(t)$ — функция Лапласа, значения которой представлены в таблице приложения в конце книги. Заметим, что значение $A = A_{\frac{1+\gamma}{2}}$, удовлетворяющее (14.6), представляет квантиль уровня $\frac{1+\gamma}{2}$ распределения $N_{0,1}$. Найдя его по таблице и подставив в (14.5), получим равенство:

$$\gamma = \mathbf{P}\left\{\left|\sqrt{n}\frac{\overline{X}-a}{\sigma}\right| < A_{\frac{1+\gamma}{2}}\right\} = \mathbf{P}\left\{-A_{\frac{1+\gamma}{2}} < \sqrt{n}\frac{a-\overline{X}}{\sigma} < A_{\frac{1+\gamma}{2}}\right\} \iff$$

$$\iff \gamma = \mathbf{P}\left\{\overline{X} - \sigma\frac{A_{\frac{1+\gamma}{2}}}{\sqrt{n}} < a < \overline{X} + \sigma\frac{A_{\frac{1+\gamma}{2}}}{\sqrt{n}}\right\}, \tag{14.7}$$

откуда искомый γ-доверительный интервал:

$$(a_-;\ a_+) = \left(\overline{X} - \sigma\frac{A_{\frac{1+\gamma}{2}}}{\sqrt{n}},\ \overline{X} + \sigma\frac{A_{\frac{1+\gamma}{2}}}{\sqrt{n}}\right).$$

Подставляя сюда конкретные данные из условия, вычисляем реализацию доверительного интервала:

$$(a_-;\ a_+) \approx \left(2,7 - 2\frac{1,96}{\sqrt{10}},\ 2,7 + 2\frac{1,96}{\sqrt{10}}\right) \approx (1,46; 3,94) \iff$$

$$\iff \mathbf{P}(1,46 < \theta < 3,94) = 0,95.$$

Замечание 14.1. *Построенный доверительный интервал оказывается симметричным относительно выборочного среднего $\overline{X}$ и имеет длину $2A\frac{\sigma}{\sqrt{n}}$, пропорциональную значению A, которое было найдено из условия* (14.5).

Пример 14.2. *Пусть $\vec{X} \Subset N_{a,\sigma^2}$, где $a \in \mathbf{R}$, $\sigma^2 > 0$ — два неизвестных параметра. Построить доверительный интервал для параметра σ^2. Вычислить реализацию доверительного интервала с уровнем доверия $\gamma = 0,9$, располагая данными: $n = 10$, $S^2 = 4$.*

Решение. Используя лемму Фишера, проще всего построить так называемый односторонний доверительный интервал. Для этого по заданной доверительной вероятности $\gamma = 0,9$ найдем такое $B > 0$, что

$$\gamma = \mathbf{P}\left\{\frac{nS^2}{\sigma^2} > B\right\} \iff \mathbf{P}\left\{\frac{nS^2}{\sigma^2} \leq B\right\} = 1 - \gamma. \tag{14.8}$$

Учитывая, что случайная величина $\frac{nS^2}{\sigma^2}$ имеет распределение χ^2_{n-1}, нетрудно видеть, что искомое B есть не что иное, как квантиль $\chi^2_{1-\gamma,n-1}$ этого распределения. Подставляя ее в (14.8) и разрешая неравенство под знаком вероятности относительно σ^2, находим доверительный интервал:

$$\gamma = \mathbf{P}\left\{\frac{nS^2}{\sigma^2} > \chi^2_{1-\gamma,n-1}\right\} \iff \gamma = \mathbf{P}\left\{\frac{nS^2}{\chi^2_{1-\gamma}, n-1} > \sigma^2\right\} \iff$$

$$\iff (\sigma^2_-;\ \sigma^2_+) = \left(0, \frac{nS^2}{\chi^2_{1-\gamma,n-1}}\right).$$

Подставляя конкретные данные, находим реализацию доверительного интервала:

$$\chi^2_{0.1,9} = 4,17; \quad (\sigma^2_-;\ \sigma^2_+) = (0;\ 8,63).$$

Чтобы построить двусторонний доверительный интервал, вместо (14.8) используем следующее уравнение:

$$\gamma = \mathbf{P}\left\{x_1 < \frac{nS^2}{\sigma^2} < x_2\right\} \iff \mathbf{P}_\theta\left\{\frac{nS^2}{x_2} < \sigma^2 < \frac{nS^2}{x_1}\right\} = \gamma, \tag{14.9}$$

где $0 < x_1 < x_2$, удовлетворяющие (14.9), находим по известному распределению χ^2_{n-1} случайной величины $\frac{nS^2}{\sigma^2}$. В общем случае эта задача не имеет единственного решения. Если обратиться к графику плотности распределения χ^2_{n-1}, представленному на рис. 14.1, то $x_1 < x_2$ следует выбирать таким образом, чтобы сумма вероятностей, представленных площадями заштрихованных областей под графиком плотности, равнялась $1 - \gamma$. Ясно, что это можно сделать многими способами.

Чтобы сделать решение однозначным, выберем $x_1 < x_2$ так, чтобы каждая из заштрихованных площадей равнялась $\frac{1-\gamma}{2}$, тогда нетрудно видеть, что x_1, x_2 выражаются через квантили распределения χ^2_{n-1}:

$$x_1 = \chi^2_{\frac{1-\gamma}{2},n-1}, \quad x_2 = \chi^2_{\frac{1+\gamma}{2},n-1}.$$

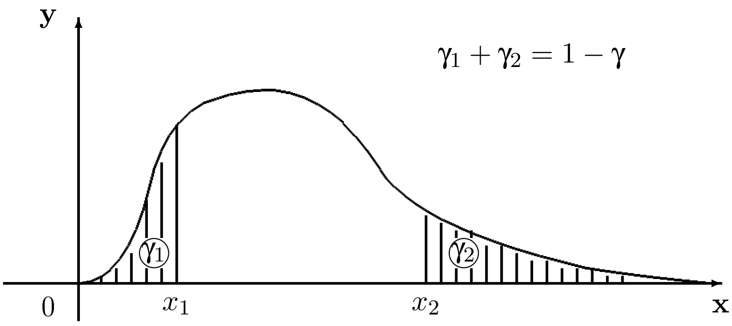


Рис. 14.1: Плотность распределения χ^2_{n-1}

Подставляя это в (14.9), находим искомый двусторонний доверительный интервал:

$$\mathbf{P}_\theta\left\{\frac{nS^2}{\chi^2_{\frac{1+\gamma}{2},n-1}}<\sigma^2<\frac{nS^2}{\chi^2_{\frac{1-\gamma}{2},n-1}}\right\}=\gamma \iff (\sigma^2_-;\,\sigma^2_+)=\left(\frac{nS^2}{\chi^2_{\frac{1+\gamma}{2},n-1}},\ \frac{nS^2}{\chi^2_{\frac{1-\gamma}{2},n-1}}\right).$$

Находя по таблице распределения хи-квадрат $\chi^2_{\frac{1-\gamma}{2},n-1}$, $\chi^2_{\frac{1+\gamma}{2},n-1}$ — квантили распределения χ^2_{n-1} для конкретных значений $\gamma=0,9$, $n=10$, и данных в условии численных значений, находим реализацию доверительного интервала:

$$\chi^2_{\frac{1+0,9}{2},9}=16,9;\ \chi^2_{\frac{1-0,9}{2},9}\approx 3,325;$$

$$(\sigma^2_-,\ \sigma^2_+)=(2,37;\ 12,03).$$

Пример 14.3. *Пусть* $\vec{X}\Subset E_\alpha$, $\alpha>0$. *Построить асимптотический доверительный интервал для параметра* α.

Решение. Так как для показательного распределения с параметром α математическое ожидание равняется $\mathbf{E}X_1=1/\alpha$, то согласно формуле 14.4

$$-A\le\frac{\sqrt{n}(\overline{X}-1/\alpha)}{S}<A.$$

Последовательно выразим

$$-\frac{AS}{\sqrt{n}}\le\overline{X}-\frac{1}{\alpha}<\frac{AS}{\sqrt{n}};$$

$$\overline{X} - \frac{AS}{\sqrt{n}} < \frac{1}{\alpha} \le \overline{X} + \frac{AS}{\sqrt{n}};$$

$$\frac{1}{\overline{X} - \dfrac{AS}{\sqrt{n}}} > \alpha \ge \frac{1}{\overline{X} + \dfrac{AS}{\sqrt{n}}}.$$

Итак, мы получили доверительный интервал $(\alpha_-;\ \alpha_+)$, где

$$\alpha_- = \frac{1}{\overline{X} + \dfrac{AS}{\sqrt{n}}}, \qquad \alpha_+ = \frac{1}{\overline{X} - \dfrac{AS}{\sqrt{n}}},$$

а константа A выбирается в соответствии с равенством (14.2).

§ 14.6. Задачи для самостоятельного решения

14.1 Пусть $\vec{X} \Subset N(\theta_1, \theta_2)$, $\theta_1 \in R$, $\theta_2 > 0$. Построить центральный доверительный интервал для параметра θ_1. Вычислить реализацию доверительного интервала с уровнем $\gamma = 0,95$, располагая данными: n=10, $\overline{X} = 2,7;\ S^2 = 4$. Сравнить с результатом примера 14.1.

14.2 Пусть $\vec{X} \Subset N(a, \theta)$, $\theta > 0$, a- известно. Построить точные доверительные интервалы (односторонний и центральный двухсторонний) на основе статистики $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2$. Вычислить реализации построенных интервалов с уровнем $\gamma = 0,9$, располагая данными: n=10, $S_1^2 = 4$. Сравнить с результатами примера 14.2.

14.3 Пусть $\vec{X} \Subset K(\theta)$, $\theta \in R$. Построить оптимальный точный доверительный интервал для параметра θ по одному наблюдению (n=1).

14.4 $\vec{X} \Subset B(p)$, $0 < p < 1$. Построить приближенные доверительные интервалы для параметра p на основе оценки $\theta^* = \overline{X}$:

а) при помощи неравенства Чебышева;

б) используя асимптотическую нормальность оценки.

14.5 $\vec{X} \Subset (\lambda)$, $\lambda > 0$. Построить асимптотический доверительный интервал для параметра λ с помощью оценки $\lambda^* = \overline{X}$.

14.6 Пусть $\vec{X} \Subset U(0, \theta)$, где $\theta > 0$. С помощью статистик $\overline{X}$ $\overline{X^2}$ построить асимптотические доверительные интервалы (соответственно (θ_1^-, θ_1^+) и (θ_2^-, θ_2^+)) уровня $1 - \varepsilon$ и показать, что случайный интервал (θ_2^-, θ_2^+) короче соответствующего (θ_1^-, θ_1^+).

Глава 15

Проверка статистических гипотез

§ 15.1. Статистические гипотезы

Пусть $\vec{X} = (X_1, X_2, ..., X_n)$ — выборка, $\vec{X} \Subset \mathbf{F}$, где $\mathbf{F}$ - полностью или частично неизвестное распределение отдельного наблюдения X_i.

Определение 15.1. *Статистической гипотезой* будем называть всякое утверждение о виде или свойствах неизвестного распределения $\mathbf{F}$.

Пример 15.1. *Пусть $\mathbf{F}$ — полностью неизвестное распределение. Примерами гипотез являются*

Н : $\mathbf{F} = \mathbf{F}_0$, *где* $\mathbf{F}_0$ *— полностью определенное распределение;*

Н : $\mathbf{F} \in \widehat{\mathbf{F}}_0$, *где* $\widehat{\mathbf{F}}_0$ *— множество распределений (например,* $\widehat{\mathbf{F}}_0 = N_{a,\sigma^2}$ *или* $\widehat{\mathbf{F}}_0 = B_p$*).*

В этих примерах наблюдения имеют распределения из некоторого одно- или двухпараметрического семейства. Но могут быть непараметрические множества, например, $\mathbf{F}_0 \in \{\mathbf{F} :\ \mathbf{E}X_i > 0\}$ *— класс распределений с положительными математическими ожиданиями.*

Пример 15.2. *Пусть $\mathbf{F}$ — частично известное распределение. Например,* $\mathbf{F} \in U_{[a;\ b]}$ *(наблюдения имеют равномерное распределение). В этом случае примеры гипотез:*

Н : $a = 0,\ b = 1$ *(распределение равномерное на* $[0;\ 1]$*);*

Н : $a = 0$ *(распределение равномерное на* $[0;\ b]$*);*

H : $a < b - 1$ *(распределение равномерное на отрезке длины более 1).*

Гипотеза называется *простой*, если она однозначно определяет распределение $\mathbf{F}$, в противном случае гипотеза называется *сложной*. В приведенных выше примерах простыми являются гипотезы:

H : $\mathbf{F} = \mathbf{F}_0$ и H : $a = 0,\ b = 1$ (последняя в случае, когда известно, что распределение равномерное на $[a;\ b]$).

Остальные гипотезы являются сложными.

Мы будем рассматривать ситуацию, когда гипотез всего две. Одну из них называют *основной*, а другую — *альтернативной*, обозначая соответственно H_0 и H_1.

§ 15.2. Статистические критерии

Определение 15.2. *Статистическим критерием* называют всякое правило, позволяющее на основании наблюдаемого выборочного вектора $\vec{X}$ принять одну из гипотез: основную или альтернативную.

При применении статистического критерия могут возникнуть ошибки двух родов. Ошибка нулевого рода состоит в том, что отвергается верная нулевая гипотеза. Ошибка первого рода — отвергается верная первая гипотеза. Вообще ошибка i-го рода состоит в том, что статистический критерий отвергает верную i-ю гипотезу.

принимаемая гипотеза	верна гипотеза H_0	верна гипотеза H_1
H_0	нет ошибки	ошибка 1-го рода
H_1	ошибка 0-го рода	нет ошибки

Критерий характеризуется вероятностями ошибок:

$$\alpha_0 = \mathbf{P}_{H_0}(H_0 \text{ отвергается}), \qquad \alpha_1 = \mathbf{P}_{H_1}(H_1 \text{ отвергается}).$$

Здесь нижний индекс у символа вероятности указывает, при выполнении какой гипотезы подсчитывается вероятность. Из всевозможных критериев надо выбирать такие, у которых вероятности ошибок по возможности малы. К сожалению, в невырожденной статистической задаче не существует критерия, для которого обе вероятности ошибок равны нулю. Как правило, чем меньше вероятность ошибки нулевого рода, тем больше вероятность ошибки первого рода.

Рассмотрим введенные понятия на следующем примере.

Пример 15.3. *Студенты группы А считают, что они играют в шахматы вдвое лучше, чем студенты группы В. В свою очередь, студенты группы В считают, что они играют в шахматы втрое лучше, чем студенты группы А. Для решения спора назначается шахматный матч между группами А и В. С каждой стороны участвуют 3 студента, выбираемые по жребию. Решено считать справедливым мнение группы, выигрывшей матч, то есть набравшей не менее 2 очков в 3 партиях. Предполагается, что ничьих нет. Найти, в чем состоят ошибки нулевого и первого рода. Вычислить вероятности этих ошибок.*

Решение. Предполагаем, что нулевая гипотеза (соответствующая мнению студентов группы А) состоит в том, что вероятность выигрыша каждого студента группы А у студента группы В вдвое больше вероятности проигрыша, то есть вероятность выигрыша равна $2/3$. Согласно первой гипотезе (мнению студентов группы В), вероятность выигрыша каждого студента группы А втрое меньше вероятности проигрыша, то есть равняется $1/4$.

Итак, проводятся три испытания схемы Бернулли с вероятностью успеха p, гипотеза $H_0: \ p = 2/3$; гипотеза $H_1: \ p = 1/4$.

Критерий (исход матча) предписывает принять гипотезу H_0, если число успехов в схеме Бернулли равняется двум или трем, а в противном случае принять гипотезу H_1.

Ошибка нулевого рода состоит в том, что критерий предписывает считать вероятность выигрыша студента первой группы равной $1/4$ в то время, как она равняется $2/3$. Ошибка первого рода описывает противоположную ситуацию: вероятность выигрыша студента первой группы равняется $1/4$, а критерий предписывает считать ее равной $2/3$.

Вычислим вероятности ошибок.

Вероятность ошибки нулевого рода α_0 — это вероятность отвергнуть верную нулевую гипотезу, то есть получить ноль или один успех в схеме Бернулли, которая предполагает 3 испытания с $p = 2/3$ в каждом. Вычислим эту вероятность на основании формулы Бернулли:

$$\alpha_0 = \mathbf{P}_{H_0}(H_0 \text{ отвергается}) =$$

$$= C_3^0(2/3)^0(1/3)^3 + C_3^1(2/3)^1(1/3)^2 = 1/27 + 2/9 \approx 0,25.$$

Вероятность ошибки первого рода α_1 — это вероятность получить два или три успеха в схеме Бернулли, которая предполагает 3 испытания с $p = 1/4$ в каждом.

$$\alpha_1 = \mathbf{P}_{H_1}(H_1 \text{ отвергается}) =$$

$$= C_3^2(1/4)^2(3/4)^1 + C_3^3(1/4)^3(3/4)^0 = 9/64 + 1/64 \approx 0,15.$$

§ 15.3. Критерии согласия

Удобно представлять статистический критерий как функцию $\delta(\vec{X})$ от выборочного вектора, принимающую два значения: H_0 и H_1. Наиболее общий подход для построения статистических критериев состоит в следующем.

Пусть $T = T(\vec{X})$ - некоторая статистика, характеризующая отклонение эмпирических данных, представленных выборкой, от теоретических, соответствующих проверяемой гипотезе H_0. Если распределение статистики $T(\vec{X})$ известно (точно или хотя бы приближенно), то для любого $\alpha > 0$ можно найти такое множество T_α значений T, для которого будет выполнено неравенство:

$$\mathbf{P}(T \in \mathrm{T}_\alpha / \mathrm{H}_0) \le \alpha. \tag{15.1}$$

Пусть $\alpha > 0$ настолько мало, что событие, имеющее вероятность, не превосходящую α, может считаться практически невозможным. Тогда статистический критерий можно задать следующим образом:

$$\delta(\vec{X}) = \begin{cases} \mathrm{H}_1, & \text{если } T(\vec{X}) \in \mathrm{T}_\alpha, \\ \mathrm{H}_0, & \text{если } T(\vec{X}) \notin \mathrm{T}_\alpha. \end{cases} \tag{15.2}$$

Это правило основано на здравом смысле: оно предписывает отвергнуть гипотезу H_0 (то есть принять H_1), если происходит событие $\{T(\vec{X}) \in \mathrm{T}_\alpha\}$, которое не должно произойти, будь гипотеза H_0 справедлива. Число $\alpha > 0$, которое фигурирует в (15.1) - (15.2), называется *уровнем критерия*, или *уровнем значимости*, статистика $T(\vec{X})$ называется *статистикой критерия*, а множество T_α - *критическим множеством.*

§ 15.4. Достигаемый уровень значимости

От статистики $T = T(\vec{X})$ требуют следующих свойств:

1) при выполнении гипотезы H_0 статистика T имеет известное распределение или, по крайней мере, сходится по распределению к некоторой случайной величине J с известным распределением;

2) при выполнении гипотезы H_1 статистика T сходится почти наверное к бесконечности с ростом объема выборки.

Для того, чтобы получить критерий уровня α, задают критическое множество в виде

$$\mathrm{T}_\alpha = \{T \geq C\},$$

где C — константа, определяемая условием

$$\mathbf{P}\{J \geq C\} = \alpha,$$

то есть $F_J(C) = 1-\alpha$. Ясно, что при таком выборе константы C вероятность ошибки нулевого рода α_0 либо равна уровню критерия α (в случае, когда статистика T при верной нулевой гипотезе распределена в точности как J), либо, по крайней мере, сходится к α с ростом объема выборки.

Сходимость статистики T почти наверное к бесконечности при выполненной первой гипотезе гарантирует *состоятельность* критерия, то есть сходимость вероятности ошибки первого рода α_1 к нулю с ростом объема выборки.

Для каждой конкретной выборки $\vec{X}$ можно найти предельное значение уровня $\alpha^* = \alpha^*(\vec{X})$, при котором гипотеза H_0 еще может быть принята. Такое значение называется *(реально) достигаемым уровнем значимости*. Как сказано в [10], α^* «имеет смысл вероятности получить худшее согласие с проверяемой гипотезой, чем реально полученное, если гипотеза H_0 верна». Поэтому чем меньше α^*, тем более это говорит против гипотезы H_0.

Достигаемый уровень значимости вычисляется с помощью распределения статистики J:

$$\alpha^* = \mathbf{P}\{J \geq T(\vec{X})\} = 1 - F_J(T(\vec{X})).$$

В терминах достигаемого уровня значимости критическая область имеет вид

$$\mathrm{T}_\alpha = \{\alpha^* \leq \alpha\},$$

то есть нулевая гипотеза отвергается на уровне α в случае, когда $\alpha^* \leq \alpha$.

