Министерство образования Республики Беларусь

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«ГРОДНЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ ЯНКИ КУПАЛЫ»

М.А. Маталыцкий, Т.В. Русилко

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебно-методическое пособие для студентов
Института последипломного образования
и студентов, получающих высшее образование
по сокращенной форме обучения

Гродно 2007

УДК 519.2 (075.8)
ББК 22.171
М33

Рецензенты: доктор физико-математических наук, профессор *Н.Н. Труш*,
кандидат физико-математических наук, доцент *А.А. Крушельницкий*.

Рекомендовано Советом факультета математики и информатики ГрГУ им. Я. Купалы

Маталыцкий, М.А.

М33 Теория вероятностей и математическая статистика : учеб.- метод. пособие / М.А. Маталыцкий, Т.В. Русилко. – Гродно : ГрГУ, 2007. – 219 с.

ISBN

В учебно-методическом пособии даются основы теории вероятностей и математической статистики. В каждом из параграфов кратко излагается теоретический материал, приведены решения типовых примеров, даны задачи для самостоятельного решения различной степени трудности. Представленные задачи могут быть использованы при составлении контрольных работ и индивидуальных домашних зданий для студентов очной и заочной форм обучения. Адресовано студентам инженерных и экономических специальностей, а также лицам, занимающимся самообразованием, инженерным работникам, которые интересуются теорией вероятностей, математической статистикой и их применениями.

УДК 519.2 (075.8)
ББК 22.171

ISBN

Предисловие

За последнее десятилетие значительно увеличился объем преподавания теории вероятностей в высших учебных заведениях. В университетах и институтах для студентов различных специальностей читается полугодовой или годовой курс теории вероятностей и математической статистики, обязательным является раздел, связанный с вероятностью и статистикой, для студентов инженерных и экономических специальностей.

Предлагаемое учебно-методическое пособие является пособием по теории вероятностей и математической статистики для слушателей Института последипломного образования (ИПО) и студентов факультета непрерывного образования, получающих высшее образование по сокращенной форме обучения. Оно написано на достаточно строгом математическом уровне, но в то же время понятном студентам-нематематикам. Примеры и задачи также подобраны соответствующим образом. Пособие поможет студентам в самостоятельной работе и выполнении контрольных работ. Каждый из параграфов пособия имеет введение, где приводятся краткие сведения о понятиях и утверждениях теории вероятностей, необходимых для решения задач, приводятся решения типовых примеров. В пособии дается большое число задач различной степени трудности; в нем представлено значительное число задач «прикладного» характера, что позволяет не только обучить студента теоретическим основам, но и привить навыки вероятностно-статистического моделирования реальных явлений.

При написании пособия был использован ряд отечественных и зарубежных учебников и задачников, приведенных в списке литературы. Некоторые из задач составлены авторами.

Выражаем благодарность рецензентам, сделавшим ряд полезных замечаний.

Глава 1
ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§1. Случайные события. Классическое определение вероятности

Определение. Возможные события, порождаемые комплексом условий, называются элементарными, если:

а) они различны (т.е. осуществление одного означает неосуществление любого другого);

б) после выполнения комплекса условия обязательно происходит одно из них.

Обозначим через $\Omega = \{\omega_1, \omega_2, ..., \omega_n, ...\}$ множество (пространство) элементарных событий.

Определение. Любое объединение элементарных событий называется случайным событием, $B \subseteq \Omega$.

Событие B осуществляется тогда, когда происходит одно из элементарных событий $\omega \in B$. В этом смысле множество может рассматриваться тоже как событие. Т.к. одно из элементарных событий происходит всегда, то и событие Ω происходит всегда, поэтому оно является достоверным событием. Событие, не содержащее ни одного элементарного события, является невозможным и обозначается .

Таким образом, мы пришли к описанию случайных событий как множеств, получающихся объединением элементарных событий. В связи с этим для определения соотношений между случайными событиями в теории вероятностей принят язык теории множеств, который приобретает своеобразную вероятностную трактовку. Поясним некоторые из них при помощи таблицы 1.

Таблица 1

Продолжение таблицы 1

Язык теории множеств	Соотношения между событиями на языке теории множеств
$B \cap C = \varnothing$	События B и C являются несовместными. Если событие C происходит, то событие B не происходит.
$C \subseteq B$	Событие C влечет за собой событие B.
$A = \Omega \setminus B$, $\left(A = \overline{B}\right)$	Событие A является дополнительным (противоположным) к событию B. Событие A происходит тогда и только тогда, когда не происходит событие B.

Определение *(классическое определение вероятности)*. Пусть множество элементарных событий состоит из конечного числа равновозможных элементарных событий $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ и пусть случайное событие A состоит из n элементарных событий: $A = \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_n}\}$, $\omega_{j_i} \in \Omega$, $i = \overline{1, n}$. Тогда вероятностью события A называется число

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

Данная формула называется формулой классической вероятности.

Пример 1.1. Монету бросают дважды. Найти вероятность того, что хотя бы один раз герб окажется наверху.

Решение. Множеством элементарных событий является множество $\Omega = \{\text{ГГ}, \text{ГЦ}, \text{ЦГ}, \text{ЦЦ}\}$. Здесь, например, ГЦ означает, что при первом бросании появился герб, а при втором – цифра. Таким образом: $N = 4$,

$$A = \{\text{ГГ}, \text{ГЦ}, \text{ЦГ}\} \Rightarrow n = 3 \Rightarrow P(\text{A}) = \frac{3}{4}.$$

Основной проблемой при решении задач с использованием формулы классической вероятности является подсчет числа способов, которыми могло произойти то или иное событие. В связи с этим такие задачи решаются, как правило, методами комбинаторики.

Часто используется следующее очевидное правило (основной принцип комбинаторики): если некий выбор A можно осуществить

m различными способами, а некоторый другой выбор можно осуществить n способами, то выбор и B (A или B) можно осуществить $mn\left(m+n\right)$ способами.

При этом классическое определение вероятности можно дать другими словами.

Определение. Рассмотрим эксперимент, имеющий N одинаково возможных исходов (любой мыслимый результат эксперимента называется элементарным событием). Предположим, что событию A благоприятствует n из этих исходов (оно состоит из n элементарных событий). Тогда справедлива формула классической вероятности.

Рассмотрим свойства классической вероятности:

1. Для любого события $P(A) \geq 0$, поскольку $n > 0$, $N \geq 0$.

2. Для достоверного события Ω , так как оно состоит из N элементарных событий.

3. Если события A и B несовместны, то $P(A \bigcup B) = P(A) + P(B)$, поскольку если событие A содержит n_1 элементарных событий, а событие B – n_2 элементарных событий, то событие $A \bigcup B$ содержит $n_1 + n_2$ элементарных событий; и, следовательно, $P(A \bigcup B) = \frac{n_1 + n_2}{N} = \frac{n_1}{N} + \frac{n_2}{N} = P(A) + P(B)$.

Остальные свойства вытекают из этих трех свойств.

4. Вероятность противоположного события $\overline{A}$ равна $P(\overline{A}) = 1 - P(A)$, так как $A \bigcup \overline{A} = \Omega$, $A \bigcap \overline{A} = \varnothing$, $1 = P(\Omega) = P(A) + P(\overline{A})$.

5. Если $A \leq B$, то , поскольку $= (B \bigcap A) \bigcup (B \bigcap \overline{A}) = A \bigcup (B \bigcap \overline{A})$, и $P(B) = P(A) + P(B \bigcap \overline{A}) \geq P(A)$.

6. Для любых событий A и B $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$.

7. Для любого события A $0 \le P(A) \le 1$, так как $P(A) + P(\overline{A}) = 1$ и $P(A) \ge 0$, $P(\overline{A}) \ge 0$.

При решении задач часто используют размещения, перестановки и сочетания. Если дано множество $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$, то размещением (сочетанием) из n элементов по называется любое упорядоченное (неупорядоченное) подмножество из k элементов множества Ω. При размещение называется перестановкой из элементов.

Пусть, например, дано множество Размещениями из трех элементов этого множества по два являются $(\omega_1, \omega_2), (\omega_1, \omega_3), (\omega_2, \omega_3), (\omega_2, \omega_1), (\omega_3, \omega_1), (\omega_3, \omega_2)$; сочетаниями: $(\omega_1, \omega_2), (\omega_1, \omega_3), (\omega_2, \omega_3)$. Два сочетания отличаются друг от друга хотя бы одним элементом, а размещения отличаются либо самими элементами, либо порядком их следования.

Число сочетаний из n элементов по вычисляется по формуле

$$\mathrm{C}_n^k = \frac{\mathrm{A}_n^k}{\mathrm{P}_k} = \frac{n!}{k!(n-k)!},$$

$\Omega = \{\omega_1, \omega_2, \omega_3\}$.

где $\mathrm{A}_n^k = \dfrac{n!}{(n-k)!} = n(n-1)...(n-k+1)$ – число размещений из n элементов по , а $\mathrm{P}_k = k!$ – число перестановок из k элементов.

Рассмотрим перестановки с повторениями. Пусть из элементов $\omega_1, \omega_2, ..., \omega_i$ образуются конечные последовательности, содержащие n членов, в которых повторяется k_1 раз, $\omega_2 - k_2$ раза, …, $\omega_i - k_i$ раз, $k_1 + k_2 + ... + k_i = n$. Такие последовательности называются перестановками с повторениями. Две перестановки считаются одинаковыми, если они совпадают порядком расположения элементов и считаются различными, если у них различный порядок расположения элементов. Число различных перестановок с повторениями равно

$$\mathrm{P}_n(k_1, k_2, ..., k_i) = \frac{n!}{k_1!k_2!...k_i!}.$$

Пример 1.2. Какова вероятность того, что из шести отмеченных чисел в карточке «спортлото» (игра 6 из 49) k чисел будут выигрышными, $k=\overline{0,6}$.

Решение. В данном примере эксперимент состоит в том, что случайным образом отмечаются 6 чисел из 49 в карточке «спортлото». Поэтому равновозможными элементарными событиями будут наборы из шести отмеченных чисел. Т.к. для определения того, произойдет или не произойдет событие A – среди отмеченных чисел k чисел выигрышные, – порядок чисел не существен, то в качестве равновозможных элементарных событий достаточно рассматривать неупорядоченные наборы 6 чисел из 49. Следовательно, число равновозможных элементарных событий равно C_{46}^{6}. Событие A состоит из наборов 6 чисел, k из которых выигрышные, а 6-k проигрышные. Набор из k выигрышных чисел можно выбрать C_{6}^{k} способами, а набор 6-k проигрышных чисел можно выбрать C_{43}^{6-k} способами. Тогда по основному принципу комбинаторики набор из k выигрышных и 6-k проигрышных чисел можно выбрать $\mathrm{C}_{6}^{k}\mathrm{C}_{43}^{6-k}$ способами, следовательно,

$$P(A)=\frac{\mathrm{C}_{6}^{k}\mathrm{C}_{43}^{6-k}}{\mathrm{C}_{49}^{6}}.$$

Например, для $k=6$ имеем $P(A)\approx\left(14\cdot 10^{6}\right)^{-1}$.

Задачи

1.1. Игральный кубик бросают дважды. Найти вероятность того, что сумма выпавших очков не превосходит 4.

1.2. Среди n лотерейных билетов   выигрышных. Наудачу взяли m билетов. Определить вероятность того, что среди них   выигрышных.

1.3. В партии, состоящей из N изделий, имеется k дефектных. В процессе приемного контроля из партии выбирается n изделий.

Найти вероятность того, что из них ровно изделий будут дефектными.

1.4. Рабочий у конвейера при сборе механизма устанавливает в него две одинаковые детали. Берет он их случайным образом из имеющихся у него 10 штук. Среди деталей находятся 2 уменьшенного размера. Механизм не будет работать, если обе установленные детали окажутся уменьшенного размера. Определить вероятность того, что механизм будет работать.

1.5. человек случайным образом рассаживаются за круглым столом $(K > 2)$. Найти вероятность того, что два фиксированных лица A и B окажутся рядом.

1.6. Определить вероятность того, что серия наудачу выбранной облигации не содержит одинаковых цифр, если номер серии может быть любым пятизначным числом, начиная с 00001.

1.7. На десяти карточках написаны буквы А, А, А, М, М, Т, Т, Е, И, К. После перестановки вынимают наугад одну карточку за другой и раскладывают их в том порядке, в каком они были вынуты. Найти вероятность того, что на карточках будет написано слово «МАТЕМАТИКА».

1.8. Из разрезной азбуки составляют слово «ЭКОНОМИКА». Затем все буквы этого слова перемешиваются и снова вкладываются в случайном порядке. Какова вероятность того, что снова получится слово «ЭКОНОМИКА».

1.9. Телефонный номер в г. Гродно – из шести цифр. Найти вероятность того, что все цифры различны.

1.10. Какова вероятность того, что четырехзначный номер случайно взятого автомобиля в г. Гродно имеет все цифры различные?

1.11. К четырехстороннему перекрестку подъехало с каждой стороны по автомобилю. Каждый автомобиль может с равной вероятностью совершить один из четырех маневров на перекрестке: развернуться и поехать обратно, поехать прямо, направо или налево. Через некоторое время все автомобили покинули перекресток. Найти вероятность следующих событий: $\{$все автомобили поедут по одной и той же улице$\}$, $B = \{$3 автомобиля поедут по одной и той же улице$\}$, $C = \{$по каждой из четырех улиц поедет один автомобиль$\}$.

1.12. Некоторые жители г. Гродно и других городов шестизначный номер троллейбусного или автобусного билета считают «счастливым», если сумма первых его трех цифр совпадает с суммой последних трех цифр. Найти вероятность получить «счастливый» билет.

1.13. В лифт двенадцатиэтажного дома на первом этаже вошли пять человек. Каждый из них с одинаковой вероятностью выходит на любом этаже, начиная со второго. Найти вероятность того, что все пассажиры выйдут: а) на одном и том же этаже; б) на восьмом этаже.

1.14. На полке в случайном порядке расставлено 20 книг, среди которых находится трехтомник Янки Купалы. Найти вероятность того, что эти тома стоят в порядке возрастания слева направо (но не обязательно рядом).

1.15. Для беспрепятственного полета над некоторой территорией самолет, приближаясь к ней, посылает по радио парольную кодовую группу, состоящую из нескольких точек и тире. Найти вероятность того, что радист, не знающий парольной группы, угадает ее, передав какую-нибудь группу наугад, если известно, что число кодовых элементов в группе (точек и тире) равно пяти.

1.16. В ящике имеется K типовых элементов замены (ТЭЗ), из них K_1 элементов 1-го типа, …, K_i элементов i-го типа, …,

элементов n-го типа; . Из ящика выбирают наугад k ТЭЗ.

Найти вероятность того, что среди них будет k_1 ТЭЗ 1-го типа, …, k_i

ТЭЗ i-го типа, …, ТЭЗ n-го типа.

1.17. За перенумерованными ПЭВМ будут работать n студентов (один студент – за одной ПЭВМ). Каждый студент выбирает любую ПЭВМ случайно и с одинаковой вероятностью. Найти вероятность того, что для работы будут выбраны ПЭВМ с номерами .

1.18. Для работы на N ПЭВМ случайным образом распределяются K студентов. Под состоянием совокупности из N ПЭВМ будем понимать вектор $k = (k_1, k_2, ..., k_N)$, где k_i – число студентов,

которые выполняют свое задание на i-й ПЭВМ, . Состояния считаются различными, если им соответствуют векторы с различными компонентами. Найти: а) число состояний сети, б) вероятности состояний, предполагая, что все состояния равновозможные.

1.19. Из 30 экзаменационных вопросов студент знает 20. Какова вероятность того, что он правильно ответит на два вопроса из двух?

1.20. Найти вероятность того, что в группе из 25 студентов найдутся, по меньшей мере, два, которые имеют общий день рождения.

1.21. Пакет из десяти различных сообщений должен быть передан по электронной почте. Сообщения передаются одно за другим произвольным образом. Определить вероятность того, что сообщение A будет передано раньше, чем сообщение B.

1.22. По линии связи в случайном порядке передают 30 букв русского алфавита. Найти вероятность того, что на ленте появится последовательность букв, которые образуют слово «МИНСК».

1.23. По N каналам связи случайным образом передают K сообщений, $N > K$. Определить вероятность того, что на каждый канал припадет не более одного сообщения.

1.24. Используя условие предыдущей задачи, найти вероятность того, что среди каналов n_0 таких, по которым не будет передано ни одного сообщения, n_1 таких, по которых будет передано только одно сообщение, …, n_K таких, по которым будет передано K сообщений;

$$\sum_{i=0}^{K} n_i = N, \quad \sum_{i=1}^{K} i n_i = K .$$

1.25. По N каналам связи, которые пронумерованы, случайным образом передаются K сообщений. Какова вероятность того, что по 1-му каналу будет передано k_1 сообщений, 2-му – $k_2, \ldots,$ N-му каналу – k_N сообщений, причем $\sum_{i=1}^{N} k_i = K$.

1.26. Принимаются кодовые комбинации, в которые входят десять цифр от 0 до 9, при этом цифры не повторяются. Какова вероятность того, что в принятой комбинации цифры образуют последовательность 9876…210?

1.27. В N ячейках случайно размещены n частиц. Чему равна вероятность того, что в -ю ячейку попало частиц?

1.28. Газ, состоящий из K молекул, находится в замкнутом сосуде. Мысленно разделим сосуд на K равных клеток и будем считать, что вероятность для каждой молекулы попасть в любую из клеток одна и та же. Какова вероятность того, что молекулы распределятся так, что в 1-й клетке будет n_1 молекул, во 2-й – n_2 молекул, …, в K -й – n_K молекул?

1.29. Из колоды карт (52 карты) наугад вынимают 3 карты. Найти вероятность того, что это будут тройка, семерка и туз.

1.30. Из колоды карт (52 карты) наугад вынимают 6 карт. Определить вероятность того, что среди этих карт: а) будет дама пик; б) - будут карты всех мастей.

1.31. Полная колода карт (52 карты) делится пополам. Найти вероятность того, что: а) число черных и красных карт в обеих пачках будет одинаковым (по 13), б) в каждой половине будет по два туза.

1.32. Из колоды в 36 карт наугад выбираются 4. Найти вероятность того, что среди них окажется хотя бы один туз.

1.33. Из колоды в 36 карт берется наугад 10 карт. Найти вероятность того, что среди них будут 8 одномастных.

1.34. В очереди, где продаются билеты по $5, стоят n человек. Какова вероятность того, что ни одному из покупателей не придется ждать сдачи, если перед началом продажи денег у кассира не было, а получение платы за каждый билет равновозможно как 5-и, так и 10-долларовыми купюрами?

1.35. Решить предыдущую задачу при условии, что перед продажей билетов у кассира было 5-долларовых купюр.

1.36. В лотерее 100 билетов, среди них один выигрыш в $50, 3 выигрыша по $25, 6 выигрышей по $10 и 15 – по $3. Найти вероятность какого-нибудь выигрыша при покупке трех лотерейных билетов. Что вероятнее: выиграть не менее $25 или не более $25 при покупке одного лотерейного билета?

1.37. В лотерее K билетов, из них m выигрышных. Найти вероятность одного выигрыша для лица, имеющего билетов.

1.38. Пусть эксперимент состоит в проведении голосования по стратегии развития компании собранием из K членов. Каждый сотрудник может голосовать «за», «против» или воздержаться от голосования. Найти число элементарных событий в Ω, если голосование является а) открытым, б) тайным. Если в процессе обсуждения сотрудники могут менять свое мнение, то сколько элементов содержит Ω, если голосование проводится дважды (двумя способами)?

§2. Геометрическое определение вероятности

Геометрическая вероятность является расширением понятия классической вероятности на случай несчетного множества элементарных событий. В случае, когда – несчетное множество, вероятность определяется не на элементарных событиях, а на их множествах.

Определение. Пусть из области G выбирается точка таким образом, что выбор точки из некоторой области A, содержащейся в G, объективно не имеет преимуществ перед выбором точки из любой другой области, содержащейся в G, с мерой, равной мере области
$\Omega \leq a$ A, какой бы формы она не была. Такой выбор называется выбором с равновозможными исходами. Пусть событие A состоит в том, что точка будет выбрана из области A. Тогда вероятностью события A называется число

$$P(A) = \frac{mesA}{mesG},$$

где $mesA$ – мера (на прямой – длина, на плоскости – площадь, в пространстве – объем) области A.

Геометрическая вероятность обладает всеми свойствами классической вероятности. Первые три свойства следуют из определения, остальные из первых трех.

Пример 2.1 *(Задача Бюффона).* На пол, построенный из досок шириной a, бросается игла длиной . Найти вероятность того, что игла пересечет линию пола.

Решение. Положение иглы относительно линий пола зададим двумя координатами: – расстояние от нижнего конца иглы до ближайшей верхней линии,   – угол между иглой и направлением линий пола, рис. 1.

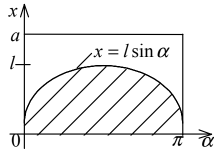

Рис. 1 Рис. 2

Независимо друг от друга они могут принять одно из значений в пределах: $0 \le x \le a$, $0 \le \alpha \le \pi$, причем каждое из этих значений равновозможно. Этим значениям соответствует точка прямоугольника со сторонами a и , рис. 2. Тогда все элементарные события можно поставить в соответствие с точками этого прямоугольника; из сказанного выше следует, что любая точка прямоугольника будет равновозможна, .

Пусть $A=$ {игла пересечет линию пола}, тогда $A = \{(x, \alpha): x \le l \sin\alpha\}$, такие точки заполняют заштрихованную область, рис. 2.

Мера в данном случае является площадью. Из геометрического определения вероятности имеем $P(A) = \dfrac{mesA}{mes\Omega} = \dfrac{\int_0^{\pi} l \sin\alpha \, d\alpha}{a\pi} = \dfrac{2l}{l\pi}$. Из этого соотношения, пользуясь тем, что частота появления события A близка к вероятности этого события, можно найти приближенное значение числа π: , где N – число бросаний иглы, $n(A)$ – число тех из них, в которых игла пересекла линию пола. Учеником Бюффона около 400 лет назад было проделано $\approx$ 6000 таких опытов и получено правильное приближенное значение числа   с точностью до четырех знаков после запятой.

Задачи

2.1. Два студента имеют одинаковую вероятность прийти к указанному месту в любой момент промежутка времени . Определить вероятность того, что время ожидания одним другого будет не больше t.

2.2. По маршруту независимо друг от друга ходят два автобуса: № – через 10 минут, №15 – через 7 минут. Студент приходит на остановку в случайный момент времени. Какова вероятность, что ему придется ждать автобуса менее трех минут.

2.3. Дано уравнение $x^2 + ax + b = 0$. Известно, что $0 \le a \le 1$, $0 \le b \le 1$, причем вероятность попадания каждого из чисел a и в какой-либо интервал отрезка $[0,1]$ пропорциональна длине интервала и не зависит от его положения относительно отрезка $[0,1]$. Найти вероятность того, что данное уравнение имеет действительные корни.

2.4. Найти вероятность того, что сумма двух наудачу взятых положительных правильных дробей не больше 1, а их произведение не больше $\dfrac{3}{16}$.

[0< a < b#

2.5. Два студента договорились встретиться в определенном месте между 19 и 20 часами. Пришедший первым ждет второго в течение 5 минут. Найти вероятность того, что встреча состоится, если каждый студент наудачу выбирает момент своего прихода в промежутке от 19 до 20 часов.

2.6. На паркетный пол случайным образом падает монета диаметром d размеры паркетных плиток $a \times b$, причем Какова вероятность того, что упавшая монета не пересечет границ паркетной плитки?

2.7. На отрезке случайным образом выбираются две точки, Какова вероятность того, что из отрезков, полученных разбиением отрезка $[0,1]$ этими точками, можно построить треугольник?

2.8. На бесконечную шахматную доску, сторона каждой клетки равна $2a$, бросают монету радиуса $r < a$. Найти вероятность того, что монета попадает внутрь одной клетки целиком.

2.9. Два танкера должны подойти к одному и тому же причалу. Время прихода обоих танкеров равновозможно в течение одних суток. Найти вероятность, что одному из танкеров придется ждать освобождения причала, если время разгрузки первого танкера – три часа, а второго – четыре часа.

2.10. Два судна плывут в тумане: одно идет вдоль пролива шириной , а другое курсирует без остановок поперек этого пролива. Скорости движения судов соответственно равны v_1 и v_2. Второе судно подает звуковые сигналы, которые слышны на расстоянии $l < L$. Определить вероятность того, что на первом судне услышат сигналы, если пересечение курсов судов равновозможно в любом месте пролива.

2.11. Катер перевозит груз с одного берега на другой, пересекая пролив за один час. Какова вероятность того, что судно, которое движется вдоль пролива, будет замечено, если с катера замечают судно в случае пересечения его курса не раньше, чем за 20 мин до пересечения с курсом катера, и не позже, чем через 20 мин после пересечения судном курса катера? Любой момент и любое место пересечения судном курса катера равновозможны. Курс судна перпендикулярен курсу лодки.

2.12. В круге радиуса наудачу выбирают точку. Вероятность попадания точки в некоторую область круга пропорциональна площади этой области. Определить вероятность того, что: а) точка находится от центра на расстоянии меньшем, чем r, $r < R$; б) меньший угол между заданным направлением и прямой, которая соединяет точку с началом координат, будет не больше, чем α.

2.13. На окружности с радиусом 1 и центром в начале координат наудачу выбирают точку. Вероятность выбора точки на некоторой дуге окружности зависит только от длины этой дуги и пропорциональна ей. Найти вероятность того, что: а) проекция точки на диаметр находится от центра на расстоянии не большем, чем ; б) расстояние от выбранной точки до точки с координатами (1,0) не больше, чем r.

2.14. Спутник Земли двигается по орбите, которая заключена между северной и 60° южной широты. Найти вероятность того, что спутник упадет выше 30° северной широты, если считать равновозможным падение спутника в любую точку поверхности Земли между указанными параллелями.

2.15. Слой воздуха толщиной H задерживает пылинки радиусом r в количестве штук в одной кубической единице. Найти вероятность того, что луч света, перпендикулярный слою, не пересечет ни одной пылинки.

2.16. Электрон вылетает из случайной точки нити накаливания и движется перпендикулярно ей. С какой вероятностью он свободно пройдет через сетку, которая окружает нить и имеет вид винтовой линии с радиусом , толщиной σ и шагом ?

2.17. Рассмотрим частицу с энергией $E = \frac{mv^2}{2}$, которая движется в случайном направлении. Пусть $(v_1, v_2, v_3)^2$ – вектор скорости частицы в некоторой системы координат. Какова вероятность того, что $\alpha \le v_1 \le \beta$?

2.18. На круглом экране радиолокатора радиусом r имеется точечное отображение объекта, которое занимает случайное положение в границах экрана, причем ни одна зона в границах не имеет преимущества перед другой. Найти вероятность того, что расстояние от точки объекта до центра экрана будет меньше, чем .