Каждый критерий согласия использует свою статистику, предназначенную для различения нулевой гипотезы и альтернативы и обладающую нужными свойствами: сходимостью к фиксированному распределению при выполнении нулевой гипотезы и сходимостью почти наверное к бесконечности при ее невыполнении.

В качестве важных примеров критериев согласия рассмотрим критерии Колмогорова и хи-квадрат Пирсона.

§ 15.5. Критерии согласия Колмогорова и χ^2 Пирсона

Рассмотрим выборку $\vec{X} \Subset \mathrm{F}$ объема n с неизвестной функцией распределения F и простую гипотезу $\mathrm{H}_0 : \mathrm{F} = \mathrm{F}_0$. Альтернативной для H_0 является сложная гипотеза $\mathrm{H}_1 : \mathrm{F} \neq \mathrm{F}_0$.

Критерий Колмогорова применяется в случае, когда функция распределения $\mathrm{F}_0(t)$ непрерывна. Рассматривается следующее расстояние между эмпирической и теоретической функциями распределения:

$$D_n = D(F_n^*, F_0) = \sup_{-\infty<t<\infty} |F_n^*(t) - F_0(t)| = \max_{-\infty<t<\infty} |F_n^*(t) - F_0(t)|.$$

В качестве статистики критерия Колмогорова выбирается это расстояние, умноженное на $\sqrt{n}$, где n — объем выборки:

$$T_n = \sqrt{n} D_n = \sqrt{n} \max_{-\infty<t<\infty} |F_n^*(t) - F_0(t)|.$$

А. Н. Колмогоров доказал следующие свойства статистики T_n:

1) если гипотеза H_0 верна, то T_n с ростом n сходится к случайной величине J с функцией распределения, называемой функцией распределения Колмогорова:

$$F_J(t) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1} e^{-2k^2t^2};$$

2) если гипотеза H_0 неверна, то T_n сходится почти наверное к $+\infty$ при $n \to \infty$.

Таким образом, достигаемый уровень значимости критерия Колмогорова равен

$$\alpha^* = 1 - F_J(T_n) = 2\sum_{k=1}^{\infty}(-1)^{k+1} e^{-2k^2T_n^2} = 2\sum_{k=1}^{\infty}(-1)^{k+1} e^{-2k^2 n D_n^2}. \qquad (15.3)$$

Отметим, что для расчетов по этой формуле нужно брать не всю бесконечную сумму, а только несколько слагаемых, при этом ошибка вычислений не превосходит последнего отброшенного слагаемого. Критерий Колмогорова отвергает гипотезу H_0 на уровне α, если $\alpha^* \le \alpha$.

Для практического вычисления статистики $D_n = D_n(\vec{X})$ можно использовать следующую формулу:

$$D_n(\vec{X}) = \max_{1 \le i \le n} \max \left(\left| F(X_{(i)}) - \frac{i}{n} \right|, \left| F(X_{(i)}) - \frac{i-1}{n} \right| \right). \tag{15.4}$$

Здесь $X_{(i)}$ — это элементы *вариационного ряда*, то есть для этих вычислений выборку следует предварительно *упорядочить по возрастанию*.

Если гипотетическая функция распределения $\mathrm{F}_0(x)$ не является непрерывной, то критерий Колмогорова неприменим. В этом случае можно воспользоваться χ^2*-критерием Пирсона*. Статистика критерия Пирсона строится после предварительного «группирования» выборочных данных. Для этого все множество S возможных значений случайных величин X_i разбивается на конечное число непересекающихся частей:

$$S = S_1 \cup S_2 \cup \dots \cup S_r, \quad S_i \cap S_j = \emptyset, i \ne j.$$

Обозначим ν_j — число элементов выборки $\vec{X}$, попавших в множество S_j, а p_j — вероятность попадания случайной величины X_i в множество S_j, вычисленная с помощью гипотетической функции распределения $\mathrm{F} = \mathrm{F}_0$. Тогда в качестве статистики критерия χ^2 рассматривают следующую предложенную Пирсоном меру отклонения эмпирического распределения от предполагаемого теоретического:

$$\chi^2(\vec{X}) = \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j}. \tag{15.5}$$

Справедлива следующая теорема, позволяющая находить распределение статистики χ^2 при больших значениях n, а стало быть, и строить статистический критерий.

Теорема 15.1. *Если гипотеза* H_0 *однозначно фиксирует вероятности* $p_1, p_2, \dots, p_r$, *где* $p_j = \mathbf{P}(X_i \in S_j)$, *то при выполнении этой гипотезы статистика* $\chi^2(\vec{X})$ *слабо сходится к распределению* χ^2_{r-1}*:*

$$\chi^2 \Longrightarrow \chi^2_{r-1}, \quad n \to \infty.$$

При невыполнении нулевой гипотезы статистика $\chi^2(\vec{X})$ *сходится почти наверное к* $+\infty$.

Для построения критерия, основанного на статистике χ^2, используем распределение χ^2_{r-1}, и по найденному значению $\chi^2(\vec{X})$ отыскиваем достигаемый уровень значимости

$$\alpha^* = 1 - F_{\chi^2_{r-1}}(\chi^2(\vec{X}))$$

по таблице 5 распределения хи-квадрат или с помощью математических пакетов. В пакете Microsoft Excel достигаемый уровень значимости вычисляется формулой

`=ХИ2РАСП(ячейка;r-1)` (15.6)

(в качестве ячейки надо подставить адрес ячейки, в которой вычислена статистика хи-квадрат, а $r-1$ — число степеней свободы).

Тогда критерий Пирсона имеет следующий вид:

$$\mathrm{H}_0 \Leftrightarrow \alpha^* > \alpha. \tag{15.7}$$

Заметим, что для практического применения рекомендуется разбиение производить таким образом, чтобы выполнялось условие $np_j \geq 10$. При нарушении этого условия нужно объединить соседние множества S_j. Вероятности p_j надо выбирать по возможности равными.

Критерий хи-квадрат часто используют для проверки сложных гипотез о принадлежности распределения к некоторому параметрическому семейству (например, к нормальному). При этом вместо известных вероятностей p_j подставляют их оценки p_j^*, полученные путем оценивания неизвестных парметров распределения. Важно понимать, что в этом случае предельное распределение статистики $\chi^2(\vec{X})$ уже не будет распределением χ^2_{r-1}, а будет близко к распределению χ^2_{r-1-s}, где s — число оцениваемых параметров ($s = 2$ для нормального распределения). Более точно, предельная функция распределения заключена между функциями распределения χ^2_{r-1-s} и χ^2_{r-1}.

Достигаемый уровень значимости α^* заключен между $1 - F_{\chi^2_{r-1}}(\chi^2(\vec{X}))$ и $1 - F_{\chi^2_{r-1-s}}(\chi^2(\vec{X}))$, где s — число оцениваемых параметров.

Для того, чтобы получить в точности распределение хи-квадрат с $r-1-s$ степенями свободы, следует оценивать неизвестные параметры методом максимального правдоподобия по *группированной* выборке, но это приводит, как правило, к сложным вычислительным процедурам.

§ 15.6. Решение типовых примеров

Пример 15.4. *Вариационный ряд выборки имеет вид* (1; 2; 3; 4; 5; 6; 7; 8; 9; 10). *Проверить гипотезу о равномерности распределения элементов выборки на отрезке от 0 до 10 с помощью критерия Колмогорова: найти реализацию достигаемого уровня значимости и сделать вывод о принятии гипотезы на уровнях 0,1 и 0,01.*

Решение. Построим на одном графике эмпирическую $F_n^*(t)$ и теоретическую $F_0(t)$ функции распределения.

Эмпирическая функция распределения — это ступенчатая функция, высота ступеньки равна $1/10$ в точках 1; $\ldots$; 10.

Теоретическая функция распределения равномерного закона на отрезке от 0 до 10 равна

$$\mathrm{F}_0(t) = \begin{cases} 0, & \text{если } t \leq 0, \\ t/10, & \text{если } 0 < t \leq 10, \\ 1, & \text{если } t > 10. \end{cases} \tag{15.8}$$

Так как функция распределения $\mathrm{F}_0(t)$ непрерывна, то можно применять критерий Колмогорова. Найдем по графику значение D_n — наибольшую по модулю разность между эмпирической и теоретической функциями распределения. Эта разность достигается в точках разрыва эмпирической функции распределения и равна $1/10$. Вычислим реализацию достигаемого уровня значимости, вспоминая, что $n = 10$: согласно (15.3),

$$\alpha^* = 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2nD_n^2} \approx$$

$$\approx 2e^{-0,2} - 2e^{-4\cdot 0,2} + 2e^{-9\cdot 0,2} - 2e^{-16\cdot 0,2} + 2e^{-25\cdot 0,2} - 2e^{-36\cdot 0,2} + 2e^{-49\cdot 0,2} \approx 0,99997.$$

Достигаемый уровень значимости оказался близким к 1; это означает, что нет оснований отвергать гипотезу о равномерности выборочных значений. Эту гипотезу следовало бы отвергнуть только в случае, когда достигаемый уровень значимости оказался бы близким к нулю.

В частности, в нашем случае выполнено неравенство $\alpha^* > 0,1$. Следовательно, гипотеза о равномерности принимается на уровне 0,1. Тем более она будет приниматься на уровне 0,01.

Пример 15.5. *Решить пример 15.4 для реализации выборки* (10; 0; 0; 10; 10; 10; 0; 0; 0; 10).

Решение. Упорядочив реализацию выборки по неубыванию, получим реализацию вариационого ряда: (0; 0; 0; 0; 0; 10; 10; 10; 10; 10). Как и в предыдущем примере, построим на одном графике эмпирическую $F_n^*(t)$ и теоретическую $F_0(t)$ функции распределения. В отличие от предыдущего примера, эмпирическая функция распределения здесь имеет всего две ступеньки в точках 0 и 10, высотой по $5/10 = 0,5$. Теоретическая функция распределения остается той же самой и определяется формулой (15.8). Значение D_n достигается в точках разрыва эмпирической функции распределения и равняется $0,5$. Вычислим реализацию достигаемого уровня значимости:

$$\alpha^* = 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2nD_n^2} \approx 2e^{-2\cdot 10\cdot 0,5^2} = 2e^{-5} \approx 0,0135.$$

Здесь мы взяли только одно слагаемое суммы, так как остальные слагаемые гораздо меньше.

В этом примере достигаемый уровень значимости оказался близким к 0, что говорит против гипотезы H_0. В частности, $\alpha^* < 0,1$, то есть гипотеза однородности отвергается на уровне 0,1. Однако она принимается на более низком уровне 0,01, так как $\alpha^* > 0,01$.

Пример 15.6. *Проверить гипотезу о равномерности на отрезке от 0 до 10 для выборок из двух предыдущих примеров с помощью критерия хи-квадрат Пирсона: найти реализации достигаемых уровней значимости и сделать выводы о принятии гипотезы на уровнях 0,1 и 0,01. Число промежутков группирования выбрать по формуле Стеджеса.*

Решение.

Согласно формуле Стеджеса (11.3), вычисляем целую часть логарифма по основанию 2 от объема выборки и прибавляем единицу:

$$r = [\log_2 n] + 1 = [\log_2 10] + 1 = 3 + 1 = 4,$$

так как $2^3 = 8 < 10 < 2^4 = 16$.

Итак, множество допустимых выборочных значений — отрезок [0; 10] — следует разбить на 4 промежутка равной длины:

$$S_1 = [0;\ 2,5),\quad S_2 = [2,5;\ 5),\quad S_3 = [5;\ 7,5),\quad S_4 = [7,5;\ 10].$$

Согласно нулевой гипотезе, распределение равномерное на отрезке от 0 до 10. Следовательно, равны вероятности попадания элемента выборки в отрезки равной длины:

$$p_1 = p_2 = p_3 = p_4 = 1/4 = 0,25.$$

Значения статистики хи-квадрат Пирсона различны для примеров 15.4 и 15.5:

1) В примере 15.4 количества элементов, попавших в каждый из промежутков, равны соответственно

$$\nu_1 = 2, \quad \nu_2 = 2, \quad \nu_3 = 3, \quad \nu_4 = 3.$$

Вычислим статистику хи-квадрат согласно формуле (15.5):

$$\chi^2(\vec{X}) = \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} =$$

$$= \frac{(2 - 10 \cdot 0,25)^2}{10 \cdot 0,25} + \frac{(2 - 10 \cdot 0,25)^2}{10 \cdot 0,25} + \frac{(3 - 10 \cdot 0,25)^2}{10 \cdot 0,25} + \frac{(3 - 10 \cdot 0,25)^2}{10 \cdot 0,25} = 0,4.$$

Найдем достигнутый уровень значимости по формуле (15.6), используя функцию ХИ2РАСП и подставляя значение 0,4 и число степеней свободы, равное $r - 1 = 4 - 1 = 3$:

$$\texttt{ХИ2РАСП(0,4;3)} \approx 0,94.$$

Итак, здесь достигнут уровень значимости 0,94, что не дает оснований отвергать гипотезу о равномерности ни на уровне $0,1 < 0,94$, ни тем более на уровне 0,01.

2) В примере 15.5 количества элементов, попавших в каждый из промежутков, принимают значения

$$\nu_1 = 5, \quad \nu_2 = 0, \quad \nu_3 = 0, \quad \nu_4 = 5.$$

Как и в пункте (1), вычислим статистику хи-квадрат и найдем достигнутый уровень значимости:

$$\chi^2(\vec{X}) = \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} =$$

$$= \frac{(5 - 10 \cdot 0,25)^2}{10 \cdot 0,25} + \frac{(0 - 10 \cdot 0,25)^2}{10 \cdot 0,25} + \frac{(5 - 10 \cdot 0,25)^2}{10 \cdot 0,25} + \frac{(0 - 10 \cdot 0,25)^2}{10 \cdot 0,25} = 10;$$

$$\texttt{ХИ2РАСП(10;3)} \approx 0,0186.$$

В этом примере достигнут низкий уровень значимости 0,0186, что дает основания отвергать гипотезу о равномерности на уровне $0,1 > 0,0186$, но не на уровне 0,01.

Отметим, что для рассмотренных примеров критерии Колмогорова и хи-квадрат Пирсона дают похожие результаты — достигнутые уровни значимости для обоих критериев оказались довольно близкими. В случае, когда основная гипотеза предполагает дискретное распределение, критерий Колмогорова неприменим, и мы будем пользоваться только критерием хи-квадрат Пирсона.

Пример 15.7. *При* 4040 *бросаниях монеты Бюффон получил* $\nu_1 = 2048$ *выпадений герба и* $\nu_2 = n - \nu_1 = 1992$ *выпадений решетки. Согласуется ли это с гипотезой о том, что монета правильная, при уровне значимости* $\alpha = 0,1$*? С каким предельным уровнем значимости может быть принята эта гипотеза?*

Решение. Можно считать, что мы имеем дело со статистической моделью $\vec{X} \Subset \mathrm{B}_p$, где неизвестен параметр p - вероятность выпадения герба. Проверяемая гипотеза $\mathrm{H}_0: \ p = 0,5$. Поскольку выборочные данные уже сгруппированы ($\nu_1 = 2048$ — число значений $X_i = 1$, ν_2 — число значений $X_i = 0$), то можем вычислить наблюдаемое значение статистики χ^2:

$$p_1 = \mathbf{P}_{H_0}(X_i = 1) = 0,5; \quad p_2 = \mathbf{P}_{H_0}(X_i = 0) = 0,5;$$

$$\frac{(\nu_1 - np_1)^2}{np_1} = \frac{(2048 - 2020)^2}{2020} = 0,285;$$

$$\frac{(\nu_2 - np_2)^2}{np_2} = \frac{(1992 - 2020)^2}{2020} = 0,388; \qquad \chi^2 = 0,285 + 0,388 = 0,673.$$

Число множеств разбиения $r = 2$, поэтому достигнутый уровень значимости

$$\texttt{ХИ2РАСП(0,673;1)} \approx 0,412.$$

Достигнутый уровень значимости довольно высок. В частности, $0,412 > 0,1$, то есть гипотеза о симметричности монеты принимается на уровне 0,1.

§ 15.7. Задачи для самостоятельного решения

15.1 При n=4000 независимых испытаний события $A1$, $A2$, $A3$, составляющие полную группу, осуществились соответственно 1905, 1015 и 1080 раз. Проверить, согласуются ли эти данные при уровне значимости 0,05 с гипотезой H_0: $p_1 = 1/2$, $p_2 = p_3 = 1/4$, где $p_j = \mathbf{P}(A_j)$. Найти достигнутый уровень значимости.

15.3 В экспериментах с селекцией гороха Мендель наблюдал частоты различных видов семян, полученных при скрещивании растений с круглыми желтыми семенами и растений с морщинистыми зелеными семенами. Эти данные и значения теоретических вероятностей по теории наследственности приведены в следующей таблице:

Семена	Частота	Вероятность
Круглые и желтые	315	9/16
Морщинистые и желтые	101	3/16
Круглые и зеленые	108	3/16
Морщинистые и зеленые	32	1/16
Σ	n=556	1

Следует проверить гипотезу H_0 о согласовании частотных данных с теоретическими вероятностями (на уровне значимости 0,1) и найти достигнутый уровень значимости.

15.3 В таблице приведены числа m_i участков равной площади 0,25 км^2 южной части Лондона, на каждый из которых приходилось по i попаданий самолетов-снарядов во время второй мировой войны. Проверить согласие опытных данных с законом распределения Пуассона, приняв за уровень значимости $\alpha = 0,05$:

i	0	1	2	3	4	5 и более	Итого
m_i	229	211	93	35	7	1	$\Sigma m_i = 576$

15.4 Крупная партия товаров может содержать долю дефектных изделий. Поставщик полагает, что эта доля составляет 3%, а покупатель — 10%. Условия поставки: если при проверке 20 случайным образом отобранных товаров обнаружено не более одного дефектного, то партия принимается на условиях поставщика, в противном случае — на условиях покупателя. Требуется определить: 1) каковы статистические гипотезы, статистика критерия, область ее значений, критическая область; 2) какое распределение имеет статистика критерия, в чем состоят ошибки первого и второго рода и каковы их вероятности.

15.5 Имеется выборка объема 1 из нормального распределения $\mathrm{N}_{a,1}$. Проверяются простые гипотезы $H_0 :\ a = 0,\ \ H_1 :\ a = 1$. Используется следующий критерий (при заданной постоянной c):

$$H_0 \Leftrightarrow X_1 \le c.$$

Вычислить, в зависимости от c, вероятности ошибок первого и второго рода.

15.6 Построить критерий, обладающий нулевыми вероятностями ошибок, для проверки гипотез $\mathrm{H}_0 : \vec{X} \Subset \mathrm{N}_{0,1}$ против $\mathrm{H}_1 : \vec{X} \Subset \Pi_\lambda$.

15.7 Пусть$\vec{X} \Subset \mathrm{N}_{a,1}$. Для проверки гипотез $\mathrm{H}_0 : a = 0$ против $\mathrm{H}_1 : a = 1$ используется следующий критерий: H_0 принимается, если $X_{(n)} < 3$, и отвергается в противном случае. Найти вероятности ошибок.

15.8 Используя конструкции доверительного интервала, построить критерий уровня ε для проверки гипотезы $\mathrm{H} : \theta = 1$, если а) $\vec{X} \Subset \mathrm{N}_{\theta,1}$; б) $\vec{X} \Subset \mathrm{N}_{1,\theta}$; в) $\vec{X} \Subset \mathrm{E}_\theta$; г) $\vec{X} \Subset \mathrm{B}_{\theta/2}$; д) $\vec{X} \Subset \Pi_\theta$.

Глава 16

Регрессионный анализ

§ 16.1. Линейная регрессия

Наблюдается случайная величина y, значения которой зависят от случайного вектора **факторов регрессии** $\overrightarrow{x} = (1, x_1, ..., x_k)$. Введем в рассмотрение вектор неизвестных параметров регрессии $\overrightarrow{\beta} = (a_0, a_1, ..., a_k)$. Будем изучать линейную регрессию

$$\mathbf{E}(y|\overrightarrow{x}) = a_0 + \sum_{i=1}^{k} a_i x_i.$$

Пусть в t-м эксперименте факторы регрессии принимают заданные значения

$$\overrightarrow{x}^{(t)} = (1, x_{1,t}, ..., x_{k,t}),$$

где $t = 1, ..., N$. После $N \geq k + 1$ экспериментов получен набор **откликов** $(y_1, ..., y_N)$:

$$\overrightarrow{y} = X^T \overrightarrow{\beta} + \overrightarrow{u},$$

с матрицей плана $X(k + 1 \times N)$

$$\begin{pmatrix} 1 & \dots & 1 \\ x_{1,1} & \dots & x_{1,N} \\ x_{2,1} & \dots & x_{2,N} \\ \vdots & \ddots & \vdots \\ x_{k,1} & \dots & x_{k,N} \end{pmatrix}.$$

Матрицу X^T — будем называть матрицей линейного преобразования, вектор $\overrightarrow{u} = (u_1, ..., u_N)$ — будем называть вектором ошибок (случайных остатков).

Сформулируем следующую теорему.

Теорема 16.1 (Гаусса-Маркова)**.** *Пусть матрица X имеет ранг $k+1$ и вектор ошибок $\overrightarrow{u}$ – состоит из независимых гауссовских случайных величин с распределением Φ_{0,σ^2} с одной и той же дисперсией. Тогда оценка полученная по методу наименьших квадратов $\widehat{\beta}$, которая минимизирует функцию*

$$S(\overrightarrow{\beta}) = (\overrightarrow{y} - X^T\overrightarrow{\beta})^T(\overrightarrow{y} - X^T\overrightarrow{\beta})$$

имеет вид

$$\widehat{\beta} = (XX^T)^{-1}X\overrightarrow{y}.$$

Ковариационная матрица оценки $\widehat{\beta}$ вычисляется по формуле

$$\mathrm{Cov}(\widehat{\beta}, \widehat{\beta}|X) = \sigma_u^2(XX^T)^{-1},$$

где $\sigma_u^2 = \frac{1}{N-k}(\overrightarrow{y} - X^T\widehat{\beta})^T(\overrightarrow{y} - X^T\widehat{\beta}) = \frac{1}{N-k}\sum_{i=1}^N \widehat{u}_i^2$ — несмещенная оценка для дисперсии случайного остатка, через $\widehat{u}_i$ мы обозначили i-ую компоненту вектора $\overrightarrow{y} - X^T\widehat{\beta}$.

В простейшем варианте задачи о линейной регрессии даны пары точек $(x_1,\ y_1), \ldots, (x_n,\ y_n)$. Задача состоит в отыскании прямой $y = kx + b$, наилучшим образом приближающей эти точки в следующем смысле: значение суммы квадратов отклонений значений y_i от соответствующих значений $kx_i + b$ достигает минимума:

$$\sum_{i=1}^{n}(y_i - kx_i - b)^2 \to \min.$$

Решение задачи дается следующими формулами:

$$k = \frac{\overline{xy} - \overline{x}\ \overline{y}}{\overline{x^2} - (\overline{x})^2};$$

$$b = \overline{y} - k\overline{x}.$$

Вероятностная постановка задачи, приводящая к тому же ответу, состоит в следующем: случайные величины y_i имеют нормальное распределение с одной и той же дисперсией и с математическим ожиданием $Kx_i + B$. Тогда k и b — это оценки параметров K, B по методу максимального правдоподобия.

§ 16.2. Критерий Дарбина-Ватсона

Рассмотрим статистику Дарбина-Ватсона

$$DW = \frac{\sum_{i=2}^{N}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{N}\widehat{u}_i^2},$$

где через $\widehat{u}_i$ обозначается i-ая компоненту вектора $\overrightarrow{y} - X^T\widehat{\beta}$ (см. теорему 16.1). На основе этой статистики строится критерий для проверки предпосылки теоремы Гаусса-Маркова о некоррелированности случайных остатков в модели линейной регрессии, а именно, проверяется основная гипотеза

$$H_1 : \mathrm{Cov}(u_i, u_j) = 0, \quad j = i - 1.$$

Этот критерий реализуется в итоге следующих шагов.

Шаг 1. По количеству уравнений наблюдений N, а также количеству $k+1$ неизвестных параметров регрессии и ошибке первого рода критерия α следует выбрать две величины d_L и d_U (для этих величин составлены таблицы, скажем для $\alpha = 0,05$ и $k = 1$, $N = 10$ значения d_L и d_U будут равны соответственно $0,88$ и $1,32$).

Шаг 2. Областью принятия основной гипотезы служит промежуток $(d_U, 4 - d_U)$.

При попадании реализации статистики DW в множество $(0, d_l]$ принимается альтернативная гипотеза:

$$H_1 : \ \mathrm{Cov}(u_i, u_j) > 0, \ j = i - 1.$$

Если же реализация статистики DW попадает в множество $[4 - d_L, 4)$, то принимается альтернативная гипотеза:

$$H_1 : \ \mathrm{Cov}(u_i, u_j) < 0, \ j = i - 1.$$

§ 16.3. Обобщенный метод наименьших квадратов

Пусть вектор случайных ошибок $\overrightarrow{u} = (u_1, u_2, ..., u_N)$ в модели линейной регрессии является гауссовским с линейно независимыми компонентами. Ковариационную матрицу $\mathrm{Cov}(\overrightarrow{u}, \overrightarrow{u})$ всегда можно представить в виде $\mathrm{Cov}(\overrightarrow{u}, \overrightarrow{u}) = \sigma_0^2\Omega$, где Ω — симметричная невырожденная матрица, а σ_0^2 — произвольно выбранная положительная константа. Оценка неизвестных параметров регрессии $\widehat{\beta}$, доставляемая процедурой обобщенного метода наименьших квадратов, может быть вычислена в процессе решения следующей системы уравнений

$$X\Omega^{-1}X^T\widehat{\beta} = X\Omega^{-1}\overrightarrow{y}.$$

Далее мы рассмотрим две модели стационарных временных рядов, а именно: модель авторегрессии первого порядка (обозначение AR(1)) и модель скользящего среднего первого порядка (обозначение MA(1)).

§ 16.4. Модель авторегрессии первого порядка, AR(1)

Определим последовательность $\{u_t\}_{t=1,2,\dots}$ случайных величин следующим образом:

$$u_t = \rho u_{t-1} + \xi_t,$$

где $\{\xi_t\}_{t=1,2,\dots}$ — последовательность независимых гауссовских случайных величин с нулевым средним и дисперсией равной σ_ξ^2, ρ – некоторая константа, такая что $|\rho| < 1$, при этом $\mathbf{E}u_1 = 0$, $\mathbf{D}u_1 = \sigma_u^2$, $\mathrm{Cov}(u_{t-1}, \xi_t) = 0$ и $(1-\rho^2)\sigma_u^2 = \sigma_\xi^2$.

Мы будем говорить, что определенный таким образом временной ряд $\{u_t\}_{t=1,2,\dots}$ соответствует модели авторегрессии первого порядка и обозначать $u_t \in \mathrm{AR}(1)$.

Рассмотрим уровни временного ряда $\{u_t\}_{t=1,2,\dots}$ в два произвольных но фиксированных момента времени $t = i$ и $t = j$. Уровни u_i и u_j являются случайными величинами, определим для них коэффициент корреляции

$$\rho_{uu}(i,j) = \mathrm{Cor}(u_i, u_j).$$

Функция $\rho_{uu}(i,j)$ называется автокорреляционной функцией, заметим, что для стационарного временного ряда она зависит только от разности $i-j$, т. е. $\rho_{uu}(i,j) = f(|i-j|)$, где f – некоторая функция. В частности для $u_t \in \mathrm{AR}(1)$ автокорреляционная функция равна $\rho_{uu}(i,j) = \rho^{|i-j|}$.

§ 16.5. Модель скользящего среднего первого порядка, MA(1)

Определим последовательность $\{u_t\}_{t=1,2,\dots}$ случайных величин

$$u_t = \gamma\xi_{t-1} + \xi_t,$$

где $\{\xi_t\}_{t=0,1,...}$ — последовательность независимых гауссовских случайных величин с нулевым средним и дисперсией, равной σ_ξ^2, γ – некоторая константа. Будем говорить, что определенный таким образом временной ряд $\{u_t\}_{t=1,2,...}$ соответствует модели скользящего первого порядка и обозначать $u_t \in \mathrm{MA}(1)$. Автокорреляционная функция $\rho_{uu}(i,j)$ в этом случае равна

$$\rho_{uu}(i,j) = \begin{cases} \left(\frac{\gamma}{1+\gamma^2}\right)^{|i-j|}, & |i-j| \le 1; \\ 0, \quad |i-j| \ge 2. \end{cases}$$

§ 16.6. Оценивание линейных регрессионных моделей с зависимыми остатками

Рассмотрим регрессионную модель

$$\begin{cases} y_t = a_0 + a_1 x_{1,t} + ... + a_k x_{k,t} + u_t; \\ u_t \in \mathrm{AR}(1), \quad t = 1,2,...,N. \end{cases}$$

Обратим внимание, что y_t, $t = 1,2,...,N$ — это набор откликов в регрессионной модели, $a_1, a_2, ..., a_k$ — неизвестные параметры регрессии, u_t, $t = 1,2,...,N$ — набор случайных остатков. Поскольку $u_t \in \mathrm{AR}(1)$ мы для всех $t = 2,3,...,N$ можем выписать следующие соотношения:

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_{1,t} - \rho x_{1,t-1}) + ... + a_k(x_{k,t} - \rho x_{k,t-1}) + \xi_t, \quad (16.1)$$

где $\{\xi_t\}_{t=2,3,...,N}$ — последовательность независимых одинаково распределенных гауссовских случайных величин.

Если бы константа ρ была известна, то модель (16.1) была бы моделью множественной регрессии, в рамках которой все предпосылки теоремы 16.1 (теоремы Гаусса-Маркова) справедливы. Так что можно было бы получить методом наименьших квадратов оценки неизвестных параметров регрессии. Однако ρ — неизвестная константа, значения которой принадлежит промежутку $[0,1)$. Алгоритм ее определения нелинейным методом наименьших квадратов принадлежит Хилдрету и Лу и состоит из следующих шагов:

Шаг 1. Задаться в промежутке $[0,1)$ набором пробных значений $\rho = \rho_i$ по правилу

$$\rho_i = (i-1)/n, \quad i = 1,...,n,$$

где n — некоторое натуральное число.
Шаг 2. При каждом значении $\rho = \rho_i$ составить в рамках модели (16.1) систему уравнений

$$\begin{cases} y_2 - \rho y_1 = a_0(1-\rho) + a_1(x_{1,2} - \rho x_{1,1}) + ... + a_k(x_{k,2} - \rho x_{k,1}) + \xi_2; \\ ... \\ y_N - \rho y_{N-1} = a_0(1-\rho) + a_1(x_{1,N} - \rho x_{1,N-1}) + ...+ \\ \qquad\qquad\qquad\qquad +a_k(x_{k,N} - \rho x_{k,N-1}) + \xi_N \end{cases}$$

и вычислить на основании этой системы оценки неизвестных параметров с помощью метода наименьших квадратов, а также оценки случайных остатков регрессии $\widehat{\xi}_j,\ j = 2, ..., N$. Далее следует вычислить сумму квадратов оценок случайных остатков

$$\sum_{j=2}^{N} \widehat{\xi}_j^2.$$

Шаг 3. Выбрать из множества пробных значений ρ_i такую величину $\widehat{\rho}$ и соответствующие ей оценки неизвестных параметров регрессии, для которой достигается минимум $\sum_{j=2}^{N} \widehat{\xi}_j^2$. Полученные оценки и будут искомыми оценками параметра ρ в модели авторегрессии, а также неизвестных параметров регрессии $a_0, a_1, ..., a_k$.

В предлагаемых задачах всюду предполагается, что $k = 1$.

§ 16.7. Задачи для самостоятельного решения

16.1 Методом наименьших квадратов найти неизвестные параметры регрессии.

$x_{1,t}$	1	2	3
y_t	5	-1	2

16.2 Методом наименьших квадратов найти неизвестные параметры регрессии.

$x_{1,t}$	1	2	3	4	5	6	7	8	9	10
y_t	5,4	10,4	14,6	16,8	20,07	26,13	30,8	35,3	41,5	43,5

16.3 Методом наименьших квадратов найти неизвестные параметры регрессии.

$x_{1,t}$	1	2	3
y_t	2	-1	2

Вычислить статистику Дарбина-Ватсона и сделать выводы, вычислить сумму квадратов остатков. Далее, предполагая, что случайные остатки подчиняются модели AR(1) с $\rho = 1/2$, вычислить также оценки параметров регрессии и сумму квадратов остатков, сравнить эту сумму квадратов с полученной ранее и сделать выводы.

16.4 Пусть последовательность случайных величин $\{u_t\}$ подчиняется модели авторегрессии первого порядка с коэффициентом корреляции между соседними уровнями, равным 1/3. Вычислить автокорреляционную функцию.

16.5 Последовательность случайных величин $\{u_t\}$ удовлетворяет модели скользящего среднего первого порядка с коэффициентом корреляции между соседними уровнями, равным 1/2. Вычислить автокорреляционную функцию.

16.6 Предполагая, что случайные остатки в модели регрессии подчиняются модели авторегрессии первого порядка, вычислить коэффициент корреляции между соседними уровнями случайных остатков (использовать алгоритм Хилдрета-Лу).

$x_{1,t}$	1	2	3	4
y_t	7,7	9,2	13	18,8

16.7 Предполагая, что случайные остатки в модели регрессии удовлетворяют модели скользящего среднего первого порядка, найти оценки неизвестных параметров регрессии.

$x_{1,t}$	1	2	3	4	6
y_t	5	−1	2	−1	3

16.8 Пусть реализация случайных остатков в модели линейной регрессии имеет вид:

$$0,27; 2,18; 0,18; -1,17; 0,68; 2,05; 0,45; -1,1; -1,74; -3,21.$$

Будем предполагать, что случайные остатки удовлетворяют модели $AR(1)$, т.е. $u_t = \rho u_{t-1} + \xi_i$, где $\{\xi_i\}$ — последовательность независимых одинаково распределенных нормальных случайных величин. Используя метод наименьших квадратов, найти оценку для ρ.

Глава 17

Марковские цепи и процессы

§ 17.1. Цепи Маркова. Эргодическая теорема

Пусть каждая из случайных величин $X_0, X_1, ...$ принимает конечное число значений $\{x_1, x_2, ..., x_m\}$, эти значения мы будем называть состояниями. Последовательность $X_0, X_1, ...$ называется цепью Маркова, если для любого момента времени n и любых состояний $x_{i_0}, ..., x_{i_n}$ выполняется

$$\mathbf{P}(X_n = x_{i_n} | X_{n-1} = x_{i_{n-1}}, ..., X_0 = x_{i_0}) = \mathbf{P}(X_n = x_{i_n} | X_{n-1} = x_{i_{n-1}}).$$

Цепь Маркова $\{X_n\}_{n \geq 0}$ называется однородной по времени с матрицей переходных вероятностей $M = (p_{ij})$, если для любого n и любых x_i и x_j

$$\mathbf{P}(X_n = x_j | X_{n-1} = x_i) = p_{ij}.$$

Обратим внимание, что для любого i

$$\sum_{j=1}^{m} p_{ij} = 1.$$

В дальнейшем будут рассматриваться только однородные цепи Маркова! Обозначим

$$p_{ij}(k) = \mathbf{P}(X(k) = x_j | X(0) = x_i)$$

вероятность перехода за k шагов из состояния x_i в состояние x_j и

$$p_j(k) = \mathbf{P}(X(k) = x_j)$$

— вероятность нахождения в состоянии x_j в момент времени k. Для дальнейшего изложения нам понадобится следующая матрица порядка m и вектор-строка размера m

$$M^{(k)} = (p_{ij}(k))$$

и

$$N^{(k)} = (p_i(k)).$$

Переходные вероятности удовлетворяют уравнениям Колмогорова-Чэпмена

$$p_{ij}(k+l) = \sum_{\alpha} p_{i\alpha}(k) p_{\alpha j}(l)$$

и

$$p_j(k+l) = \sum_{\alpha} p_{\alpha}(k) p_{\alpha j}(l).$$

В матричной форме эти уравнения можно переписать таким образом

$$M^{(k+l)} = M^{(k)} M^{(l)}$$

и

$$N^{(k+l)} = N^{(k)} M^{(l)}.$$

Следующая теорема описывает широкий класс марковских цепей, обладающих так называемым свойством эргодичности: пределы $\pi_j = \lim_{n\to\infty} p_{ij}(n)$ не только существуют и не зависят от i, но и образуют распределение вероятностей ($\pi_j \geq 0, \quad \sum_{j=1}^{m} \pi_j = 1$), но и таковы, что $\pi_j > 0$ для всех j.

Теорема 17.1 (эргодическая теорема)**.** *Пусть $M = (p_{ij})$ — матрица переходных вероятностей марковской цепи с конечным множеством состояний $\{x_1, ..., x_m\}$.*
Если найдется n_0 такое, что

$$\min_{i,j} p_{ij}(n_0) > 0,$$

то существуют числа $\pi_1, ..., \pi_m$ такие, что

$$\pi_j > 0, \qquad \sum_{j=1}^{m} \pi_j = 1$$

и для любого i

$$\pi_j = \lim_{n\to\infty} p_{ij}(n).$$

При этом числа $(\pi_1, ..., \pi_m)$ удовлетворяют системе уравнений

$$\pi_j = \sum_{\alpha} \pi_{\alpha} p_{\alpha j}, \quad j = 1, ..., m.$$

§ 17.2. Марковские процессы с непрерывным временем и конечным числом состояний

Поведение системы с возможными состояниями $S_0, S_1, ..., S_m$ может быть описано случайной функцией $X(t)$, принимающей значение k, если в момент t система находилась в состоянии S_k. Если переход системы из одного состояния в другое возможен в любой момент времени t, а вероятности $p_{ik}(t,\tau)$ перехода системы из состояния S_i в момент времени t в состояние S_k в момент времени τ, $\tau \geq t$ не зависят от поведения системы до момента времени t, то $X(t)$ является марковским случайным процессом с конечным числом состояний. Процесс называется однородным, если

$$p_{ik}(t,\tau) = p_{ik}(t-\tau, 0).$$

Вероятности $P_k(t)$ нахождения системы в состоянии S_k в момент времени t определяются из системы уравнений

$$\frac{dP_k(t)}{dt} = \sum_{j=0}^{m} \lambda_{jk}(t) P_j(t), \quad k = 0, 1, 2, ..., m,$$

где $\lambda_{jk}(t)$ — интенсивность перехода в момент t из состояния S_j в S_k. В случае однородных марковских процессов интенсивности перехода из одного состояния в другое не зависят от времени, т.е. для всех t выполняется равенство

$$\lambda_{jk}(t) = \lambda_{jk}.$$

Если начальное состояние S_i задано, то

$$P_k(0) = \begin{cases} 0, & i \neq k \\ 1, & i = k. \end{cases}$$

При однородном марковском процессе с сообщающимися состояниями существуют независимые от номеров исходного состояния пределы

$$\lim_{t\to\infty} p_{ik}(t) = \lim_{t\to\infty} P_k(t) = P_k, \quad k = 0, 1, 2, ..., m.$$

Предельные вероятности P_k находятся из системы алгебраических уравнений

$$\begin{cases} \sum_{j=0}^{m} \lambda_{jk} P_j = 0, & k = 0, 1, 2, ..., m; \\ \sum_{j=0}^{m} P_j = 1. \end{cases}$$

§ 17.3. Процессы размножения и гибели

Однородный марковский процесс с бесконечным, но счетным числом состояний $S_0, S_1, ...$ называется процессом размножения (рождения) и гибели, если только для соседних состояний вероятности перехода положительны, то есть $\lambda_{ij} = 0$ во всех случаях, когда j не равняется ни $i+1$, ни $i-1$. Обозначим $\lambda_i = \lambda_{ii+1}$, $\mu_j = \lambda_{jj-1}$. Процессы размножения и гибели могут «убегать на бесконечность» — в этом случае $X(t)$ сходится с вероятностью единица к бесконечности при $t \to \infty$. Такое поведение процесса называется неэргодическим. Напротив, для эргодического процесса выполнено

$$\lim_{t\to\infty} p_{ik}(t) = \lim_{t\to\infty} P_k(t) = P_k, \quad k = 0, 1, 2, ...,$$

и $\sum_{n=0}^{\infty} P_n = 1$.

Условие эргодичности вместе с формулой для вычисления стационарных вероятностей содержится в следующей теореме.

Теорема 17.2. *Процесс размножения и гибели эргодичен в том и только том случае, когда сумма ряда*

$$\gamma = 1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \ldots + \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n} + \ldots$$

конечна. Стационарные вероятности равны

$$P_0 = \frac{1}{\gamma}, \quad P_n = \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n} P_0, \quad n = 1,\ 2, \ldots.$$

§ 17.4. Задачи для самостоятельного решения

17.1 Найти матрицу переходных вероятностей для марковской цепи, описывающей следующий эксперимент. Пусть бросают монету, причем вероятность выпадения решки равна p. Определим X_n как разность между числом выпадений решки и числом выпадения герба после n бросаний монеты.

17.2 Найти матрицу переходных вероятностей для марковской цепи, описывающей следующий эксперимент. Бросают игральную кость. Положим X_n равным наибольшему из чисел, выпадающих в первых n бросаниях.

17.3 Устройство S состоит из двух узлов, каждый из которых в ходе работы устройства может выйти из строя. Возможны состояния системы: S_1 – оба узла работают; S_2 – первый узел отказал, второй работает; S_3 –

второй узел отказал, первый работает; S_4 – оба узла отказали. Построить граф состояний и переходов системы. Выбрать интенсивности переходов между состояниями системы и записать соответствующую систему дифференциальных уравнений.