2.19. По радиолокатору в течение промежутка времени $(0, T)$

$(0, T-\tau)$ $\frac{a}{2}$, $d < a < \sqrt{l^2 + d^2}$

передаются два сигнала длительностью $\tau < T$ и каждый с одинаковой возможностью начинается в любой момент интервала .
Когда сигналы перекрывают друг друга хотя бы частично, они оба искажаются и приняты быть не могут. Найти вероятность того, что сигналы будут приняты.

2.20. Самолет с радиолокационной станцией, дальность действия которой L, в районе площадью s осуществляет поиск подводной лодки со скоростью . Лодка может всплыть в любой точке района на время . Найти вероятность обнаружения подводной лодки радиолокатором, если время t невелико.

2.21. Имеются две параллельные линии связи длиной , расстояния между которыми $d < l$. Известно, что на каждой линии где-то есть разрыв, но неизвестно, в каком месте. Найти вероятность того, что расстояние между точками разрыва не больше, чем .

2.22. Панорамный приемник периодически с постоянной скоростью проходит полный диапазон частот (f_1, f_2), где возможно появление сигнала, за которым установлено наблюдение. Полоса пропускания приемника определяется допустимой расстройкой относительно сигнала $\pm\Delta f$. Считая сигнал импульсным (выявленной точкой как на оси времени, так и на оси частот), появление его равновозможным в любой момент и в любой точке интервала , найти вероятность обнаружения сигнала.

2.23. Используя условие предыдущей задачи, найти вероятность пеленга передатчика, если известна частота сигнала и то, что антенна пеленгатора вращается равномерно с углом раскрытия диафрагмы антенны $\alpha = 20^\circ$.

2.24. Поезда метро идут в данном направлении с интервалом 2 мин. Какова вероятность того, что пассажиру доведется ждать поезда не больше, чем 30 с?

2.25. Концентрация доходов различных социальных групп изображается кривой Лоренца. Пусть наудачу выбираются социальные слои, суммарная доля x которых от всего населения изменяется в интервале а суммарный относительный доход y изменяется соответственно в интервале Найти вероятность события, состоящего в том, что наудачу выбранная часть населения будет иметь относительный доход, удовлетворяющий соотношению

2.26. Состояние работы банка за сутки характеризуется суммарной величиной d_1 вкладов от индивидуальных вкладчиков и не зависящей от нее величиной d_2 вкладов от фирм. Работа банка оценивается его правлением как успешная, если $d_1 + d_2 > 0$ и выполняется пропорция вкладов: $d_1 + d_2 > d$, где $d > 0$ – заданный коэффициент. Предполагая равновероятность значений $d_i \in \left[\underline{d}_i, \overline{d}_i\right]$, $i = 1,2,$ вычислить вероятность того, что итоги работы банка в течение суток успешны.

§3. Условная вероятность и независимость событий

Определение. Предположим, что $P(B)>0$. Условной вероятностью события A при условии, что произошло событие B, называется отношение вероятностей $P(A\cap B)$ и $P(B)$ и обозначается $P(A/B)$, т.е.

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

Она обладает следующими свойствами:

а) если $B\subseteq A$, то $P(A/B)=1$, т.к. в этом случае $A\cap B=B$;

б) если $A\cap B=\varnothing$, то $P((A\cup B)/C)=P(A/C)+P(B/C)$, т.к. $(A\cap C)\cap(B\cap C)=\varnothing, P((A\cup B)\cap C)=P(A\cap C)+P(B\cap C)$;

в) $P(\overline{A}/B)=1-P(A/B)$, т.к. $1=P(\Omega/B)=P((A\cup\overline{A})/B)=P(A/B)+$ $+P(\overline{A}/B)$.

Пример 3.1. Бросается игральный кубик. Какова вероятность того, что выпало число очков большее трех (событие A), если известно, что выпала четная грань (событие B)?

Решение. Событию B соответствует выпадение чисел 2, 4, 6; событию A – выпадение чисел 4, 5, 6; событию $A\cap B$ – 4, 6. Поэтому

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{2}{6}:\frac{3}{6}=\frac{2}{3}.$$

Определение. Два события A и B называются независимыми, если $P(A\cap B)=P(A)P(B)$.

Определение. События $A_1, A_2, ..., A_n$ называются независимыми в совокупности, если для любых наборов индексов $\{i_1, i_2, ..., i_m\}$ таких, что $1\le i_1<i_2<...<i_m\le n$, $2\le m\le n$, имеет место равенство

$$P\left(A_{i_1}\cap A_{i_2}\cap ...\cap A_{i_m}\right)=\prod_{k=1}^{m}P\left(A_{i_k}\right).$$

Если это равенство справедливо только для случая, когда $m = 2$, то события называются попарно независимыми.

Пример 3.2. Бросаются две монеты. Пусть событие A состоит в выпадении герба на первой монете, событие B состоит в выпадении цифры на второй монете; событие C – монеты выпадут разными сторонами. В этом случае:

$$\Omega = \{\text{ГЦ}, \text{ГГ}, \text{ЦГ}, \text{ЦЦ}\},\ A = \{\text{ГЦ}, \text{ГГ}\},\ B = \{\text{ГЦ}, \text{ЦЦ}\},\ C = \{\text{ГЦ}, \text{ЦГ}\},$$

$$AB = \{\text{ГЦ}\};\ P(AB) = \frac{1}{4} = P(A)P(B);\ P(AC) = \frac{1}{4} = P(A)P(C);$$

$$P(BC) = \frac{1}{4} = P(B)P(C);\ P(ABC) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8},$$

т.е. A, B, C – попарно независимы (но независимости в совокупности здесь нет).

Следует отметить, что на практике независимость событий проверяется не из определения, а исходя из условий эксперимента: можно показать, что если события связаны с независимыми экспериментами, то и сами события будут независимыми.

Справедливо следующее утверждение, известное как теорема умножения вероятностей. Пусть для событий A_k, $k = 1, 2, ..., n$; $P\left(\bigcap_{k=1}^{m} A_k\right) > 0$ для всех $1 \le m \le n$, тогда

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1)\prod_{m=1}^{n-1} P\left(A_{m+1} / \bigcap_{k=1}^{m} A_k\right).$$

Пример 3.3. Абонент забыл последнюю цифру номера телефона и поэтому набирает ее наугад. Определить вероятность того, что ему придется звонить не более чем три раза.

Решение. Обозначим через A_i событие, заключающееся в том, что абонент звонит i-й раз и ему соответствует неудача, . Тогда имеем:

$$P(A_1) = 1 - \frac{1}{10} = \frac{9}{10}, P(A_2/A_1) = \frac{8}{9}, P(A_3/A_1A_2) = \frac{7}{8}.$$

Искомая вероятность равна:

$$1 - P(A_1 A_2 A_3) = 1 - P(A_1) P(A_2/A_1) P(A_3/A_1 A_2) = 0{,}3 .$$

Задачи

3.1. Студент пришел на экзамен, зная лишь 20 из 25 вопросов программы. Экзаменатор задал студенту 3 вопроса. Используя понятие условной вероятности, найти вероятность того, что студент знает все эти вопросы.

3.2. Вероятность попадания в первую мишень для стрелка равна $\frac{2}{3}$. Если при первом выстреле зафиксировано попадание, то стрелок получает право на второй выстрел по другой мишени. Вероятность поражения обеих мишеней при двух выстрелах равна 0,5. Определите вероятность поражения второй мишени.

3.3. Сколько нужно взять чисел из таблицы случайных целых чисел, чтобы с вероятностью не менее 0,9 быть уверенным, что среди них хотя бы одно число четное?

3.4. Для некоторой местности среднее число ясных дней в июле
m равно 25. Найти вероятность того, что первые два дня в июле будут ясными.

3.5. В обществе из $2n$ человек одинаковое число мужчин и женщин. Места за столом занимают наудачу. Определить вероятность того, что два лица одного пола не займут места рядом.

3.6. Имеется группа из k космических объектов, каждый из которых независимо от других обнаруживается радиолокационной станцией с вероятностью p. За группой объектов ведут наблюдение независимо друг от друга   радиолокационных станций. Найти вероятность того, что не все объекты, входящие в группу, будут обнаружены.

3.7. Определить вероятность того, что выбранное наудачу изделие является первосортным, если известно, что 5 % всей продукции является браком, а 75 % не бракованных изделий удовлетворяют требованиям первого сорта.

3.8. Партия из ста деталей подвергается выборочному контролю. Условием непригодности всей партии является наличие хотя бы одной бракованной детали среди пяти проверяемых. Какова вероятность для данной партии быть не принятой, если она содержит 5 % неисправных деталей?

3.9. Производится испытание прибора. При каждом испытании прибор выходит из строя с вероятностью . После первого выхода из строя прибор ремонтируется; после второго – признается негодным. Найти вероятность того, что прибор окончательно выйдет из строя в точности при -м испытании.

3.10. Вероятность выхода из строя k -го блока ЭВМ за время T равна $p_k, k = 1, 2, .., n$. Определить вероятность выхода из n блоков ЭВМ, если работа всех блоков взаимно независима.

3.11. ЭВМ, в которой подозревается дефект, подвергается тестированию с целью локализации дефекта. Для этого применяется последовательно тестов. При обнаружении дефекта тестирование прекращается. Вероятность локализации дефекта при первом тесте равна ; условная вероятность локализации дефекта при втором тесте (если при первом он не был локализован) равна p_2 ; условная вероятность локализации дефекта на i -м тесте (если при первых он не был локализован) равна p_i , $i = 1, 2, .., n$. Найти вероятности следующих событий: а) проведено не менее трех тестов; б) проведено не более трех тестов; в) дефект локализован в точности при четвертом тесте; г) дефект не локализован после тестов; д) проведены все n тестов.

3.12. Каждая буква слова «ЭЛЕКТРОНИКА» написана на отдельной карточке, которые тщательно перемешаны. Последовательно вынимают четыре карточки. Какова вероятность получить слово «КИНО»?

3.13. Вероятность того, что некоторое устройство космического корабля испортится, равна . Сколько запасных устройств нужно иметь на корабле, чтобы обеспечить вероятность правильной работы не меньшую, чем ?

3.14. Измерительное устройство состоит из двух приборов. Вероятность исправной работы k -го прибора за рассматриваемый промежуток времени равна $1 - \alpha_k, k = 1, 2$. Найти вероятность того, что оба прибора будут работать: а) если известно, что поломки в них возникают независимо, б) если ничего не известно о зависимости между поломками этих приборов.

3.15. Проведены три независимых измерения некоторой физической величины. Вероятность того, что при одном измерении ошибка превысит заданную точность, равна p. Определить вероятность того, что только в одном из измерений ошибка превысит заданную точность.

3.16. Во время стрельбы ракетами по цели попадания отдельных ракет независимы и вероятность попадания каждой ракеты равна . Каждая ракета поражает цель с вероятностью . Стрельба ведется до поражения цели или до израсходования всего боезапаса; количество ракет $n > 2$. Найти вероятность того, что не весь боезапас будет израсходован.

3.17. Сообщение, которое передают по каналу связи, состоит из n знаков. При передаче каждый знак искажается независимо от других с вероятностью . Для надежности сообщение дублируют, т.е. повторяют раз. Какова вероятность того, что хотя бы одно из переданных сообщений не будет искажено полностью.

p p_1

3.18. Для того чтобы найти специальную книгу, студент решил обойти три библиотеки. Наличие книги в фонде библиотеки одинаково равновероятно, и если книга есть, то одинаково вероятно, занята она другим читателем или нет. Что более вероятно – возьмет студент книгу или нет, если библиотеки комплектуются независимо одна от другой.

3.19. На железнодорожном вокзале пассажир воспользовался автоматической камерой хранения багажа, шифр который состоит из одной буквы русского алфавита и трехзначного цифрового кода. Пассажир набрал шифр, запер сейф, но, возвратившись, забыл свой шифровой набор. Найти вероятность событий: $A = \{$сейф открывается при первой попытке$\}$, $B = \{$сейф открывается после k попыток$\}$.

3.20. Дворцовый чеканщик кладет в каждый ящик вместительностью n монет одну фальшивую. Король подозревает чеканщика и подвергает проверке монеты, взяв наудачу по одной в каждом ящике. Какова вероятность того, что чеканщик не будет разоблачен?

3.21. Прибор состоит из блоков, которые выходят из строя независимо один от другого. Надежность (вероятность безотказной работы) каждого блока равна . Найти надежность прибора для случаев, которые изображены на рис. 3.

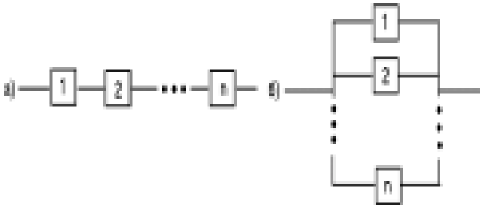

Рис. 3

3.22. Система управления состоит из четырех узлов, рис.4. Вероятности их безотказной работы соответственно равны: 0,7; 0,6; 0,8; 0,9. Вычислить вероятность безотказной работы всей системы управления.

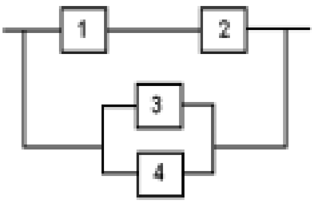

Рис. 4

3.23. Электрическая цепь составлена из пяти элементов, рис. 5. При выходе из строя любого элемента цепь в месте его включения разрывается. Вероятность выхода из строя за данный период элемента под номером , равна $p_k, k = \overline{1,5}$. Предполагается, что элементы выходят из строя независимо друг от друга. Найти вероятность того, что на рассматриваемом периоде по цепи будет проходить ток.

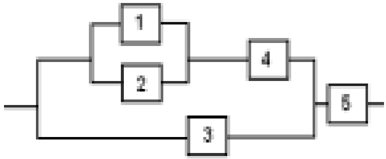

Рис. 5

3.24. Прибор состоит из узлов. Вероятность безотказной работы -го узла равна . Для работы прибора требуется безотказная работа всех его узлов. При вычислении вероятности R отказа прибора вероятности $p_i, i = 1, 2, ..., n$, приближенно заменяют их средней арифметической:

$p_i, i = 1, 2, ..., n$

$$\overline{p} = \frac{1}{n}\sum_{i=1}^{n} p_i \ .$$

Будет ли при этом вычисленное приближенное значение $\overline{R}$ вероятности R больше или меньше истинного R ?

3.25. Упрощенная схема контроля качества изделий состоит из двух независимых проверок. В результате k -й проверки изделие, удовлетворяющее стандарту, отбраковывается с вероятностью β_k , а бракованное изделие принимается с вероятностью $\alpha_k, k = 1, 2$. Изделие принимается, если оно прошло обе проверки. Найти вероятности следующих событий: а) будет принято бракованное изделие, б) изделие, удовлетворяющее стандарту, будет отбраковано.

3.26. Уникальный прибор, от которого требуется очень большая надежность, состоит из k деталей $D_1, D_2, ..., D_k$. Перед сборкой каждую деталь всесторонне проверяют и, если она окажется высококачественной, включают в прибор, а если нет – заменяют запасным экземпляром, который также проверяют. Сборщик имеет в наличии запас деталей каждого типа: n_i экземпляров детали D_i, $i = \overline{1,k}, \sum_{i=1}^{k} n_i = n$. Если запасных деталей не хватает, сборка откладывается. Вероятность того, что отдельный экземпляр детали D_i окажется высококачественным, равна P_i и не зависит от качества других экземпляров. Найти вероятности следующих событий: A ={имеющегося запаса деталей достаточно для сборки прибора}, B ={при данном запасе деталей сборщик может собрать прибор, и хотя бы одна деталь любого типа останется в запасе}.

3.27. Рабочий обязан поддерживать функционирование автоматической линии, состоящей из трех станков, работающих независимо друг от друга. Вероятность того, что в течение смены первой станок не потребует наладки, равна 0,9, для второго – 0,8, для третьего – 0,85. Какова вероятность того, что в течении смены: а) линия не потребует наладки; б) все три станка потребуют наладки; в) какой-нибудь один станок потребует наладки; г) хотя бы один станок потребует наладки.

3.28. Баллотируются два кандидата, причем за первого в урну опущено n бюллетеней, за второго бюллетеней, . Какова вероятность того, что на протяжении всего времени подсчета бюллетеней количество подсчитанных голосов, которые отданы за первого кандидата, будет больше числа голосов, отданных за второго?

§4. Формула полной вероятности и формула Байеса

Определение. Совокупность событий называется полной группой событий, если выполнены следующие условия:

а) $\bigcup_{k=1}^{n} A_k = \Omega$; б) $A_i \cap A_k = \varnothing$, $i \neq k$; в) $P(A_k) > 0$, $k = \overline{1,n}$.

Если $A_1, A_2, \ldots, A_n$ образуют полную группу событий, то для любого события

$$B = B \bigcap \Omega = B \bigcap \left(\bigcup_{k=1}^{n} A_k \right) = \bigcup_{k=1}^{n} \left(B \bigcap A_k \right),$$

и поэтому

$$P(B) = P\left(\bigcup_{k=1}^{n} (B \cap A_k) \right) = \sum_{k=1}^{n} P(B \cap A_k) = \sum_{k=1}^{n} P(A_k) P(B/A_k).$$

Формула

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

называется формулой полной вероятности для случайных событий.

Если $P(B) > 0$, то

$$P(A_k \bigcap B) = P(B) P(A_k/B) = P(A_k) P(B/A_k).$$

Отсюда следует, что

$$P(A_k/B) = \frac{P(A_k) P(B/A_k)}{\sum\limits_{i=1}^{n} P(A_i) P(B/A_i)}, \quad k = \overline{1,n},$$

$A_1,\ A_2, ...,\ A_n$ т.е. имеет место формула Байеса для случайных событий:

$$P(A_k/B) = \frac{P(A_k) P(B/A_k)}{\sum\limits_{i=1}^{n} P(A_i) P(B/A_i)}, \quad k = \overline{1,n}.$$

При решении задач удобно применять следующую формулировку: если событие B может происходить только с одним из событий $A_1,\ A_2, ...,\ A_n$, образующих полную группу событий, то при $P(B) > 0$ справедливы формула полной вероятности и формула Байеса.

Формула Байеса используется также в следующем случае. Пусть событие B может происходить в различных условиях, относительно которых выдвигается n гипотез . Предположим, что известны вероятности $P(A_i)$ и $P(B/A_i)$, $i = \overline{1, n}$. Далее проводиться эксперимент, в результате которого происходит событие B. Это должно вызывать переоценку вероятностей гипотез. Появляются вероятности $P(A_i / B)$. $i = \overline{1, n}$, которые вычисляются по формуле Байеса

Пример 4.1. Для контроля продукции из трех партий деталей взята на проверку одна деталь. Какова вероятность выявления бракованной продукции, если в одной партии 2/3 деталей – бракованные, а в двух других все доброкачественные.

Решение. Пусть $B = \{$взятая деталь – бракованная$\}$, $A_k = \{$деталь берется из k-й партии$\}$, $k = \overline{1, 3}$. Тогда

$$P(A_k) = \frac{1}{3}, k = \overline{1, 3}; \; P(B/A_1) = \frac{2}{3}, \; P(B/A_2) = P(B/A_3) = 0$$

и поэтому $P(B) = \sum_{k=1}^{3} P(A_k) P(B/A_k) = \frac{2}{9}$.

Пример 4.2. Прибор состоит из двух узлов; работа каждого узла необходима для работы прибора в целом. Надежность (вероятность безотказной работы в течение времени t первого узла равна , второго p_2. Прибор испытывался в течение времени t, в результате чего обнаружено, что он вышел из строя (отказал). Найти вероятность того, что отказал только первый узел, а второй исправен.

Решение. Пусть $= \{$оба узла исправны$\}$, $A_2 = \{$первый узел отказал, а второй исправен$\}$, $A_3 = \{$первый узел исправен, а второй отказал$\}$, $A_4 = \{$оба узла отказали$\}$. Эти события образуют полную группу событий. Найдем их вероятности.

$$P(A_1) = p_1 p_2, \qquad , \qquad ,$$

.

Поскольку наблюдалось событие $\{$прибор отказал$\}$, то

$$P(B/A_1) = 0, \; P(B/A_2) = P(B/A_3) = P(B/A_4) = 1.$$

По формуле Байеса находим:

$$P(A_2/B) = \frac{(1 - p_1) p_2}{(1 - p_1) p_2 + p_1 (1 - p_2) + (1 - p_1)(1 - p_2)} = \frac{(1 - p_1) p_2}{1 - p_1 p_2}.$$

Задачи

4.1. Среди N экзаменационных билетов n «счастливых». Студенты подходят за билетами один за другим. У кого больше вероятность взять счастливый билет: у того, кто подошел первым, или у того, кто подошел вторым? Какова вероятность взять «счастливый» билет у последнего студента?

4.2. 15 экзаменационных билетов содержат по 2 вопроса, которые не повторяются. Экзаменующийся может ответить только на 25 вопросов. Определить вероятность того, что экзамен будет сдан, если для этого достаточно ответить на два вопроса из одного билета или на один вопрос из первого билета и на указанный дополнительный вопрос из другого билета.

4.3. В техникуме студентов, из которых человек учатся k -й год. Среди двух наудачу выбранных студентов оказалось, что один из них учится больше второго. Какова вероятность того, что этот студент учится третий год?

4.4. Из чисел $1,\ 2, ...,\ n$ одно за другим выбирают наугад два числа. Какова вероятность того, что разность между первым выбранным числом и вторым будет не меньше m ?

$M_k,\ k=1, 2, 3$

4.5. Водитель автомобиля, оказавшись в неизвестной местности, пытается наудачу попасть из пункта в пункт B . Оценить шансы водителя (вероятность) попасть в пункт B , если возвращения запрещены, а схема дорог приведена на рис. 6.

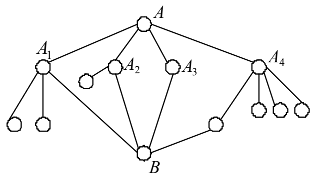

Рис. 6

4.6. Вероятность того, что письмо находится в столе, равна ,
причем оно с равной вероятностью может находиться в любом из восьми ящиков стола. Проверили семь ящиков – письма не нашли. Какова вероятность того, что письмо в восьмом ящике?

4.7. В группе 10 студентов. Трое подготовились к экзамену на оценку «отлично», четверо на «хорошо», двое на «удовлетворительно», один на «неудовлетворительно». В экзаменационных билетах 20 вопросов. Отличник знает ответ на все вопросы, хороший студент – на 16 вопросов, посредственный – на 10, плохой – на 5. Вызванный студент ответил на все три вопроса. Найти вероятность, что он

а) отличник,

б) плохой студент.

4.8. Вероятности того, что во время работы ЭВМ произойдет сбой в процессоре, в памяти, в остальных устройствах, относятся как 3:2:5. Вероятности обнаружения сбоя в процессоре, в памяти и в остальных устройствах соответственно равны 0,8; 0,9; 0,9. Найти вероятность того, что возникший в ЭВМ сбой будет обнаружен.

4.9. Студент приходит в лабораторию, в которой находятся ПЭВМ, для выполнения лабораторной работы. Вероятности того, что ПЭВМ заражены вирусом, равны соответственно . Для работы студент выбирает наугад одну из ПЭВМ. При выполнении работы на исправной ПЭВМ (без вирусов) студент ошибается в среднем в k % случаев. Найти вероятность того, что студент правильно выполнит лабораторную работу.

4.10. К серверу подключены N ПЭВМ. Вероятность того, что запросы, поступающие на сервер из одной ПЭВМ в момент времени t, прекратятся до момента , равна . Если в момент времени t из ПЭВМ на сервер не поступают запросы, то вероятность того, что они начнут поступать до момента , равна независимо от работы других ПЭВМ. Составить дифференциальные уравнения, которым будут удовлетворять $P_n(t)$ – вероятности того, что в момент t на сервер поступают запросы из ПЭВМ. Найти стационарное решение (при) этих уравнений.

4.11. Вероятность поступления на телефонную станцию вызовов за промежуток времени t равна . Будем считать количество вызовов за любые два соседних промежутка времени независимыми. Определить вероятность $P_{2t}(k)$ поступления s вызовов за промежуток времени .

4.12. Производится посадка самолета не аэродром. Если позволяет погода, летчик сажает самолет, наблюдая за аэродромом визуально. В этом случае вероятность благополучной посадки равна p_1. Если аэродром затянут низкой облачностью, летчик сажает самолет вслепую по приборам. Надежность (вероятность безотказной работы) приборов слепой посадки равна P. Если приборы слепой посадки сработали нормально, то самолет садится благополучно с той же вероятностью p_1, что и при визуальной посадке. Если же приборы слепой посадки не сработали, то летчик может благополучно посадить самолет с вероятностью $p_2 < p_1$. Найти полную вероятность благополучной посадки самолета, если известно, что в k % всех случаев посадки аэродром затянут низкой облачностью.

4.13. В условии предыдущей задачи известно, что самолет приземлился благополучно. Найти вероятность того, что летчик пользовался приборами слепой посадки.

4.14. Расследуются причины авиационной катастрофы, о которых можно сделать четыре гипотезы H_1, H_2, H_3, H_4. По статистике, их вероятности равны соответственно 0,2; 0,4; 0,3; 0,1. Выявлено, что в ходе катастрофы произошло возгорание горючего. Условные вероятности этого события при гипотезах H_1, H_2, H_3, H_4, согласно той же статистике, составляют 0,9;0;0,2; 0,3. Найти апостериорные вероятности гипотез.

4.15. Для поиска пропавшей подводной лодки выбрано 10 самолетов, и каждый из них можно использовать для поиска в одном из двух возможных районов, где лодка может находиться с вероятностями 0,8 и 0,2 соответственно. Как нужно распределить самолеты по районам поиска, чтобы вероятность обнаружения лодки была наибольшей, если каждый самолет выявляет ее в районе поиска с вероятностью 0,2 и осуществляет поиски независимо от других? Найти вероятность обнаружения лодки при оптимальной процедуре поисков.

4.16. Объект, за которым ведется наблюдение, может находиться в одном из состояний A_1 или A_2, априорные вероятности которых равны p_1 и p_2 соответственно. Есть два источника информации о состоянии объекта: из первого известно, что объект находится в состоянии A_1, из другого – что в состоянии A_2. Из первого источника правильные сведения о состоянии объекта поступают в 90 %, а из второго – в 60 % случаев. На основе анализа донесений найти апостериорные вероятности состояний A_1 и A_2.

4.17. Перед опытом о его условиях можно сделать n независимых гипотез , которые образуют полную группу с априорными вероятностями $p_1, p_2, ..., p_n$. В результате опыта известно, что имела место некоторая гипотеза из группы $H_1, H_2, ..., H_k$, $k < n$, а остальные гипотезы невозможны: , $H_{k+1} \bigcup H_{k+2} \bigcup ... \bigcup H_n = \varnothing$. Найти апостериорные вероятности гипотез.