17.4 По двум линиям связи в один пункт поступают два независимых простейших потока телеграмм. Найти вероятность того. что за время t в пункт приема придет n телеграмм, если параметры составляющих потоков равны λ_1 и λ_2.

17.5 Частицы, вылетающие из радиоактивного вещества в процессе его распада, образуют простейший поток с параметром λ. Каждая частица независимо от другой регистрируется счетчиком с вероятностью p. Определить вероятность того, что за время t будет зарегистрировано ровно n частиц.

17.6 Пусть матрица переходов для цепи Маркова имеет вид:

а) $\begin{pmatrix} 0 & 1 \\ 0,25 & 0,75 \end{pmatrix}$, б) $\begin{pmatrix} 0 & 1 \\ 0,85 & 0,15 \end{pmatrix}$, в) $\begin{pmatrix} 1 & 0 \\ 0,25 & 0,75 \end{pmatrix}$,
г) $\begin{pmatrix} 0 & 1/3 & 2/3 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$.

Проверить выполнение эргодической теоремы для этой цепи. Если эргодическая теорема выполняется найти соответствующее стационарное распределение. Нарисовать граф состояний и переходов цепи Маркова.

17.7 Двое рабочих обслуживают три автоматических станка, которые при нормальной работе не требуют их вмешательства. Каждый станок останавливается вследствие неполадок с интенсивностью один станок в 30 мин. Среднее время, которое тратится рабочим на ремонт одного станка, составляет 15 минут. Найти предельную вероятность того, что два станка не работают.

17.8 Поток поступления неисправной аппаратуры в мастерскую гарантийного ремонта является простейшим с параметром $\lambda = 10$ ед./час. Продолжительность ремонта является случайной величиной, имеющей показательное распределение с параметром $\mu = 5$ ед./час. Определить среднее время, проходящее от момента прихода неисправной аппаратуры до начала ремонта, если в мастерской четверо рабочих, каждый из которых ремонтирует только один прибор.

17.9 На коммутатор, имеющий четыре внешние линии связи, поступает в среднем за час 60 требований на связь. Продолжительность переговоров в среднем составляет 4 мин. Определить абсолютную пропускную способность системы; среднее число занятых линий; вероятность отказа абоненту.

Глава 18

Типовой расчет

Вариант 1

1. Наугад выбраны два числа. События A и B соответственно означают, что выбрано хотя бы одно нечетное число и хотя бы одно четное число. Что означают события AB и $A \cup B$?

2. Из корзины с пятью красными, четырьмя зелеными и тремя желтыми яблоками берутся (без возвращения) наудачу три яблока. С какой вероятностью среди этих трех яблок: а) ровно два зеленых, б) хотя бы одно красное и хотя бы одно желтое яблоко?

3. Молодой человек договорился встретиться с девушкой между 9 и 10 часами и обещал ждать её до 10 часов. Девушка обещала ждать его 10 минут, если придет раньше. Найти вероятность того, что они встретятся. Предполагается, что моменты их прихода равновероятны в течение часа.

4. При передаче текста в среднем 5 % букв искажается и принимается неверно. Передано слово из 6 букв. Какова вероятность того, что все буквы слова будут приняты правильно? Предполагается, что буквы искажаются независимо друг от друга.

5. В тире имеется 6 одинаковых на вид ружей. Вероятность попадания в мишень для двух из них по 0,9, для трех по 0,8 и для одного 0,3. Какова вероятность того, что стрелок попадет в мишень, если он выбирает ружье наудачу? Какова вероятность того, что было выбрано ружье, для которого вероятность попадания 0,3, при условии, что стрелок попал в мишень?

6. Вероятность попадания в мишень равна 0,6 при каждом выстреле. Стрельба ведется одиночными выстрелами до первого попадания, пока не будет израсходован боезапас. Найти ряд распределения, математическое ожидание и дисперсию числа произведенных выстрелов, если боезапас составляет 3 единицы. Построить график функции распределения.

7. Случайная величина ξ имеет треугольное распределение. Плотность распределения равна

$$f(x) = \begin{cases} Ax & \text{при } 0 \le x \le \theta; \\ 0 & \text{при } x \notin [0;\ \theta]. \end{cases}$$

Найти коэффициент A, математическое ожидание и стандартное отклонение. Найти вероятность того, что $\xi > \theta/2$. Начертить графики плотности распределения и функции распределения.

8. Составить таблицу совместного распределения числа выпавших единиц и числа выпавших шестерок при одном подбрасывании игральной кости. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y имеет вид: $f(x, y) = c$, если $x, y \ \geq 0$, $x + 3y \leq 3$ и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Участник лотереи бросает игральную кость 10 раз. Участник получает ценный приз, если сумма очков больше 50. Оценить вероятность получения ценного приза.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 2$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \theta x^{-(\theta+1)} & \text{при } x > 1; \\ 0 & \text{при } x \le 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

0,78	1,26	1,58	2,11	0,01	1,35	2,05	0,76	1,65	1,61	0,12	2,03	1,07	1,10
3,06	0,38	0,64	1,63	0,54	2,65	0,82	1,21	0,73	1,99	2,44	0,93	0,47	0,88

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне доверия 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,46	0,68	0,59	1,97	1,03	0,62	0,89	1,93	0,88
1,66	1,34	1,99	0,59	0,00	0,46	1,48	1,35	1,74

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4
Y	3	-1	2	-2

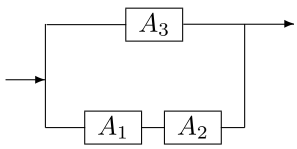
Вариант 2

1. Из таблицы случайных чисел наудачу взято одно число. Событие A — выбранное число делится на 5; событие B — данное число оканчивается нулем. Что означают события $A \setminus B$ и $A\overline{B}$?

2. В квадрат с вершинами $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$ наудачу брошена точка. Пусть (X,Y) — ее координаты. Найти: $\mathbf{P}(\max\{X+3Y, Y\} \le 1/2)$.

3. Бросают 3 игральные кости. Какова вероятность того, что на них выпадет разное число очков?

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя каждого элемента A_k равна 0,02. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Одинаковые детали поступают на сборку с трех автоматов. Первый автомат дает 20 %, второй 30 %, третий 50 % всех деталей, необходимых для сборки. Брак в продукции первого автомата составляет 2,5 %, второго — 2 %, третьего — 2,5 %. Найти вероятность поступления на сборку бракованной детали. Найти вероятность того, что оказавшаяся бракованной деталь изготовлена на первом автомате.

6. По мишени одновременно стреляют три стрелка, вероятности попаданий которых равны соответственно 0,55, 0,6 и 0,65. Найти ряд распределения, математическое ожидание и дисперсию числа попаданий в мишень. Построить график функции распределения.

7. Распределение Парето приближенно описывает распределение доходов физических лиц. Плотность распределения равна

$$f(x) = \begin{cases} \frac{A}{x^{\alpha}} & \text{при } x \geq \theta; \\ 0 & \text{при } x < \theta. \end{cases}$$

Здесь $\alpha > 2$, $\theta > 0$ — параметры распределения, A — нормирующая константа. Найти константу A. Вычислить значение параметра α, при котором математическое ожидание превосходит значение параметра θ в 3 раза.

8. Подбрасываются три симметричных монеты. Составить таблицу совместного распределения количеств выпавших гербов на трех монетах и на первых двух монетах. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c$, если $x, y \geq 0$, $x + 2y \leq 2$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Время ожидания поезда метро за одну поездку имеет равномерное распределение на отрезке от 0 до 5 минут. Оценить вероятность того, что суммарное время ожидания за 30 поездок окажется меньше 1,5 часов.

11. Для выборки $(X_1, X_2, \ldots, X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по второму моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{2x}{\theta} e^{-x^2/\theta} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.
2,05 0,80 0,32 3,31 1,12 3,29 3,87 2,65 2,01 2,65 1,19 -0,85 4,07 1,23
3,38 5,17 1,51 2,20 5,41 1,22 1,89 2,02 3,17 -1,02 2,73 1,10 3,87

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,24 1,25 0,87 0,54 0,48 1,20 1,79 0,62 0,75 0,55
0,46 1,02 1,71 1,91 0,83 0,99 1,46 1,09 0,94

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	3	5	7
Y	3	-1	2	-2

Вариант 3

1. Событие A — хотя бы одно из имеющихся четырех изделий бракованное, событие B — бракованных изделий среди них не менее двух. Что означают противоположные события $\overline{A}$ и $\overline{B}$?

2. В квадрат с вершинами $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$ наудачу брошена точка. Пусть (X,Y) — ее координаты. Найти: $\mathbf{P}(\max\{2X,Y\} < 1/3)$?

3. В студенческой группе 15 юношей и 10 девушек. На университетский праздничный бал группа получила только 2 пригласительных билета, которые разыгрываются по жребию. Какова вероятность того, что на бал попадет разнополая пара?

4. Рабочий обслуживает 4 станка, работающих независимо друг от друга. Вероятность того, что в течение часа станок не потребует внимания рабочего, равна для первого станка 0,6, для второго — 0,8, для третьего — 0,9, для четвертого — 0,7. Найти вероятность того, что хотя бы один из станков в течение часа не потребует внимания рабочего.

5. Станок обрабатывает 2 вида деталей A и B, причем время работы распределяется между ними в соотношении 1:4. При обработке детали вида A он работает с максимальной для него нагрузкой в течение 70 % времени, при обработке детали вида B — 50 % времени. В случайный момент времени станок работал с максимальной нагрузкой. Определить вероятность того, что в это время он обрабатывал деталь вида A; вида B.

6. Вероятность попадания баскетбольного мяча в кольцо при бросании начинающим спортсменом равна $1/4$. Мяч бросают до первого попадания, но дают не более 4 попыток. Найти ряд распределения, математическое ожидание и дисперсию числа промахов. Построить график функции распределения.

7. Скорость пешехода на дистанции в 1 км является случайной величиной, равномерно распределенной на отрезке от 2 км/ч до 4 км/ч. Найти математическое ожидание и стандартное отклонение времени, затрачен-

ного на преодоление дистанции. Найти вероятность того, что это время превысит 24 минуты.

8. В группе из 25 студентов только двое изучали в школе модальную логику, и именно они получили оценку «5» на экзамене. Из остальных студентов 10 человек получили оценку «4», 10 человек — оценку «3», и 3 студента получили «двойки». Составить таблицу совместного распределения оценки на экзамене и индикатора изучения модальной логики для выбранного наудачу студента. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x,y) = c(x^2 + y)$, если $0 \le x \le 1$, $0 \le y \le 1$ и 0 иначе. Найти константу c и $\rho(X,Y)$.

10. Число опечаток на странице книги имеет распределение Пуассона с параметром 2. Найти пределы, в которых с вероятностью 0,9 лежит число опечаток в книге из 400 страниц.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по третьему моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{3x^2}{\theta} e^{-x^3/\theta} & \text{при } x > 0; \\ 0 & \text{при } x \le 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

7,61 3,33 7,35 3,05 2,54 1,91 1,77 2,92 5,95 2,31 0,27 5,12 6,60 -1,58
5,42 5,67 6,28 -0,09 2,74 2,45 1,11 6,97 -1,59 -1,41 2,69 4,99 7,24 1,75
4,76

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,49 1,42 0,61 2,00 0,59 0,58 1,18 1,58 1,01 0,57
0,05 0,25 0,17 1,30 0,52 0,91 0,84 1,66 1,24

14. Оценить параметры нормальной регрессии Y на X по двумерной

выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4
Y	-3	-1	2	-2

Вариант 4

1. Монета подбрасывается три раза подряд. Построить пространство элементарных исходов Ω. Описать событие A, состоящее в том, что выпало не менее двух гербов.

2. В компании из трех человек решили сделать друг другу подарки, для чего каждый принес подарок. Все подарки сложили вместе, перемешали и случайно распределили среди участников. Найти вероятность, что хотя бы один подарок вернется к своему владельцу.

3. На линейке длиной 20 см случайно сделаны две насечки. Какова вероятность того, что первая окажется дальше от начала не менее, чем на 5 см по сравнению со второй?

4. Детали проходят три операции обработки. На каждой из операций может возникнуть брак независимо от остальных операций с вероятностями 0,02, 0,03 и 0,035 соответственно. Найти вероятность получения небракованной детали.

5. Вероятность того, что изделие удовлетворяет стандарту, равна 0,95. На заводе принята система из трех независимых испытаний, каждое из которых изделие, удовлетворяющее стандарту, проходит с вероятностью 0,8, а неудовлетворяющее — с вероятностью 0,3. Какова вероятность того, что наудачу взятое изделие выдержит испытания? Какова вероятность того, что изделие, выдержавшее испытания, удовлетворяет стандарту?

6. Вероятность изготовления нестандартного изделия при налаженном технологическом процессе постоянна и равна $1/5$. Для проверки изделий отдел технического контроля берет из партии изделия одно за другим, но не более 3 изделий. При обнаружении нестандартного изделия вся партия задерживается. Найти ряд распределения, математическое ожидание и дисперсию числа изделий, проверяемых в каждой партии. Построить график функции распределения.

7. Мощность W, выделяемая на сопротивлении R, вычисляется по закону $W = RI^2$, где I — сила тока. Предполагается, что сила тока распределена равномерно на отрезке от 1 до 2 ампер. Найти плотность распределе-

ния и математическое ожидание мощности, выделяемой на сопротивлении в 1000 Ом. Найти вероятность того, что мощность превысит 2 кВт.

8. В научном отделе 3 лаборатории. В первой лаборатории 4 сотрудника и 2 исследовательских проекта, во второй 6 сотрудников и 1 проект, в третьей — 3 сотрудника и 2 проекта. Найти совместное распределение числа сотрудников и числа проектов в выбранной наудачу лаборатории. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x,y) = c(x+y^2)$, если $0 \le x \le 1$, $0 \le y \le 1$ и 0 иначе. Найти константу c и $\rho(X,Y)$.

10. Маршрут разбит на 900 участков. Погрешность измерений длины каждого из них распределена по нормальному закону с нулевым средним и стандартным отклонением 5 метров. Найти, в каких пределах лежит суммарная погрешность с вероятностью 0,95.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{2\theta\sqrt{x}} e^{-\sqrt{x}/\theta} & \text{при } x > 0; \\ 0 & \text{при } x \le 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

-1,70 -0,72 -4,46 -3,24 2,42 -1,70 -1,24 -0,07 6,20 2,67 1,80 0,26 9,61 2,51 1,44 -3,65 5,50 4,17 -2,06 7,48 2,60 7,61 2,54 9,77 9,67 7,36 7,86 11,22 3,38

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,60 1,44 0,70 1,33 0,74 0,61 1,03 1,25 0,85
0,81 1,04 0,76 0,80 1,55 1,61 0,82 1,70 1,63

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую ли-

нейной регрессии.

X	2	4	6	8
Y	3	-1	2	-2

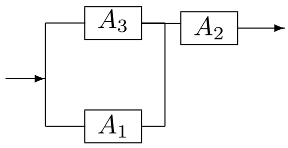
Вариант 5

1. Игральная кость подбрасывается два раза подряд. Описать пространство элементарных исходов Ω. Описать событие A, состоящее в том, что хотя бы один раз выпала единица, событие B, состоящее в том, что сумма очков при первом и втором подбрасывании нечетна.

2. В шахматном турнире участвуют 10 человек, которые разбиваются на пары по жребию. Какова вероятность того, что два самых сильных шахматиста попадут в одну пару?

3. В круг единичного радиуса наудачу брошены пять точек. С какой вероятностью расстояние от границы круга до ближайшей точки окажется не меньше $1/3$?

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_k равна 0,1. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь не будет пропускать ток.

5. На сборку поступают детали с двух автоматов. Первый автомат дает 80 %, а второй 20 % всех деталей, необходимых для сборки. Брак в продукции первого автомата составляет 1 %, а второго — 4 %. Деталь, изготовленная автоматом, оказалась бракованной. Какова вероятность того, что она изготовлена на первом автомате?

6. Вероятность отказа сервера при каждом из независимых подключений с помощью модема равна 0,3. Попытки подключения производятся до установления связи. Найти ряд распределения, математическое ожидание и дисперсию числа произведенных попыток подключения, если число попыток ограничено четырьмя. Построить график функции распределения.

7. Закон Рэлея с плотностью распределения

$$f(x) = \begin{cases} Axe^{-x^2/(2\sigma^2)} & \text{при } x \geq 0; \\ 0 & \text{при } x < 0 \end{cases}$$

в ряде случаев описывает распределение срока службы электронной аппаратуры. Найти коэффициент A, математическое ожидание и дисперсию. (Рекомендуется использовать таблицы определенных интегралов).

8. На 4 карточках написаны цифры от 1 до 4. Найти совместное распределение числа, написанного на выбранной наудачу карточке, и индикатора того, что это число четное. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y имеет вид: $f(x, y) = c$, если $x, y \ \geq 0$, $x + 4y \leq 3$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество воды, расходуемое жителями одной квартиры в сутки, имеет показательное распределение со средним значением 100 литров. Найти, какого количества воды достаточно с вероятностью 0,98 для удовлетворения потребностей жильцов 250000 квартир.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

4,83 -1,10 13,11 10,84 9,45 8,56 7,87 7,34 -4,06 3,48 4,70 7,13 -1,08
4,53 13,56 2,66 7,29 9,41 11,86 9,54 10,86 2,50 -2,84 11,21 8,93

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,32 0,49 1,12 1,98 0,25 1,52 0,52 0,03 1,10
1,59 0,27 1,30 1,79 1,93 0,23 1,84 1,04

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4
Y	1	-1	2	-2

Вариант 6

1. Пусть A, B, C — произвольные события. найти выражения для событий, состоящих в том, что из A, B и C произошло хотя бы два события.

2. Шесть книг на полке расставлены случайным образом. Найти вероятность того, что две определенные книги окажутся рядом (в любом порядке).

3. Два лица A и B имеют одинаковую вероятность прийти к указанному месту в любой момент времени между 12 и 13 часами. Лицо A ждет другого в течение 10 минут после чего уходит; лицо B ждет другого в течение 15 минут. Найти вероятность того, что A и B встретятся.

4. Вероятность попадания в мишень при одном выстреле равна 0,7. По мишени стреляют одиночными выстрелами до первого попадания, после чего стрельбу прекращают. Найти вероятность того, что будет сделано не более трех выстрелов.

5. Студент выучил к экзамену только 20 вопросов из 30. Для сдачи экзамена достаточно ответить на два из трех разных вопросов. Какова вероятность того, что экзамен будет сдан? Какова вероятность того, что студент ответил на все три вопроса, если известно, что он сдал экзамен?

6. Пользователь компьютера забыл пароль и перебирает наудачу 4 возможных. После трех неудачных попыток компьютер блокируется. Найти ряд распределения, математическое ожидание и дисперсию числа попыток. Построить график функции распределения.

7. Скорость V молекул газа имеет плотность распределения

$$f(v) = \begin{cases} \frac{v^2}{\sigma^3}\sqrt{\frac{2}{\pi}}e^{-v^2/(2\sigma^2)} & \text{при } v \geq 0; \\ 0 & \text{при } v < 0 \end{cases}$$

(распределение Максвелла). Определить математическое ожидание V. (Можно использовать таблицы определенных интегралов).

8. В двух из четырех комнат температура 20 градусов, а влажность 80 процентов. В третьей комнате температура 25 градусов, а влажность 90 процентов. В четвертой комнате температура 20 градусов, а влажность 90 процентов. Найти совместное распределение температуры и влажности в выбранной наудачу комнате. Найти коэффициент корреляции между температурой и влажностью.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = cx(x + y)$, если $0 \le x \le 1$, $0 \le y \le 1$ и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Участник лотереи бросает 6 шаров, каждый из которых может попасть в лузы с номерами от 1 до 6. Участник получает ценный приз, если сумма очков меньше 12. Оценить вероятность получения ценного приза.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по третьему моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{3\sqrt{x}}{2\theta} e^{-x\sqrt{x}/\theta} & \text{при } x > 0; \\ 0 & \text{при } x \le 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.
5,16 6,70 2,88 9,09 -2,06 6,25 6,46 4,25 16,16 7,07 1,35 13,58 7,96 14,64
-2,14 10,81 2,50 2,24 -1,04 5,31 11,93 16,20 7,49 -5,21 5,90 5,63 7,26

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,46 0,79 0,64 1,06 0,42 0,69 1,65 0,45 0,43
1,48 0,44 0,97 1,49 0,46 1,29 0,37 0,45

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4
Y	6	-1	2	-8

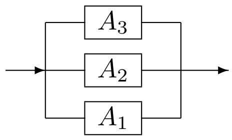
Вариант 7

1. Рабочий изготовил три детали. Пусть событие A_i состоит в том, что i-ая изготовленная им деталь имеет дефект. Записать событие, заключающееся в том, что ровно одна деталь имеет дефект.