4.18. Прибор может принадлежать к одной из трех партий с вероятностями p_1, p_2, p_3, где $p_1 = p_3 = 0{,}25$, $p_2 = 0{,}5$. Вероятности того, что прибор будет работать заданное число часов, равны для этих партий соответственно 0,1, 0,2 и 0,4. Определить вероятность того, что этот прибор проработает заданное число часов.

4.19. Апробируется прибор, состоящий из двух блоков, которые выходят из строя независимо один от другого; работа каждого блока безусловно необходима для работы прибора в целом. Надежность (вероятность исправной работы на протяжении времени t) равна для них соответственно и p_2. По истечении времени t выяснилось, что прибор вышел из строя. Найти с учетом этого вероятности наступления следующих событий: = {вышел из строя только первый блок}, A_2 = {вышел из строя только второй блок}, A_3 = {вышли из строя оба блока}.

4.20. Во время испытаний было установлено, что вероятность безотказной работы прибора при отсутствии повреждений равна 0,99, при перегреве – 0,95, при вибрации – 0,9, при вибрации и перегреве – 0,8. Найти вероятность P_1 отказа этого прибора во время работы в жарких странах (вероятность перегрева – 0,2, вибрации – 0,1) и вероятность P_2 отказа во время работы в передвигающейся лаборатории (вероятность перегрева – 0,1, вибрации – 0,3), если считать перегрев и вибрацию независимыми событиями.

4.21. В условиях задачи 4.20 найти границы, в которых могут лежать вероятности P_1 и P_2, если отказаться от предложения о независимости перегрева и вибрации.

4.22. По каналу связи передается одна из последовательностей букв (команда) AAAA, BBBB, CCCC с вероятностями p_1, p_2, p_3 $(p_1+p_2+p_3=1)$. Каждая независимая буква принимается правильно с вероятностью α и с вероятностями и $\frac{1}{2}(1-\alpha)$ принимается за каждую из двух других букв. Предполагается, что буквы искажаются независимо друг от друга. Найти вероятность того, что было передано AAAA, если принято ABCA.

4.23. По каналу связи передают символы A, B, C с вероятностями 0,4; 0,3; 0,3 соответственно. Вероятность искажения символа равна 0,4, и все искажения равновероятны. Для увеличения надежности каждый символ повторяют четыре раза. На выходе восприняли

n_1 p_i, $i=\overline{1,5}$ $\frac{1}{2}(1-\alpha)$

последовательность BACB. Какова вероятность того, что передали AAAA, BBBB, CCCC?

4.24. Сообщение может передаваться по каждому из n каналов связи, находящихся в разных состояниях: каналов – в отличном состоянии, n_2 – в хорошем, n_3 – в посредственном и n_4 – в плохом, $n_1+n_2+n_3+n_4=n$. Вероятности правильной передачи сообщения для различных типов каналов равны соответственно p_1, p_2, p_3, p_4. Для увеличения точности сообщения его передают два раза по двум разным каналам, которые выбирают наугад. Найти вероятность того, что хотя бы по одному из каналов оно будет передано правильно.

4.25. Есть пять каналов связи, передача сообщений по которым распределяется случайным образом с равной вероятностью. Вероятность искажения сообщения при его передаче по i-му каналу равна . Выбран некоторый канал и по нему переданы $n-1$ сообщений; ни одно из них не исказилось. Найти вероятность того, что n - е сообщение, переданное по тому же самому каналу, не будет искажено.

4.26. Передача сигналов происходит с вероятностями , в одном из трех режимов. В каждом из них сигнал доходит до адресата неискаженным помехами с вероятностями соответственно p_1, p_2, p_3. Передача трех сигналов происходила в одном из трех режимов, в каком – неизвестно. Найти апостериорные вероятности того, что передача происходила в первом, втором и третьем режимах.

4.27. На вход радиолокационного устройства с вероятностью p поступает смесь полезного сигнала с помехой, а с вероятностью – только одна помеха. Когда поступает полезный сигнал с помехой, то устройство регистрирует наличие какого-либо сигнала с вероятностью p_1, если поступает только помеха – с вероятностью p_2. Известно, что устройство зарегистрировало присутствие какого-то сигнала. Найти вероятность того, что в его составе присутствует полезный сигнал.

4.28. Радиолокационная станция ведет наблюдение за объектом, который может применять либо не применять помехи. Если объект не применяет помех, то за один цикл наблюдения станция обнаруживает его с вероятностью p_0, если применяет – с вероятностью $p_1 < p_0$. Вероятность того, что на протяжении цикла будут применяться помехи, равна p и не зависит от того, как и когда применялись помехи во время остальных циклов. Определить вероятность того, что объект будет обнаружен хотя бы один раз за циклов наблюдения.

4.29. На наблюдательной станции установлены четыре радиолокатора различных конструкций. Вероятность обнаружения целей с помощью первого локатора равна 0,86, второго – 0,9, третьего – 0,92, четвертого – 0,95. Наблюдатель наугад включает один из локаторов. Какова вероятность обнаружения цели?

4.30. Самолет, вылетающий на задание, производит радиопомехи, которые с вероятностью 0,5 «забивают» радиосредства системы противовоздушной обороны (ПВО). Если радиосредства «забиты», самолет подлетает к объекту необстрелянным, производит стрельбу ракетами и поражает объект с вероятностью 0,9. Если радиосредства системы ПВО «не забиты», то самолет подвергается обстрелу и сбивается с вероятность 0,8. Найти вероятность того, что объект будет разрушен.

4.31. Цель, по которой ведется стрельба, с вероятностью находится в пункте , а с вероятностью $(1-p)$ – в пункте B. В распоряжении стреляющего есть N ракет, каждая из которых поражает цель с вероятностью P независимо от других. Какое количество ракет нужно выпустить по пункту A для того, чтобы поразить цель с максимальной вероятностью?

4.32. Противник может применять ракеты трех типов A, B, C с вероятностями: $P(A)=0{,}3;\ P(B)=0{,}6;\ P(C)=0{,}1$. Вероятности сбить ракету этих типов, равны соответственно, 0,6; 0,8; 0,9. Известно, что противник применил две ракеты одного типа. Определить вероятность того, что обе ракеты будут сбиты.

4.33. Вероятность размножения бактерии в течение времени $(t, t+\Delta t)$ равна $\beta\Delta t+o(\Delta t)$, $\Delta t\to 0$. Процесс размножения каждой бактерии протекает независимо от других бактерий и ее поведения до момента t. В начальный момент в банке было бактерий. Найти вероятность того, что в момент в банке будет бактерий.

4.34. Частица блуждает по целым точкам отрезка , причем движется направо с вероятностью p и налево с вероятностью . Определить вероятность того, что она достигнет правого конца, если в начальный момент находится в точке $n\in[a,b]$.

4.35. Некоторая деталь производится на двух заводах. Известно, что объем продукции первого завода в n раз превышает объем продукции второго завода. Доля брака на первом заводе , на втором P_2. Наугад взятая деталь оказалась бракованной. Какова вероятность того, что эта деталь выпущена первым заводом?

4.36. Трое сотрудников фирмы выдают соответственно 30 %, 50 % и 20 % всей продукции изделий, производимых фирмой. У первого брак составляет 2 %, у второго – 5 %, у третьего – 1 %. Какова вероятность, что случайно выбранное изделие фирмы дефектно?

4.37. В условиях предыдущей задачи известно, что случайно выбранное изделие оказалось дефектным. Найти вероятность, что оно было сделано соответственно первым, вторым и третьим сотрудником фирмы.

4.38. Отдел входного технического контроля (ОТК) проверяет взятые наугад изделия из партии, содержащей, по данным поставщика, a изделий первого сорта и изделий второго сорта. ОТК считает возможным количество изделий первого сорта в размере $a-2,\ a-1,\ a,\ a+1$ с вероятностями соответственно $p_i,\ \ i=\overline{1,4}$. Проверка первых m изделий обнаружила, что все они второго сорта. С какой вероятностью ОТК может утверждать, что партия содержит изделий второго сорта больше, чем b?

4.39. На технический контроль качества предъявляется партия из 1000 деталей, в которой 200 деталей изготовлено на заводе A, 300 деталей – на заводе B, остальные – на заводе C. Доля брака зависит от завода-изготовителя и оставляет для завода A и B 15 %, а для завода C – 30 %. Найти вероятность того, что наудачу извлеченная деталь окажется отличного качества.

4.40. Вероятность того, что некоторое устройство перестанет функционировать на протяжении времени $(t,t+\Delta t)$ равна $\lambda\Delta t+o(\Delta t),\ \Delta\ \to 0.$ Какова вероятность того, что оно проработает до момента t, если отказ его после момента не зависит от функционирования до момента s?

4.41. Среди женщин-избирателей 70 % поддерживают кандидата от партии , а среди мужчин-избирателей – 60 %. Используя данные переписи, согласно которым доля женщин среди избирателей составляет 55 %, оценить вероятность победы на выборах кандидата от партии A.

4.42. Исследуется динамика курсов валют A и B (по отношению к некоторой валюте C) с целью прогнозирования. Статистика валютных торгов показывает, что курс валюты B возрастает: в 80 % случаев, если вырос курс A; в 30 % случаев, если снизился курс A; в 50 % случаев, если курс A не изменился. Предполагая, что все три исходные гипотезы об изменении курса A равновозможны, оценить вероятности этих гипотез, если известно, что на последних торгах курс валюты вырос.

§5. Схема независимых испытаний Бернулли

Определение. Испытанием (экспериментом, опытом) называется последовательность из двух актов: 1) создание комплекса условий, 2) наблюдение появившегося события. Испытания называются независимыми, если наблюдаемые события являются независимыми.

Определение. Независимыми испытаниями Бернулли называются такие испытания, для которых вероятности появления событий в каждом испытании одинаковы и не меняются от испытания к испытанию.

Нас будет интересовать следующая задача. Пусть производятся n испытаний Бернулли. В каждом испытании возможно появления события с вероятностью p и невозможно с вероятностью . Нужно определить $P_n(m)$ – вероятность того, что в n испытаниях событие появляется ровно m раз.

Результат испытаний удобно описать набором букв длиной n, который состоит из букв и B: $\omega = (A\ A\ B \ldots\ A\ B\ A)$, где буква A означает, что в испытании появилось событие A, а B – что в испытании появилось противоположное событие $\overline{A}$. Каждый набор интересующих нас исходов содержит m букв и $n-m$ букв , поэтому все такие исходы имеют одинаковую вероятность $p^m q^{n-m}$. Разные наборы отличаются только размещением букв A и B, поскольку число случаев, в которых появляется событие A, фиксировано. Размещение букв A и B однозначно определяется выбором m элементов из , что можно сделать способами. Поэтому

$$P_n(m) = \mathrm{C}_n^m p^m q^{n-m} = \mathrm{C}_n^m p^m (1-p)^{n-m}, \ m = \overline{0, n}.$$

Эта формула называется формулой Бернулли. Очевидно, что

$$\sum_{m=0}^{n} P_n(m) = 1 .$$

Пример 5.1. В течение смены, которая длится время t, эксплуатируется ПЭВМ. Каждая ПЭВМ имеет надежность (вероятность безотказной работы) и выходит из строя независимо от других.

Найти вероятность того, что инженер-электроник, вызванный по окончанию времени t для ремонта неисправных ПЭВМ, справится со своей задачей за время , если на ремонт каждой неисправной ПЭВМ ему требуется время .

Решение. Событие A равносильно тому, что число вышедших из строя ПЭВМ меньше, чем $l = [\tau / \tau_0]$, где $[\tau / \tau_0]$ – наибольшее целое число, которое меньше либо равно τ / τ_0. Поэтому

$$P(A) = \sum_{m=0}^{l} C_n^m (1-p)^m p^{n-m} .$$

Когда число испытаний велико, для вычисления $P_n(m)$ можно пользоваться приближенными формулами, которые вытекают из предельной теоремы Пуассона и локальной предельной теоремы Муавра – Лапласа.

В частности, имеет место ***предельная теорема Пуассона:*** если $n \to \infty$, $p \to 0$, так, что $np \to \lambda$, $0 < \lambda < \infty$, то

$$P_n(m) \xrightarrow[n \to \infty]{} \frac{\lambda^m}{m!} e^{-\lambda} .$$

Она выполняется потому, что если положить $np = \lambda_n$, то

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

$$= \frac{n(n-1)\ldots(n-m+1)}{m!} \left(\frac{\lambda_n}{n}\right)^m \left(1 - \frac{\lambda_n}{n}\right)^{n-m} =$$

$$= \frac{\lambda_n^m}{m!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \ldots \left(1 - \frac{m-1}{n}\right)\left(1 - \frac{\lambda_n}{n}\right)^{-m} .$$

Далее путем предельного перехода при $n \to \infty$ получим утверждение теоремы.

Приближенная формула, которая следует из этой теоремы, имеет вид (при больших и малых)

.

Она применяется при решении задач, в основном, когда $\lambda_n = np \le 10$.

Пример 5.2. Вероятность появления события в каждом испытании равна 0,001. Проведено 5000 испытаний. Найти вероятность, что событие в них произойдет не менее двух раз.

Решение. $\lambda_n = np = 5000 \cdot 0{,}001 = 5 < 10.$ Искомая вероятность равна

$$P\{m \geq 2\} = \sum_{m=2}^{n} P_n(m) = 1 - P_n(0) - P_n(1) = 1 - P_{5000}(0) - P_{5000}(1) =$$

$$= 1 - e^{-5} - 5e^{-5} = 1 - 6e^{-5} \approx 0{,}9596.$$

Заметим, что в данном примере $p \approx 0$, а найденная вероятность $P\{m \geq 2\} \approx 1$.

При $n \to \infty$ имеет место также ***локальная предельная теорема Муавра – Лапласа:***

$$\frac{P_n(m)}{\frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{x_m^2}{2}}} \xrightarrow[n\to\infty]{} 1,$$

где $x_m = \dfrac{m - np}{\sqrt{np(1-p)}}$, $-\infty < a \leq x_m \leq b < +\infty.$ Из нее при больших n вытекает следующая приближенная формула

$$P_n(m) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{x_m^2}{2}}.$$

На практике ею обычно пользуются, когда $\lambda_n = np > 10$. Она даёт хорошие приближения при $p \approx \dfrac{1}{2}$ и часто используется, когда $n > 100$, $np(1-p) > 20$.

Интегральная предельная теорема Муавра – Лапласа имеет вид:

$$P\left(a \le \frac{m-np}{\sqrt{np(1-p)}} \le b\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}}\,dt.$$

Для того чтобы показать, что она имеет место, можно воспользоваться предыдущей теоремой, из которой следует: $\forall\ \varepsilon > 0\ \exists n_\varepsilon$, такое что $\forall\, n \ge n_\varepsilon$

$$\left|\frac{P_n(m)}{\frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-\frac{x_m^2}{2}}}\right| < \varepsilon,$$

т.е.

$$(1-\varepsilon)\frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}} \frac{1}{\sqrt{np(1-p)}} < P_n(m) < (1+\varepsilon)\frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}} \frac{1}{\sqrt{np(1-p)}}.$$

Пусть $-\infty < a \le b < +\infty, \sum^* = \sum\limits_{m: a \le \frac{m-np}{\sqrt{np(1-p)}} \le b}$, тогда, учитывая, что

$$\sum\nolimits^* P_n(m) = P\left(a \le \frac{m-np}{\sqrt{np(1-p)}} \le b\right),\ \Delta x_m = x_{m+1} - x_m = \frac{1}{\sqrt{np(1-p)}},$$

можно записать

$$(1-\varepsilon)\frac{1}{\sqrt{2\pi}}\sum\nolimits^* e^{-\frac{x_m^2}{2}}\Delta x_m < P\left(a \le \frac{m-np}{\sqrt{np(1-p)}} \le b\right) < (1+\varepsilon)\frac{1}{\sqrt{2\pi}}\sum\nolimits^* e^{-\frac{x_m^2}{2}}\Delta x_m.$$

Если $n \to \infty$,то , поэтому при $n \to \infty$:

Из данной теоремы вытекает приближенная формула

$$P\left(a \le \frac{m-np}{\sqrt{np(1-p)}} \le b\right) \approx \Phi(b) - \Phi(a),$$

где $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$ – интеграл Лапласа; $\Phi(0) = 0$; $\Phi(-x) = -\Phi(x)$; $\Phi(x) \approx 0{,}5,\ x \ge 5.$

Таблица значений функции $\Phi(x)$ приведена в Приложении 2.

Пример 5.3. Театр, вмещающий 1000 зрителей, имеет два входа. У каждого входа имеется свой гардероб. Сколько мест должно быть в каждом из гардеробов, чтобы в среднем в 99 случаях из 100 все зрители могли раздеться в гардеробе того входа, через который они вошли? Предполагается, что зрители приходят парами и каждая пара независимо от других выбирает с вероятностью 0,5 любой из входов.

A

$$(1-\varepsilon)\frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-\frac{t^2}{2}} dt \le P\left(a \le \frac{m-np}{\sqrt{np(1-p)}} \le b\right) \le (1+\varepsilon)\frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-\frac{t^2}{2}} dt$$

Решение. Пусть число мест в каждом гардеробе равно N, для его нахождения составим уравнение. Занумеруем гардеробы номерами 1 и 2. Выбор зрителями того или иного гардероба можно рассматривать как испытание Бернулли, в каждом из которых определенная пара с вероятностью 0,5 выбирает гардероб, например, №1. По условию задачи, $n = 500$, $p = 0{,}5$. Пусть событие A состоит в том, что зрители разденутся в гардеробе того входа, куда они зашли, m – число пар зрителей, выбравших гардероб №1. Событие будет происходить, если $500 - \frac{N}{2} \le m \le \frac{N}{2}$. По условию, $P(A) = 0{,}99$. Поэтому

$$0{,}99 = P(A) = P\left(500 - \frac{N}{2} \le m \le \frac{N}{2}\right) =$$

$$= P\left(\frac{500 - \frac{N}{2} - np}{\sqrt{np(1-p)}} \le \frac{m - np}{\sqrt{np(1-p)}} \le \frac{\frac{N}{2} - np}{\sqrt{np(1-p)}} \right) \approx$$

$$\approx \Phi\left(\frac{\frac{N}{2} - 250}{\sqrt{125}} \right) - \Phi\left(\frac{250 - \frac{N}{2}}{\sqrt{125}} \right) = 2\Phi\left(\frac{\frac{N}{2} - 250}{\sqrt{125}} \right),$$

т.е. $\Phi\left(\frac{\frac{N}{2} - 250}{\sqrt{125}} \right) \approx 0{,}495.$

С помощью таблицы для функции $\Phi(x)$ находим $\Phi(2{,}56) \approx 0{,}495$, таким образом, $\frac{\frac{N}{2} - 250}{\sqrt{125}} \approx 2{,}56$, откуда следует, что $N \approx 556$.

Из интеграла предельной теоремы Муавра – Лапласа получаем

$$P\left(\left| \frac{m}{n} - p \right| \le \varepsilon \right) = P\left(-\varepsilon \le \frac{m - np}{n} \le \varepsilon \right) =$$

$$= P\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \le \frac{m - np}{\sqrt{np(1-p)}} \le \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \right) \xrightarrow[n \to \infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1,$$

т.е. $P\left(\left| \frac{m}{n} - p \right| \le \varepsilon \right) \xrightarrow[n \to \infty]{} 1 \quad \forall \varepsilon > 0$. Последнее соотношение носит название *закона больших чисел в форме Бернулли*. Из него следует, что при большом числе испытаний частота появления события почти не отличается от вероятности этого события.

З а д а ч и

5.1. При проведении зачета с помощью ЭВМ студенту предлагается 5 вопросов. Вероятность, что студент правильно ответит на один вопрос, равна 0,5. Для получения зачета студенту необходимо правильно ответить не менее чем на 3 вопроса. Найти вероятность получения зачета.

5.2. Какова вероятность, что в группе, состоящей из 30 студентов, никто не родился в сентябре?

5.3. На лекции по теории вероятностей присутствуют 50 человек. Найти вероятность того, что k человек из присутствующих родились 14 июня и l человек родились 23 ноября. Считать, что вероятность рождения в фиксированный день одна и та же для всех дней года. Решить задачу при $k = 1$, $l = 2$. Найти вероятность того, что число родившихся 14 июня и 23 ноября не больше двух.

5.4. Рыбак забросил спиннинг 100 раз. Какова вероятность того, что он поймал хотя бы одну рыбу, если одна рыба приходится в среднем на 200 забрасываний.

5.5. Вероятность поражения мишени при одном выстреле равна 0,8. Найти вероятность того, что при 100 выстрелах мишень будет поражена 75 раз.

5.6. Визуальное наблюдение искусственного спутника Земли возможно в данном пункте с вероятностью 0,1 каждый раз, когда он пролетает над этим пунктом. Сколько раз должен пролететь спутник над пунктом наблюдения, чтобы с вероятностью не меньшей 0,9975 удалось сделать над ним не менее пяти наблюдений?

5.7. Попытки наблюдать спутник (см. предыдущую задачу) проводятся 100 раз. Найти практически достоверный диапазон числа успешных наблюдений.

5.8. В поселке 2500 жителей. Каждый из них примерно 6 раз в месяц ездит на поезде в город, выбирая дни поездок по случайным мотивам независимо от остальных. Какой наименьшей вместимостью должен обладать поезд, чтобы он переполнялся в среднем не чаще одного раза в 100 дней (поезд ходит раз в сутки)?

5.9. В одном из матчей на первенство мира по шахматам ничьи не учитывались, и игра шла до тех пор, пока один из участников матча не набирал 6 очков (выигрыш – 1 очко, проигрыш и ничья – 0 очков). Считая участников матча одинаковыми по силе, а результаты отдельных игр независимыми, найти вероятность того, что при таких правилах в момент окончания матча проигравший набирает 3 очка.

5.10. Вероятность появления события в одном опыте равна p. С какой вероятностью можно утверждать, что частота появления этого события в опытах будет лежать в пределах от до β ? Решить задачу при $n = 100$, $p = 0{,}3$, $\alpha = 0{,}2$, .

5.11. Вероятность появления события в каждом из независимых опытов равна . Найти положительное число такое, что с вероятностью абсолютная величина отклонения частоты появления события от его вероятности будет не больше ε. Решить задачу при

5.12. Сколько нужно провести независимых опытов, чтобы с вероятностью P событие A, вероятность появления которого в одном опыте равна p, наблюдалось не менее, чем раз? Решить задачу при .

5.13. Сколько нужно провести независимых опытов, чтобы с вероятностью P утверждать, что частота интересующего нас события будет отличаться от вероятности этого события, равной p, не больше, чем на ? Решить задачу .

5.14. Каждую секунду с вероятностью p по дороге проезжает автомашина. Пешеходу, для того чтобы перейти дорогу, нужно 3 с. Какова вероятность того, что пешеход, подошедший к дороге, будет ждать возможности перехода: а) 3 с; б) 4 с; в) 5 с?

5.15. Вероятность столкновения космического корабля с метеоритом в течение часа полета равна 0,001. Найти практически возможные границы числа столкновений с метеоритами на протяжении трех месяцев полета – с 1 июня по 31 августа, если вероятность практической возможности принимается в данном случае равной 0,9995.

5.16. Аппаратура состоит из элементов. Вероятность выхода из строя одного элемента за наблюдаемое время равна и не зависит от состояния других элементов. Найти вероятность выхода из строя: а) равно элементов, б) не меньше, чем m элементов, в) не больше, чем m элементов. Решить задачу, когда 1) ; 2) $n = 500$, $p = 0{,}002$, $m = 2$.

5.17. Электрическая цепь состоит из n последовательно включенных лампочек. Определить вероятность того, что при повышении напряжения в сети выше номинального произойдет разрыв цепи, если вероятность того, что лампочка перегорит, равна . Решить задачу для: а) , б) .

5.18. Найти наивероятнейшее число отрицательных и положительных ошибок и соответствующую вероятность при четырех изменениях некоторой величины, если в каждом из измерении вероятность получить положительную ошибку равна , а отрицательную – $\frac{1}{3}$.

5.19. Линия связи, имеющая k каналов, связывает два города, где есть n абонентов, каждый из которых пользуется для этого телефоном в среднем мин в час. Найти вероятность безотказного обслуживания абонентов. Решить задачу для $n = 1000$, $k = 130$, $l = 6$.

5.20. Телефонная станция , которая обслуживает 20000 абонентов, должна соединять их с другой станцией B. Какое наименьшее число линий должно связывать A и B, чтобы в 99 % вызовов нашлась свободная линия. Имеется в виду, что в течение наиболее напряженного часа каждый абонент разговаривает в среднем 2 мин.

$=20, pp=0{,}01$ $\frac{\ }{3}$

5.21. По каналу связи передаются n сообщений. Каждое из них независимо от других с вероятностью искажается помехами. Найти вероятности следующих событий: {из n сообщений искажаются помехами}, {искажается не более половины всех передаваемых сообщений}.

5.22. Для увеличения надежности передачи важного сообщения, которое состоит из n символов, каждый из передаваемых символов дублируется раз. В качестве воспринимаемого символа в пункте приема принимается тот, который продублирован не меньше раз из m. Когда символ в пункте приема повторяется меньше чем раз, то такой символ считается искаженным. Вероятность правильной передачи каждого символа одинакова и не зависит от того, как передаются другие символы. Найти вероятности следующих событий: $A =$ {отдельный передаваемый символ в сообщении будет правильно воспринят в пункте приема}, $B =$ {все сообщение будет правильно воспринято в пункте приема}, $C =$ {в сообщении искажаются не больше m символов}.

5.23. В течение часа фирма принимает в среднем сообщений по электронной почте, обработкой которых занимается специальный сотрудник. Какова вероятность того, что за m минут на фирму не поступит ни одного сообщения? Решить задачу, когда: а) ; б) $k = 60,\ m = 5$.

5.24. Железнодорожный состав состоит из n вагонов, каждый из которых с вероятностью имеет дефект. Все вагоны осматривают независимо друг от друга два мастера; первый из них обнаруживает дефект (если он имеется) с вероятностью , второй – с вероятностью p_2. Какова вероятность отправления в рейс состава, в котором имеется хотя бы один дефектный вагон? Решить задачу для $n = 150,\ p_1 = 0{,}95,\ p_2 = 0{,}9$.

5.25. При установившемся технологическом процессе 80 % всех изготавливаемых заводом изделий выпускается высшим сортом. Приемщик наугад берет 100 изделий. Чему равна вероятность того, что среди них изделий высшего сорта окажется от 60 до 80 штук?

5.26. Завод изготовляет изделия, каждое из которых с вероятностью r (независимо от других) оказывается дефектным. При осмотре дефект, если он имеется, обнаруживается с вероятностью . Для контроля из продукции завода выбираются изделий. Найти вероятность, что хотя бы в одном из них обнаружен дефект. Решить задачу при а) ; б) .