2. Один школьник, желая подшутить над своими товарищами, собрал в гардеробе все пальто, а потом развесил их в случайном порядке. Какова вероятность, что каждое пальто снова попало на прежнее место, если в гардеробе шесть крючков и на них висело шесть пальто.

3. На отрезке AB наудачу выбираются две точки M и N. Какова вероятность того, что точка M окажется ближе к точке N, чем к точке A?

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_1 равна 0,1, остальных элементов A_k — по 0,04. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Прибор состоит из двух независимо работающих блоков, вероятности отказа которых за смену равны соответственно 0,05 и 0,08. Вероятность выхода из строя прибора при отказе одного из блоков равна 0,8; при отказе обоих блоков — 1. Определить вероятность выхода прибора из строя за смену. Найти вероятность того, что отказали оба блока, если известно, что прибор вышел из строя.

6. При игре с автоматом в случае выигрыша игрок получает 10 рублей. Вероятность выигрыша составляет 0,3. Найти сумму x рублей, которую игрок бросает в автомат и теряет в случае проигрыша, если математическое ожидание выигрыша равно минус 2 рублям. (В случае проигрыша сумма выигрыша считается отрицательным числом, равным сумме проигрыша, взятой со знаком «минус».) Найти ряд распределения и дисперсию суммы выигрыша. Построить график функции распределения.

7. Плотность распределения вероятностей случайной величины ξ имеет вид $f(x) = Ae^{-2|x|}$ (распределение Лапласа). Найти коэффициент A, вычислить математическое ожидание и стандартное отклонение. Найти вероятность того, что случайная величина ξ примет значение, большее 1.

8. В трех из четырех аудиторий по 20 студентов и уровень шума 80 децибелл, а в четвертой аудитории нет студентов и уровень шума 20 децибелл. Найти совместное распределение числа студентов и уровня шума в выбранной наудачу аудитории. Найти коэффициент корреляции между числом студентов и уровнем шума.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c(x + y)$, если $0 \le x \le 1$, $0 \le y \le 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество 10-копеечных монет, необходимое для выдачи каждой сдачи в кассе, принимает значения от 0 до 4 с равными вероятностями. Найти, сколько должно быть 10-копеечных монет в кассе, чтобы с вероятностью 0,9 их хватило на 2500 выдач сдачи.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $0 < \theta < 1$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{\theta} x^{-(\theta+1)/\theta} & \text{при } x > 1; \\ 0 & \text{при } x \le 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.
1,29 12,70 10,80 -10,19 4,32 12,02 13,68 3,75 -0,90 2,94 15,07 2,08 16,22 13,42 1,55 -6,05 15,70 12,35 13,94 -0,56 24,10 7,45 3,60 -0,24 16,84 6,13 -5,28 3,00 10,04

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,28 1,13 1,78 0,65 0,55 1,02 0,88 0,76 0,57
1,71 0,62 1,69 0,15 0,23 1,99 1,53 1,91 1,57

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	3	-1	2	-2	-4

Вариант 8

1. Наудачу брошены три монеты. Описать события: A — хотя бы на одной выпала решка, B — хотя бы на двух выпал орел, а также событие AB.

2. Номер лотерейного билета состоит из 3 цифр. Какова вероятность того, что все цифры взятого наудачу билета окажутся различными?

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что длина хотя бы одной из первых двух получившихся частей не превосходит $2/3$.

4. По мишени по одному разу стреляют 3 стрелка. Вероятность попадания для первого равна 0,5, для второго — 0,6, для третьего — 0,7. Найти вероятность ровно двух попаданий.

5. В семи урнах содержится по 2 белых и 2 черных шара, а в трех урнах по 7 белых и 3 черных шара. Какова вероятность, что из урны, взятой наудачу, будет извлечен белый шар? Найти вероятность, что шар извлечен из урны с 7 белыми и 3 черными шарами, если он оказался белым.

6. Прибор состоит из трех малонадежных элементов. Отказы элементов за некоторый период времени независимы, а их вероятности равны соответственно 0,1; 0,2; 0,3. Найти ряд распределения, математическое ожидание и дисперсию числа отказавших элементов. Построить график функции распределения.

7. Точка M движется по оси Ox по закону $x = at^2$. В случайный момент времени, равномерно распределенный на отрезке $[0;\ 1]$, наблюдается положение ξ точки M. Найти плотность распределения, математическое ожидание и стандартное отклонение случайной величины ξ.

8. Четыре поезда метро, уходящие с интервалом в 4 минуты, увезли по 200 пассажиров. Четыре поезда, уходящие с интервалом в 6 минут, увезли по 100 пассажиров. Два поезда, уходящие с интервалом в 8 минут, увезли по 50 пассажиров. Найти совместное распределение числа пассажиров и интервала движения для выбранного наудачу поезда. Найти коэффициент корреляции.

9. Двумерная плотность распределения случайных величин X, Y имеет вид: $f(x, y) = c$, если $x, y \ \geq 0$, $x + 2y \leq 3$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество бракованных изделий в коробке имеет распределение Пуассона с параметром 3. Найти вероятность того, что в 25 коробках менее

100 бракованных изделий.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 3$ по второму моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} (\theta - 1)x^{-\theta} & \text{при } x > 1; \\ 0 & \text{при } x \leq 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

14,92 7,48 2,82 22,84 7,49 8,98 13,84 14,17 7,07 9,69 -8,35 12,77 14,93 5,81 8,62 11,22 3,85 2,86 9,52 15,93 9,43 19,48 19,19 12,20 19,40 12,09 8,47 6,79

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке $[0;\ 2]$. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,91 0,91 0,30 1,34 0,61 1,12 1,00 0,53 1,58 0,62
0,41 0,89 1,20 1,51 0,78 1,44 0,46 0,69 1,33

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	2	3	4	5
Y	3	-1	2	-2

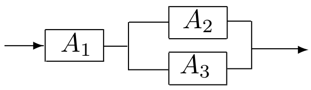
Вариант 9

1. Из колоды карт в 52 листа наудачу вынимаются три карты (без возвращения). Описать пространство элементарных исходов, а также событие, состоящее в том, что среди этих трех карт окажется ровно один туз.

2. В бригаде 3 рабочих. Какова вероятность того, что по крайней мере двое из них родились в один и тот же день недели? Считать, что вероятности родиться в каждый из дней одинаковы.

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что длина каждой из трех получившихся частей не превосходит 3/4.

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_2 равна 0,01, остальных элементов A_k — по 0,1. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Телеграфное сообщение состоит из сигналов «точка» и «тире». Известно, что среди передаваемых сигналов «точка» и «тире» встречаются в отношении 3:2. Из-за помех искажается в среднем 25 % сигналов «точка» и 20 % сигналов «тире», причем «точка» искажается в «тире», а «тире» в «точку». Найти вероятность искажения сигнала. Определить вероятность того, что передавали «тире», если известно, что приняли «точку».

6. Два игрока играют в шахматы на деньги. Известно, что в среднем из 4 партий одну выигрывает первый игрок, одна заканчивается вничью, и две выигрывает второй игрок. В случае проигрыша первый игрок платит второму 5 рублей. Сколько он должен получать в случае выигрыша, чтобы математическое ожидание его выигрыша равнялось нулю? Найти ряд распределения и дисперсию суммы выигрыша (отрицательная сумма выигрыша — это сумма проигрыша, взятая со знаком «минус»). Построить график функции распределения.

7. Случайная величина ξ имеет плотность распределения

$$f(x) = \begin{cases} \frac{A}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)} & \text{при } x \in [a;\ \infty); \\ 0 & \text{при} x < a \end{cases}$$

Найти нормирующую константу A, вычислить математическое ожидание. Построить график плотности распределения при $a = \sigma = 1$.

8. В подъезде 15 однокомнатных квартир площадью по 50 кв. м., 10 двухкомнатных квартир по 70 кв. м. и 5 трехкомнатных квартир по 80

кв. м. Для выбранной наудачу квартиры найти совместное распределение числа комнат и площади. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c(x + y)$, если $0 \le x \le 2$, $0 \le y \le 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Суммарное время работы машины складывается из 10 000 интервалов времени, каждый из которых измеряется со стандартным отклонением в 1 минуту. Найти вероятность того, что фактическое время работы отличается от измеренного больше, чем на 1 час.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{x}{\theta^2} e^{-x/\theta} & \text{при } x > 0; \\ 0 & \text{при } x \le 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

22,59 -2,61 11,87 1,37 5,92 -5,10 5,38 14,71 7,55 3,91 1,23 8,50 -5,58 -1,97 17,93 9,42 11,99 9,39 4,78 5,43 9,40 8,68 2,20 7,15 14,78 14,77 -15,16

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,49 1,96 0,64 0,76 0,01 0,82 0,23 0,82 1,96
1,28 1,49 1,07 1,92 0,17 1,68 1,01 0,48

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	0	1	2	3
Y	3	-3	2	-2

Вариант 10

1. Брошены две игральные кости. Пусть событие A состоит в том, что выпавшая сумма очков нечетна, а событие B — в том, что хотя бы на одной из костей выпала тройка. Описать события $\overline{AB}$ и $A\overline{B}$.

2. Из полного набора костей домино наудачу берутся пять костей. Найти вероятность того, что среди них будет хотя бы одна с шестеркой.

3. На линейке наудачу поставлены 2 точки. Какова вероятность того, что расстояние между ними окажется больше половины длины линейки?

4. Два стрелка поочередно стреляют по одной и той же мишени. У каждого стрелка 2 патрона. При первом попадании стрельба прекращается. Вероятность попадания в мишень при одном выстреле для первого стрелка 0,3, для второго — 0,4. Найти вероятность того, что оба стрелка израсходуют весь свой боезапас.

5. Первое орудие 2-орудийной батареи пристреляно так, что вероятность попадания для него равна $3/11$. Для второго орудия она равна $1/5$. Батарея дала залп по цели. Найти вероятность того, что цель поражена. Найти вероятность того, что первое орудие попало в цель, если известно, что цель была поражена. Для поражения цели достаточно одного попадания.

6. Вероятность приема отдельного сигнала равна 0,15. Радиосигнал передается 4 раза. Найти ряд распределения, математическое ожидание и дисперсию числа принятых сигналов. Построить график функции распределения. Найти вероятность того, что принятых сигналов будет не меньше 2, но не больше 3.

7. Радиус круга является случайной величиной, равномерно распределенной на отрезке [0;1]. Найти плотность распределения, функцию распределения, математическое ожидание и дисперсию площади круга. Найти вероятность того, что площадь превосходит $\pi/16$. Начертить графики плотности распределения и функции распределения.

8. В отделе работает один сотрудник с двумя высшими образованиями, автор 6 изобретений, четыре сотрудника с высшим образованием, каждый из которых является автором одного изобретения, и четыре сотрудника без высшего образовани, на счету которых изобретений нет. Для выбранного наудачу сотрудника найти совместное распределение количества изобретений и высших образований. Вычислить коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y зада-

ется формулой $f(x, y) = c(x + y)$, если $x, y \geq 0$, $x + y \leq 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Время ожидания автобуса пассажиром имеет показательное распределение со средним значением 10 минут. Найти пределы, в которых с вероятностью 0,8 лежит суммарное время, затраченное на ожидание автобуса за 48 поездок.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{x^2}{2\theta^3} e^{-x/\theta} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

9,36 19,48 3,89 4,45 15,11 15,90 24,94 1,72 3,25 -3,77 12,17 10,08 14,36
9,39 1,27 7,89 8,68 1,59 10,57 3,21 -6,11 15,61 10,82 1,68 5,63 6,79
20,27 -2,15

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,46 1,43 0,40 1,23 1,40 0,76 1,09 1,65 1,32
1,24 1,39 0,81 0,39 0,76 1,14 1,24 1,69 1,58

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	-1	0	1	2
Y	-3	-1	2	-5

Вариант 11

1. События: A — хотя бы один из трех проверяемых приборов бракованный, B — все приборы доброкачественные. Что означают события: $A \cup B$ и AB?

2. В ящике лежат 3 черных и 3 белых шара. Найти вероятность того, что при последовательном случайном извлечении шаров из ящика сначала вынут все белые шары.

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что длина каждой из трех получившихся частей не меньше $2/3$.

4. Вероятность изготовления некачественной детали равна 0,2. Найти вероятность того, что из 4 деталей найдется хотя бы одна качественная.

5. Запрос абонента автоматически с равными вероятностями направляется на один из двух серверов. Вероятность возниконовения сбоя в работе первого сервера равна 0,1, второго — 0,01. Какова вероятность того, что запрос будет обслужен без сбоя? Какова вероятность того, что абонент обслуживался на первом сервере, если известно, что он был обслужен без сбоя?

6. Вероятность попадания в мишень равна 0,8 при каждом выстреле. Стрельба ведется одиночными выстрелами до первого попадания, пока не будет израсходован боезапас. Найти ряд распределения, математическое ожидание и дисперсию числа произведенных выстрелов, если боезапас составляет 4 единицы. Построить график функции распределения.

7. Точку бросают наудачу в шар радиуса R. Случайная величина ξ — расстояние от точки до центра шара. Найти функцию распределения, плотность распределения, математическое ожидание и стандартное отклонение случайной величины ξ. Найти вероятность того, что ξ примет значение, большее половины радиуса шара. Начертить графики плотности распределения и функции распределения.

8. Составить таблицу совместного распределения числа выпавших двоек и числа выпавших четных чисел при одном подбрасывании игральной кости. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c(x + y)$, если $x, y \geq 0$, $x + y \leq 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Участник лотереи бросает игральную кость 20 раз. Участник получает ценный приз, если сумма очков больше 90. Оценить вероятность получения ценного приза.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 3$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} (\theta - 1)x^{-\theta} & \text{при } x > 1; \\ 0 & \text{при } x \le 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

1,87 1,26 1,58 2,11 0,01 1,35 2,05 0,76 1,65 1,61 0,12 2,03 1,07 1,10
3,06 0,38 0,64 1,63 0,54 2,65 0,82 1,21 0,73 1,99 2,44 0,93 0,47 0,88

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке $[0;\ 2]$. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,46 0,68 0,59 1,97 1,03 0,62 0,89 1,93 0,88
1,66 1,34 1,99 0,59 0,00 0,46 1,48 1,35 1,74

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4
Y	-3	-1	2	-4

Вариант 12

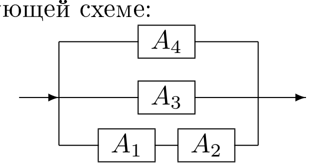
1. Игральная кость бросается n раз, $n \ge 6$. Пусть событие A_i означает, что грань i выпадет хотя бы раз, а событие A — что среди граней найдется такая, которая не выпадет ни разу. Описать событие A, используя

операции над событиями.

2. Бросают 4 игральные кости. Какова вероятность того, что на них выпадут только «5» и «6»?

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что длина максимальной части из трех получившихся частей не превосходит 4/5.

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя каждого элемента A_k равна 0,02. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Одинаковые детали поступают на сборку с трех заводов. Первый завод дает 10 %, второй 40 %, третий 50 % всех деталей, необходимых для сборки. Брак в продукции первого завода составляет 2 %, второго — 3 %, третьего — 4 %. Найти вероятность поступления на сборку бракованной детали. Найти вероятность того, что оказавшаяся бракованной деталь изготовлена на первом заводе.

6. Для трех саженцев вероятности успешно вынести пересадку равны 0,7, 0,8 и 0,85. Найти ряд распределения, математическое ожидание и дисперсию числа вынесших пересадку саженцев. Построить график функции распределения.

7. Распределение Парето приближенно описывает распределение доходов физических лиц. Плотность распределения равна

$$f(x) = \begin{cases} \frac{A}{x^{\alpha+1}} & \text{при } x \geq \theta; \\ 0 & \text{при } x < \theta. \end{cases}$$

Здесь $\alpha > 1$, $\theta > 0$ — параметры распределения, A — нормирующая константа. Найти константу A. Вычислить значение параметра α, при котором математическое ожидание превосходит значение параметра θ в 10 раз.

8. Подбрасываются три симметричных монеты. Составить таблицу совместного распределения количеств выпавших гербов на первых двух моне-

тах и на последних двух монетах. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c(x + y + xy)$, если $0 \leq x \leq 1$, $0 \leq y \leq 1$ и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Время ожидания троллейбуса за одну поездку имеет равномерное распределение на отрезке от 0 до 15 минут. Оценить вероятность того, что суммарное время ожидания за 10 поездок окажется меньше 1,5 часов.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по второму моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{2x}{\theta^2} e^{-x^2/\theta^2} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

3,50 0,80 0,32 3,31 1,12 3,29 3,87 2,65 2,01 2,65 1,19 -0,85 4,07 1,23
3,38 5,17 1,51 2,20 5,41 1,22 1,89 2,02 3,17 -1,02 2,73 1,10 3,87

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,42 1,25 0,87 0,54 0,48 1,20 1,79 0,62 0,75 0,55
0,46 1,02 1,71 1,91 0,83 0,99 1,46 1,09 0,94

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	0	2	4	6
Y	-3	-1	2	-4

Вариант 13

1. Найти случайное событие X из равенства

$$\overline{X + A} + \overline{X + \overline{A}} = B.$$

2. В студенческой группе 10 юношей и 15 девушек. На университетский праздничный бал группа получила только 3 пригласительных билета, которые разыгрываются по жребию. Какова вероятность того, что на бал попадут три девушки?

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что длина минимальной части из трех получившихся частей не превосходит $4/5$.

4. Системный администратор обслуживает 4 сервера, работающих независимо друг от друга. Вероятность того, что в течение рабочего дня сервер не потребует внимания администратора, равна для первого и второго сервера 0,8, для третьего и четвертого — 0,10. Найти вероятность того, что хотя бы один из серверов не потребует внимания администратора.

5. Фирма распространяет 2 вида рекламных листовок A и B, причем количества листовок двух видов находятся в соотношении 2:3. На листовку вида A положительно реагируют 20 % получателей, на листовку вида B — 10 % получателей. Найти вероятность положительной реакции получателя листовки. Найти вероятность того, что получена листовка вида A, если известно, что реакция была положительной.

6. Вероятность попадания баскетбольного мяча в кольцо при бросании начинающим спортсменом равна $1/5$. Мяч бросают до первого попадания, но дают не более 5 попыток. Найти ряд распределения, математическое ожидание и дисперсию числа промахов. Построить график функции распределения.

7. Скорость автомобиля на дистанции в 100 км является случайной величиной, равномерно распределенной на отрезке от 40 км/ч до 80 км/ч. Найти математическое ожидание и стандартное отклонение времени, затраченного на преодоление дистанции. Найти вероятность того, что это время превысит 2 часа.

8. В группе из 20 студентов только двое пропустили более половины занятий, и именно они получили оценку «2» на экзамене. Из остальных студентов 5 человек получили оценку «5», 10 человек — оценку «4», и 3

студента получили «тройки». Составить таблицу совместного распределения оценки на экзамене и индикатора пропуска более половины занятий для выбранного наудачу студента. Найти коэффициент корреляции.

9. Двумерная плотность распределения случайных величин X, Y имеет вид: $f(x,y) = c$, если $x^2 + y^2 \leq 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Число сериалов, просматриваемых за день выбранным наудачу студентом, имеет распределение Пуассона с параметром 0,5. Найти пределы, в которых с вероятностью 0,8 лежит число просмотров сериалов за день студентами группы из 20 человек.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по третьему моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{3x^2}{\theta^2} e^{-x^3/\theta^2} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

8,16 3,33 7,35 3,05 2,54 1,91 1,77 2,92 5,95 2,31 0,27 5,12 6,60 -1,58 5,42 5,67 6,28 -0,09 2,74 2,45 1,11 6,97 -1,59 -1,41 2,69 4,99 7,24 1,75 4,76

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,94 1,42 0,61 2,00 0,59 0,58 1,18 1,58 1,01 0,57
0,05 0,25 0,17 1,30 0,52 0,91 0,84 1,66 1,24

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4
Y	-3	-1	-2	-4

Вариант 14

1. Пусть A, B и C — события. Каков смысл равенств: $ABC = A$ и $A \cup B \cup C = A$? Привести примеры.

2. Случайно выбранная кость домино оказалась не дублем. Найти вероятность того, что вторую также взятую наудачу кость можно приставить к первой.

3. Встречные поезда приходят на станцию в случайные моменты времени в течение суток. Один поезд стоит на станции 30 минут, другой 40 минут. Найти вероятность встречи поездов на станции.

4. Предназначенный к печати текст проверяется сначала автором, затем корректором. Автор находит в среднем 80 % допущенных в тексте опечаток, корректор — 90 %. Найти вероятность того, что будут исправлены все 4 содержащиеся в первоначальном тексте опечатки.