5.27. Имеется партия изделий. Каждая из них независимо от других может оказаться дефектным с вероятностью 0,2. Из партии берут 10 изделий и проверяют их на годность. Если число дефектных изделий при этом не более 1, то партию принимают, в противном случае подвергают сплошному контролю. Какова вероятность того, что партия будет принята?

5.28. В лотерее 40000 билетов, ценные выигрыши выпадают на 3 билета. Определить: а) вероятность получения хотя бы одного выигрыша на 1000 билетов, б) сколько необходимо приобрести билетов, чтобы вероятность получения ценного выигрыша была не менее 0,5.

5.29. Транспортная фирма занимается перевозкой изделий со склада в магазин. Вероятность того, что при перевозке изделия будет повреждено ровно 0,002. Фирме необходимо перевести 1000 изделий. Найти вероятность того, что магазин получит: а) хотя бы одно поврежденное изделие; б) менее двух поврежденных изделий; в) 3 % поврежденных изделий. Какова вероятность наиболее вероятного числа поврежденных изделий в наудачу выбранных пяти контейнерах (в одном контейнере – 20 изделий)?

5.30. Для нового офиса фирма приобрела n новых персональных компьютеров. В течение определенного периода времени каждый компьютер может выйти из строя с вероятностью . Устранение неисправностей в компьютерах занимается фирма . В конце данного периода фирма A обращается к услугам фирмы B и платит ей за ремонт каждого неисправного компьютера сумму $\$d$. Какова вероятность того, что фирме A по истечении этого периода придется заплатить фирме B сумму: а) менее $\$D$, б) не менее $\$D$. Решить задачу для $n = 50,\ p = 0{,}01,\ d = 40,\ D = 1000$.

5.31. Что вероятнее выиграть у брокера одинаковой квалификации: а) три сделки из четырех или пять из восьми, б) не менее трех сделок из четырех или не менее пяти сделок из восьми, в) не более n
$2n+1$ из сделок или более n из того же числа, г) не более n из
сделок или более n из того же числа.

5.32. Товаровед исследует 50 образцов некоторого товара. Производитель этого товара указывает, что процент брака составляет 15 %. Найти наивероятнейшее число образцов, которые товаровед признает как годные.

5.33. При наступлении кризиса сбыта продукции фирма не терпит убытков с вероятностью , полностью терпит банкротство с вероятностью p_2 и несет серьезные издержки с вероятностью $p_3 = 1 - p_1 - p_2$. Две серии серьезных издержек приводят к полному банкротству фирмы. Найти вероятность того, что при наступлении n признаков сбыта фирма не обанкротится.

5.34. Пункт нужно связать компьютерной связью с 10 абонентами пункта B . Каждый абонент занимает канал связи 12 минут в час. Вызовы любых двух абонентов независимы. Какое минимальное число каналов необходимо для того, чтобы можно было в любой момент времени с вероятностью 0,99 обслужить всех абонентов?

5.35. Вероятность появления фальшивой банкноты в банке равна $p = 10^{-8}$. В течение рабочей недели банк оперирует с $n = 7.5 \cdot 10^{8}$ банкнотами. Оценить вероятности встретить в ходе обработки 0; 1; 2; 3 фальшивые банкноты.

5.36. Для определения доли избирателей, поддерживающих кандидата A, производится выборочное обследование. Определить объем выборки, при которой с вероятностью, не меньшей 0,99, погрешность составит менее 0,005.

5.37. Страховая компания заключила договор со спортсменом-теннисистом на 365 дней, предусматривающий выплату страхового возмещения клиенту в случае травмы специального вида. Из предыдущей практики известно, что вероятность получения такой травмы теннисистом в любой фиксированный день равна 0,00037. Найти вероятность того, что в течение срока договора: а) не произойдет ни одного страхового случая; б) произойдет один страховой случай; в) произойдет два страховых случая.

5.38. Портфель страховой компании состоит из 1000 договоров, заключенных в начале года и действующих в течение текущего года. При наступлении страхового случая по каждому их договоров компания обязуется выплатить 2000 у.е. Вероятность наступления страхового события по каждому из договоров равна 0,05 и не зависит от наступления страховых событий по другим контрактам. Каков должен быть совокупный размер резерва страховой компании для того, чтобы с вероятностью 0,99 она могла бы удовлетворить требования, возникающие по указанным договорам.

§6. Одномерные случайные величины

Определение. Пусть $\Omega = \{\omega\}$ – множество элементарных событий. Случайной величиной (СВ) называется функция $\xi(\omega)$, определённая на множестве , принимающая числовые значения и такая, что для любого действительного определена вероятность

.

Пример 6.1. Пусть Ω – множество студентов на факультете. Каждый отдельный студент – элемент . Определим на элементах ω функцию которая принимает значения, равные году рождения студента, который является элементом . Таким образом определенная функция является случайной величиной (имеется в виду, что, кроме , задана вероятностная функция $P(\omega)$).

Пример 6.2. По промежуткам времени безотказной работы приборы делятся на несколько типов, например, первый, второй, третий. Пусть $\Omega = \{\omega\}$ – множество значений, которые могут принимать промежутки времени безотказной работы прибора, а $\xi(\omega)$ – номер типа, который присваивается прибору с промежутком безотказной работы . Тогда являются СВ.

Определение. Функция

называется функцией распределения (ф.р.) случайной величины $\xi(\omega)$.

Теорема (свойства ф.р.). 1) – монотонно неубывающая функция; 2) $\lim\limits_{x \to -\infty} F_\xi(x) = 0$, $\lim\limits_{x \to +\infty} F_\xi(x) = 1$; 3) $F_\xi(x)$ – непрерывная справа функция $\lim\limits_{x_k \to x} F_\xi(x_k)$

Доказательство.

1) Поскольку $\{\omega : \xi(\omega) \le x_1\} \le \{\omega : \xi(\omega) \le x_2\}$ для $x_1 \le x_2$, то

$P\{\omega : \xi(\omega) \le x_1\} \le P\{\omega : \xi(\omega) \le x_2\} \Rightarrow F_\xi(x_1) \le F_\xi(x_2)$.

2) $\{\omega : \xi(\omega) \le x\} \xrightarrow[x \to -\infty]{} \varnothing \;\Rightarrow$ по аксиоме непрерывности ; $\{\omega : \xi(\omega) \le x\} \xrightarrow[x \to +\infty]{} \Omega \;\Rightarrow$ по аксиоме непрерывности

3) Пусть $x_1 > x_2 > \ldots$, $\lim\limits_{k \to \infty} x_k = x$, $\{\omega : \xi(\omega) \le x_k\} = A_k$, тогда $A = \bigcap\limits_k A_k$

$\Rightarrow$ по аксиоме непрерывности .

Рассмотрим случай, когда множество Ω состоит из дискретного множества элементарных событий (счетного или конечного) . Пусть $x_k = \xi(\omega_k)$, $\xi(\omega) \in \{x_1, x_2, \ldots, x_k, \ldots\}$,

$A_k = \{\omega : \xi(\omega) = x_k\}$, $k = 1, 2, \dots$. Случайные величины, которые могут принимать только конечное или счетное множество значений, называются дискретными. Для их описания удобно пользоваться набором вероятностей $\{p_1, p_2, \dots, p_k, \dots\}$, где $p_k = P(A_k) = P\{\omega : \xi(\omega) = x_k\}$, который называется распределением вероятностей дискретной СВ $\xi(\omega)$. Поскольку ,

$A_i \cap A_j = \varnothing$, $i \neq j$, то (условие

нормировки). Совокупность

$$\begin{matrix} x_1, x_2, \dots, x_k, \dots \\ p_1, p_2, \dots, p_k, \dots \end{matrix}$$

называется дискретным законом (рядом) распределения вероятностей.

Установим связь между распределением вероятностей и функцией распределения:

$$F_\xi(x) = P\{\omega : \xi(\omega) \le x\} = P\left(\bigcup_{\{k : x_k \le x\}} \{\omega : \xi(\omega) = x_k\}\right) = \sum_{\{k : x_k \le x\}} P\{\omega : \xi(\omega) = x_k\} = \sum_{\{k : x_k \le x\}} p_k,$$

$$p_k = F_\xi(x_k) - F_\xi(x_{k-1}), \text{ если считать, что } F_\xi(x_0) = 0.$$

В связи с тем, что свойства СВ полностью определяются свойствами их функций распределения, их принято классифицировать по характеру этих функций.

I. Дискретные СВ (ф.р.). В этом случае множество значений $\xi(\omega)$: $x_1, x_2, \dots, x_k, \dots$ – счётно либо конечно, $F_\xi(x) = \sum_{\{k : x_k \le x\}} p_k$; ф.р. обладает, кроме основных, следующими свойствами: 1) $F_\xi(x)$ имеет конечное или счетное множество точек разрыва первого рода, 2) если x – точка непрерывности , то $\exists$ и $\dfrac{dF_\xi(x)}{dx} = 0$.

Примеры дискретных распределений CB:

1) СВ имеет распределение Бернулли, если $\xi(\omega)\in\{0,1\}$,

$$p_k = P\{\omega:\xi(\omega)=k\} = p^k(1-p)^{1-k},\ k=0,1,\qquad ,$$

$$F_\xi(x)=\begin{cases}0,\ x\le 0,\\ 1-p,\ 0<x\le 1,\\ 1,\ x>1;\end{cases}$$

2) СВ $\xi(\omega)$ имеет биномиальное распределение, если ,

$$p_k = P\{\omega:\xi(\omega)=k\} = C_n^k p^k(1-p)^{n-k},\ k=0,1,2,\dots,n,\qquad ,$$

$$F_\xi(x)=\begin{cases}\sum\limits_{k=0}^{l} C_n^k p^k(1-p)^{n-k},\ l<x\le l+1,\\ 1,\ x>n,\\ 0,\ x\le 0,\end{cases}$$

$\xi(\omega)\in\{0,1,2,\dots,n\}$

отметим, что, как следует из формулы Бернулли, число появлений события в n независимых испытаниях Бернулли имеет биномиальное распределение;

3) СВ имеет геометрическое распределение, если $\xi(\omega)\in\{0,1,2,\dots k,\dots\}$,

$$p_k = P\{\omega:\xi(\omega)=k\} = p(1-p)^k,\ k=0,1,2,\dots,\qquad ,$$

$$F_\xi(x)=\begin{cases}0,\ x\le 0\\ \sum\limits_{k=0}^{l} p(1-p)^k, l<x\le l+1\end{cases} = \begin{cases}0,\ x\le 0,\\ 1-(1-p)^l, l<x\le l+1;\end{cases}$$

4) CB $\xi(\omega)$ имеет распределение Пуассона с параметром $\lambda>0$, если

,

$$p_k = P\{\omega:\xi(\omega)=k\} = \frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\dots,$$

$$F_\xi(x) = \begin{cases} 0, x \le 0, \\ e^{-\lambda} \sum_{k=0}^{l} \dfrac{\lambda^k}{k!}, l < x \le l+1. \end{cases}$$

Дадим интерпретацию некоторых из этих СВ. Предположим, что студент идет сдавать зачет. На некоторые вопросы по сдаваемому предмету он знает ответы, а на остальные – нет. Поэтому событие, заключающееся в том, что он получит зачет, является случайным. Определим СВ следующим образом: если зачет сдан, то $\xi = 1$, если нет, то . Таким образом определенная СВ является бернуллиевой, параметр в том случае соответствует относительному числу вопросов, на которые студент знает ответ. Пусть студенту необходимо сдать зачетов и он делает по одной попытке получить каждый из этих зачетов. Определим СВ как число зачетов, которые получит студент. Такая СВ будет биноминальной. Число студентов, которых успел выслушать преподаватель на зачете за фиксированный интервал времени, а также число заданий, которые выполняет ЭВМ за фиксированный промежуток времени, являются СВ, распределенными по закону Пуассона с соответственно определенными параметрами .

Пример 6.4. Производятся последовательные испытания трех приборов на надежность. Каждый следующий прибор испытывается только в том случае, если предыдущий оказался надежным. Составить таблицу распределения и найти ф.р. случайного числа испытанных приборов, если вероятность надежности каждого прибора равна .

Решение. СВ ξ, описывающая число испытанных приборов, имеет распределение вероятностей

поэтому таблица распределения имеет вид

k	1	2	3
p_k	0,1	$0{,}9 \cdot 0{,}1$	$(0{,}9)^2 \cdot 0{,}1$

а ф.р.

$$F_{\xi}(x)=\begin{cases}0,\ x\le 1,\\ \sum\limits_{i=1}^{k} p_i,\ k<x\le k+1,\\ 1,\ x\ge 3.\end{cases}$$

II. Непрерывные СВ (СВ с абсолютно непрерывными ф.р.).

В этом случае $F_{\xi}(x)-$ непрерывная функция и $F_{\xi}(x)=\int\limits_{-\infty}^{x} p_{\xi}(t)dt$. Ясно, что $p_{\xi}(x)=\dfrac{dF_{\xi}(x)}{dx}$ в точках существования производной. Функция $p_{\xi}(x)$ называется плотностью распределения вероятностей СВ $\xi(\omega)$. Она обладает следующими свойствами:

1) $p_{\xi}(x)\geq 0$ (как производная неубывающей функции);

2) $\int\limits_{-\infty}^{\infty} p_{\xi}(x)dx=1$ условие нормировки, (следует из свойства ф.р. $F_{\xi}(+\infty)=1$);

3) кроме того, $P\{\omega: x_1<\xi(\omega)\le x_2\}=F_{\xi}(x_2)-F_{\xi}(x_1)=\int\limits_{x_1}^{x_2} p_{\xi}(t)dt.$

Рассмотрим примеры непрерывных СВ:

1) СВ $\xi(\omega)$ имеет равномерное распределение на отрезке , если

$$p_{\xi}(x)=\begin{cases}\dfrac{1}{b-a},x\in[a,b]\\ 0,x\notin[a,b]\end{cases},\quad F_{\xi}(x)=\begin{cases}0,x<a,\\ \dfrac{x-a}{b-a},a<x\le b,\\ 1,x>b;\end{cases}$$

2) СВ $\xi(\omega)$ имеет показательное (экспоненциальное) распределение с параметром если

$$, \ F_\xi(x) = \begin{cases} 0, x < 0 \\ 1 - e^{-\lambda x}, x \geq 0 \end{cases},$$

3) СВ $\xi(\omega)$ имеет нормальное распределение с параметрами ,
$\xi(\omega) \sim$, если

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \ F_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt;$$

4) СВ $\xi(\omega)$ имеет распределение Коши с параметром , если

$$p_\xi(x) = \frac{a}{\pi(a^2 + x^2)}, \ F_\xi(x) = \frac{1}{\pi} arctg \frac{x}{a}.$$

В частности, интервалы времени между соседними автомобилями на дорогах являются экспоненциально распределенными СВ с соответствующими параметрами $\lambda > 0$. Если – СВ, имеющая экспоненциальное распределение, то

$$P(\xi > x + \tau / \xi > \tau) = \frac{P(\xi > x + \tau, \xi > \tau)}{P(\xi > \tau)} = \frac{P(\xi > x + \tau)}{P(\xi > \tau)} = \frac{1 - P(\xi \leq x + \tau)}{1 - P(\xi \leq \tau)} =$$

$$= \frac{1 - 1 - e^{-\lambda(x+\tau)}}{1 - 1 - e^{-\lambda\tau}} = e^{-\lambda x} = 1 - F_\xi(x) = 1 - P(\xi \leq x) =$$

$$= P(\xi > x) \Rightarrow P(\xi - \tau \leq x / \xi > \tau) = P(\xi \leq x),$$

следовательно, см. рис. 7, СВ $\xi - \tau$ имеет то же самое экспоненциальное распределение, как и СВ .

Рис. 7

Пример 6.5. Проверим, что функция является

плотностью распределения вероятностей.

Решение. Ясно, что $f(x) > 0$. Нужно еще проверить условие нормировки. Это можно сделать несколькими способами: а) пользуясь равенством , где $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$ – гамма-функция; б) возводя интеграл $\int_{-\infty}^{\infty} f(x)dx$ в квадрат и переходя в двойном интеграле к полярным координатам.

Рассмотрим первый способ:

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

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-\frac{x^2}{2}} \left(\frac{x^2}{2}\right)^{-\frac{1}{2}} d\left(\frac{x^2}{2}\right) = \left[\frac{x^2}{2} = t\right] =$$

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

Второй способ дает:

$$\left[\int_{-\infty}^{\infty} f(x)dx\right]^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy = [x = r\cos\varphi, y = r\sin\varphi] =$$

$$= \frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_0^\infty e^{-\frac{r^2}{2}} r dr = 1,$$

откуда следует тот же самый результат. Плотность $p(x)$ называют плотностью стандартного нормального распределения, у которого $a = 0$, $\sigma = 1$.

Теорема Лебега. Любую ф.р. однозначно можно представить в виде (разложение Лебега)

$$F_\xi(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x),$$

где $a_i \ge 0$, $i = \overline{1,3}$, $a_1 + a_2 + a_3 = 1$, $F_1(x)$, $F_2(x)$, $F_3(x)$ – дискретная, абсолютно непрерывная и сингулярная ф.р. соответственно.

Пусть $\xi(\omega)$ – СВ с абсолютно непрерывной ф.р. и плотностью распределения $p_\xi(x)$, $y = f(x)$ – непрерывная возрастающая функция. Тогда для $\eta(\omega) = f(\xi(\omega))$

$$F_\eta(y) = P\{\eta \le y\} = P\{f(\xi) \le y\} = P\{\xi \le f^{-1}(y)\} = F_\xi(f^{-1}(y)),$$

и если, кроме того, $f(x)$ – дифференцируемая функция, то

$$p_\eta(y) = p_\xi\left(f^{-1}(y)\right)\frac{df^{-1}(y)}{dy}.$$

Если же функция $f(x)$ – убывающая, то

$$F_\eta(y) = 1 - F_\xi(f^{-1}(y)), \; p_\eta(y) = -p_\xi(f^{-1}(y))\frac{df^{-1}(y)}{dy}.$$

Пример 6.6. Случайная величина ξ имеет показательное распределение с параметром . Найти плотность распределения СВ

Решение:

$$F_\eta(y) = P\left\{\sqrt{\xi} \le y\right\} = P\left\{\xi \le y^2\right\} = 1 - e^{-\lambda y^2},$$

$$p_\eta(y) = F_\eta'(y) = 2\lambda y e^{-\lambda y^2}, \; y \ge 0.$$

Пример 6.7. Пусть СВ ξ равномерно распределена на отрезке [0,1]. Найти ф.р. и плотности распределения СВ: а)

б) $\eta_2 = \sqrt{\xi}$, в) $\eta_3 = \sin^2\left(\frac{\pi\xi}{2}\right)$.

Решение:

а) $F_{\eta_1}(x) = P\{\xi^2 \le x\} = P\{\xi \le \sqrt{x}\} = \sqrt{x}$, $0 \le x \le 1$,

$$p_{\eta_1}(x) = F'_{\eta_1}(x) = \frac{1}{2\sqrt{x}},\ 0 < x \le 1;$$

б) $F_{\eta_2}(x) = P\{\sqrt{\xi} \le x\} = P\{\xi \le x^2\} = x^2$,

$$p_{\eta_2}(x) = F'_{\eta_2}(x) = 2x,\ 0 \le x \le 1;$$

в) $F_{\eta_3}(x) = P\left\{\sin^2\left(\frac{\pi\xi}{2}\right) \le x\right\} = P\left\{\sin\left(\frac{\pi\xi}{2}\right) \le \sqrt{x}\right\} =$

$$= P\left\{\xi \le \frac{2}{\pi}\arcsin\sqrt{x}\right\} \quad p_{\eta_3}(x) = F'_{\eta_3}(x) = \frac{1}{\pi\sqrt{x(1-x)}},\ 0 < x < 1.$$

Пример 6.8. Найти плотности распределения СВ: а) $\eta_1 = \xi^3$, б) $\eta_2 = |\xi|$, если известна плотность распределения СВ ξ.

Решение.

а) $p_{\eta_1}(y) = p_\xi(\sqrt[3]{y})(\sqrt[3]{y})' = \frac{1}{3}\sqrt[3]{y^2}\, p_\xi(\sqrt[3]{y})$;

б) $F_{\eta_2}(y) = P(\eta_2 \le y) = P(-y \le \xi \le y) = \int\limits_{-y}^{y} p_\xi(x)dx =$

$$= \int\limits_{0}^{y}[p_\xi(x) + p_\xi(-x)]dx,\ y > 0,$$

откуда следует, что $p_{\eta_2}(y) = p_\xi(y) + p_\xi(-y)$, $y > 0$, согласно определению плотности.

Пример 6.9. При проведении математических экспериментов на ЭВМ поведение построенной модели многократно наблюдают при различных случайных исходных условиях. Такой способ исследования называется методом статистических испытаний или методом «Монте-Карло». При этом возникает задача получения случайных чисел, распределенных по любому какому угодно заданному закону. В ЭВМ эта задача решается при помощи функционального преобразования случайных чисел, распределенных равномерно в интервале [0,1], методы получения которых хорошо разработаны. Это делается следующим образом.

Пусть СВ $\xi(\omega)$ равномерно распределена на интервале [0,1]. Надо найти такое преобразование $f(x)$, чтобы СВ $\eta(\omega) = f(\xi(\omega))$ имела заданную ф.р. $F(y)$. Т.к. $0 \le F(y) \le 1$, выберем $f(x)$ в виде $f(x) = F^{-1}(x)$. Рассмотрим СВ $\eta(\omega) = F^{-1}(\xi(\omega))$. Для нее $p_\eta(y) =$ $= p_\xi(F(y))\left|\dfrac{\partial F(y)}{\partial y}\right|$, но т.к. $\xi(\omega)$ равномерно распределена на интервале [0,1], то , и мы получаем $p_\eta(y) = \dfrac{\partial F(y)}{\partial y}$ (знак модуля здесь можно снять, т.к. $F(y)$ – неубывающая функция). Таким образом, $F_\eta(y) = F(y)$, что и требовалось доказать.

Модой дискретной СВ называется ее наиболее вероятное значение, модой непрерывной СВ ξ – значение аргумента , при котором ее плотность распределения максимальна. Медианой СВ ξ называется значение аргумента , при котором .

Задачи

6.1. Плотность распределения вероятностей СВ ξ имеет вид

Найти: а) константу С, б) плотность распределения вероятностей СВ

$\eta = \dfrac{1}{\xi}$, в) $P\{4 < \eta \le 9\}$.

6.2. Дана плотность распределения СВ

$$p_\xi(x) = \begin{cases} 0, \ x < 0, \ x > 2, \\ a\left(4x - x^3\right), 0 \le x \le 2. \end{cases}$$

Найти a, $F_\xi(x)$, $P\{-2 < \xi \le 1\}$.

6.3. СВ ξ имеет показательное распределение с параметром .

Найти плотности распределения СВ: а) б) $\eta_2 = \dfrac{1}{\lambda}\ln\xi$; в) $\eta_3 = 1 - e^{-\lambda\xi}$; г) $\eta_4 = e^{-\xi}$.

6.4. СВ ξ равномерно распределена на отрезке . Найти плотности распределения СВ: а) $\eta_1 = 2\xi + 1$; б) $\eta_2 = -\ln(1-\xi)$; в) $\eta_3 = tg\left(\pi\left(\xi - \dfrac{1}{2}\right)\right)$.

6.5. СВ ξ распределена по нормальному закону с параметрами , σ^2. Найти плотность распределения СВ $\eta = \dfrac{1}{\xi}$.

6.6. В ячейке ЭВМ записано n-разрядное двоичное число; каждый знак этого числа, независимо от остальных, принимает с равной вероятностью значение 0 или 1. СВ – число знаков «1» в записи двоичного числа. Найти распределение СВ ξ и вероятности

, $P\{\xi < m\}$.

6.7. Времена между двумя сбоями ЭВМ распределены по показательному закону с параметром λ. Решение определенной задачи требует безотказной работы машины в течении времени . Если за время произошел сбой, то задачу приходится решать заново. Сбой обнаруживается только через время τ после начала решения задачи.

Рассматривается СВ – время, за которое задача будет решена. Найти ее закон распределения.

6.8. При работе ЭВМ в случайные моменты возникают неисправности. Время работы ЭВМ до первой неисправности распределено по показательному закону с параметром λ. При возникновении неисправности она мгновенно обнаруживается, и ЭВМ поступает в ремонт. Ремонт продолжается время , после чего ЭВМ снова включается в работу. Найти плотность $p_\xi(t)$ и ф. р. $F_\xi(x)$ промежутка времени ξ между двумя соседними неисправностями. Найти вероятность .

6.9. СВ ξ распределена по нормальному закону с параметром . Задан интервал $(\alpha, \beta]$, не включающий начала координат. При каком значении σ вероятность попадания случайной величины в интервал достигает максимума?

6.10. СВ ξ имеет распределение Пуассона. Найти вероятности случайных событий: $\{\xi$ принимает четное значение$\}$, $\{\xi$ принимает нечетное значение$\}$.

6.11. Интервалы времени безотказной работы ЭВМ имеют показательное распределение с параметром Найти вероятность безотказной работы ЭВМ в течении времени $2T$.

6.12. Плотность распределения СВ равна $p_\xi(x) = ax^2e^{-kx}$, $k > 0$, $0 \le x < +\infty$. Найти: а) коэффициент a; б) ф.р. этой СВ; в) вероятность попадания этой СВ в интервал .

6.13. Пусть $\xi \sim N(0,1)$. Что больше: или $P\{1 < \xi \le 2\}$?

6.14. Показать, что функция $p(x)=\frac{\lambda}{n!}(x\lambda)^n e^{-\lambda x}$, λ, , является плотностью вероятности некоторой СВ ξ и найти вероятность попадания СВ ξ в интервал при $n=2$.

6.15. СВ ξ имеет ф.р. Найти ф.р. СВ $\eta=\frac{1}{2}\left(\xi+|\xi|\right)$.

6.16. Дискретная СВ ξ характеризуется законом распределения

ξ	-1	0	[illegible]
P	0,4	0,3	0,3

Найти закон распределения СВ $\eta=\xi^2+1$, $\theta=|\xi|$.

6.17. Ф.р. Вейбулла

$$F_\xi(x)=\begin{cases}1-e^{-\lambda x^m}, \ x\ge 0,\\ 0, x<0,\end{cases}$$

в некоторых случаях характеризует время службы элементов электронной аппаратуры ξ. Найти моду СВ ξ.

6.18. Случайное время простоя радиоэлектронной аппаратуры в некоторых случаях имеет плотность распределения

$$p_\xi(x)=\frac{M}{\sqrt{2\pi}x\sigma}e^{-\frac{(\lg x-\lg x_0)^2}{2\sigma^2}},$$

где $M=\lg e\approx 0,4343$ (логарифмический нормальный закон распределения). Найти: а) модуль распределения при $x_0=1$, $\sigma=\sqrt{5M}$; б) ф.р. $F_\xi(x)$.