5. Вероятность того, что изделие удовлетворяет стандарту, равна 0,8. На заводе принята система из трех независимых испытаний, каждое из которых изделие, удовлетворяющее стандарту, проходит с вероятностью 0,9, а неудовлетворяющее — с вероятностью 0,3. Какова вероятность того, что наудачу взятое изделие выдержит испытания? Какова вероятность того, что изделие, выдержавшее испытания, удовлетворяет стандарту?

6. Вероятность успешного соединения компьютера с сервером равна 0,6. Попытки соединения производятся до установления соединения, но не более 6 попыток. Найти ряд распределения, математическое ожидание и дисперсию числа попыток соединения. Построить график функции распределения.

7. Мощность W, выделяемая на сопротивлении R, вычисляется по закону $W = U^2/R$, где U — напряжение в сети. Предполагается, что напряжение — случайная величина, распределенная равномерно на отрезке от 200 до 250 вольт. Найти плотность распределения и математическое ожидание мощности, выделяемой на сопротивлении в 100 Ом. Найти вероятность того, что мощность превысит 500 Вт.

8. В офисе 4 комнаты. В первой комнате 2 сотрудника, а компьютеров нет, во второй 4 компьютера и 1 сотрудник, в остальных двух по 2 компьютера и по 2 сотрудника. Найти совместное распределение числа сотрудников и числа компьютеров в выбранной наудачу комнате. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c(x^2 + y^2)$, если $0 \le x \le 1$, $0 \le y \le 1$, и 0 иначе.

Найти константу c и $\rho(X, Y)$.

10. Взвешивают груз, находящийся в 200 мешках. Погрешность измерений веса каждого из них распределена по нормальному закону с нулевым средним и стандартным отклонением 100 грамм. Найти вероятность того, что суммарная погрешность по абсолютной величине меньше 1 кг.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{2\theta^2\sqrt{x}} e^{-\sqrt{x}/\theta^2} & \text{при } x > 0; \\ 0 & \text{при } x \le 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

1,07 -0,72 -4,46 -3,24 2,42 -1,70 -1,24 -0,07 6,20 2,67 1,80 0,26 9,61 2,51 1,44 -3,65 5,50 4,17 -2,06 7,48 2,60 7,61 2,54 9,77 9,67 7,36 7,86 11,22 3,38

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке $[0;\ 2]$. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,06 1,44 0,70 1,33 0,74 0,61 1,03 1,25 0,85
0,81 1,04 0,76 0,80 1,55 1,61 0,82 1,70 1,63

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4
Y	-3	-1	6	-4

Вариант 15

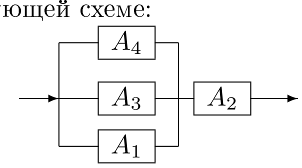
1. Брошены три монеты. Описать события
$A = \{$выпало не больше двух гербов и по крайней мере одна решка$\}$ и

B = {выпал по крайней мере один герб и хотя бы одна решка}. Описать также события AB, $A\overline{B}$.

2. В шахматном турнире участвуют 16 человек, которые разбиваются на пары по жребию и играют по олимпийской системе (проигравший выбывает из игры, ничьих нет). Какова вероятность того, что второй по силе шахматист не попадет в финал?

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что сумма длин первых двух частей не превосходит длины последней части.

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_k равна 0,1. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь не будет пропускать ток.

5. На сборку поступают детали с трех автоматов. Первый автомат дает 50 %, а второй 30 % всех деталей, необходимых для сборки. Брак в продукции первого автомата составляет 1 %, второго — 2 %, а третьего — 4 %. Деталь, изготовленная автоматом, оказалась бракованной. Какова вероятность того, что она изготовлена на первом автомате?

6. Вероятность отказа сервера при каждом из независимых подключений с помощью модема равна 0,2. Попытки подключения производятся до установления связи. Найти ряд распределения, математическое ожидание и дисперсию числа произведенных попыток подключения, если число попыток ограничено пятью. Построить график функции распределения.

7. Закон Эрланга с плотностью распределения

$$f(x) = \begin{cases} Ax^2 e^{-\alpha x} & \text{при } x \geq 0; \\ 0 & \text{при } x < 0 \end{cases}$$

описывает распределение времени прибытия двух вызовов в пуассоновском потоке. Найти коэффициент A, математическое ожидание и дисперсию. (Рекомендуется использовать таблицы определенных интегралов). Построить график плотности распределения.

8. На 5 карточках написаны цифры от 1 до 5. Найти совместное распределение числа, написанного на выбранной наудачу карточке, и индикатора того, что это число нечетное. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y имеет вид: $f(x, y) = c$, если $x^2 + y^2 \leq 4$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество воды, расходуемое жителями одной квартиры в сутки, имеет показательное распределение со средним значением 200 литров. Найти, с какой вероятностью для удовлетворения потребностей жильцов 500 квартир будет достаточно 12 000 литров воды.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{\theta^2} e^{-x/\theta^2} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

5,38 -1,10 13,11 10,84 9,45 8,56 7,87 7,34 -4,06 3,48 4,70 7,13 -1,08
4,53 13,56 2,66 7,29 9,41 11,86 9,54 10,86 2,50 -2,84 11,21 8,93

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,23 0,49 1,12 1,98 0,25 1,52 0,52 0,03 1,10
1,59 0,27 1,30 1,79 1,93 0,23 1,84 1,04

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	-3	-1	2	-4	4

Вариант 16

1. Случайная точка A наудачу выбирается в прямоугольнике со сторонами 1 и 2. Описать событие, означающее, что расстояние от A до каждой стороны прямоугольника не превосходит 1.

2. На полке в случайном порядке расставлены 8 книг, в том числе двухтомник Мандельштама. Найти вероятность того, что один из томов Мандельштама окажется у правого края полки, а другой — у левого.

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что сумма длин последних двух частей не превосходит длины первой части.

4. Вероятность попадания в мишень при одном выстреле равна 0,9. По мишени стреляют одиночными выстрелами до первого попадания, после чего стрельбу прекращают. Найти вероятность того, что будет сделано более трех выстрелов.

5. Студент выучил к экзамену только 30 вопросов из 40. Для сдачи экзамена достаточно ответить на два из четырех разных вопросов. Какова вероятность того, что экзамен будет сдан? Какова вероятность того, что студент ответил на все четыре вопроса, если известно, что он сдал экзамен?

6. Пользователь компьютера забыл пароль и перебирает наудачу 6 возможных. После трех неудачных попыток компьютер блокируется. Найти ряд распределения, математическое ожидание и дисперсию числа попыток. Построить график функции распределения.

7. Время достижения стандартным броуновским движением уровня a имеет плотность распределения

$$f(t) = \begin{cases} At^{-3/2}e^{-a^2/(2t)} & \text{при } t \geq 0; \\ 0 & \text{при } t < 0. \end{cases}$$

Найти нормирующую константу A. Доказать, что математическое ожидание времени достижения не существует. (Сделать замену $a/\sqrt{t} = y$. Можно использовать таблицы определенных интегралов).

8. В течение трех дней недели температура была 30 градусов, а влажность 60 процентов. В течение других трех дней температура 20 градусов, а влажность 90 процентов, а в последний день 10 градусов и 100 процентов. Найти совместное распределение температуры и влажности в выбранный наудачу день. Найти коэффициент корреляции между температурой и влажностью.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x,y) = cxy$, если $0 \le x \le 1$, $0 \le y \le 1$, и 0 иначе. Найти константу c и $\rho(X,Y)$.

10. Участник лотереи бросает 5 шаров, каждый из которых может попасть в лузы с номерами от 1 до 6. Участник получает ценный приз, если сумма очков больше 23. Оценить вероятность получения ценного приза.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по третьему моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{3\sqrt{x}}{2\theta^2} e^{-x\sqrt{x}/\theta^2} & \text{при } x > 0; \\ 0 & \text{при } x \le 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.
4,61 6,70 2,88 9,09 -2,06 6,25 6,46 4,25 16,16 7,07 1,35 13,58 7,96 14,64
-2,14 10,81 2,50 2,24 -1,04 5,31 11,93 16,20 7,49 -5,21 5,90 5,63 7,26

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,64 0,79 0,64 1,06 0,42 0,69 1,65 0,45 0,43
1,48 0,44 0,97 1,49 0,46 1,29 0,37 0,45

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	0	2	4	6	8
Y	-3	-1	2	-4	4

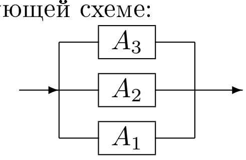
Вариант 17

1. Случайная точка A наудачу выбирается в прямоугольнике со сторонами 1 и 2. Описать событие, означающее, что расстояние от A до ближайшей стороны прямоугольника не превосходит $1/2$.

2. Из колоды карт в 36 листов вынимаются три карты. Найти вероятность того, что среди них окажутся хотя бы две красные карты.

3. На отрезке AB наудачу выбираются две точки M и N. Какова вероятность того, что точка M окажется по крайней мере втрое ближе к точке N, чем к точке A?

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_1 равна 0,1, остальных элементов A_k — по 0,04. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Прибор состоит из трех независимо работающих блоков, вероятности отказа которых за смену равны соответственно 0,01, 0,05 и 0,08. Вероятность выхода из строя прибора при отказе одного из блоков равна 0,5; при отказе двух блоков — 0,8, при отказе всех трех блоков — 1. Определить вероятность выхода прибора из строя за смену. Найти вероятность того, что отказали все три блока, если известно, что прибор вышел из строя.

6. При игре с автоматом игрок получает 50 рублей с вероятностью 0,1, 10 рублей с вероятностью 0,3. Найти сумму x рублей, которую игрок бросает в автомат и теряет в случае проигрыша, если математическое ожидание выигрыша равно минус 2 рублям. (В случае проигрыша сумма выигрыша считается отрицательным числом, равным сумме проигрыша, взятой со знаком «минус».) Найти ряд распределения и дисперсию суммы выигрыша. Построить график функции распределения.

7. Плотность распределения вероятностей случайной величины ξ имеет вид

$$f(x) = \begin{cases} \dfrac{A}{1+\left(\dfrac{x}{\theta}\right)^2} & \text{при } |x| \le \theta; \\ 0 & \text{при } |x| > \theta \end{cases}$$

(усеченное распределение Коши). Найти коэффициент A, вычислить математическое ожидание и стандартное отклонение. Найти вероятность того, что случайная величина ξ примет значение, большее $\theta/\sqrt{3}$.

8. В двух из четырех аудиторий по 20 студентов и уровень шума 60 децибелл, в третьей 10 студентов и уровень шума 50 децибелл, а в четвертой аудитории нет студентов и уровень шума 20 децибелл. Найти совместное распределение числа студентов и уровня шума в выбранной наудачу аудитории. Найти коэффициент корреляции между числом студентов и уровнем шума.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x,y) = cx$, если $x, y \ \geq 0$, $x + 2y \leq 2$, и 0 иначе. Найти константу c и $\rho(X,Y)$.

10. Количество 10-копеечных монет, необходимое для выдачи каждой сдачи в кассе, принимает значения от 0 до 4 с равными вероятностями. Найти, с какой вероятностью на 100 выдач сдачи будет достаточно 220 10-копеечных монет .

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $1 < \theta < 2$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{\theta-1} x^{-\theta/(\theta-1)} & \text{при } x > 1; \\ 0 & \text{при } x \leq 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.
2,92 12,70 10,80 -10,19 4,32 12,02 13,68 3,75 -0,90 2,94 15,07 2,08 16,22 13,42 1,55 -6,05 15,70 12,35 13,94 -0,56 24,10 7,45 3,60 -0,24 16,84 6,13 -5,28 3,00 10,04

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке $[0;\ 2]$. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,82 1,13 1,78 0,65 0,55 1,02 0,88 0,76 0,57
1,71 0,62 1,69 0,15 0,23 1,99 1,53 1,91 1,57

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	3	-1	2	-4	4

Вариант 18

1. Брошены три игральные кости. Описать событие, означающее, что хотя бы на одной кости появилась единица и не более чем на двух выпали двойки.

2. Номер лотерейного билета состоит из 6 цифр. Какова вероятность того, что хотя бы две цифры взятого наудачу билета совпадают?

3. Стержень единичной длины AB разломан в двух наудачу выбранных точках X и Y. С какой вероятностью расстояние между этими точками не превзойдет максимального из двух отрезков AX или AY.

4. По мишени по одному разу стреляют 4 стрелка. Вероятность попадания для первого равна 0,5, для второго — 0,6, для третьего — 0,7, для четвертого — 0,9. Найти вероятность ровно двух попаданий.

5. В семи урнах содержится по 3 белых и 2 черных шара, а в трех урнах по 7 белых и 3 черных шара. Какова вероятность, что из урны, взятой наудачу, будет извлечен белый шар? Найти вероятность, что шар извлечен из урны с 7 белыми и 3 черными шарами, если он оказался белым.

6. Прибор состоит из трех малонадежных элементов. Отказы элементов за некоторый период времени независимы, а их вероятности равны соответственно 0,2; 0,3; 0,4. Найти ряд распределения, математическое ожидание и дисперсию числа отказавших элементов. Построить график функции распределения.

7. Точка M движется по оси Ox по закону $x = vt - at^2$. В случайный момент времени, равномерно распределенный на отрезке $[0;\ T]$, наблюдается координата ξ точки M. Найти математическое ожидание и стандартное отклонение случайной величины ξ. При $v = 10$ м/с, $a = 10$ м/с^2, $T = 3$ с найти вероятность того, что $\xi > 0$.

8. Четыре автобуса, уходящие с интервалом в 5 минут, увезли по 20 пассажиров. Два автобуса, уходящие с интервалом в 10 минут, увезли по 30 пассажиров. Два автобуса, уходящие с интервалом в 15 минут, увезли по

35 пассажиров. Найти совместное распределение числа пассажиров и интервала движения для выбранного наудачу автобуса. Найти коэффициент корреляции.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c(x^2 + y^2 + xy)$, если $0 \le x \le 1$, $0 \le y \le 1$ и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество бракованных изделий в коробке имеет распределение Пуассона с параметром 2. Найти вероятность того, что в 16 коробках более 40 бракованных изделий.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 2$ по второму моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \theta x^{-(\theta+1)} & \text{при } x > 1; \\ 0 & \text{при } x \le 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

13,29 7,48 2,82 22,84 7,49 8,98 13,84 14,17 7,07 9,69 -8,35 12,77 14,93 5,81 8,62 11,22 3,85 2,86 9,52 15,93 9,43 19,48 19,19 12,20 19,40 12,09 8,47 6,79

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,19 0,91 0,30 1,34 0,61 1,12 1,00 0,53 1,58 0,62
0,41 0,89 1,20 1,51 0,78 1,44 0,46 0,69 1,33

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	3	5	7	9
Y	-3	-1	2	-4	4

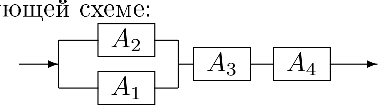
Вариант 19

1. Некто написал n адресатам письма, в каждый конверт вложил по одному письму и затем наудачу написал на каждом конверте один из n адресов. Пусть событие A_i состоит в том, что i-е письмо попало в свой конверт. Описать событие, заключающееся в том, что ровно одно письмо попало в свой конверт.

2. В бригаде 4 рабочих. Какова вероятность того, что по крайней мере трое из них родились в один и тот же день недели? Считать, что вероятности родиться в каждый из дней одинаковы.

3. Стержень единичной длины AB разломан в двух наудачу выбранных точках X и Y. С какой вероятностью расстояние между этими точками не превзойдет длины отрезка AX.

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_2 равна 0,01, остальных элементов A_k — по 0,1. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Телеграфное сообщение состоит из сигналов «точка» и «тире». Известно, что среди передаваемых сигналов «точка» и «тире» встречаются в отношении 7:3. Из-за помех искажается в среднем 25 % сигналов «точка» и 20 % сигналов «тире», причем «точка» искажается в «тире», а «тире» в «точку». Найти вероятность искажения сигнала. Определить вероятность того, что передавали «точку», если известно, что приняли «тире».

6. Два игрока играют в шахматы на деньги. Известно, что в среднем из 5 партий одну выигрывает первый игрок, две заканчиваются вничью, и две выигрывает второй игрок. В случае проигрыша первый игрок платит второму 50 рублей. Сколько он должен получать в случае выигрыша, чтобы математическое ожидание его выигрыша равнялось нулю? Найти ряд распределения и дисперсию суммы выигрыша (отрицательная сумма выигрыша — это сумма проигрыша, взятая со знаком «минус»). Построить график функции распределения.

7. Случайная величина ξ имеет плотность распределения

$$f(x) = \begin{cases} \frac{A}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)} & \text{при } |x-a| \le 2\sigma; \\ 0 & \text{при } |x-a| > 2\sigma. \end{cases}$$

Найти нормирующую константу A, вычислить математическое ожидание. Построить график плотности распределения при $a = \sigma = 1$.

8. В подъезде 5 однокомнатных квартир площадью по 40 кв. м., 10 двухкомнатных квартир по 60 кв. м., 10 трехкомнатных квартир по 70 кв. м. и 5 четырехкомнатных по 90 кв. м. Для выбранной наудачу квартиры найти совместное распределение числа комнат и площади. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c$, если $|x| + |y| \leq 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Суммарное время работы машины складывается из 1000 интервалов времени, каждый из которых измеряется со стандартным отклонением в 10 минут. Найти вероятность того, что фактическое время работы отличается от измеренного больше, чем на 10 часов.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{x}{\theta^4} e^{-x/\theta^2} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

23,95 -2,61 11,87 1,37 5,92 -5,10 5,38 14,71 7,55 3,91 1,23 8,50 -5,58 -1,97 17,93 9,42 11,99 9,39 4,78 5,43 9,40 8,68 2,20 7,15 14,78 14,77 -15,16

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке $[0;\ 2]$. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,94 1,96 0,64 0,76 0,01 0,82 0,23 0,82 1,96
1,28 1,49 1,07 1,92 0,17 1,68 1,01 0,48

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую ли-

нейной регрессии.

X	1	2	3	4	5
Y	-3	-1	2	-4	8

Вариант 20

1. Некто написал n адресатам письма, в каждый конверт вложил по одному письму и затем наудачу написал на каждом конверте один из n адресов. Пусть событие A_i состоит в том, что i-е письмо попало в свой конверт. Описать событие, заключающееся в том, что ровно два письма попало в свой конверт.

2. Два раза подбрасываются две монеты. Найти вероятность того, что во второй раз выпадет столько же гербов, сколько и в первый.

3. На линейке наудачу поставлены 2 точки. Какова вероятность того, что расстояние между ними окажется больше четверти длины линейки?

4. Три стрелка поочередно стреляют по одной и той же мишени. У каждого стрелка 2 патрона. При первом попадании стрельба прекращается. Вероятность попадания в мишень при одном выстреле для первого стрелка 0,3, для второго — 0,4, для третьего — 0,6. Найти вероятность того, что все стрелки израсходуют весь свой боезапас.

5. Первое орудие 3-орудийной батареи пристреляно так, что вероятность попадания для него равна $3/11$. Для второго и третьего орудия она равна $1/5$. Батарея дала залп по цели. Найти вероятность того, что цель поражена. Найти вероятность того, что первое орудие попало в цель, если известно, что цель была поражена. Для поражения цели достаточно одного попадания.

6. Вероятность приема отдельного сигнала равна 0,05. Радиосигнал передается 5 раз. Найти ряд распределения, математическое ожидание и дисперсию числа принятых сигналов. Построить график функции распределения. Найти вероятность того, что принятых сигналов будет не меньше 2, но не больше 3.

7. Катет равнобедренного прямоугольного треугольника является случайной величиной, равномерно распределенной на отрезке [0;1]. Найти

плотность распределения, функцию распределения, математическое ожидание и дисперсию площади треугольника. Найти вероятность того, что площадь превосходит $1/8$. Начертить графики плотности распределения и функции распределения.

8. В отделе работает один сотрудник с двумя высшими образованиями возрастом 30 лет, два сотрудника с высшим образованием возрастом по 50 лет и два сотрудника без высшего образования возрастом по 20 лет. Для выбранного наудачу сотрудника найти совместное распределение возраста и количества высших образований. Вычислить коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x,y) = c$, если $x \geq 0$, $|x| + |y| \leq 1$, и 0 иначе. Найти константу c и $\rho(X,Y)$.

10. Время ожидания автобуса пассажиром имеет показательное распределение со средним значением 8 минут. Найти количество поездок, за которое суммарное время, затраченное на ожидание автобуса, не превысит 5 часов с вероятностью 0,9.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{x^2}{2\theta^6} e^{-x/\theta^2} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

10,63 19,48 3,89 4,45 15,11 15,90 24,94 1,72 3,25 -3,77 12,17 10,08 14,36 9,39 1,27 7,89 8,68 1,59 10,57 3,21 -6,11 15,61 10,82 1,68 5,63 6,79 20,27 -2,15

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,64 1,43 0,40 1,23 1,40 0,76 1,09 1,65 1,32
1,24 1,39 0,81 0,39 0,76 1,14 1,24 1,69 1,58

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	-8	-1	2	-4	8

Вариант 21

1. Бросаются две игральные кости. Пусть событие A состоит в том, что выпавшая сумма очков нечетна, а событие B — в том, что хотя бы на одной из костей выпала тройка. Описать события AB и $A\overline{B}$.

2. В ящике 5 красных и 4 синих пуговиц. Какова вероятность того, что из четырех наудачу вынутых пуговиц хотя бы две будут одноцветными.

3. Молодой человек договорился встретиться с девушкой между 9 и 10 часами и обещал ждать её до 10 часов. Девушка обещала ждать его 10 минут, если придет раньше. Найти вероятность того, что они встретятся. Предполагается, что моменты их прихода равновероятны в течение часа.

4. При передаче сообщений в среднем 20 % писем не доходят до получателя. Найти вероятность того, что из 6 писем более половины на будет получено адресатами.

5. В пункте проката имеется 6 одинаковых на вид велосипедов. Вероятность поломки для двух из них по 0,1, для трех по 0,2 и для одного 0,7. Какова вероятность того, что велосипед сломается, если его выбирают наудачу? Какова вероятность того, что был выбран велосипед, для которого вероятность поломки 0,7, при условии, что он сломался?

6. Вероятность попадания в мишень равна 0,4 при каждом выстреле. Стрельба ведется одиночными выстрелами до первого попадания, пока не будет израсходован боезапас. Найти ряд распределения, математическое ожидание и дисперсию числа произведенных выстрелов, если боезапас составляет 5 единиц. Построить график функции распределения.

7. Случайная величина ξ — координата точки, совершающей колебательные движения по закону $x = a\sin(\omega t)$, и наблюдаемой в случайный момент времени T, равномерно распределенный на периоде колебаний $[0;\ 2\pi/\omega]$. Найти математическое ожидание и стандартное отклонение случайной величины ξ. Найти вероятность того, что $\xi > a/2$.

8. Составить таблицу совместного распределения числа выпавших четных и нечетных чисел при одном подбрасывании игральной кости. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c$, если $y \geq 0$, $|x| + |y| \leq 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Оценить, сколько раз нужно бросить игральную кость, чтобы сумма выпавших очков превысила 300 с вероятностью не менее 0,92.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 4$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} (\theta - 2)x^{-\theta+1} & \text{при } x > 1; \\ 0 & \text{при } x \leq 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

0,87 1,26 1,58 2,11 0,01 1,35 2,05 0,76 1,65 1,61 0,12 2,03 1,07 1,10
3,06 0,38 0,64 1,63 0,54 2,65 0,82 1,21 0,73 1,99 2,44 0,93 0,47 0,88

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке $[0;\ 2]$. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,46 0,68 0,59 1,97 1,03 0,62 0,89 1,93 0,88
1,66 1,34 1,99 0,59 0,00 0,46 1,48 1,35 1,74

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	-3	-1	8	-4	4

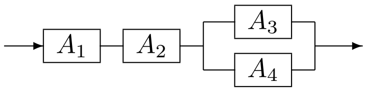
Вариант 22

1. Прибор состоит из двух блоков первого типа и трех блоков второго типа. События: $A_k, k = 1, 2$ — исправен k-й блок первого типа, $B_j, j = 1, 2, 3$ — исправен j-й блок второго типа. Прибор исправен, если исправны хотя бы один блок первого типа и не менее двух блоков второго типа. Выразить C, означающее исправность прибора, через A_k и B_j.

2. Бросают 4 игральные кости. Какова вероятность того, что хотя бы на двух из них выпадет одинаковое число очков?

3. Стержень единичной длины AB разломан в двух наудачу выбранных точках X и Y. С какой вероятностью расстояние между этими точками не превзойдет длины отрезка BY?

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя каждого элемента A_k равна 0,02. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Одинаковые детали поступают на сборку с трех автоматов. Первый автомат дает 25 %, второй 30 %, третий 45 % всех деталей, необходимых для сборки. Брак в продукции первого автомата составляет 2,5 %, второго — 2 %, третьего — 3 %. Найти вероятность поступления на сборку небракованной детали. Найти вероятность того, что оказавшаяся небракованной деталь изготовлена на первом автомате.

6. По мишени одновременно стреляют три стрелка, вероятности попаданий которых равны соответственно 0,4, 0,7 и 0,9. Найти ряд распределения, математическое ожидание и дисперсию числа попаданий в мишень. Построить график функции распределения.

7. Максимальный нуль стандартного броуновского движения на [0; 1] имеет координату ξ с функцией распределения

$$F(x) = \begin{cases} A\arcsin\sqrt{x} & \text{при } x \in [0;\ 1]; \\ 1 & \text{при } x > 1. \end{cases}$$

Найти константу A. Построить графики функции распределения и плотности распределения случайной величины ξ.

8. Подбрасываются три симметричных монеты. Составить таблицу совместного распределения количеств выпавших гербов на первой монете и на трех монетах. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c(x^3 + y^3 + xy^2)$, если $0 \le x \le 1$, $0 \le y \le 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Время ожидания поезда метро за одну поездку имеет равномерное распределение на отрезке от 0 до 5 минут. Оценить число поездок, в течение которых суммарное время ожидания окажется меньше 1 часа с вероятностью 0,96.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по второму моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{2x}{\theta^3} e^{-x^2/\theta^3} & \text{при } x > 0; \\ 0 & \text{при } x \le 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

2,50 0,80 0,32 3,31 1,12 3,29 3,87 2,65 2,01 2,65 1,19 -0,85 4,07 1,23
3,38 5,17 1,51 2,20 5,41 1,22 1,89 2,02 3,17 -1,02 2,73 1,10 3,87

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,42 1,25 0,87 0,54 0,48 1,20 1,79 0,62 0,75 0,55
0,46 1,02 1,71 1,91 0,83 0,99 1,46 1,09 0,94

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	$-3,5$	$-1,4$	$2,2$	$-4,6$	$4,1$

Вариант 23

1. Судно имеет одно рулевое устройство, четыре котла и две турбины. Событие A означает исправность рулевого устройства, $B_k, k = 1, 2, 3, 4$ — исправность k-го котла, а $C_j, j = 1, 2$ — исправность j-й турбины. Событие D — судно управляемое, что будет в том и только в том случае, когда исправны рулевое устройство, хотя бы один котел и хотя бы одна турбина. Выразить D через A, B_k и C_j.

2. В студенческой группе 10 юношей и 15 девушек. На университетский праздничный бал группа получила 5 пригласительных билетов, которые разыгрываются по жребию. Какова вероятность того, что на бал попадет хотя бы одна девушка?

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что длина хотя бы одной из трех получившихся частей не превосходит $2/3$.

4. На трех телеканалах часть времени занята рекламой: на первом — 60 % времени, на втором — 40 %, на местном — 30 %. Найти вероятность того, что в случайный момент времени нет рекламы хотя бы на одном из каналов.

5. Станок обрабатывает 2 вида деталей A и B, причем время работы распределяется между ними в соотношении 2:3. При обработке детали вида A он работает с максимальной для него нагрузкой в течение 60 % времени, при обработке детали вида B — 90 % времени. В случайный момент времени станок работал с максимальной нагрузкой. Определить вероятность того, что в это время он обрабатывал деталь вида A; вида B.

6. Вероятность попадания баскетбольного мяча в кольцо при бросании начинающим спортсменом равна $1/9$. Мяч бросают до первого попадания, но дают не более 6 попыток. Найти ряд распределения, математическое ожидание и дисперсию числа промахов. Построить график функции распределения.

7. Сила, действующая на электрон в электрическом поле, вычисляется по формуле $F = k/r^2$, где r — расстояние от анода — случайная величина, распределенная равномерно на $[R;\ 2R]$. Найти математическое ожидание и стандартное отклонение силы F. Найти вероятность того, что эта сила превысит $k/(2R^2)$.

8. В группе из 20 студентов только двое изучали в школе французский язык, и именно они получили оценку «4» на экзамене. Из остальных студентов 10 человек получили оценку «3», 5 человек — оценку «3», и 3

студента получили «двойки». Составить таблицу совместного распределения оценки на экзамене и индикатора изучения французского языка для выбранного наудачу студента. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x,y) = c$, если $x \geq 0$, $|x| + |y| \leq 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Число опечаток на странице книги имеет распределение Пуассона с параметром 0,5. Найти, сколько должно быть страниц в книге, чтобы число опечаток в ней не превысило 200 с вероятностью 0,75.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по третьему моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{3x^2}{\theta^3} e^{-x^3/\theta^3} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

7,16 3,33 7,35 3,05 2,54 1,91 1,77 2,92 5,95 2,31 0,27 5,12 6,60 -1,58
5,42 5,67 6,28 -0,09 2,74 2,45 1,11 6,97 -1,59 -1,41 2,69 4,99 7,24 1,75
4,76

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,94 1,42 0,61 2,00 0,59 0,58 1,18 1,58 1,01 0,57
0,05 0,25 0,17 1,30 0,52 0,91 0,84 1,66 1,24

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	$-3,5$	-1	$2,5$	-4	4

Вариант 24

1. Машинно-котельная установка состоит из двух котлов и одной машины. Событие A — исправна машина, событие $B_k, k = 1, 2$ — исправен k-й котел. Событие C означает работоспособность машинно-котельной установки, что будет в том и только в том случае, если исправна машина и хотя бы один котел. Выразить события C и $\overline{C}$ через A и B_k.

2. В компании из десяти человек решили сделать друг другу подарки, для чего каждый принес подарок. Все подарки сложили вместе, перемешали и случайно распределили среди участников. Найти вероятность того, что три конкретных человека получат свой собственный подарок.

3. На линейке длиной 20 см случайно сделаны две насечки. Какова вероятность того, что растояние от первой насечки до начала линейки превосходит расстояние от второй насечки до начала линейки более чем на 15 см?

4. Радиосигнал передается последовательно через 3 ретранслятора. На каждом ретрансляторе может возникнуть помеха независимо от остальных ретрансляторов с вероятностями 0,02, 0,03 и 0,04 соответственно. Найти вероятность получения радиосигнала без помехи.

5. Вероятность того, что изделие удовлетворяет стандарту, равна 0,95. На заводе принята система из четырех независимых испытаний, каждое из которых изделие, удовлетворяющее стандарту, проходит с вероятностью 0,9, а неудовлетворяющее — с вероятностью 0,4. Какова вероятность того, что наудачу взятое изделие выдержит испытания? Какова вероятность того, что изделие, выдержавшее испытания, удовлетворяет стандарту?

6. Вероятность изготовления нестандартного изделия при налаженном технологическом процессе постоянна и равна $1/9$. Для проверки изделий отдел технического контроля берет из партии изделия одно за другим, но не более 5 изделий. При обнаружении нестандартного изделия вся партия задерживается. Найти ряд распределения, математическое ожидание и дисперсию числа изделий, проверяемых в каждой партии. Построить график функции распределения.

7. Высота H, которой достигает брошенный вверх мяч, определяется по формуле $H = v^2/(2g)$, где v — скорость, с которой брошен мяч, g — ускорение свободного падения, которое примем равным 10 м/с^2. Предполагается, что v — случайная величина, распределенная равномерно на отрезке от 10 до 20 м/с. Найти плотность распределения и математическое ожидание высоты, достигнутой мячом. Найти вероятность того, что высота превысит 15 м.

8. В научном отделе 3 лаборатории. В первой лаборатории 6 сотрудников и 2 исследовательских проекта, во второй 8 сотрудников и 1 проект, в третьей — 4 сотрудника и 2 проекта. Найти совместное распределение числа сотрудников и числа проектов в выбранной наудачу лаборатории. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c$, если $-1 \leq x \leq 1$, $0 \leq y \leq x^2$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Погрешность измерений длины каждого из участков маршрута распределена по нормальному закону с нулевым средним и стандартным отклонением 5 метров. Найти, на сколько участков можно разбить маршрут, чтобы суммарная погрешность не превосходила по модулю 100 метров с вероятностью 0,95.

11. Для выборки $(X_1, X_2, \ldots, X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{2\theta^3\sqrt{x}} e^{-\sqrt{x}/\theta^3} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

-1,07 -0,72 -4,46 -3,24 2,42 -1,70 -1,24 -0,07 6,20 2,67 1,80 0,26 9,61 2,51 1,44 -3,65 5,50 4,17 -2,06 7,48 2,60 7,61 2,54 9,77 9,67 7,36 7,86 11,22 3,38

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,06 1,44 0,70 1,33 0,74 0,61 1,03 1,25 0,85
0,81 1,04 0,76 0,80 1,55 1,61 0,82 1,70 1,63

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	-3	-1	2	-4	0

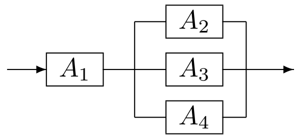
Вариант 25

1. Брошены четыре монеты. Пусть событие A состоит в том, что по крайней мере на двух монетах выпал герб, а событие B — в том, что хотя бы на двух монетах выпала решка. Описать события AB, $\overline{A}B$, $\overline{AB}$.

2. В шахматном матче участвуют 4 пары шахматистов. Вероятность ничьей в каждой партии равна $1/4$. Найти вероятность того, что в матче будет хотя бы одна ничья.

3. На отрезке единичной длины наудачу поставлены две точки, в результате чего этот отрезок оказывается разделенным на три части. Определить вероятность того, что длина первых двух частей не превосходит $3/5$, длина же последней части больше $1/2$.

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_k равна 0,1. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь не будет пропускать ток.

5. На сборку поступают детали с четырех автоматов. Первый и второй автоматы дают по 40 %, а третий и четвертый по 10 % всех деталей, необходимых для сборки. Брак в продукции первого и второго автомата составляет 1 %, а третьего и четвертого — 4 %. Деталь, изготовленная автоматом, оказалась бракованной. Какова вероятность того, что она изготовлена на первом автомате?

6. Вероятность отказа сервера при каждом из независимых подключений с помощью модема равна 0,2. Попытки подключения производятся до установления связи. Найти ряд распределения, математическое ожидание и дисперсию числа произведенных попыток подключения, если число попыток ограничено шестью. Построить график функции распределения.

7. Закон Эрланга с плотностью распределения

$$f(x) = \begin{cases} Ax^2 e^{-\alpha x} & \text{при } x \geq 0; \\ 0 & \text{при } x < 0 \end{cases}$$

описывает время ожидания прихода трех вызовов в пуассоновском потоке. Найти коэффициент A, математическое ожидание и дисперсию. (Рекомендуется использовать таблицы определенных интегралов).

8. На 8 карточках написаны цифры от 1 до 9. Найти совместное распределение числа, написанного на выбранной наудачу карточке, и индикатора того, что это число больше трех. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c$, если $0 \leq x \leq 1$, $0 \leq y \leq x^2$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество воды, расходуемое жителями одной квартиры в сутки,8 имеет показательное распределение со средним значением 100 литров. Найти, для какого количества квартир достаточно 100 000 литров воды с вероятностью 0,94.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{\theta^3} e^{-x/\theta^3} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

4,38 -1,10 13,11 10,84 9,45 8,56 7,87 7,34 -4,06 3,48 4,70 7,13 -1,08
4,53 13,56 2,66 7,29 9,41 11,86 9,54 10,86 2,50 -2,84 11,21 8,93

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,23 0,49 1,12 1,98 0,25 1,52 0,52 0,03 1,10

1,59 0,27 1,30 1,79 1,93 0,23 1,84 1,04

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	0	2	3	4	6
Y	-3	-1	2	-4	4

Вариант 26

1. На отрезке $[0,1]$ наудачу ставятся две точки. Построить подходящее пространство элементарных исходов Ω и описать событие A, означающее, что вторая точка ближе к правому концу отрезка $[0,1]$, чем к левому, и событие B, означающее, что расстояние между двумя точками меньше половины длины отрезка, а также событие AB.

2. Трое женщин и трое мужчин садятся случайным образом за круглый стол. Найти вероятность того, что мужчины и женщины за столом будут чередоваться.

3. Случайная точка A наудачу выбирается в прямоугольнике со сторонами 1 и 2. найти вероятность того, что расстояние от А до ближайшей стороны прямоугольника не превосходит $1/3$.

4. Вероятность установления соединения с сервером при каждой попытке равна 0,9. Найти вероятность того, что соединение будет установлено не раньше четвертой попытки.

5. Студент выучил к зачету только 10 вопросов из 30. Для получения зачета достаточно ответить на два из четырех разных вопросов. Какова вероятность того, что зачет будет получен? Какова вероятность того, что студент ответил не менее чем на три вопроса, если известно, что он получил зачет?

6. Пользователь компьютера забыл пароль и перебирает наудачу 5 возможных. После четырех неудачных попыток компьютер блокируется. Найти ряд распределения, математическое ожидание и дисперсию числа попыток. Построить график функции распределения.

7. Случайная величина ξ имеет стандартное логарифмически нормальное распределение, если $\xi = e^{\eta}$, где η имеет стандартное нормальное распределение. Найти плотность распределения и математическое ожидание случайной величины ξ. Найти вероятность того, что $\xi > 1$.

8. В двух из четырех комнат температура 25 градусов, а влажность 80 процентов. В третьей комнате температура 20 градусов, а влажность 90 процентов. В четвертой комнате температура 25 градусов, а влажность 90 процентов. Найти совместное распределение температуры и влажности в выбранной наудачу комнате. Найти коэффициент корреляции между температурой и влажностью.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x,y) = c$, если $0 \leq x \leq 1$, $0 \leq y \leq \sqrt{x}$, и 0 иначе. Найти константу c и $\rho(X,Y)$.

10. Участник лотереи бросает несколько шаров, каждый из которых может попасть в лузы с номерами от 1 до 6. Участник получает ценный приз, если сумма очков меньше 12. Найти, при каком числе шаров вероятность получения ценного приза будет меньше 0,01.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по третьему моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{3\sqrt{x}}{2\theta^3} e^{-x\sqrt{x}/\theta^3} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

5,61 6,70 2,88 9,09 -2,06 6,25 6,46 4,25 16,16 7,07 1,35 13,58 7,96 14,64
-2,14 10,81 2,50 2,24 -1,04 5,31 11,93 16,20 7,49 -5,21 5,90 5,63 7,26

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,64 0,79 0,64 1,06 0,42 0,69 1,65 0,45 0,43
1,48 0,44 0,97 1,49 0,46 1,29 0,37 0,45

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	-1	2	3	4	6
Y	-3	-1	2	-4	4

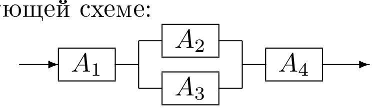
Вариант 27

1. Из множества супружеских пар выбирается одна пара. Событие $A =$ {Мужу больше 25 лет}, событие $B =$ {Муж старше жены}, событие $C =$ {Жене больше 25 лет}.
Выяснить смысл событий: ABC, $A \backslash AB$, $A\overline{B}C$.

2. Собрались вместе три незнакомых человека. Найти вероятность, что хотя бы у двух из них совпадают дни рождения.

3. На отрезке AB наудачу выбираются две точки M и N. Какова вероятность того, что точка M окажется по крайней мере вдвое ближе к точке A, чем к точке N?

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_1 равна 0,1, остальных элементов A_k — по 0,04. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Прибор состоит из четырех независимо работающих блоков, вероятности отказа которых за смену равны соответственно 0,01, 0,02, 0,03 и 0,04. Вероятность выхода из строя прибора при отказе одного из блоков равна 0,8; при отказе более чем одного блока — 1. Определить вероятность выхода прибора из строя за смену. Найти вероятность того, что отказал один блок, если известно, что прибор вышел из строя.

6. При игре с автоматом в случае выигрыша игрок получает 10 рублей. Для участия в игре игрок бросает в автомат 5 рублей. Найти вероятность выигрыша, если математическое ожидание выигрыша равно минус 2 рублям. (В случае проигрыша сумма выигрыша считается отрицательным числом, равным сумме проигрыша, взятой со знаком «минус».) Найти ряд распределения и дисперсию суммы выигрыша. Построить график функции распределения.

7. Плотность распределения вероятностей случайной величины ξ имеет вид $f(x) = Ae^{-|x-a|}$ (распределение Лапласа). Найти коэффициент A, вычислить математическое ожидание и стандартное отклонение. Найти вероятность того, что случайная величина ξ примет значение, большее $2a$.