6.19. СВ R – расстояние от места попадания до центра мишени – распределена по закону Релея, т.е. ее плотность распределения

$$p(x)=\begin{cases}axe^{-\alpha^2x^2}, x>0,\\ 0, x\le 0.\end{cases}$$

Найти: а) коэффициент a; б) вероятность того, что окажется меньше, чем мода.

6.20. На электронное реле воздействует случайное напряжение, распределенное по закону Релея с параметрами $a = \dfrac{1}{\sigma^2}$, $\gamma = \dfrac{1}{\sqrt{2\sigma}}$.

Какова вероятность схемы сработать, если электронное реле срабатывает каждый раз, когда напряжение на его входе больше 2 В?

6.21. Случайные ошибки измерений дальности до неподвижной цели подчинены нормальному закону с параметрами $a = 100$ м и 10 м. Определить вероятность того, что измеренное значение дальности отклонится от действительного не больше, чем на 15 м.

6.22. Закон распределения ошибок при измерении радиуса круга нормальный с параметрами 1000, 0,25. Найти закон распределения ошибок при вычислении длины окружности, площади круга.

6.23. В счетчике Гейгера – Мюллера для подсчета космических частиц частица, попавшая в счетчик, вызывает разряд, длящийся время τ. Попавшие в этот промежуток времени в счетчик новые частицы счетчиком не регистрируются. Считая, что распределение числа частиц, попавших в счетчик, подчиняется закону Пуассона, т.е. вероятность попадания в счетчик частиц за время t равна

, $k = 0, 1, 2, \ldots$, найти вероятность того, что счетчик

за время t сосчитает все попавшие в него частицы.

6.24. Закон ошибок при наблюдении температуры выражен по шкале Фаренгейта формулой для плотности вероятности

.

Написать этот закон, приспособив его к шкале Цельсия.

6.25. Угол сноса самолета α определяется формулой

, где β – угол ветра – равномерно распределен в интер-

вале , u – скорость ветра, – воздушная скорость самолета. Найти плотность вероятности угла сноса самолета, если 20 м/с, 1200 км/ч.

6.26. В группе из 5 изделий имеется одно бракованное. Чтобы его обнаружить, выбирают наугад одно изделие за другим и каждое вынутое проверяют. Построить закон распределения и ф.р. числа проверенных изделий.

6.27. Из партии 15 изделий, среди которых имеются две бракованные, выбраны случайным образом 3 изделия для проверки их качества. Построить закон распределения и ф.р. числа бракованных изделий.

6.28. Независимые опыты продолжаются до первого положительного исхода, после чего они прекращаются. Найти для случайного числа опытов: а) ряд распределения; б) наивероятнейшее число опытов, если вероятность успешного исхода в каждом опыте равна 0,5.

6.29. На пути движения автомобиля шесть светофоров, каждый из них либо разрешает, либо запрещает дальнейшее движение автомобиля с вероятностью 0,5. Составить ряд распределения и построить ф.р. числа светофоров, пройденным автомобилем до первой остановки.

6.30. Известно, что при определенных параметрах динамических систем может наступить резонанс. Пусть параметр a является СВ, следующей нормальному закону $N(a,\sigma^2)$. Найти вероятность того, что значение параметра удалено от точек резонанса более чем на расстоянии d, где $d \le \frac{l}{2}$, а точки резонанса равны nl, $n \in Z$

6.31. Бюджетная прямая спроса потребителя на два товара и Y подвергается изменению вследствие изменения цены на товар Y. Предполагая, что изменение зависит от угла наклона α, равномерно распределенного в промежутке $\left[0,\frac{\pi}{2}\right]$, найти ф.р. и плотность распределения величины полного расходования дохода потребителя на товар Y, если бюджетная прямая проходит через точку $B(1,0)$.

§7. Многомерные случайные величины. Независимость и функциональные преобразования случайных величин

Пусть на одном и том же множестве элементарных событий $\Omega = \{\omega\}$ заданы СВ $\xi_1(\omega)$, $\xi_2(\omega), \ldots,$ $\xi_n(\omega)$. Вектор $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots,$ $\xi_n(\omega))$ называется n-мерной СВ, или случайным вектором.

Функция n аргументов

называется n-мерной ф.р. n-мерной СВ.

Пример 7.1. Пусть – СВ, равномерно распределенная на отрезке $[0,1]$, $\xi_2(\omega) = \xi_1^2(\omega)$. Тогда $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega))$ – двумерная СВ. Найдем ее ф.р.

$$F_{\xi_1 \xi_2}(x_1, x_2) = P\{\omega : \xi_1(\omega) \le x_1, \xi_2(\omega) \le x_2\} = P\{\omega : \xi_1(\omega) \le x_1, \xi_1^2(\omega) \le x_2\}$$

При $x_1 < 0$, $x_2 < 0$ это выражение равно 0, а при $x_1 \ge 0$, $x_2 \ge 0$

$$F_{\xi_1 \xi_2}(x_1, x_2) = P\{\omega : \xi_1(\omega) \le x_1, \xi_1(\omega) \le \sqrt{x_2}\} = P\{\omega : \xi_1(\omega) \le \min(x_1, \sqrt{x_2})\} =$$

$$= \begin{cases} \min\left(x_1, \sqrt{x_2}\right), \text{если} \min\left(x_1, \sqrt{x_2}\right) \le 1, \\ 1, \ \text{если} \min\left(x_1, \sqrt{x_2}\right) > 1. \end{cases}$$

Ф.р. многомерной СВ имеет следующие свойства:

1) $F_\xi(x)$ является неубывающей по всем аргументам;

2) $\lim\limits_{x_k \to -\infty} F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = 0$;

3) $\lim\limits_{x_k \to +\infty,\, k = \overline{1,n}} F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = 1$;

4) $F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)$ непрерывна справа по всем аргументам;

5) условие согласованности:

$$\lim_{x_k \to +\infty} F_{\xi_1 \, \xi_{k-1} \, \xi_k \, \xi_{k+1} \ldots \xi_n}(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n) =$$

$$= F_{\xi_1 \, \xi_{k-1} \, \xi_{k+1} \ldots \xi_n}(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)$$

Ф.р. меньшей размерности, которая получается из ф.р. большей размерности, если применить для нее условие согласованности, называется маргинальной;

6) отметим, что для того, чтобы некоторая функция $F(x_1, ..., x_n)$ была ф.р. n-мерной , недостаточно, чтобы для нее были выполнены условия 1) – 5). Необходимо также выполнение еще одного условия. Пусть $a_k < b_k$, $A_k = \{x : a_k < x \le b_k\}$, $k = \overline{1, n}$. Нетрудно видеть, что

$$P\{\omega : \xi_1(\omega) \le x_1, ..., \xi_{k-1}(\omega) \le x_{k-1}, a_k < \xi_k(\omega) \le b_k, \xi_{k+1}(\omega) \le x_{k+1}, ...,$$

$$\xi_n(\omega) \le x_n\} = F_{\xi_1 ... \xi_n}(x_1, ..., x_{k-1}, b_k, x_{k+1}, ..., x_n) -$$

$$- F_{\xi_1 ... \xi_n}(x_1, ..., x_{k-1}, a_k, x_{k+1}, ..., x_n) = \Delta_k\ F_{\xi_1 ... \xi_n}(x_1, ..., x_n).$$

а также $P\{\omega : \xi_k(\omega) \in A_k, k = \overline{1, n}\} = \Delta_1 \Delta_2 ... \Delta_n\ F_{\xi_1 ... \xi_n}(x_1, ..., x_n)$. Отсюда ясно, что для многомерной ф.р. должно выполняться условие

$$\Delta_1 \Delta_2 ... \Delta_n\ F_{\xi_1 ... \xi_n}(x_1, ..., x_n) \ge 0.$$

Это условие не следует из свойств 1) – 5). Покажем это на примере.

СВ

Пример 7.2. Пусть $n = 2$,

$$F_{\xi_1 \xi_2}(x_1, x_2) = \begin{cases} 0, x_1 < 0, x_2 < 0, x_1 + x_2 < 1, \\ 1, \text{в остальных случаях.} \end{cases}$$

Для такой функции, что легко проверить, выполняются свойства 1) – 5). Пусть $A_k = \left\{\omega : \frac{1}{3} < \xi_k(\omega) \le 1\right\}$, тогда

$$P\{\omega : \xi_k(\omega) \in A_k, k = 1, 2\} = \Delta_1 \Delta_2\ F_{\xi_1 \xi_2}(x_1, x_2) =$$

$$= \Delta_1 \left[F_{\xi_1 \xi_2}(x_1, 1) - F_{\xi_1 \xi_2}\left(x_1, \frac{1}{3}\right) \right] =$$

$$= F_{\xi_1 \xi_2}(1, 1) - F_{\xi_1 \xi_2}\left(\frac{1}{3}, 1\right) - F_{\xi_1 \xi_2}\left(1, \frac{1}{3}\right) + F_{\xi_1 \xi_2}\left(\frac{1}{3}, \frac{1}{3}\right) = 1 - 1 - 1 + 0 = -1.$$

Таким образом получили, что если $F_{\xi_1 \xi_2}(x_1, x_2)$ – ф.р., то найденная вероятность – отрицательная. Это невозможно, значит, выполнение условий 1) – 5) является недостаточным для того, чтобы $F_{\xi_1\xi_2}(x_1, x_2)$ была ф.р.

Так же, как и в одномерном случае, $F_{\xi}(x)$ относится к дискретному типу, если каждая из СВ $\xi_k(\omega)$, $k = \overline{1, n}$, принимает значения из счетного или конечного множества. Дискретную СВ $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega))$ удобно описывать распределением вероятностей

$$P_{\xi}(x) = P\{\omega : \xi_1(\omega) = x_1, \xi_2(\omega) = x_2, ..., \xi_n(\omega) = x_n\}, \ \sum_x P_{\xi}(x) = 1,$$

$$\text{где } x = (x_1, x_2, ..., x_n).$$

$F_{\xi}(x)$ относится к абсолютно непрерывному типу распределения, если ее можно представить в виде

$$F_{\xi}(x) = F_{\xi_1 ... \xi_n}(x_1, ..., x_n) = \int_{-\infty}^{x_1} dt_1 \ ... \int_{-\infty}^{x_n} p_{\xi_1 ... \xi_n}(t_1, ..., t_n) dt_n .$$

Здесь

$$\frac{\partial^n F_{\xi_1 ... \xi_n}(x_1, ..., x_n)}{\partial x_1 ... \partial x_n} = p_{\xi_1 ... \xi_n}(x_1, ..., x_n),$$

$$p_{\xi_1 ... \xi_n}(x_1, ..., x_n) \geq 0, \ \int_{-\infty}^{+\infty} dx_1 \ ... \int_{-\infty}^{+\infty} p_{\xi_1 ... \xi_n}(x_1, ..., x_n) dx_n = 1 .$$

Функция $p_{\xi}(x) = p_{\xi_1 ... \xi_n}(x_1, ..., x_n)$, обладающая перечисленными свойствами, называется плотностью распределения вероятностей многомерной СВ $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega))$ (совместной плотностью вероятностей величин $\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega)$. Из условия согласованности для $F_{\xi}(x)$ вытекает следующее свойство для совместной плотности вероятностей

$$\int_{-\infty}^{+\infty} p_{\xi_1 \dots \xi_{k-1} \xi_k \xi_{k+1} \dots \xi_n}(x_1, \dots x_{k-1}, x_k, x_{k+1}, \dots, x_n) dx_k =$$

$$= p_{\xi_1 \dots \xi_{k-1} \xi_{k+1} \dots \xi_n}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n).$$

Плотность распределения, стоящая справа, называется маргинальной по отношению к исходной. Справедлива также следующая важная формула:

$$P\{\omega : (\xi_1(\omega), \dots, \xi_n(\omega)) \in G\} = \int \dots \int_{G} p_{\xi_1 \dots \xi_n}(x_1, \dots x_n) dx_1 \dots dx_n ,$$

например,

$$P\{\omega : \xi_1^2(\omega) + \xi_2^2(\omega) \le Z\} = \iint_{\{x_1^2 + x_2^2 \le Z\}} p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 .$$

Рассмотрим примеры многомерных распределений.

1. Полиномиальное распределение имеет дискретная СВ $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$, где

$$\xi_k(\omega) \in \{0, 1, 2, \dots, N\}, \ \sum_{k=1}^{n} \xi_k(\omega) = N ,$$

$$P_\xi(x) = P_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) = N! \prod_{k=1}^{n} \frac{p_k^{x_k}}{x_k!}, \ 0 \le x_k \le N, \ k = \overline{1, n}, \ \sum_{k=1}^{n} x_k = N ,$$

$$0 < p_k < 1, \ k = \overline{1, n}, \ \sum_{k=1}^{n} p_k = 1 .$$

2. Многомерное нормальное распределение имеет СВ $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$, непрерывного типа, для которой

$$p_\xi(x) = p_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) = (2\pi)^{-\frac{\pi}{2}} \sqrt{|G|} \exp\left\{ -\frac{1}{2} \sum_{i,j=1}^{n} (x_i - a_i) g_{ij} (x_j - a_j) \right\},$$

где $G = \|g_{ij}\|_{n \times n}$ – неотрицательно определенная матрица, $|G|$ – ее определитель.

Пример 7.3. Пусть $(\xi(\omega), \eta(\omega))$ – двумерная СВ, имеющая нормальное распределение,

$$p_{\xi\eta}(x,y)=\frac{\sqrt{|G|}}{2\pi}exp\left\{-\frac{1}{2}\begin{pmatrix}x-a\\ y-b\end{pmatrix}^T G\begin{pmatrix}x-a\\ y-b\end{pmatrix}\right\}, G=\| g_{ik}\|_{2\times 2}.$$

Найдем плотности распределения СВ $\xi(\omega)$ и $\eta(\omega)$:

$$p_{\xi}(x)=\int_{-\infty}^{\infty}p_{\xi\eta}(x,y)dy=$$

$$=\sqrt{\frac{1}{2\pi}\left|g_{11}-\frac{g_{12}g_{21}}{g_{22}}\right|}exp\left\{-\frac{1}{2}\left(g_{11}-\frac{(g_{12}+g_{21})^2}{4g_{22}}\right)(x-a)^2\right\},$$

$$p_{\eta}(x)=\int_{-\infty}^{\infty}p_{\xi\eta}(x,y)dy=$$

$$=\sqrt{\frac{1}{\sqrt{2\pi}}\left|g_{22}-\frac{g_{12}g_{21}}{g_{11}}\right|}exp\left\{-\frac{1}{2}\left(g_{22}-\frac{(g_{12}+g_{21})^2}{4g_{11}}\right)(y-b)^2\right\}.$$

Из условий нормировки $\int_{-\infty}^{\infty}p_{\xi}(x)dx=\int_{-\infty}^{\infty}p_{\eta}(y)dy=1$ для элементов матрицы G вытекает следующее требование: $g_{12}=g_{21}$, $g_{11}g_{22}-g_{12}g_{21}=|G|>0$, т.е. матрица G должна быть симметричной и положительно определенной.

Пусть $\xi_k(\omega)$, $k=\overline{1,n}$, – СВ, определённые на вероятностном множестве элементарных событий $\Omega=\{\omega\}$, $A_k=\{\omega:\xi_k(\omega)\in B_k\}$, где B_k – множество на числовой прямой R, $B_k\subseteq R$.

Определение. СВ $\xi_k(\omega)$, $k=\overline{1,n}$, называются независимыми в совокупности, если, как бы ни выбирались множества B_k, $k=\overline{1,n}$, случайные события A_k, $k=\overline{1,n}$, являются независимыми в совокупности, т.е.

$$P\left(\bigcap_{i=1}^{m}A_{j_i}\right)=\prod_{i=1}^{m}P\left(A_{j_i}\right),\ \forall m\le n,\qquad .$$

Определение. СВ $\xi_k(\omega)$, $k=\overline{1,n}$, называются независимыми парами, если $\forall 1 \le i < k \le n$ и для любых множеств B_i и B_k события A_i и A_k являются независимыми, т.е. $P\left(A_i \bigcap A_k\right)=P\left(A_i\right)P\left(A_k\right)$.

Ясно, что СВ, независимые в совокупности, являются независимыми парами.

Приведём два критерия независимости СВ в совокупности.

Теорема. Для того чтобы СВ $\xi_k(\omega)$, $k=\overline{1,n}$, были независимыми в совокупности, необходимо и достаточно, чтобы $\forall$, $k=\overline{1,n}$,

$$F_{\xi_1 \ldots \xi_n}\left(x_1, \ldots, x_n\right)=\prod_{k=1}^{n} F_{\xi_k}\left(x_k\right).$$

Если $F_{\xi_1 \ldots \xi_n}\left(x_1, \ldots, x_n\right)$ – абсолютно непрерывная функция, то имеет место следующее утверждение.

Теорема. Для того чтобы СВ $\xi_k(\omega)$, $k=\overline{1,n}$, были независимыми в совокупности, необходимо и достаточно, чтобы

x_k

$$p_{\xi_1 \ldots \xi_n}\left(x_1, \ldots, x_n\right)=\prod_{k=1}^{n} p_{\xi_k}\left(x_k\right).$$

Доказательство следует из предыдущей теоремы и определения абсолютно непрерывной ф.р. Докажем, например, необходимость

$$F_{\xi_1 \ldots \xi_n}\left(x_1, \ldots, x_n\right)=\int_{-\infty}^{x_1} dt_1 \ldots \int_{-\infty}^{x_n} p_{\xi_1 \ldots \xi_n}\left(t_1, \ldots, t_n\right) dt_n=\prod_{k=1}^{n} F_{\xi_k}\left(x_k\right)=$$

$$=\prod_{k=1}^{n} \int_{-\infty}^{x_k} p_{\xi_k}\left(t_k\right) dt_k=\int_{-\infty}^{x_1} dt_1 \ldots \int_{-\infty}^{x_n} \prod_{k=1}^{n} p_{\xi_k}\left(t_k\right) dt_k .$$

Поскольку это выполняется $\forall$, x_2, …, x_n, то отсюда следует, что

$$p_{\xi_1 \ldots \xi_n}\left(t_1, \ldots, t_n\right)=\prod_{k=1}^{n} p_{\xi_k}\left(t_k\right).$$

Пример 7.4. Из примера 7.3 следует, что, для того чтобы нормально распределённые СВ $\xi(\omega)$ и $\eta(\omega)$ были независимыми, необходимо и достаточно, чтобы диагональные элементы матрицы G были равны нулю, т.е. $q_{12} = q_{21} = 0$. Приведем аналогичное утверждение для дискретных СВ.

Теорема. Пусть $\xi_1(\omega)$, $\xi_2(\omega)$, …, $\xi_n(\omega)$ – СВ, каждая из которых может принимать не более, чем счетное число значений. Они являются независимыми тогда и только тогда, когда $\forall$, x_2, …, x_n

$$P\{\omega : \xi_1(\omega) = x_1, \xi_2(\omega) = x_2, ..., \xi_n(\omega) = x_n\} = \prod_{k=1}^{n} P\{\omega : \xi_k(\omega) = x_k\}.$$

Рассмотрим следующую задачу. Пусть $f_k(x_1, x_2, ..., x_n)$, $k = \overline{1,m}$ – некоторые функции, определенные на R^n. Определим СВ

$$\eta_k(\omega) = f_k(\xi(\omega)) = f_k(\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega)),\ k = \overline{1,m},$$

где $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega))$ – n-мерная СВ. =СВ

$\eta_2(\omega), ..., \eta_m(\omega))$ является m-мерной. Пусть и $F_\eta(y)$ – ф.р. соответственно СВ $\xi(\omega)$ и $\eta(\omega)$. Нужно выразить $F_\eta(y)$ через $F_\xi(x)$ и систему функций $f_k(x)$, $k = \overline{1,m}$.

Допустим, что ф.р. $F_\xi(x)$ – абсолютно непрерывна. Рассмотрим следующие случаи.

1) Пусть $m = n$ и все функции , $k = \overline{1,n}$, являются дифференцируемыми и функционально независимыми, для последнего достаточно, чтобы

$$\det A = \det\|a_{ik}\| = \det\left\|\frac{\partial f_i(x)}{\partial x_k}\right\| \neq 0 \quad \forall x.$$

В данном случае будем иметь:

$$= \int\limits_{\{f_k(x_1,...,x_n) \le y_k,\, k=\overline{1,n}\}} dx_1 ... \int p_\xi(x_1, ..., x_n) dx_n.$$

Сделаем замену переменных $f_k(x_1, ..., x_n) = z_k$, $k = \overline{1,n}$, тогда

$$x_k = q_k(z_1,\dots,z_n),\ \ f_k\left[q_1(z_1,\dots,z_n),\dots,q_n(z_1,\dots,z_n)\right] = z_k\ ,$$
$$q\left(f_1(x_1,\dots,x_n),\dots,f_k(x_1,\dots,x_n)\right) = x_k\ ,\ \ k = \overline{1,n}\ ,$$

и поэтому

$$F_\eta(y) = \int\limits_{\left\{z_k \le y_k,\, k=\overline{1,n}\right\}} dq_1(z_1,\dots z_n)\int p_\xi\left(q_1(z_1,\dots,z_n),\dots,q_n(z_1,\dots,z_n)\right) dq_n(z_1,\dots,z_n) =$$

$$= \int\limits_{-\infty}^{y_1} dz_1 \dots \int\limits_{-\infty}^{y_n} p_\xi\left(q_1(z_1,\dots,z_n),\dots,q_n(z_1,\dots,z_n)\right)\left|\frac{\partial(q_1,\dots,q_n)}{\partial(z_1,\dots,z_n)}\right| dz_n\ ,$$

где $\left|\dfrac{\partial(q_1,\dots,q_n)}{\partial(z_1,\dots,z_n)}\right|$ – якобиан невырожденного преобразования $x_k = q_k(z_1,\dots,z_n)$, $k = \overline{1,n}$, т.е. определитель $(n \times n)$ – матрицы G с элементами $q_{ik} = \dfrac{\partial q_i(z_1,\dots,z_n)}{\partial z_k}$. Таким образом, $F_\eta(y)$ также абсолютно непрерывная ф.р., и плотность распределения СВ $\eta(\omega)$ равна

$f_k(x)$ $rang\left\|\dfrac{\partial f_i(x)}{\partial x_k}\right\| = m$

$$p_\eta(y_1,\dots,y_n) = p_\xi\left(q_1(y_1,\dots,y_n),\dots,q_n(y_1,\dots,y_n)\right)\left|\frac{\partial(q_1,\dots,q_n)}{\partial(y_1,\dots,y_n)}\right|.$$

Данное выражение можно записать в более краткой форме:

$$p_\eta(y) = p_\xi\left(f^{-1}(y)\right)\left|\frac{\partial f^{-1}(y)}{\partial(y)}\right|,$$

где $f(y) = \left(f_1(y),\dots,f_n(y)\right) = \left(f_1(y_1,\dots,y_n),\dots,f_n(y_1,\dots,y_n)\right)$.

2) Пусть теперь $m < n$ и $\forall x$.

В этом случае систему функций , $k = \overline{1,m}$, можно дополнить $(n-m)$ функциями $f_{m+j}(x)$, $1 \le j \le n-m$, так, чтобы они были непрерывно дифференцируемыми. Новая система функций определит n-мерную СВ . Тогда из условия согласованности и предыдущего случая будем иметь:

$$p_\eta(y_1,\dots,y_m)=\int\limits_{-\infty}^{\infty}dy_{m+1}\dots\int\limits_{-\infty}^{\infty}p_{\bar{\eta}}(y_1,\dots,y_n)dy_n=$$

$$=\int\limits_{-\infty}^{\infty}dy_{m+1}\dots\int\limits_{-\infty}^{\infty}p_\xi(q_1(y_1,\dots,y_n),\dots,q_n(y_1,\dots,y_n))\left|\frac{\partial(q_1,\dots,q_n)}{\partial(y_1,\dots,y_n)}\right|dy_n\,.$$

***Пример** 7.5.* Пусть у нас есть двумерная СВ

$$\xi(\omega)=(\xi_1(\omega),\xi_2(\omega)).$$

Образуем СВ $\eta(\omega)=\xi_1(\omega)+\xi_2(\omega)$ и найдем ее плотность распределения. В данном случае $n=2$, $m=1$, $f_1(x_1,x_2)=x_1+x_2=y_1$. Т.к. $m<n$, , то нужно ввести еще одну функцию, выберем $f_2(x_1,x_2)=x_2=y_2$. Тогда обратное преобразование определяется функциями $x_1=q_1(y_1,y_2)=y_1-y_2$, $x_2=q_2(y_1,y_2)=y_2$. Поэтому

$$\left|\frac{\partial(q_1,q_2)}{\partial(y_1,y_2)}\right|=\begin{vmatrix}1 & -1\\ 0 & 1\end{vmatrix}=1 \text{ и } p_\eta(y)=\int\limits_{-\infty}^{\infty}p_\xi(y-y_2,y_2)dy_2\,.$$

Если $\xi_1(\omega)$ и $\xi_2(\omega)$ – независимые СВ, то

$$p_\eta(y)=\int\limits_{-\infty}^{\infty}p_{\xi_1}(y-y_2)p_{\xi_2}(y_2)dy_2\,.$$

Эта формула известна в литературе как свертка для плотностей распределения вероятностей и обозначается $p_\eta=p_{\xi_1}*p_{\xi_2}$.

***Пример** 7.6.* Пусть $\xi_1(\omega)$, $\xi_2(\omega)$, …, $\xi_n(\omega)$ – независимые СВ, имеющие экспоненциальное распределение с параметром λ, т.е.

Найдем распределение СВ $\eta_n(\omega)=\xi_1(\omega)+\xi_2(\omega)+\ldots+\xi_n(\omega)$.

Пусть $n = 2$. Имеем

$$p_{\eta_2} = \int_0^y \lambda e^{-\lambda(y-y_2)} \lambda e^{-\lambda y_2} dy_2 = \lambda^2 e^{-\lambda y} \int_0^y dy_2 = \lambda^2 y e^{-\lambda y}, \ y \ge 0 .$$

Предположим, что для некоторого $n \ge 1$ справедлива формула

$$p_{\eta_n} = \lambda \frac{(\lambda y)^{n-1}}{(n-1)!} e^{-\lambda y}, \ y \ge 0 .$$

Докажем, что она справедлива также и для $n+1$. По той же формуле для плотности суммы двух независимых СВ $\eta_{n+1}(\omega) = \eta_n(\omega) + \xi_{n+1}(\omega)$ получаем

$$p_{\eta_{n+1}}(y) = \int_0^y \lambda \frac{[\lambda(y-y_2)]^{n-1}}{(n-1)!} e^{-\lambda(y-y_2)} \lambda e^{-\lambda y_2} dy_2 =$$

$$= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda y} \int_0^y (y-y_2)^{n-1} dy_2 = \lambda \frac{(\lambda y)^n}{n!} e^{-\lambda y}, \ y \ge 0 .$$

$\mu\ \xi_1(\omega)$ $P_{\xi_1}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Распределение с плотностью $p_{\eta_n}(y)$ называется распределение Эрланга n-го порядка.

Если СВ , $\xi_2(\omega)$ независимы и дискретны, то формула для распределения вероятностей их суммы $\eta(\omega) = \xi_1(\omega) + \xi_2(\omega)$ имеет вид

$$P_\eta(y) = P\{\omega : \xi_1(\omega) + \xi_2(\omega) = y\} = \sum_{\{y_1, y_2 : y_1 + y_2 = y\}} P\{\omega : \xi_1(\omega) = y_1, \xi_2(\omega) = y_2\} =$$

$$= \sum_{\{y_1, y_2 : y_1 + y_2 = y\}} P\{\omega : \xi_1(\omega) = y_1\} P\{\omega : \xi_2(\omega) = y_2\} = \sum_{\{y_1, y_2 : y_1 + y_2 = y\}} P_{\xi_1}(y_1) P_{\xi_2}(y_2) =$$

$$= \sum_{\{y_2\}} P_{\xi_1}(y - y_2) P_{\xi_2}(y_2).$$

***Пример** 7.7.* Пусть $\xi_1(\omega)$, $\xi_2(\omega)$ независимы и имеют пуассоновские распределения с параметрами λ, :

$$, \ P_{\xi_2}(l) = \frac{\mu^l}{l!} e^{-\mu}, \ k, \ l = 0,1,2,\ldots .$$

Покажем, что их сумма $\xi_1(\omega)+\xi_2(\omega)=\eta(\omega)$ имеет пуассоновское распределение с параметром $\lambda+\mu$:

$$=\frac{(\lambda+\mu)^k}{k!}e^{-(\lambda+\mu)},$$

поскольку справедлива формула бинома Ньютона

$$(\lambda+\mu)^k=\sum_{i=0}^{k}C_k^i\mu^i\lambda^{k-i}\ .$$

***Пример** 7.8.* Пусть имеем двумерную СВ

$$\xi(\omega)=(\xi_1(\omega),\xi_2(\omega))$$

непрерывного типа, для которой известна плотность распределения $p_\xi(x_1,x_2)$. Необходимо найти $p_\eta(y)$, где

$$\eta(\omega)=\frac{\xi_1(\omega)}{\xi_2(\omega)}.$$

В данном случае опять $n=2$, $m=1$. При этом

$$f_1(x_1,x_2)=\frac{x_1}{x_2}=y_1$$

и введем функцию $f_2(x_1,x_2)=x_2=y_2$. Обратное преобразование имеет вид: $x_1=q_1(y_1,y_2)=y_1y_2$, $x_2=q_2(y_1,y_2)=y_2$. Якобиан этого преобразования равен

$$\left|\frac{\partial(q_1,q_2)}{\partial(y_1,y_2)}\right|=\begin{vmatrix}y_2 & y_1\\ 0 & 1\end{vmatrix}=y_2,$$

поэтому $p_{\bar\eta}(y_1,y_2)=p_\xi(y_1y_2,y_2)|y_2|$, и, таким образом,

$$p_\eta(y)=\int_{-\infty}^{\infty}p_\xi(yy_2,y_2)|y_2|dy_2\ .$$

Если СВ $\xi_1(\omega)$ и $\xi_2(\omega)$ независимы, то

$$p_\eta(y) = \int_0^\infty |y_2| p_{\xi_1}(yy_2) p_{\xi_2}(y_2) dy_2 =$$

$$= \int_0^\infty y_2 p_{\xi_1}(yy_2) p_{\xi_2}(y_2) dy_2 - \int_{-\infty}^0 y_2 p_{\xi_1}(yy_2) p_{\xi_2}(y_2) dy_2 .$$

Используя этот результат, можно найти также плотность распределения произведения $\eta(\omega) = \xi_1(\omega)\xi_2(\omega)$:

$$p_\eta(y) = \int_{-\infty}^\infty p_\xi\left(y_2, \frac{y}{y_2}\right) \frac{1}{|y_2|} dy_2 ,$$

а если $\xi_1(\omega)$, $\xi_2(\omega)$ независимы, то

$$p_\eta(y) = \int_{-\infty}^\infty \frac{1}{|y_2|} p_{\xi_1}(y_2) p_{\xi_2}\left(\frac{y}{y_2}\right) dy_2 .$$

Задачи

η $p_\xi(x)$

$$p_{\xi\eta}(x, y) = \frac{a}{1 + x^2 + y^2 + x^2 y^2}$$

7.1. Дана плотность распределения вероятностей двумерной СВ

$$p_{\xi_1 \xi_2}(x, y) = \begin{cases} \frac{1}{2}\sin(x + y), 0 \le x \le \frac{\pi}{2}, 0 \le y \le \frac{\pi}{2}, \\ 0, \text{ в остальных случаях.} \end{cases}$$

Найти ф.р. $F_{\xi_1 \xi_2}(x, y)$.

7.2. Совместная плотность распределения случайных величин ξ и   имеет вид:

.

Найти коэффициент a ,   , $p_\eta(y)$; определить вероятность попадания случайной точки (ξ, η) в пределы квадрата с центром в начале координат, стороны которого параллельны осям координат и имеют длину, равную 2.

7.3. СВ (ξ_1, ξ_2) имеет плотность распределения

$$p_{\xi_1 \xi_2}(x, y) = \frac{a}{\pi^2 (3 + x^2)(1 + y^2)}.$$

Найти: а) величину а; б) ф.р. $F_{\xi_1 \xi_2}(x, y)$; в) вероятность попадания (ξ_1, ξ_2) в квадрат, который ограничен прямыми $x = 0$, $y = 0$, $x = 1$, $y = 1$.

7.4. Случайный вектор $\xi = (\xi_1, \xi_2, \xi_3)$ имеет плотность распределения

$$p_\xi(x, y, z) = \frac{a}{1 + x^2 + y^2 + z^2 + x^2 y^2 + x^2 z^2 + y^2 z^2 + x^2 y^2 z^2}.$$

Найти коэффициент a.

7.5. Случайный вектор имеет равномерное распределение внутри цилиндра

$$p_{\xi_1 \xi_2 \xi_3}(x_1, x_2, x_3) = \begin{cases} (2\pi r^2 h)^{-1}, x_1^2 + x_2^2 \le r^2, |x_3| \le h, \\ 0, x_1^2 + x_2^2 > r^2 \text{ или } |x_3| > h. \end{cases}$$

Определить плотности распределения отдельных компонентов ξ_1, ξ_2, ξ_3.

7.6. Пусть $0 < a \le 1$ и

$$p(x, y) = \begin{cases} [(1 + ax)(1 + ay) - a] e^{x - y - axy}, x > 0, y > 0, \\ 0, \text{ в остальных случаях.} \end{cases}$$

Доказать, что $p(x, y)$ – двумерная плотность распределения вероятностей, и найти маргинальные плотности распределения.

7.7. Пусть $u(x)$ – нечетная непрерывная функция на прямой, которая принимает значения, равные нулю, вне интервала $[-1, 1]$, причем $|u(x)| < \dfrac{1}{\sqrt{2\pi e}}$. Доказать, что функция

$p(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} + u(x)u(y)$ является двумерной плотностью распределения, отличающегося от нормального, но маргинальные распределения – нормальны.

7.8. Плотность распределения случайного вектора $\xi = (\xi_1, \xi_2)$ является равномерной внутри круга радиуса r с центром в начале координат. Написать ее выражение и выражения для плотностей распределения отдельных его компонент.

7.9. Студент и студентка договорились встретиться между 19 и 20 ч, условившись не ждать друг друга более 10 мин. Предположим, что моменты их прихода к месту встречи равномерно распределены между 19 и 20 ч. Найти вероятность встречи.

7.10. Закон распределения системы двух случайных величин определяется таблицей

$y_i \backslash z_i$	20	40	60
10	3λ	2λ	λ
20	λ	4λ	2λ
30	0	2λ	5λ

Найти λ. Составить ряд распределения для каждой из случайных величин и .

7.11. Передаются два сообщения, каждое из которых может быть независимо от другого либо искажено, либо не искажено. Вероятность искажения для первого сообщения равна , для второго – p_2 Рассматривается система двух случайных величин (ξ_1, ξ_2), определяемых следующим образом

$$\xi_1 = \begin{cases} 1, \text{если первое сообщение искажено,} \\ 0, \text{если первое сообщение не искажено,} \end{cases}$$

$$\xi_2 = \begin{cases} 1, \text{если второе сообщение искажено,} \\ 0, \text{если второе сообщение не искажено.} \end{cases}$$

Найти совместное распределение пары случайных величин (ξ_1, ξ_2). Найти совместную функцию распределения $F_{\xi_1 \xi_2}(x_1, x_2)$.

7.12. Каким условиям должны удовлетворять числа a , , c для того, чтобы при подходящем выборе нормировочной константы функция

$$Ae^{-(ax^2+2bx+c)}$$

являлась плотностью распределения вероятностей на плоскости?

7.13. Состояние замкнутой компьютерной сети, состоящей из систем (такими системами могут быть серверы, компьютеры пользователей и т.д.), описывается вектором

$$k=(k_1,k_2,...,k_n),\ \sum_{i=1}^{n}k_i=K,$$

где k_i – число заданий (запросов, сообщений) в i-й системе, – число заданий, обрабатываемых в сети. Распределение вероятностей состояний сети имеет вид:

$$P(k)=G\prod_{i=1}^{n}\left(\frac{e_i}{\mu_i}\right)^{k_i},$$

где μ_i – интенсивность обработки заданий в i-й системе, удовлетворяют системе уравнений

$$e_i=\sum_{j=1}^{n}e_j p_{ji},\ i=\overline{1,n},$$

p_{ji} – вероятность перехода задания после обработки из j-й системы в -ю, – нормировочная константа, определяемая из условия нормировки

$$\sum_{k\in D(k)}P(k)=1,\ D(K)=\left\{k\ /\ k_i\geq 0,\ i=\overline{1,n},\ \sum_{i=1}^{n}k_i=K\right\}.$$

Найти вероятности состояний сети в случае: а) $K=n=2$, $p_{12}=p_{21}=1$, $\mu_1=1$; $\mu_2=2$, б) $K=n=3$, $p_{12}=p_{13}=1$, $p_{21}=p_{31}=1$, $\mu_1=100$, $\mu_2=\mu_3=1$.

7.14. В условиях задач 7.4, 7.5 установить, является или нет СВ ξ_1, ξ_2, ξ_3 зависимыми.

7.15. Двумерная СВ $\xi = (\xi_1, \xi_2)$ задана плотностью распределения

$$p_{\xi_1\xi_2}(\xi_1, \xi_2) = \begin{cases} \dfrac{1}{6\pi}, \text{внутри эллипса } \dfrac{x_1^2}{9} + \dfrac{x_2^2}{4} \le 1, \\ 0, \text{вне эллипса.} \end{cases}$$

Зависимы или нет СВ ξ_1 и ξ_2?

7.16. Случайный вектор (ξ, η) с неотрицательными компонентами имеет ф.р. $F_{\xi\eta}(x, y) = 1 - e^{-\alpha x} - e^{-\beta y} + e^{-\alpha x - \beta y}$ $(\alpha > 0,\ \beta > 0)$. Являются ли зависимыми его компоненты?

7.17. Система СВ (ξ_1, ξ_2) распределена равномерно с постоянной плотностью внутри квадрата со стороной a. Написать выражения для плотностей $p_{\xi_1\xi_2}(x_1, x_2)$, $p_{\xi_1}(x_1)$, $p_{\xi_2}(x_2)$ и определить, являются ли СВ ξ_1 и ξ_2 независимыми.

7.18. Распределение СВ $\xi = (\xi_1, \xi_2)$ определяется формулами

$$P\{\xi_1 = 0, \xi_2 = 1\} = P\{\xi_1 = 1, \xi_2 = 0\} =$$
$$= P\{\xi_1 = 0, \xi_2 = -1\} = P\{\xi_1 = -1, \xi_2 = 0\} = 0{,}25.$$

Являются ли СВ ξ_1, ξ_2 независимыми?

7.19. Двумерная СВ задана таблицей

$\xi_1 \backslash \xi_2$	y_1	y_2	y_3
x_1	0.15	0.12	0.09
x_2	0	0.35	0.21
x_3	0	0	0.08

Зависимы ли СВ ξ_1 и ξ_2?

7.20. Пусть СВ ξ и   независимы и равномерно распределены на отрезке $[0,1]$. Найти вероятность того, что действительны корни квадратного уравнения $x^2 + \xi x + \eta = 0$.

7.21. Пусть ξ и – независимые СВ с одинаковой плотностью распределения . Найти плотность распределения суммы $\xi+\eta$.

7.22. Найти плотность распределения суммы независимых СВ ξ и , если имеет равномерное распределение на отрезке , а η – равномерное распределение на отрезке , $a<b$, , .

7.23. Пусть и ξ_2 – независимые СВ, ξ_1 имеет показательное распределение с параметром λ, равномерно распределена на отрезке $[a,b]$. Найти плотность распределения СВ $\xi_1+\xi_2$, $\xi_1-\xi_2$.

7.24. Доказать, что сумма $\xi_1+\xi_2$ независимых нормально распределенных СВ ξ_1 и ξ_2 с параметрами соответственно (a_1,τ_1^2), (a_2,τ_2^2) нормально распределена с параметрами $(a_1+a_2,\tau_1^2+\tau_2^2)$.

7.25. СВ ξ_i, $i=\overline{1,n}$, независимы и имеют нормальное распределение с параметрами соответственно (a_i,τ_i^2)**.** Показать, что СВ $\xi=\xi_1+\ldots+\xi_n$ имеет нормальное распределение с параметрами $a=a_1+\ldots+a_n$, $\tau=\tau_1^2+\ldots+\tau_n^2$.

7.26. Найти ф.р. произведения независимых СВ ξ_1 и ξ_2 по их ф.р. $F_{\xi_1}(x)$ и $F_{\xi_2}(x)$.

7.27. СВ ξ принимает значения , , 2 с вероятностями соответственно 0.25, 0.5, 0.25, а СВ η, независимая от , принимает значения и 1 с вероятностью 0.5. Найти распределение СВ $\xi+\eta$.

7.28. СВ ξ и независимы и имеют равномерное распределение на отрезке . Найти плотности распределения СВ $\xi+\eta$,

, , .

7.29. Пусть СВ ξ_1 и ξ_2 независимы и каждая имеет показательное распределение с параметром λ. Показать, что: а) СВ равномерно распределена на отрезке $[0,\ 1]$, б) СВ $\dfrac{\xi_1}{\xi_1+\xi_2}$ и $\xi_1+\xi_2$ независимы.

7.30. Студент при поездке в университет пользуется двумя автобусами; первого ему приходится ожидать не более 5 минут, второго – не более 10 минут. Считая время ожидания ξ и η автобусов независимыми случайными величинами, распределенными равномерно соответственно в интервалах и $[0,\ 10]$, найти плотность распределения суммарного ожидания $\xi+\eta$.

7.31. Решить предыдущую задачу в случае, когда ξ и распределены по показательному закону соответственно с параметрами и .

7.32. Для выполнения некоторой работы необходимо выполнить последовательно две операции. Время выполнения первой операции имеет равномерное распределение на отрезке $[1,\ 3]$, время выполнения второй операции t_2 – СВ, равномерно распределенная на отрезке $[2,\ 5]$. Найти распределение времени выполнения всей работы t_1+t_2, если t_1 и t_2 – независимые CB.

7.33. СВ ξ_i доходов фирмы за i-й рабочий день имеет гамма-распределение с параметром , $i=\overline{1,n}.$ Найти плотность распределения среднего дохода фирмы $\eta=\dfrac{1}{n}\left(\xi_1+\ldots+\xi_n\right)$ за один рабочий день.

7.34. Пусть ξ – случайное число изделий. Каждое изделие с вероятностью p является бракованным. Обозначим через число бракованных изделий, через ξ_2 – число не бракованных изделий. Показать, что СВ ξ_1 и ξ_2 независимы тогда и только тогда, когда ξ имеет распределение Пуассона.

7.35. Совместное распределение $p_{ij} = P\{\xi_1 = a_i,\ \xi_2 = a_j\},\ i,j = \overline{1,3},$ случайных доходов фирмы ξ_1, ξ_2 в течении двух последовательных рабочих дней задано таблицей $\left(a_i \in \{-1000,\ 0,\ 1000\}\right)$:

ξ_2 \ ξ_1	-1000	0	1000
-1000	0.2	0.1	0.3
0	0.1	0.2	0.2
1000	0.1	0.2	0.3

Найти: а) одномерные распределения $p_i = P\{\xi_1 = a_i\},\ p_i = P\{\xi_2 = a_j\}$; б) распределение среднего дохода $\xi = \frac{1}{2}(\xi_1 + \xi_2)$; в) совместное распределение среднего дохода ξ и прироста дохода

§8. Математическое ожидание случайных величин

Перед тем, как дать формальное определение математического ожидания, рассмотрим пример.

Пример 8.1. *(среднее значение СВ).* Пусть $\xi(\omega)$ – СВ со значениями на числовой оси R, заданная на множестве элементарных событий $\Omega = \{\omega\}$. Нужно найти среднее значение СВ – значение, которое СВ принимает в среднем.

Решение. Выберем сначала единицу измерения ε значений СВ и рассмотрим множество точек $A_{k,\varepsilon} = \{\omega : k\varepsilon < \xi(\omega) \le (k+1)\varepsilon\}$, т.е. $A_{k,\varepsilon}$ – множество точек множества элементарных событий Ω, в которых $\xi(\omega)$ принимает значения от $k\varepsilon$ до $(k+1)\varepsilon$, $k = 1,2,\dots$. На множестве рассмотрим последовательность СВ $\xi_\varepsilon(\omega)$, которые зависят от параметра ε, положив для $\omega \in A_{k,\varepsilon}$, $k = 1, 2, \dots$.

В качестве среднего значения СВ естественно рассматривать величину

$$\lim_{\varepsilon\to 0}\sum_k k\varepsilon P\{\omega : k\varepsilon < \xi(\omega) \le (k+1)\varepsilon\},$$

которую принято называть математическим ожиданием случайной величины $\xi(\omega)$ (средним значением СВ $\xi(\omega)$).

Определение. Возьмем в качестве ε величину . Таким образом, математическим ожиданием СВ $\xi = \xi(\omega)$, заданной на множестве $\Omega = \{\omega\}$, называется число, равное

$$M\xi = \lim_{n\to\infty}\sum_{k=-\infty}^{\infty}\frac{k}{n}P\left\{\omega :\ \frac{k}{n} < \xi(\omega) \le \frac{k+1}{n}\right\} =$$

$$= \lim_{n\to\infty}\sum_{k=-\infty}^{\infty}\frac{k}{n}\left[F_\xi\left(\frac{k+1}{n}\right) - F_\xi\left(\frac{k}{n}\right)\right] = \int_X x dF_\xi(x),$$

$\xi(\omega)$ $\int_X x dF_\xi(x)$ n

где $F_\xi(x)$ – ф.р. СВ ξ, называется интегралом Стилтьеса.

Из этого определения имеются следствия:

а) если $\xi(\omega)$ – дискретная СВ, которая принимает значения из множества $\{x_1, x_2, ..., x_k, ...\}$, то

$$M\xi = \sum_{k=1} x_k P\{\omega : \xi(\omega) = x_k\} = \sum_{k=1} x_k p_k\ ;$$

б) если $\xi(\omega)$ имеет абсолютно непрерывное распределение (плотность $p_\xi(x)$), то

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

в) математическое ожидание $M\xi$ существует, если $M|\xi| < +\infty$.

Пример 8.2. Найдем математическое ожидание СВ ξ, распределенной по закону Пуассона.

Решение. Для такой СВ , $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$,

$k = 0, 1, 2, \ldots$.
Поэтому

.

Пример 8.3. Пусть ξ принимает значения с вероятностями $p_k = \frac{1}{2^k}$, $k = 1, 2, \ldots$. Попарно просуммировав последовательные значения , получим $x_{2k-1} p_{2k-1} + x_{2k} p_{2k} = 0$, $k = 1, 2, \ldots$. На первый взгляд может показаться, что . Однако

$$\sum_{\{k\,:\,x_k > 0\}} x_k p_k = \sum_{k=1}^{\infty} 2^{2k} 2^{-2k} = \infty, \quad \sum_{\{k\,:\,x_k < 0\}} x_k p_k = -\sum_{n=1}^{\infty} 2^{2k-1} 2^{1-2k} = -\infty,$$

поэтому $M\xi$ не существует.

Пример 8.4. Найдем математическое ожидание СВ ξ, равномерно распределенной на отрезке .

Решение. Данная СВ непрерывного типа. Учитывая для нее вид плотности распределения, будем иметь

$$M\xi = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2},$$

т.е. математическое ожидание равно середине отрезка $[a,b]$.

Свойства математического ожидания следуют из свойств интегралов и рядов. Рассмотрим основные из них (будем предполагать, что соответствующие математические ожидания, входящие в приведенные в данных свойствах соотношения, существуют).

1. Если $P(\omega : \xi(\omega) \geq 0) = 1$, то $M\xi \geq 0$.

2. $M(c\xi) = cM\xi$, где c – константа, .

3. Если , то $M\xi \geq M\eta$.

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

5. $|M\xi| \leq M|\xi|$.

6. Пусть $I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$, $A \in$ F. Такая СВ называется индикатором случайного события A. Тогда

$$MI_A(\omega) = 1 \cdot P\{\omega \in A\} + 0 \cdot P\{\omega \notin A\} = P(A),$$

т.е. математическое ожидание такой СВ равно вероятности события A.

7. Пусть ξ и – независимые СВ, тогда .

8. Пусть ξ – СВ, – действительная функция. Тогда $g(\xi)$ также является СВ и, если $M|g(\xi)| < +\infty$, то

$$Mg(\xi) = \int_R g(x) dF_\xi(x),$$

в частности,

$$Mg(\xi) = \sum_{k=1}^{\infty} g(x_k) p_k, \text{ если } \xi \text{ – дискретная СВ},$$

$Mg(\xi) = \int g(x) p_\xi(x) dx$, если ξ – непрерывная СВ.

Пример 8.5. Правомерно ли следующее рассуждение: «От общежития до университета расстояние равно 1 км, студент ходит в среднем со скоростью 5 км/ч, следовательно, в среднем на дорогу у него будет уходить 12 мин»?

Решение. Прежде всего уточним формулировку задачи: имеется в виду, что скорость – случайная величина, математическое ожидание которой равно 5 км/ч. Рассуждение является неправомерным, т.к.

из свойства 8 следует, что, вообще говоря, .

Для многомерной СВ $\xi = (\xi_1, \xi_2, ..., \xi_n)$ математическое ожидание определяется в виде вектора

$$M\xi = (m_1, m_2, ..., m_n), \text{ где } m_k = M\xi_k,\ k = \overline{1, n}.$$

Если, например, СВ ξ – дискретного типа и $\xi_k \in \{0, 1, 2, ..., N\}$, $k = \overline{1, n}$, $\sum\limits_{k=1}^{n} \xi_k(\omega) = N \quad \forall \omega \in \Omega$, то

$$m_k = \sum_{x_1=0}^{N} \sum_{x_2=0}^{N-x_1} \dots \sum_{x_n=0}^{N-x_1-x_2-...-x_{n-1}} x_k P\{\xi_1 = x_1, \xi_2 = x_2, ..., \xi_n = x_n\},\ k = \overline{1, n}.$$

Если СВ $\xi = (\xi_1, \xi_2, ..., \xi_n)$ непрерывного типа и $g(x_1, x_2, ..., x_n)$ – измеримая функция, то

$$Mg(\xi_1, \xi_2, ..., \xi_n) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, x_2, ..., x_n) p_{\xi_1 \xi_2 ... \xi_n}(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n,$$

где $p_{\xi_1 \xi_2 ... \xi_n}(x_1, x_2, ..., x_n)$ – плотность распределения n-мерной СВ .
В частности,

.

Рассмотрим неравенства, справедливые для математических ожиданий.

1) **Неравенство Чебышева.** Пусть $f(x)$ – неотрицательная, монотонно неубывающая борелевская функция, определенная на интервале $[0, +\infty)$ со значениями в R. Тогда для любой СВ ξ и

,

в частности,

$$P\{|\xi| \geq \varepsilon\} \leq \frac{M|\xi|^k}{\varepsilon^k} \quad \forall k \geq 0 \ –$$

это неравенство называется неравенством Маркова. При $k = 1$ имеем:

$$P\{|\xi| \geq \varepsilon\} \leq \frac{M|\xi|}{\varepsilon}.$$

2) **Неравенство Иенсена.** Напомним, что функция $g(x)$ со значениями в R, заданная на интервале $I \subset R$, называется выпуклой на нем вниз (вверх), если для любого найдется число $\lambda(x_0)$ такое, что для всех $x \in I$:

$$(g(x_0) + \lambda(x_0)(x - x_0) \geq g(x)).$$

Т.к. $y = g(x_0) + \lambda(x_0)(x - x_0)$ является уравнением прямой, которая проходит через точку $(x_0, g(x_0))$ с угловым коэффициентом $\lambda(x_0)$, то выпуклость вниз (вверх) означает, что в любой точке $(x_0, g(x_0))$ можно провести касательную прямую так, что график функции $g(x)$ лежит выше (ниже) этой прямой.

Пусть $g(x)$ – выпуклая вниз (вверх) борелевская функция на R. Тогда, если $M|\xi| < +\infty$, то выполняется неравенство Иенсена

$$g(M\xi) \leq Mg(\xi) \quad (g(M\xi) \geq Mg(\xi)),$$

в частности,

$$(M\xi)^2 \leq M\xi^2, \ |M\xi|^r \leq M|\xi|^r, \ r \geq 1.$$

Поясним, почему неравенство Иенсена имеет место, например, для выпуклой вниз функции $g(x)$. Для такой функции

$$g(x_0) + \lambda(x_0)(x - x_0) \leq g(x), \ x \in R.$$

Положим , $x = \xi$, тогда

$$g(M\xi) + \lambda(M\xi)(\xi - M\xi) \leq g(\xi).$$

Взяв в этом неравенстве математическое ожидание в левой и правой части, получим требуемое неравенство.