8. В трех из четырех аудиторий по 20 студентов и уровень шума 60 децибелл, а в четвертой аудитории нет студентов и уровень шума 20 де-

цибелл. Найти совместное распределение числа студентов и уровня шума в выбранной наудачу аудитории. Найти коэффициент корреляции между числом студентов и уровнем шума.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c$, если $0 \le x \le 1$, $0 \le y \le x^3$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество 10-копеечных монет, необходимое для выдачи каждой сдачи в кассе, принимает значения от 0 до 4 с равными вероятностями. В кассе в начале рабочего дня находится 2500 10-копеечных монет. Найти, для какого количества покупателей получение сдачи гарантировано с вероятностью 0,8.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $2 < \theta < 3$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{1}{\theta-2} x^{-(\theta-1)/(\theta-2)} & \text{при } x > 1; \\ 0 & \text{при } x \le 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

1,92 12,70 10,80 -10,19 4,32 12,02 13,68 3,75 -0,90 2,94 15,07 2,08 16,22 13,42 1,55 -6,05 15,70 12,35 13,94 -0,56 24,10 7,45 3,60 -0,24 16,84 6,13 -5,28 3,00 10,04

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,82 1,13 1,78 0,65 0,55 1,02 0,88 0,76 0,57
1,71 0,62 1,69 0,15 0,23 1,99 1,53 1,91 1,57

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	-9	-1	2	-4	4

Вариант 28

1. Брошены две игральные кости. Пусть событие A состоит в том, что выпавшая сумма очков нечетна, событие B — в том, что хотя бы на одной из костей выпала единица, событие C — в том, что хотя бы на одной кости выпала двойка. Описать события: ABC, $A\overline{B}C$, $\overline{A}BC$.

2. Номер лотерейного билета состоит из 8 цифр. Какова вероятность того, что первые четыре цифры четные, а последние четыре — нечетные?

3. Случайная точка A наудачу выбирается в прямоугольнике со сторонами 1 и 2. найти вероятность того, что расстояние от А до каждой диагонали прямоугольника не превосходит 1/3.

4. Интервал движения между автобусами маршрута А — 5 минут, маршрута Б — 6 минут, маршрута В — 10 минут. Пассажир приходит на остановку в случайный момент времени. Какова вероятность того, что хотя бы один автобус придет в течение 2 минут после прихода пассажира?

5. В девяти урнах содержится по 4 белых и 2 черных шара, а в одной урне 9 белых и 1 черный шар. Какова вероятность, что из урны, взятой наудачу, будет извлечен черный шар? Найти вероятность, что шар извлечен из урны с 9 белыми и 1 черным шаром, если он оказался черным.

6. Прибор состоит из четырех малонадежных элементов. Отказы элементов за некоторый период времени независимы, а их вероятности равны соответственно 0,1; 0,1; 0,2; 0,2. Найти ряд распределения, математическое ожидание и дисперсию числа отказавших элементов. Построить график функции распределения.

7. Точка M движется по оси Ox по закону $x = ae^t$. В случайный момент времени, равномерно распределенный на отрезке $[0;\ T]$, наблюдается положение ξ точки M. Найти плотность распределения, математическое ожидание и стандартное отклонение случайной величины ξ.

8. Четыре поезда метро, уходящие с интервалом в 4 минуты, увезли по 200 пассажиров. Четыре поезда, уходящие с интервалом в 6 минут, увезли по 300 пассажиров. Два поезда, уходящие с интервалом в 8 минут, увезли по 100 пассажиров. Найти совместное распределение числа пассажиров и интервала движения для выбранного наудачу поезда. Найти коэффициент корреляции.

9. Двумерная плотность распределения случайных величин X, Y задается формулой $f(x, y) = c$, если $y \geq 0$, $|x| + y \leq 1$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Количество бракованных изделий в коробке имеет распределение

Пуассона с параметром 4. Найти максимальное число коробок такое, чтобы вероятность найти в них более 200 бракованных изделий была меньше 0,04.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 4$ по второму моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} (\theta - 2)x^{-\theta+1} & \text{при } x > 1; \\ 0 & \text{при } x \leq 1. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.
14,29 7,48 2,82 22,84 7,49 8,98 13,84 14,17 7,07 9,69 -8,35 12,77 14,93 5,81 8,62 11,22 3,85 2,86 9,52 15,93 9,43 19,48 19,19 12,20 19,40 12,09 8,47 6,79

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,19 0,91 0,30 1,34 0,61 1,12 1,00 0,53 1,58 0,62
0,41 0,89 1,20 1,51 0,78 1,44 0,46 0,69 1,33

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	-1	2	3	6	7
Y	-3	-1	2	-4	4

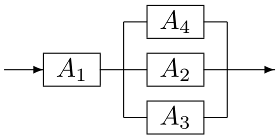
Вариант 29

1. Может ли сумма двух событий A и B совпадать с их произведением? Привести соответствующие примеры.

2. В бригаде 4 рабочих. Какова вероятность того, что по крайней мере двое из них родились в один и тот же месяц? Считать, что вероятности родиться в каждый месяц одинаковы.

3. Случайная точка A наудачу выбирается в прямоугольном треугольнике с катетами 1 и 2. Найти вероятность того, что расстояние от A до ближайшей стороны треугольника не превосходит 1/3.

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_2 равна 0,01, остальных элементов A_k — по 0,1. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Телеграфное сообщение состоит из сигналов «точка» и «тире». Известно, что среди передаваемых сигналов «точка» и «тире» встречаются в отношении 11:10. Из-за помех искажается в среднем 30 % сигналов «точка» и 20 % сигналов «тире», причем «точка» искажается в «тире», а «тире» в «точку». Найти вероятность искажения сигнала. Определить вероятность того, что сигнал не был искажен, если известно, что приняли «точку».

6. Два игрока играют в шахматы на деньги. Известно, что в среднем из 10 партий три выигрывает первый игрок, три заканчиваются вничью, и четыре выигрывает второй игрок. В случае проигрыша первый игрок платит второму 30 рублей. Сколько он должен получать в случае выигрыша, чтобы математическое ожидание его выигрыша равнялось нулю? Найти ряд распределения и дисперсию суммы выигрыша (отрицательная сумма выигрыша — это сумма проигрыша, взятая со знаком «минус»). Построить график функции распределения.

7. Случайная величина ξ имеет плотность распределения

$$f(x) = \begin{cases} \frac{A}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)} & \text{при } |x-a| > 2\sigma; \\ 0 & \text{при} |x-a| \le 2\sigma. \end{cases}$$

Найти нормирующую константу A, вычислить математическое ожидание. Построить график плотности распределения при $a = \sigma = 1$.

8. В подъезде 5 однокомнатных квартир площадью по 40 кв. м., 10 двухкомнатных квартир по 60 кв. м. и 5 трехкомнатных квартир по 70 кв. м. Для выбранной наудачу квартиры найти совместное распределение числа комнат и площади. Найти коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y имеет вид: $f(x,y) = c$, если $x, y \ \geq 0$, $x + 3y \leq 3$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Суммарное время работы машины складывается из интервалов времени, каждый из которых измеряется со стандартным отклонением в 1 минуту. Найти максимальное число интервалов времени такое, чтобы фактическое время работы отличалось от измеренного не больше, чем на 2 часа, с вероятностью 0,95.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{x}{\theta^6} e^{-x/\theta^3} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

22,95 -2,61 11,87 1,37 5,92 -5,10 5,38 14,71 7,55 3,91 1,23 8,50 -5,58 -1,97 17,93 9,42 11,99 9,39 4,78 5,43 9,40 8,68 2,20 7,15 14,78 14,77 -15,16

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке $[0;\ 2]$. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

1,94 1,96 0,64 0,76 0,01 0,82 0,23 0,82 1,96
1,28 1,49 1,07 1,92 0,17 1,68 1,01 0,48

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	-3	-1	-10	-4	4

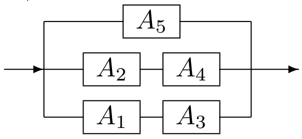
Вариант 30

1. Может ли разность двух событий совпадать с их произведением? Привести примеры.

2. В чулане лежит три пары ботинок. Случайно выбираются три ботинка. Чему равна вероятность того, что среди них не будет ни одной пары?

3. На линейке наудачу поставлены 2 точки. Какова вероятность того, что расстояние между ними окажется меньше трети длины линейки?

4. Электрическая цепь состоит из элементов A_k, соединенных по следующей схеме:



Вероятность выхода из строя элемента A_2 равна 0,01, остальных элементов A_k — по 0,1. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что цепь будет пропускать ток.

5. Первое орудие 4-орудийной батареи пристреляно так, что вероятность попадания для него равна $1/2$. Для остальных орудий она равна $2/5$. Батарея дала залп по цели. Найти вероятность того, что цель поражена. Найти вероятность того, что первое орудие попало в цель, если известно, что цель была поражена. Для поражения цели достаточно одного попадания.

6. Вероятность приема отдельного сигнала равна 0,3. Радиосигнал передается 6 раз. Найти ряд распределения, математическое ожидание и дисперсию числа принятых сигналов. Построить график функции распределения. Найти вероятность того, что принятых сигналов будет не меньше 2, но не больше 4.

7. Диаметр круга является случайной величиной, равномерно распределенной на отрезке [0; d]. Найти плотность распределения, функцию распределения, математическое ожидание и дисперсию площади круга. Найти вероятность того, что площадь превосходит $\pi d^2/32$. Начертить графики плотности распределения и функции распределения.

8. В отделе работает один сотрудник с двумя высшими образованиями по 13-му разряду, два сотрудника с высшим образованием по 12-му разряду, и шесть сотрудников без высшего образования по 10-му разряду. Для выбранного наудачу сотрудника найти совместное распределение разряда

и количества высших образований. Вычислить коэффициент корреляции между ними.

9. Двумерная плотность распределения случайных величин X, Y имеет вид: $f(x, y) = c$, если $x, y \ \geq 0$, $3x + y \leq 3$, и 0 иначе. Найти константу c и $\rho(X, Y)$.

10. Время ожидания автобуса пассажиром имеет показательное распределение со средним значением 9 минут. Найти число поездок, для которого суммарное время ожидания автобуса превысит 3 часа с вероятностью не более 0,2.

11. Для выборки $(X_1,\ X_2,\ \ldots,\ X_n)$ из распределения с плотностью распределения $f(x)$ найти оценки параметра $\theta > 0$ по первому моменту и методом максимального правдоподобия. Проверить состоятельность полученных оценок. Плотность распределения равна

$$f(x) = \begin{cases} \frac{x^2}{2\theta^9} e^{-x/\theta^3} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0. \end{cases}$$

12. Дана выборка из нормального распределения с неизвестными параметрами. Найти оценки параметров распределения. Подставляя вместо неизвестных параметров их точечные оценки, записать выражение для оценки плотности распределения. Построить на одном графике гистограмму с шагом, равным среднеквадратическому (стандартному) отклонению, и график оценки плотности распределения.

9,63 19,48 3,89 4,45 15,11 15,90 24,94 1,72 3,25 -3,77 12,17 10,08 14,36 9,39 1,27 7,89 8,68 1,59 10,57 3,21 -6,11 15,61 10,82 1,68 5,63 6,79 20,27 -2,15

13. По критерию Колмогорова проверить гипотезу о том, что выборка имеет равномерное распределение на отрезке [0; 2]. Сделать вывод о том, принимается ли эта гипотеза на уровне значимости 0,1; на уровне значимости 0,01; на уровне значимости 0,001.

0,64 1,43 0,40 1,23 1,40 0,76 1,09 1,65 1,32
1,24 1,39 0,81 0,39 0,76 1,14 1,24 1,69 1,58

14. Оценить параметры нормальной регрессии Y на X по двумерной выборке. Изобразить на чертеже точки двумерной выборки и прямую линейной регрессии.

X	1	2	3	4	5
Y	3	1	2	4	6

Приложение. Таблицы

Т а б л и ц а 1. **Нормальное распределение.**

Значения функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ и функции $\overline{\Phi}(x) = \Phi(-x) = 1 - \Phi(x)$. [1]

x	$\Phi(-x)$	$\Phi(x)$
4,75	0,000001	0,999999
4,26	0,00001	0,99999
3,72	0,0001	0,9999
3,09	0,001	0,999
2,58	0,005	0,995
2,33	0,01	0,99
2,05	0,02	0,98
1,96	0,025	0,975
1,88	0,03	0,97
1,75	0,04	0,96
1,64	0,05	0,95
1,28	0,1	0,9
0,84	0,2	0,8
0,52	0,3	0,7
0,25	0,4	0,6
0,00	0,5	0,5

[1] Для $x > 4,75$ можно использовать аппроксимацию $\overline{\Phi}(x) \sim \frac{e^{-x^2/2}}{x\sqrt{2\pi}}$.

Т а б л и ц а 2. Распределение $\chi^2(n)$. Квантили распределения:

$$p = \int_0^{\chi^2_{p,n}} k_n(x)dx = \frac{1}{2^{n/2}\Gamma(n/2)} \cdot \int_0^{\chi^2_{p,n}} x^{n/2-1}e^{-x/2}dx$$

n \ p	0,1	0,3	0,5	0,7	0,9	0,95	0.999	0,9999
1	0,016	0,148	0,455	1,07	2,71	3,84	6,63	10,8
2	0,211	0,713	1,39	2,41	4,61	5,99	9,21	13,8
3	0,584	1,42	2,37	3,67	6,25	7,82	11,3	16,3
4	1,06	2,20	3,36	4,88	7,78	9,49	13,3	18,5
5	1,61	3,00	4,35	6,06	9,24	11,1	15,1	20,5
6	2,20	3,83	5,35	7,23	10,6	12,6	16,8	22,5
7	2,83	4,67	6,35	8,38	12,0	14,1	18,5	24,3
8	3,49	5,53	7,34	9,52	13,4	15,5	20,1	26,1
9	4,17	6,39	8,34	10,7	14,7	16,9	21,7	27,9
10	4,87	7,27	9,34	11,8	16,0	18,3	23,2	29,6
11	5,58	8,15	10,3	12,9	17,3	19,7	24,7	31,3
12	6,30	9,03	11,3	14,0	18,5	21,0	26,2	32,9
13	7,04	9,93	12,3	15,52	13,4	15,5	20,1	26,1
14	7,79	10,08	13,3	16,2	21,1	23,7	29,1	36,1
15	8,55	11,7	14,3	17,3	22,3	25,0	30,6	37,7
16	9,31	12,6	15,3	18,4	23,5	26,3	32,0	39,3
17	10,09	13,5	16,3	19,5	24,8	27,6	33,4	40,8
18	10,9	14,4	17,3	20,6	26,0	28,9	34,8	42,3
19	11,7	15,4	18,3	21,7	27,2	30,1	36,2	43,8
20	12,4	16,3	19,3	22,8	28,4	31,4	37,6	45,3
21	13,2	17,2	20,3	23,9	29,6	32,7	38,9	46,8
22	14,0	18,1	21,3	24,9	30,8	33,9	40,3	48,3
23	14,8	19,0	22,3	26,0	32,0	35,2	41,6	49,7
24	15,7	19,9	23,3	27,1	33,2	36,4	43,0	51,2
25	16,5	20,9	24,3	28,2	34,3	37,7	44,3	52,6
26	17,3	21,8	25,3	29,2	35,6	38,9	45,6	54,1
27	18,1	22,7	26,3	30,3	36,7	40,1	47,0	55,5
28	18,9	23,6	27,3	31,4	37,9	41,3	48,3	56,9
29	19,8	24,6	28,3	32,5	39,1	42,6	49,6	58,3
30	20,6	25,5	29,3	33,5	40,3	43,8	50,9	59,7

Т а б л и ц а 3. Распределение Стьюдента S(n)
Значения функции $t_{\gamma,n}$

$$\frac{1+\gamma}{2} = \int\limits_{-\infty}^{t_{\gamma,n}} s_n(x)dx = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \cdot \int\limits_{-\infty}^{t_{\gamma,n}} (1+\frac{x^2}{n})^{-(n+1)/2}dx$$

$n \diagdown \gamma$	0,9	0,95	0,98	0,99
1	6,314	12,706	31,821	63,657
2	2,920	4,303	6,965	9,925
3	2,353	3,182	4,541	5,841
4	2,132	2,776	3,747	4,604
5	2,015	2,571	3,365	4,032
6	1,943	2,447	3,143	3,707
7	1,895	2,365	2,998	3,499
8	1,860	2,306	2,896	3,355
9	1,833	2,262	2,821	3,250
10	1,812	2,228	2,764	3,169
12	1,782	2,179	2,681	3,055
14	1,761	2,145	2,625	2,977
16	1,746	2,120	2,584	2,921
18	1,734	2,101	2,552	2,878
20	1,725	2,086	2,528	2,845
22	1,717	2,074	2,508	2,819
24	1,711	2,064	2,492	2,797
26	1,706	2,056	2,479	2,779
28	1,701	2,048	2,467	2,763
30	1,697	2,042	2,457	2,750
∞	1,645	1,960	2,326	2,576

Список литературы

[1] *Аркашов Н.С., Бородихин В.М., Ковалевский А.П.* Высшая математика. — Новосибирск: НГТУ, 2008. — Т. 4.2: Теория вероятностей и математическая статистика. — 228с.

[2] *Боровков А.А.* Теория вероятностей. — М.: Эдиториал УРСС, 1999. — 470с.

[3] *Боровков А.А.* Математическая статистика. — Новосибирск: Наука, 1997. — 772с.

[4] *Бородин А.Н.* Элементарный курс теории вероятностей и математической статистики. — СПб., 1999. — 223с.

[5] *Бородихин В.М.* Теория вероятностей и математическая статистика: Практикум. — Новосибирск, 2000. — Ч. 1. — 159с.

[6] *Бородихин В.М.* Теория вероятностей и математическая статистика: Практикум. — Новосибирск, 2001. — Ч. 2. — 105с.

[7] *Бородихин В.М., Ковалевский А.П.* Высшая математика. — Новосибирск: НГТУ, 2005. — Т. 4.2: Теория вероятностей и математическая статистика. — 256с.

[8] *Гнеденко Б.В.* Теория вероятностей. — М.: Наука, 1969. — 400с.

[9] *Емельянов Г.В., Скитович В.П.* Задачник по теории вероятностей и математической статистике. — Л.: Изд-во Ленинградского ун-та, 1967. — 332с.

[10] *Ермаков С.М., Михайлов Г.А.* Статистическое моделирование. — М.: Наука, 1982.

[11] *Зубков А.М., Севастьянов Б.А., Чистяков В.П.* Сборник задач по теории вероятностей. — Л.: Наука, 1989. — 320с.

[12] *Ивченко Г.И., Медведев Ю.И.* Математическая статистика. — М.: Высшая школа, 1984. — 248с.

[13] *Ивченко Г.И., Медведев Ю.И., Чистяков А.В.* Сборник задач по математической статистике. — М.: Высшая школа, 1989. — 255с.

[14] *Коршунов Д.А., Фосс С.Г., Эйсымонт И.М.* Сборник задач и упражнений по теории вероятностей. — СПб., 2004. — 192с.

[15] *Коршунов Д.А., Чернова Н.И.* Сборник задач и упражнений по математической статистике. — Новосибирск, 2001. — 120с.

[16] *Лотов В.И.* Теория вероятностей и математическая статистика. — Новосибирск: НГУ, 2006. — 128с.

[17] *Свешников А.А. и др.* Сборник задач по теории вероятностей, математической статистике и теории случайных функций. — М., 1970. — 656с.

[18] *Чистяков В.П.* Курс теории вероятностей. — М.: Наука, 1987. — 240с.

[19] *Чернова Н.И.* Теория вероятностей. — Новосибирск: НГУ, 2007. — 160с.

[20] *Чернова Н.И.* Математическая статистика. — Новосибирск: НГУ, 2007. — 148с.

Интернет-источники (PDF)

1. Лотов В.И. Лекции по теории вероятностей и математической статистике.
http://www.nsu.ru/mmf/tvims/lotov/tv&ms_ff.pdf

2. Коршунов Д.А., Фосс С.Г. Сборник задач и упражнений по теории вероятностей.
http://www.math.nsc.ru/LBRT/v1/dima/ExerciseProbability2.pdf

3. Коршунов Д.А., Чернова Н.И. Сборник задач и упражнений по математической статистике.
http://www.math.nsc.ru/LBRT/v1/dima/ExerciseStatistics2.pdf

Интернет-источники (HTML)

1. Чернова Н.И. Лекции по теории вероятностей.
http://www.nsu.ru/mmf/tvims/chernova/tv/index.html

2. Чернова Н.И. Лекции по математической статистике.
http://www.nsu.ru/mmf/tvims/chernova/ms/lec/ms.html

Аркашов Николай Сергеевич
Ковалевский Артем Павлович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И СЛУЧАЙНЫЕ ПРОЦЕССЫ

Учебное пособие

В авторской редакции
Выпускающий редактор *И.П. Брованова*
Дизайн обложки *А.В. Ладыжская*
Компьютерная верстка *Н.С. Аркашов, А.П. Ковалевский*

Подписано в печать 14.01.2014. Формат 60×84 1/16. Бумага офсетная. Тираж 200 экз.
Уч.-изд. л. 13,95. Печ. л. 15,0. Изд. №351/13. Заказ №138. Цена договорная.

Отпечатано в типографии
Новосибирского государственного технического университета
630073, г. Новосибирск, пр. К. Маркса, 20