3) **Неравенство Гёльдера.** Пусть $\frac{1}{p}+\frac{1}{q}=1$, $p>0$, $q>0$, тогда

$$M|\xi\eta|\le\left(M|\xi|^p\right)^{\frac{1}{p}}\left(M|\eta|^q\right)^{\frac{1}{q}}.$$

Если в неравенстве Гёльдера положить $\xi=|\xi|^s$, $\eta=1$, ,

$\frac{1}{q}=1-\frac{s}{t}$, где $0<s<t$, то получим неравенство Ляпунова:

$$\left(M|\xi|^s\right)^{\frac{1}{s}}\le\left(M|\xi|^t\right)^{\frac{1}{t}}.$$

При $p=q=2$ из неравенства Гёльдера получаем неравенство Коши – Буняковского (Шварца):

$$M|\xi\eta|\le\sqrt{M\xi^2}\sqrt{M\eta^2}.$$

Пример 8.6. Показать, что

$$P\left\{\left|\sum_{i=1}^{n}\xi_i\right|\ge\varepsilon\right\}\le\frac{1}{\varepsilon}\sum_{i=1}^{n}M|\xi_i|.$$

Решение. Применив неравенство Чебышева и свойство 5 математического ожидания, получим

$$P\left\{\left|\sum_{i=1}^{n}\xi_i\right|\ge\varepsilon\right\}\le\frac{1}{\varepsilon}M\left|\sum_{i=1}^{n}\xi_i\right|\le\frac{1}{\varepsilon}\sum_{i=1}^{n}M|\xi_i|.$$

Приведем практический пример, иллюстрирующий, как можно использовать понятие математического ожидания.

Пример 8.7. Продавец получает N единиц товара в день и стремится заказать число N таким образом, чтобы максимизировать ожидаемую прибыль. Найти оптимальный параметр N^*, если число покупателей товара в данный день следует закону Пуассона с параметром $\lambda>0$. Прибыль, получаемая от единицы проданного товара, равна , убыток от нереализованной единицы товара – , c – убыток, если покупатель желает приобрести товар, но их запас исчерпан.

Решение. Пусть – число покупателей в данный день, тогда чистая прибыль продавца имеем вид

Ожидаемая прибыль является математическим ожиданием $G_N = Mg_N(\xi)$. Изменение ожидаемой прибыли при добавлении еще одной единицы товара равно

$$G_{N+1} - G_N = M\left[g_{N+1}(\xi) - g_N(\xi)\right] = M\left[-b + (a+b+c)u(\xi - N)\right],$$

где $u(x) = \begin{cases} 1, x > 0 \\ 0, x \le 0 \end{cases}$ – функция Хевисайда. Поскольку $Mu(\eta) = 1 \cdot P\{\eta > 0\} +$

$+ 0 \cdot P\{\eta \le 0\} = P\{\eta > 0\}$, то приращение ожидаемой прибыли равно

$$G_{N+1} - G_N = -b + (a+b+c)Mu(\xi - N) = -b + (a+b+c)P\{\xi > N\}.$$

Для нахождения $\max\limits_N G_N$ достаточно решить уравнение

$G_N = G_{N+1}$, откуда получаем

ξ

$$g_n(\xi) = \begin{cases} a\xi - b(N - \xi), \text{ если } \xi \le N, \\ aN - c(\xi - N), \text{ если } \xi > N. \end{cases}$$

$$P\{\xi > N\} = \frac{b}{a+b+c}, \text{ т.е. } 1 - F_\xi(N) = \frac{b}{a+b+c}.$$

Т.к. для закона Пуассона

$$F_\xi(N) = \sum_{k=0}^{N} \frac{\lambda^k}{k!} e^{-\lambda},$$

то уравнение для нахождения оптимального N^* имеет вид:

$$\sum_{k=0}^{N^*} \frac{\lambda^k}{k!} e^{-\lambda} = \frac{b}{a+b+c} + 1$$

или в неявной форме

$$N^* = F_\xi^{-1}\left(\frac{b}{a+b+c} + 1\right).$$

Числовое значение данного выражения можно найти, воспользовавшись таблицей распределения Пуассона.

Задачи

8.1. Найти математическое ожидание СВ, распределенной по нормальному закону с параметрами a , .

8.2. Существует ли математическое ожидание СВ ξ, имеющей распределение Коши? Найти .

8.3. Пусть ξ – СВ, равномерно распределенная на отрезке . Найти $M\cos^2(\pi\xi)$, $M\sin^2(\pi\xi)$.

8.4. Найти математические ожидания дискретных СВ, имеющих: а) распределение Бернулли, б) биномиальное распределение, в) геометрическое распределение.

8.5. Найти $M\left(\dfrac{1}{a\xi+b}\right)$, где ξ – СВ, имеющая распределение Пуассона, , – константы.

8.6. СВ ξ имеет бета-распределение с плотностью

где

$B(a,b)=\int\limits_0^1 x^{a-1}(1-x)^{b-1}dx=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $a,b>0$, – бета-функция Эйлера.

Найти $M\xi$.

8.7. Пусть ξ – неотрицательная СВ с конечным математическим ожиданием, имеющая ф.р. . Доказать, что

$$M\xi=\int\limits_0^\infty\left(1-F_\xi(x)\right)dx.$$

8.8. СВ ξ – неотрицательная целочисленная величина с конечным математическим ожиданием. Показать, что

.

8.9. Пусть ξ и – независимые одинаково распределенные СВ. Найти . Являются ли независимыми СВ $(\xi-\eta)$ и $(\xi+\eta)$?

8.10. При условии задачи 7.2. найти $M\xi$, $M\eta$.

8.11. Найти $M\xi_i$, $i=\overline{1,3}$, в задаче 7.4.

8.12. При условии задачи 7.1 найти $M\xi_1$, $M\xi_2$.

8.13. Найти среднее значение квадрата расстояния между двумя точками, выбранными наугад на любой из сторон прямоугольника.

8.14. Интервалы времени между движением автомашин на дороге имеют показательное распределение с параметром λ. Найти интенсивность потока автомашин на дороге.

8.15. Имеется сетевой график планирования управления, согласно которому момент начала какой-то работы представляет собой максимальное время окончания двух обеспечивающих работ , ξ_2 (моменты окончания этих работ). Случайные величины ξ_1, ξ_2 независимы и имеют плотности $p_{\xi_1}(x_1)$ и $p_{\xi_2}(x_2)$. Найти среднее значение случайной величины η.

8.16. Техническое устройство состоит из узлов. Каждый узел

$M[(\xi-\eta)(\xi+\eta)]$

может выходить из строя независимо от других. Время исправной работы -го узла распределено по показательному закону с параметром . Каждый узел, оказавшийся неисправным, немедленно заменяется новым и поступает в ремонт. Ремонт i-го узла продолжается случайное время, распределенное по показательному закону с параметром . Устройство работает в течении времени t. Определить: а) среднее число узлов, которые придется заменить; б) среднее время , которое будет затрачено на ремонт вышедших из строя узлов.

8.17. В результате испытаний прибор может быть отнесен к классу I с вероятностью p_1, к классу II с вероятностью p_2, или быть забракованным с вероятностью $p_3=1-p_1-p_2$. Испытание проходят n приборов. Определить распределение вероятностей различного числа приборов классов I и II, их средние значения.

8.18. Из десяти изделий, среди которых два бракованных, случайным образом выбирают два для проверки. Найти среднее значение числа бракованных изделий.

8.19. Среди 7 приборов 3 неисправных. Наугад берут 4 прибора и проверяют их. Найти среднее значение числа приборов, которые при этом будут работать исправно.

8.20. Вероятность изготовления нестандартной детали равна 0,1. Из партии контролер случайным образом берет деталь и проверяет ее качество. Если она оказывается нестандартной, дальнейшие испытания прекращаются, а партия задерживается. Если деталь окажется стандартной, то контролер берет следующую и т.д. Всего проверяет он не более 5 деталей. Найти математическое ожидание – числа проверяемых стандартных деталей.

8.21. Уровень весеннего паводка на реке является случайной величиной ξ с функцией распределения . Плотина рассчитана так, чтобы выдерживать паводок уровня не выше z. Предполагая, что уровни паводков в разные годы независимы и одинаково распределены, найти минимальное значение z, при котором вероятность разрушения плотины паводком за 100 лет будет не больше .

8.22. Производится ряд попыток включить двигатель. Каждая попытка занимает время τ и заканчивается успехом (включением двигателя) независимо от других с вероятностью . Найти распределение общего времени T, которое потребуется для запуска двигателя и его среднее значение.

8.23. Радиолокационная станция ведет слежение за областью пространства, где находится N объектов. За один цикл обзора i-й объект независимо от других обнаруживается с вероятностью , $i=\overline{1, N}$. За время наблюдения осуществляется n циклов обзора. Найти среднее число объектов, которые будут обнаружены.

8.24. Наблюдаемый объект на круглом экране радиолокатора отображается светящейся точкой. Будем считать, что точка может занимать на экране любое положение. Диаметр экрана равен . Найти среднее значение расстояния от точки до центра экрана.

8.25. Написано n писем, но адреса на конвертах написаны в случайном порядке. Пусть – число писем, которые будут получены теми адресами, которым они предназначались. Показать, что $M\xi_n = 1$.

8.26. В N телефонных автоматах ведутся разговоры. Длительность разговора, измеряемого в секундах, имеет геометрическое распределение с математическим ожиданием μ. Найти среднее время ожидания до первого освобождения телефона-автомата.

8.27. Стреляют три раза по мишени. Вероятности попадания в каждом выстреле равны соответственно ; 0.2; 0.3. Найти среднее число попаданий.

8.28. Скорость молекул газа ξ является СВ, распределенной по закону Максвелла:

Найти среднюю скорость молекул.

8.29. Из теории броуновского движения известно, что если частица в момент времени $t = 0$ находится на расстоянии x_0 от отражающей стенки, то вероятность того, что в момент $t > 0$ она будет находиться от стенки на расстоянии между x и , равняется , где

$$p(x) = \frac{1}{2\sqrt{\pi D t}}\left[e^{-\frac{(x-x_0)^2}{4Dt}} + e^{-\frac{(x+x_0)^2}{4Dt}}\right].$$

Найти среднее значение перемещения частицы за время t.

$$p_\xi(x) = \begin{cases} \frac{4h^3}{\sqrt{\pi}} x^2 e^{-h^2x^2}, x \geq 0, \\ 0, x < 0. \end{cases}$$

8.30. Предположим, что вам необходимо выбрать работу из двух предлагаемых работ (по первой работе оплата сдельная, по второй – повременная). Известно, что на первой работе доход с одинаковой вероятностью составит \$200 при хорошей распродаже и \$100 при скромной. На второй работе ставка \$151, но если компания обанкротится (вероятность этого равна), то вы получите пособие размером \$51. Какую работу вы бы предпочли?

8.31. По маршруту ходит N автобусов. У водителя каждого из них было k билетов. Эти автобусы вместе перевезли n пассажиров. Найти среднее число пассажиров, которым не досталось билетов, если каждый пассажир независимо от остальных может сесть в любой из автобусов с одинаковой вероятностью.

8.32. В лотерее имеется билетов и разыгрывается m_1 выигрышей стоимостью c_1, m_2 выигрышей стоимостью $c_2, \ldots, m_n$ – стоимостью c_n. Какую стоимость лотерейного билета следует установить, чтобы средний выигрыш составлял 50 % его стоимости?

8.33. В партии имеется n изделий, каждое из которых независимо от остальных с вероятностью удовлетворяет стандарту, а с вероятностью не удовлетворяет ему. Изделия проходят проверку, описанную в задаче 3.33. За каждое изделие, удовлетворяющее стандарту, фирма-изготовитель получает a руб. премии; за изделие, прошедшее проверку, но не удовлетворяющее стандарту, – штраф руб.; за изделие, не прошедшее проверку, – штраф c руб. Найти среднюю прибыль фирмы, полученную за партию из изделий.

8.34. Обследуется крупный пакет акций из штук. Известно, что каждая акция с вероятностью p обесценивается. Обследование происходит путем анализа экономического проекта, в котором участвуют акции пакета. Применяют два способа обследования:

а) обследовать каждую из акций;

б) вести обследование по группам из n, , акций, причем если пакет неубыточный, то считают, что все акции данной группы растут в цене; если же наоборот (это происходит, если хотя бы одна акция группы обесценилась), то переходят к сплошному анализу акций из данной группы.

Определить: 1) какой способ выгоднее в смысле минимального среднего числа анализов, 2) при каком $n = n^*$ для обследования группы акций потребуется в среднем наименьшее число анализов.

8.35. Пусть бизнесмен имеет убыток $\alpha\tau$, если попадет к месту встречи ранее намеченного срока на время . Положим, что время , необходимое, чтобы попасть к месту встречи, является экспоненциально распределённой СВ с параметром . Предполагая, что бизнесмен отправляется к месту встречи за время до назначенного срока, доказать, что величина , минимизирующая ожидаемые потери, определяется из уравнения

$$P\{\xi \le t^*\} = F_\xi(t^*) = \frac{b}{a+b} .$$

§9. Другие числовые характеристики случайных величин

В данном параграфе рассмотрим другие различные числовые характеристики СВ, которые являются математическими ожиданиями определенных функций от СВ.

Моменты СВ

Определение. Начальным моментом порядка N СВ ξ называется , а центральным моментом порядка $N - \mu_N = M(\xi - M\xi)^N$.

Определение. В случае многомерных СВ смешанным начальным моментом порядка N СВ $\xi = (\xi_1, \xi_2, ..., \xi_n)$ называется $M\left[\xi_1^{k_1}\xi_2^{k_2} \cdot ... \cdot \xi_n^{k_n}\right]$, где $\sum_{i=1}^{n} k_i = N$, k_i – целые неотрицательные числа.

Смешанным центральным моментом порядка N называется

$$M\left[(\xi_1 - M\xi_1)^{k_1}(\xi_2 - M\xi_2)^{k_2}...(\xi_n - M\xi_n)^{k_n}\right].$$

$m_N = M\xi^N$, σ^2, $p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $m_n = M\xi^n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2}} dx$, $n = 2k$

Пример 9.1. Найдем моменты СВ, имеющей нормальное распределение с параметрами a, .

Сделаем это вначале для случая $a = 0$, $\sigma = 1$, плотность распределения СВ при этом имеет вид

$$, -\infty < x < +\infty,$$

поэтому

$$.$$

При нечетных n этот интеграл равен нулю как интеграл нечетной функции по симметричному промежутку интегрирования. Таким образом, остается найти при четных n. Пусть , тогда

$$M\xi^2=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2e^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^2e^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}\int_0^{\infty}xd\left(-e^{-\frac{x^2}{2}}\right)=$$

$$=-\sqrt{\frac{2}{\pi}}xe^{-\frac{x^2}{2}}\Bigg|_0^{\infty}+\sqrt{\frac{2}{\pi}}\int_0^{\infty}e^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}\int_0^{\infty}e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx=\int_{-\infty}^{\infty}p_\xi(x)dx=1;$$

при этом мы использовали предел $\lim\limits_{x\to\infty}x\,e^{-\frac{x^2}{2}}dx=0$, который легко проверяется по правилу Лопиталя. Аналогично, проинтегрировав интеграл, выражающий $M\xi^n$, по частям, получим

$$M\xi^n=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^ne^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^ne^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^{n-1}d\left(-e^{-\frac{x^2}{2}}\right)=$$

$$=-\sqrt{\frac{2}{\pi}}x^{n-1}e^{-\frac{x^2}{2}}\Bigg|_0^{\infty}+(n-1)\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^{n-2}e^{-\frac{x^2}{2}}dx=$$

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

Итак, доказана рекуррентная формула

$$M\xi^n=(n-1)M\xi^{n-2},\ M\xi=0.$$

Отсюда

$$M\xi^n=(n-1)M\xi^{n-2}=(n-1)(n-3)M\xi^{n-4}=\ldots=$$
$$=(n-1)(n-3)\cdot\ldots\cdot3M\xi^2=(n-1)(n-3)\ldots3\cdot1.$$

Положим $n=2k$. Тогда

$$= \frac{(2k-1)!}{2^{k-1}(k-1)!}.$$

Таким образом,

$$m_{2k} = M\xi^{2k} = \frac{(2k-1)!}{2^{k-1}(k-1)!},\ k = 1, 2, \ldots .$$

Найдем теперь центральные моменты распределения СВ η с плотностью

$$,\ -\infty < x < +\infty .$$

Если ввести СВ по формуле , то имеем $p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$,

а из формулы $\eta = a + \sigma\xi$ находим

ξ $\eta = a + \sigma\xi$ $D\xi = M(\xi - M\xi)^2$ $(n-1)(n-3)\cdot\ldots\cdot 1$ $p_\eta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ $(2k-1)(2k-3)\ldots 1 = \frac{(2k-1)!}{[2(k-1)][2(k-2)]\cdot\ldots\cdot 2} = \frac{(2k-1)!}{2^{k-1}(k-1)!}.$ $M\eta = a + \sigma M\xi = a$

Рассмотрим центральные моменты СВ η любых порядков. Т.к. , то

$$M(\eta - M\eta)^n = M(\sigma\,\xi)^n = \sigma^n M\xi^n = \begin{cases} 0,\ n = 2k-1, \\ \frac{\sigma^{2k}(2k-1)!}{2^{k-1}(k-1)!},\ n = 2k, \end{cases} \quad k = 1, 2, \ldots .$$

Дисперсия и ее свойства

Определение. Дисперсией СВ ξ называется ее центральный момент второго порядка:

,

$\sqrt{D\xi}$ называется средним квадратичным отклонением СВ ξ.

Дисперсия служит характеристикой (хотя и не полной) рассеяния значений СВ от ее среднего значения.

Из определения следует, что

.

Рассмотрим свойства дисперсии.

1. $D\xi \geq 0$.

2. $D\xi = 0$ только в том случае, когда $P(\xi = const) = 1$.

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

4. $D(c+\xi) = M[c+\xi - M(c+\xi)]^2 = M(\xi - M\xi)^2 = D\xi$.

5. Пусть ξ и   – независимые СВ, тогда

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

поскольку для независимых СВ $M(\xi\eta) = M\xi M\eta$.

6. Если в неравенстве Чебышева для математического ожидания в качестве $f(x)$ взять x^2, а в качестве СВ взять $(\xi - M\xi)$ (или, что то же самое, в неравенство Маркова вместо ξ подставить), то получим неравенство Чебышева для дисперсии:

$$P\{|\xi - M\xi| \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2} \quad \forall \varepsilon > 0.$$

Пример 9.2. Из предыдущего примера следует, что дисперсия СВ , имеющей нормальное распределение с параметрами , , равна σ^2, т.е. $D\eta = r^2$.

Пример 9.3. Найдем дисперсию СВ ξ, равномерно распределенной на отрезке .

Решение. Для такой СВ $M\xi = \dfrac{a+b}{2}$ (см. пример 8.6). Найдем начальный момент второго порядка:

$$m_2 = M\xi^2 = \int_{-\infty}^{\infty} x^2 p_\xi(x)dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^3 + ab + b^3}{3}.$$

Поэтому

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

Пример 9.4. (*правило «трех сигм»*)**.** Используя неравенство Чебышева, оценить вероятность того, что любая СВ ξ отклонится от своего среднего значения менее чем на три средних квадратичных отклонения этой величины.

Решение. Из неравенства Чебышева следует

$$P\{|\xi - M\xi| < \varepsilon\} \geq 1 - \frac{D\xi}{\varepsilon^2}.$$

В нашем случае $\varepsilon = 3\sigma$, где $\sigma = \sqrt{D\xi}$ – среднее квадратичное отклонение. Поэтому $P\{|\xi - M\xi| < 3\sigma\} \geq 1 - \dfrac{\sigma^2}{9\sigma^2}$, т.е.

$$P\{|\xi - M\xi| < 3\sigma\} \geq \frac{8}{9}.$$

Таким образом, полученная вероятность не меньше, чем $\dfrac{8}{9}$.

Ковариация и ее свойства.

Определение. Ковариацией СВ ξ и η называется

$$\operatorname{cov}(\xi,\eta) = M\left(\overset{0}{\xi}\cdot\overset{0}{\eta}\right) = M[(\xi - M\xi)(\eta - M\eta)],$$

здесь $\overset{0}{\xi} = \xi - M\xi$, $\overset{0}{\eta} = \eta - M\eta$ – центрированные СВ по отношению к СВ ξ и η.

Рассмотрим свойства ковариации.

1. .

2. Если СВ ξ и независимы, то , это следует из определения ковариации и свойства 6 математического ожидания (см. предыдущий параграф).

3. Для произвольных СВ ξ и

,

что следует из определения дисперсии и ковариации.

4. Пусть $\xi_1 = x\xi - \eta$, где x принимает действительные значения. Рассмотрим дисперсию СВ

$$D\xi_1 = M\left[x\xi - \eta - M(x\xi - \eta)\right]^2 = M\left[(x\xi - M(x\xi)) - (\eta - M\eta)\right]^2 =$$
$$= x^2 M(\xi - M\xi)^2 - 2xM\left[(\xi - M\xi)(\eta - M\eta)\right] + M(\eta - M\eta)^2 =$$
$$= x^2 D\xi - 2x\operatorname{cov}(\xi, \eta) + D\eta.$$

Это выражение является квадратным трехчленом относительно x. Поскольку то дискриминант $\left[2\operatorname{cov}(\xi, \eta)\right]^2 - 4D\xi\, D\eta$ должен быть меньше либо равен нулю, т.е. должно выполняться неравенство

$$\left[\operatorname{cov}(\xi, \eta)\right]^2 \le D\xi D\eta,$$

т.е.

$$\left|\operatorname{cov}(\xi, \eta)\right| \le \sqrt{D\xi\, D\eta},$$

или

$$-\sqrt{D\xi\, D\eta} \le \operatorname{cov}(\xi, \eta) \le \sqrt{D\xi\, D\eta}.$$

Коэффициент корреляции и его свойства. Нормированной СВ

по отношению к СВ ξ называется СВ ; очевидно, что

$M\xi^0 = 0$, $D\xi^0 = 1$.

Определение. Коэффициентом корреляции СВ ξ и называется

.

Он характеризует меру степени зависимости между СВ ξ и .

СВ и называются некоррелированными, если $\left(\operatorname{cov}(\xi,\eta)=0\right)$.

Пример 9.4. Приведем два примера, показывающие, что из равенства нулю коэффициента корреляции двух случайных величин не следует их независимость.

Рассмотрим две непрерывные СВ ξ и . Пусть СВ имеет плотность распределения , $-\infty < x < +\infty$, а . Тогда

$$\operatorname{cov}(\xi,\eta)=M(\xi\eta)-M\xi M\eta=M\xi^3-M\xi M\xi^2=$$

$$=A\int_{-\infty}^{\infty}x^3e^{-x^2}dx-A\int_{-\infty}^{\infty}xe^{-x^2}dx\,M\xi^2=0-0M\xi^2=0,$$

т.к. подынтегральные функции в интегралах – нечетные, и, следовательно, $\operatorname{cov}(\xi,\eta)=r(\xi,\eta)=0$, т.е. СВ ξ и некоррелированы, в то время как они связаны функциональной зависимостью .

Рассмотрим свойства коэффициента корреляции.

1. $|r(\xi,\eta)|\le 1$. Оно выполняется, поскольку

$$0\le D\left(\xi^0\pm\eta^0\right)=M\left(\xi^0\pm\eta^0\right)^2=M\left(\xi^0\right)^2+M\left(\eta^0\right)^2\pm 2M\left(\xi^0\eta^0\right)=$$

$$=2\left[1\pm M\left(\xi^0\eta^0\right)\right]\Rightarrow \qquad \Rightarrow \qquad .$$

2. $r(\xi,\eta)=\pm 1$ тогда и только тогда, когда ξ и связаны линейной зависимостью.

Достаточность следует из того, что, если , то

$$r(\xi,\eta)=\frac{M[(\xi-M\xi)(\eta-M\eta)]}{\sqrt{D\xi}\,\sqrt{D\eta}}=\frac{aM(\xi-M\xi)^2}{|a|D\xi}=\frac{a}{|a|}=sign\,a=\begin{cases}1, & a>0,\\ -1, & a<0.\end{cases}$$

3. Если ξ и – независимые СВ, то . Обратное утверждение не обязательно, но оно будет справедливо, когда ξ и – нормально распределенные СВ.

Важной характеристикой -мерной СВ является ковариационная матрица , где $K_{ij} = M\left[\left(\xi_i - M\xi_i\right)\left(\xi_j - M\xi_j\right)\right]$ – ковариационный момент СВ ξ_i и ξ_j; ясно, что $K_{ij} = K_{ji}$, $K_{ii} = D\xi_i$, $i,j = \overline{1,n}$. Нормированной ковариационной (корреляционной) матрицей R называется матрица, элементами которой являются коэффициенты корреляции СВ ξ_i и ξ_j:

$$R = \left\|r_{ij}\right\|_{n\times n},\ r_{ij} = \frac{K_{ij}}{\sqrt{D\xi_i}\sqrt{D\xi_j}},\ r_{ij} = r_{ji},\ r_{ii} = 1,\ i, j = \overline{1, n}.$$

Энтропия. Количество информации. Часто числовыми характеристиками СВ являются функционалы, определяющие различие между двумя распределениями вероятностей. Такое различие необходимо знать, например, если, кроме факта зависимости между двумя СВ, нужно знать, насколько велика эта зависимость. Рассмотрим случай абсолютно непрерывных распределений. В качестве величины, измеряющей степень зависимости двух СВ, можно использовать расстояние между распределениями $p_{\xi\eta}(t,\tau)$ и $p_\xi(t)p_\eta(\tau)$. Расстояние между двумя распределениями измеряется различными способами. Одним из них является так называемое количество информации Шеннона.

Определение. Энтропией называется функционал

$$H(\xi) = M\left[\ln p_\xi(\xi)\right] = \int_{-\infty}^{\infty} \ln\left(p_\xi(x)\right) p_\xi(x)\,dx.$$

В физике – это мера беспорядка (неопределенности); чем больше неопределенность, тем больше энтропия.

Следует отметить, что ξ может быть и многомерной СВ, т.е. может быть, что .

Определение. Количеством информации Шеннона называется величина

$$I(\xi,\eta) = -H(\xi) - H(\eta) + H(\xi,\eta).$$

В теории информации эта величина означает, что СВ ξ содержит информации о СВ η.

Т.к.

$$p_\xi(x)=\int_{-\infty}^{\infty} p_{\xi\eta}(x,y)dy,$$

$$H(\xi)=\int_{-\infty}^{\infty}\ln\bigl(p_\xi(x)\bigr)p_\xi(x)dx=\int_{-\infty}^{\infty}\ln p_\xi(x)\int_{-\infty}^{\infty}p_{\xi\eta}(x,y)dydx=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ln\bigl(p_\xi(x)\bigr)p_{\xi\eta}(x,y)dydx,\ H(\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ln\bigl(p_\eta(y)\bigr)p_{\xi\eta}(x,y)dydx,$$

то

$$I(\xi,\eta)=\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}\ln\left[\frac{p_{\xi\eta}(x,y)}{p_\xi(x)p_\eta(y)}\right]dy.$$

Если ξ и η – независимые СВ, то $I(\xi,\eta)=0$, т.е. никакой информации СВ ξ не содержит о СВ η. В этом случае $H(\xi,\eta)=H(\xi)+H(\eta)$.

Задачи

9.1. Дана плотность распределения СВ ξ

$$p_\xi(x)=\begin{cases}\frac{3}{2}x^2,\ 0\le x\le 1,\\ \frac{3}{2}(2-x)^2,\ 1<x\le 2,\\ 0,\ x<0,\ x>2.\end{cases}$$

Найти начальные и центральные моменты первых четырех порядков.

9.2. Дано распределение дискретной СВ ξ

Найти начальные и центральные моменты первых четырех порядков.

9.3. Найти среднее квадратичное отклонение СВ, заданной законом распределения

ξ	3	5	7	9
P	0,4	0,3	0,2	0,1

9.4. СВ ξ имеет плотность распределения

Найти дисперсию СВ ξ.

9.5. Найти начальный момент -го порядка СВ, равномерно распределенной на отрезке .

9.6. Найти начальный момент n-го порядка СВ, имеющей показательное распределение.

9.7. Найти дисперсию и моменты дискретных СВ, имеющих: а) распределение Бернулли, б) биномиальное распределение, в) распределение Пуассона, г) геометрическое распределение.

9.8. Найти дисперсию СВ в задаче 8.22.

9.9. Найти дисперсию СВ ξ, имеющей распределение Максвелла, см. задачу 8.29.

9.10. Показать, что функция вида

где $a,\ \alpha > 0,\ s = 1,2,3$, обладает свойствами плотности распределения.

Определить параметры и , исходя из заданного математического ожидания , и найти дисперсию. Заметим, что СВ, имеющая плотность распределения , распределена по закону Релея, а СВ, имеющая плотность распределения $f_2(x)$, – по закону Максвелла.

9.11. Найти дисперсию СВ ξ, распределенной по логарифмически нормальному закону, плотность вероятностей для которого имеет вид

$$p_\xi(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}x\sigma}e^{-\frac{(\ln x-a)^2}{2\sigma^2}}, & x>0,\\ 0, & x\le 0,\end{cases}$$

где a – любое действительное число, а – любое положительное действительное число.

Замечание. А.Н. Колмогорав показал, что логарифмически нормальному закону распределения подчинены размеры частиц при дроблении.

9.12. Ребро куба измерено приближенно, причем . Рассматривая длину ребра куба как СВ , распределенную равномерно в интервале , найти дисперсию объема куба.

9.13. Точка брошена наудачу внутрь круга радиуса r. Вероятность попадания точки в любую область, расположенную внутри круга, пропорциональна площади этой области. Найти ф.р. и дисперсию расстояния от точки до центра круга.

9.14. Найти дисперсию числа бракованных изделий в задаче 8.18.

9.15. Найти дисперсию числа исправно работающих приборов в задаче 8.19.

9.16. Найти дисперсию числа попаданий в задаче 8.28.

9.17. На летящий объект действуют независимо один от другого два случайных фактора и ξ_2, которые распределены нормально соответственно с параметрами (a_1, σ_1^2), (a_2, σ_2^2). Найти $D(\xi_1+\xi_2)$.

9.18. Ошибка измерений некоторых величин при одном способе равна 2ξ, где ξ – нормально распределенная СВ с параметрами , $\sigma=5$. При другом способе ошибка является суммой двух независимых нормально распределенных СВ , причем $M\xi_1 = M\xi_2 = 0$, $\sigma_{\xi_1} = \sigma_{\xi_2} = 5$. Какой способ измерений лучше?

9.19. Объект А из бесконечности движется по направлению к объекту . Максимальные дальности выявления один одного для этих объектов ξ_1 и ξ_2 являются независимыми нормально распределенными СВ соответственно с параметрами $\left(a_1, \sigma_1^2\right)$, $\left(a_2, \sigma_2^2\right)$. Найти вероятность того, что объект А выявит объект первым и дисперсии СВ ξ_j, $i = 1{,}2$.

9.20. Дискретная СВ задана законом распределения

Используя неравенство Чебышева для дисперсии, оценить вероятность того, что $|\xi - M\xi| < 0{,}2$.

9.21. Устройство состоит из 10 независимых работающих элементов. Вероятность отказа каждого элемента за время t равна 0,05. Используя неравенство Чебышева, оценить вероятность того, что абсолютная величина разности между числом отказавших элементов и средним числом отказов за время t окажется меньше трех.

9.22. Случайная ошибка измерительного прибора имеет дисперсию 16 . Систематическая ошибка прибора отсутствует. Оценить вероятность того, что при измерении ошибка превысит по модулю 6 *мВ*.

Замечание. Если измерения не содержат систематической ошибки, то отклонение полученных измеренных значений некоторой величины от ее истинного значения объясняется чисто случайными погрешностями и можно предположить, что $M\xi = 0$.

9.23. СВ ξ является ошибкой измерения некоторого расстояния. При измерении допускается систематическая ошибка в сторону завышения на 1,2 м; среднее квадратичное отклонение ошибки равно 0,8 м. Оценить вероятность того, что отклонение измеренного значения от истинного не превысит по абсолютной величине 1,6 м.

8.24. СВ ξ и независимы и

$$P(\eta=1)=P(\eta=-1)=0{,}25.$$

Будут ли СВ $\xi\eta$ и : а) независимыми, б) некоррелированными?

9.25. Плотность распределения СВ равна

$$p_{\xi_1\xi_2}(x_1,x_2)=\begin{cases}\cos x_1\cos x_2,\ x_1,x_2\in\left[0,\dfrac{\pi}{2}\right],\\ 0,\ \text{при любых других значениях}\ x_1,x_2.\end{cases}$$

Найти коэффициент корреляции СВ ξ_1 и ξ_2.

9.26. СВ ξ и независимы и имеют одинаковое распределение с математическим ожиданием и дисперсией . Найти коэффициент корреляции СВ $\xi_1=\alpha\xi+\beta\eta$, $\xi_2=\alpha\xi-\beta\eta$.

$P(\xi=-1)=P(\eta=0)=0{,}5,$

9.27. Случайный вектор (ξ,η) равномерно распределен в квадрате со стороной, равной единице, и диагоналями, совпадающими с осями координат. Найти коэффициент корреляции СВ ξ и .

9.28. По некоторой цели делают три независимые выстрела. Вероятность попадания в цель при одном выстреле равна . Рассмотрим две СВ: – число попаданий в цель, ξ_2 – число промахов. Составить таблицу распределения совокупность (ξ_1,ξ_2). Найти cov .

9.29. Дана ковариационная матрица совокупности СВ (ξ_1,ξ_2,ξ_3)

$$K=\begin{pmatrix}1 & 4 & -2\\ 4 & 4 & 3\\ -2 & 3 & 9\end{pmatrix}$$

Найти σ_{ξ_i}, $i=\overline{1,3}$, $r(\xi_1,\xi_2)$, $r(\xi_1,\xi_3)$, $r(\xi_2,\xi_3)$.

9.30. Производят четыре независимых измерения одной и той же величины. Результаты измерений следующие: ξ_1, ξ_2, ξ_3, ξ_4, причем $M\xi_i = a$, $D\xi_i = \sigma^2$, $i = \overline{1,4}$. Рассмотрим разности между соседними измерениями $\eta_1 = \xi_2 - \xi_1$, $\eta_2 = \xi_3 - \xi_2$, $\eta_3 = \xi_4 - \xi_3$. Найти $D\eta_i$, $i = \overline{1,3}$, и корреляционную матрицу.

9.31. Найти энтропию одномерной и многомерной СВ, распределенных по нормальному закону.

9.32. Вычислить количество информации Шеннона, которая содержится в скалярной нормально распределенной СВ ξ о другой скалярной , также имеющей нормальное распределение.

9.33. Среднее значение скорости ветра у земли в данном пункте равно 10 км/ч. С помощью неравенства Чебышева оценить вероятность того, что в этом пункте скорость ветра при данном наблюдении не превысит 80 км/ч.

9.34. Среднее квадратичное отклонение ошибки измерения курса самолета . Среднее значение ошибки измерения равно нулю. Оценить вероятность того, что ошибка при данном измерении курса самолета превзойдет 4°.

9.35. Среднее потребление электроэнергии за июнь месяц населением одного из микрорайонов города равно $36 \cdot 10^4$ кВт·ч. Оценить вероятность того, что потребление электроэнергии в июне текущего года превзойдет 10^6 кВт·ч.

9.36. Среднее число отказов ПЭВМ после года эксплуатации равно 5. Оценить вероятность того, что после окончания года число отказов в учебной лаборатории, в которой находятся 12 ПЭВМ, будет меньше 20.

9.37. Светофор на перекрестке работает в трех режимах: 1 мин горит зеленый свет, 0,2 мин – красный, 0,1 – желтый и т.д. Водитель подъезжает к перекрестку в случайный момент времени. Найти: а) вероятность того, что он проедет перекресток без остановки, б) ф.р. времени ожидания у перекрестка, в) числовые характеристики времени ожидания.

9.38. Бизнесмен решает задачу выбора одного из двух поставщиков, анализируя вариабельность цены на сырье. Для первого поставщика цена на сырье описывается СВ ξ_1, равномерно распределенной на отрезке $[10,14]$, а для второго – СВ ξ_2, имеющей нормальное распределение с параметрами $a=12$, $\sigma^2=\frac{4}{3}$. Бизнесмен намеревается выбрать того поставщика, который вносит меньшую неопределенность (энтропию) при одинаковой средней цене. Какое решение примет бизнесмен?

§10. Закон больших чисел. Центральная предельная теорема

С законом больших чисел в форме Бернулли мы познакомились в § 5. Из приведенного там утверждения следует, что доля успешных испытаний из n независимых испытаний Бернулли при приближается к вероятности одного успешного испытания (частота появления события стремится к вероятности данного события).

Определение. Говорят, что для последовательности СВ $\{\xi_k(\omega), k=1,2,...\}$ выполняется закон больших чисел, если для любого $\varepsilon>0$

$$\lim_{n\to\infty} P\left\{\omega: \left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k(\omega)-a_k)\right|\ge\varepsilon\right\}=0,$$

где $a_k=M_{\xi_k}$, $k=1,2,...$,или, что то же самое,

$$\lim_{n\to\infty} P\left\{\omega: \left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k(\omega)-a_k)\right|<\varepsilon\right\}=1.$$

Пример 10.1. *(закон больших чисел в форме Чебышева).* Пусть СВ $\xi_1(\omega)$, $\xi_2(\omega)$,..., независимы и имеют одинаковые математические ожидания $M\xi_k(\omega)=a$ и дисперсии $D\xi_k(\omega)=\sigma^2$. Тогда для них имеет место закон больших чисел, т.е.

$$\lim_{n\to\infty} P\left\{\omega : \left|\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) - a\right| < \varepsilon\right\} = 1 .$$

Это вытекает из неравенства Чебышева для дисперсий, из которого следует, что

$$P\left\{\omega : \left|\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) - a\right| \geq \varepsilon\right\} \leq \frac{D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega)\right)}{\varepsilon^2} = \frac{1}{n^2}\cdot\frac{n\sigma^2}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} .$$

Отметим, что аналогичный результат имеет место и в случае, когда СВ $\xi_1(\omega)$, $\xi_2(\omega)$,... независимы, одинаково распределены и имеют конечное математическое ожидание (теорема Хинчина).

Смысл закона больших чисел состоит в том, что среднее арифметическое СВ $\xi_1(\omega)$, $\xi_2(\omega)$,..., $\xi_n(\omega)$ при достаточно большом n будет с большой вероятностью почти не отличаться от ,

или, в частности, от a .

В экспериментальных науках среднее арифметическое

результатов x_1 , x_2 ,..., x_n измерений некоторой величины a рассматривают как более точное приближение к истинному значению этой величины по сравнению с отдельным измерением. Вероятностная модель измерений дается последовательностью СВ , $\xi_2(\omega)$,..., $\xi_n(\omega)$. Если измерения не содержат систематической ошибки, т.е. отклонение $x_i - a$ объясняется чисто случайными погрешностями, то следует предположить, что $M(\xi_k(\omega) - a) = 0$, т.е. $M\xi_k(\omega) = a$, $k = \overline{1,n}$. Если к тому же измерения независимы и одинаково точны, т.е. $D\xi_k(\omega) = \sigma^2$, $k = \overline{1,n}$, то закон больших чисел позволяет объяснить экспериментально наблюдаемую закономерность о стабилизации с ростом n средних арифметических вблизи a .

Качество приближения к a оценки при конечных n можно характеризовать неравенством Чебышева

,

оценивающим степень конкретизации распределения вероятностей СВ $S(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ вокруг точки a. Таким образам, точность приближения можно оценивать одним числом – дисперсией DS.

С помощью закона больших чисел в форме Чебышева можно легко доказать утверждение для схемы Бернулли, приведенное в начале параграфа. Именно, пусть

$$\xi_k(\omega) = \begin{cases} 1, \text{если } k-e \text{ испытание успешно}, \\ 0, \text{если } k-e \text{ испытание неудачно}, \end{cases}$$

p – вероятность успеха в отдельном испытании. Тогда –

$M\xi_k(\omega)=p$, $S(x_1, \ldots, x_n) = \bar{x} \approx a$, $\sum_{k=1}^{n} \xi_k$, $\frac{1}{n}\sum_{k=1}^{n}\xi_k$, $P\{\omega: |S(\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega)) - a| \geq \varepsilon\} \leq \frac{DS}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$

число успехов в n испытаниях, а – доля успехов в n испытаниях. Поскольку и $\xi_k^2(\omega) = \xi_k(\omega)$, то $M\xi_k^2(\omega) = p$, и поэтому $D\xi_k(\omega) = p - p^2$. Легко проверить, что $p - p^2 \leq \frac{1}{4}$, таким образом,

$$P\left\{\omega : \left| \frac{1}{n}\sum_{k=1}^{n} \xi_k(\omega) - p \right| \geq \varepsilon \right\} \xrightarrow[n\to\infty]{} 0.$$

Приведем результат, связанный с выполнением закона больших чисел в более общих ситуациях.

Теорема. Для того, чтобы для последовательности СВ $\{\xi_k(\omega), k = 1, 2, \ldots\}$ (они могут быть и зависимыми) выполнялся закон больших чисел, необходимо и достаточно, чтобы

$$\lim_{n\to\infty} M\left\{\frac{\left[\sum_{k=1}^{n}\left(\xi_k(\omega)-a_k\right)\right]^2}{n^2+\left[\sum_{k=1}^{n}\left(\xi_k(\omega)-a_k\right)\right]^2}\right\}=0\,.$$

Рассмотрим следствия из этой теоремы.

Теорема Маркова. Если

$$\lim_{n\to\infty}\frac{D\left(\sum_{k=1}^{n}\xi_k(\omega)\right)}{n^2}=0\,,$$

то имеет место закон больших чисел.

Это вытекает из соотношения

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k(\omega)\right)=M\eta_n^2(\omega)\geq M\left(\frac{\eta_n^2(\omega)}{1+\eta_n^2(\omega)}\right).$$

Теорема Чебышева. Если последовательность независимых СВ $\{\xi_k(\omega), k=1,2,...\}$ (не обязательно одинаково распределенных) такова, что $D\xi_k(\omega)\leq c$, $k=1,2,...$, где – некоторая константа, то имеет место закон больших чисел.

Это следует из теоремы Маркова и независимости СВ, поскольку

.

Пример 10.2. Пусть $\{\xi_k(\omega), k=1,2,...\}$ – последовательность независимых СВ и

$$P\{\omega:\xi_k(\omega)=\pm 2^k\}=2^{-(2k+1)},\ P\{\omega:\xi_k(\omega)=0\}=1-2^{-2k}\,.$$

В данном случае $M\xi_k(\omega)=\frac{2^k}{2^{2k+1}}-\frac{2^k}{2^{2k+1}}=\frac{1}{2}-\frac{1}{2}=0$, $M\xi_k^2(\omega)=\frac{2^{2k}}{2^{2k+1}}=\frac{1}{2}$,

поэтому $D\xi_k(\omega)=\frac{1}{2}$, и из теоремы Чебышева следует, что эта последовательность удовлетворяет закону больших чисел.

Далее рассмотрим другую группу предельных теорем теории вероятностей, устанавливающих сходимость законов распределения для последовательностей сумм случайных величин. Эта группа теорем ввиду ее особой важности как для теории, так и для приложений носит название центральной предельной теоремы. Две из таких теорем, а именно, теоремы Муавра – Лапласа, мы рассматривали в § 5. Они являются частными случаями более общих теорем, которые мы ниже рассмотрим.

Теорема *(Линдеберга)*. Если последовательность независимых СВ $\xi_1(\omega),\ \xi_2(\omega),\ldots$ при любых $\tau > 0$ удовлетворяет условию Линдеберга

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int\limits_{|x-a_k|>\tau B_n}(x-a_k)^2\,dF_{\xi_k}(x)=0,$$

где $F_{\xi_k}(x)$ – ф.р. СВ $\xi_k(\omega)$, $a_k = M\xi_k(\omega)$, $B_n^2 = D\left(\sum\limits_{k=1}^{n}\xi_k(\omega)\right)$, то при $n\to\infty$ равномерно относительно x

$$P\left\{\frac{1}{B_n}\sum_{k=1}^{n}\left(\xi_k(\omega)-a_k\right)<x\right\}\xrightarrow[n\to\infty]{}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt.$$

Теорему Линдеберга используют в основном в теоретических работах, требующих большой общности рассмотрения. В приложениях чаще применяют теорему Ляпунова, условия которой проверяются более эффективно.

Теорема Ляпунова. Если для последовательности независимых СВ $\xi_1(\omega),\ \xi_2(\omega),\ldots$ можно подобрать такое $\delta > 0$, что

$$\lim_{n\to\infty}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}M\left|\xi_k(\omega)-a_k\right|^{2+\delta}=0,$$

то справедлива центральная предельная теорема.

Для ее доказательства достаточно показать, что из условия Ляпунова вытекает условие Линдеберга. Если $|x-a_k| > \tau B_n$, то

$$\frac{|x-a_k|^{\delta}}{\tau^{\delta}B_n^{\delta}}>1,$$

поэтому

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int\limits_{|x-a_k|>\tau B_n}(x-a_k)^2\,dF_{\xi_k}(x)\le\frac{1}{\tau^{\delta}B_n^{2+\delta}}\sum_{k=1}^{n}\int\limits_{|x-a_k|>\tau B_n}|x-a_k|^{2+\delta}\,dF_{\xi_k}(x)\le$$

$$\le\frac{1}{\tau^{\delta}B_n^{2+\delta}}\sum_{k=1}^{n}\int\limits_{-\infty}^{\infty}|x-a_k|^{2+\delta}\,dF_{\xi_k}(x)=\frac{1}{\tau^{\delta}B_n^{2+\delta}}\sum_{k=1}^{n}M|\xi_k(\omega)-a_k|^{2+\delta}\xrightarrow[n\to\infty]{}0,$$

т.е. действительно из условия Ляпунова следует условие Линдеберга.

Пример 10.3. Пусть СВ $\xi_1(\omega),\ \xi_2(\omega),\ldots$ независимы и

$$M\xi_k(\omega)=0\,,\ M\xi_k^2=\sigma_k^2,\ 0<\sigma_k^2<\infty,\ M|\xi_k(\omega)|^3<\infty,\ k=1,\,2,\,\ldots.$$

Тогда . Предположим, что выполнено условие Ляпунова

$$\frac{1}{B_n^3}\sum_{k=1}^{n}M|\xi_k(\omega)|^3\xrightarrow[n\to\infty]{}0\,.$$

Воспользуемся неравенством Ляпунова для математического ожидания

$$\left(M|\xi_k(\omega)|^s\right)^{\frac{1}{s}}\le\left(M|\xi_k(\omega)|^t\right)^{\frac{1}{t}},\ 0<s<t\,,$$

из которого вытекает, что

$$\sigma_k^3=\left(M\xi_k^2(\omega)\right)^{\frac{3}{2}}\le M|\xi_k(\omega)|^3,$$

и поэтому

$$\frac{1}{B_n^3}\max_{k\le n}\sigma_k^3\le\frac{1}{B_n^3}\sum_{k=1}^{n}\sigma_k^3\xrightarrow[n\to\infty]{}0\,.$$

Но т.к. $\sigma_k<B_n$, $k=\overline{1,\,n}$, то

$$\frac{1}{B_n^2}\max_{k\le n}\sigma_k^2\xrightarrow[n\to\infty]{}0\,,$$

а из этого соотношения следует условие Линдеберга, т.е. в данном случае справедлива центральная предельная теорема.

Нужно отметить, что если все $\xi_k(\omega)$, $k=1,\,2,\,..$, независимы и одинаково распределены, $D\xi_k(\omega) = \sigma^2 < \infty$, $k = 1, 2, \ldots$, то условие Линдеберга также выполняется и

$$P\left\{\omega : \frac{\sum_{k=1}^{n} \xi_k(\omega) - na}{\sigma\sqrt{n}} < x\right\} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

Практический смысл центральной предельной теоремы состоит в следующем: если некоторый процесс происходит под воздействием большого числа независимо действующих случайных факторов, каждый из которых лишь очень мало изменяет ход процесса, то распределение суммарного действия этих случайных факторов можно очень близко аппроксимировать нормальным законом.

Пример 10.4. Нагрузка потребительской сети (телефонной, информационной, электрической и т.п.) в данный момент времени является результатом суммирования большого числа элементарных нагрузок, вносимых индивидуальными потребителями. Поскольку случайности, определяющие поведение потребителей, естественно считать независимыми, то имеет место суммирование большого числа независимых СВ, пренебрежимо мало влияющих на сумму. Различные статистические исследования подтверждают нормальный закон нагрузки.

Пример 10.5. Всякое измерение неизбежно сопряжено с погрешностями; систематические погрешности можно устранить, случайные же ошибки измерения полностью никогда не могут быть устранены. Случайная погрешность измерения вызывается многими причинами, каждая из которых лишь незначительно влияет на результат. Каждая из причин порождает свою, так называемую элементарную, погрешность измерения. Например, при взвешивании некоторого тела на точных весах случайная ошибка в определении его веса складывается из элементарных погрешностей, вызываемых атмосферными причинами (колебаниями плотности температуры, влажности воздуха, воздушными потоками); попаданием на чаши весов пылинок; неточностями, допущенными измерителем при снятии показаний со шкалы весов; незначительными вибрациями основания весов (которые в свою очередь могут вызываться многими причинами) и т.п. Реально наблюдаемая случайная погрешность измерения есть *сумма элементарных* погрешностей. Так как количество элементарных погрешностей велико, и роль каждой из них в образовании случайной погрешности измерения мала, то в силу теоремы Ляпунова случайная погрешность измерения должна быть распределена приближенно *по нормальному закону.*

Опыт показывает, что наблюдающиеся распределения вероятностей случайных ошибок измерения очень хорошо согласуются с нормальным законом. Итак, при *прямых измерениях случайная погрешность измерения распределена по закону, близкому к нормальному.*

Пример 10.6. Дано 5000 независимых одинаково распределенных СВ с дисперсией 50. Найти вероятность того, что среднее арифметическое этих СВ отклонится от своего математического ожидания не более, чем на $\Delta = 0.,12$.

Решение. Пусть , $D\xi_k(\omega) = \sigma^2$. Тогда искомая вероятность равна

$$P = P\left\{\omega : \left|\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) - a\right| \le \Delta\right\} = P\left\{\omega : \frac{\sigma}{\sqrt{n}}\left|\frac{\sum_{k=1}^{n}\xi_k(\omega) - a}{\sigma\sqrt{n}}\right| \le \Delta\right\} =$$

$$= P\left\{\omega : \left|\frac{\sum_{k=1}^{n}\xi_k(\omega) - a}{\sigma\sqrt{n}}\right| \le \frac{\sqrt{n}}{\sigma}\Delta\right\} \approx \Phi\left(\frac{\sqrt{n}}{\sigma}\Delta\right) - \Phi\left(-\frac{\sqrt{n}}{\sigma}\Delta\right) = 2\Phi\left(\frac{\sqrt{n}}{\sigma}\Delta\right),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$. Подставляя значения $n = 5000$, $\sigma = \sqrt{50}$, $\Delta = 0,12$, получаем .

Задачи

10.1. Дана последовательность независимых СВ $\{\xi_k(\omega), k = 1, 2, \ldots\}$, причем

$$\xi_k(\omega) = \begin{cases} -\sqrt{k} \ \text{ с вероятностью } \dfrac{1}{k}, \\ \sqrt{k} \ \text{ с вероятностью } \dfrac{1}{k}, \\ 0 \ \text{ с вероятностью } 1 - \dfrac{2}{k}. \end{cases}$$

Выполняется ли для нее закон больших чисел?

10.2. Дана последовательность независимых СВ $\{\xi_k(\omega), k = 1, 2, \ldots\}$, при этом $M\xi_k(\omega) = 0$, $D\xi_k(\omega) = k^{\alpha}$, $\alpha < 1$. Выполняется ли для нее закон больших чисел?

10.3. Независимые СВ , $\xi_2(\omega)$,... имеют нормальное распределение и $M\xi_k(\omega) = 0$, $D\xi_k(\omega) = ck^{\alpha}$, $c > 0$, $\alpha \geq 0$. При каких последовательность этих величин удовлетворяет закону больших чисел?

10.4. СВ принимает значения

$$-k, \quad ,\ldots, \ -1, \quad , \ 1, \ldots, \ (k\text{-}1), \ k,$$

при этом

$$P\{\omega : \xi_k(\omega) = 0\} = 1 - \frac{2}{3}\left(1 + \frac{1}{2^3} + \ldots + \frac{1}{k^3}\right), \ P\{\omega : |\xi_k(\omega)| = i\} = \frac{1}{2i^3}, \ i = \overline{1, k}.$$

Удовлетворяет ли последовательность $\{\xi_k(\omega), k = 1, 2, \ldots\}$ закону больших чисел?

10.5. СВ $\xi_k(\omega)$ принимает значения $-k$, , k, при этом

$$P\{\omega : \xi_k(\omega) = k\} = \frac{1}{k^3}, \ k = 1, 2, \ldots.$$

Удовлетворяет ли последовательность закону больших чисел?

10.6. Последовательность независимых СВ $\{\xi_k(\omega), k = 1, 2, \ldots\}$ задана законом распределения

$\xi_k(\omega)$	a	$-a$
P	$\frac{n}{2n+1}$	$\frac{n+1}{2n+1}$

Удовлетворяет ли она закону больших чисел?

10.7. Дана последовательность независимых СВ $\xi_1(\omega)$, $\xi_2(\omega)$,..., при этом

$$F_{\xi_k}(x) = \frac{1}{2} + \frac{1}{\rho} arctg \frac{x}{a}, \ k = 1, 2, \ldots .$$

Можно ли применить к этой последовательности теорему Хинчина?

10.8. Пусть – последовательность независимых СВ, причем $\xi_k(\omega)$ принимает значения 2^k и -2^k с вероятностями $\frac{1}{2}$. Применим ли к этой последовательности закон больших чисел?

10.9. Пусть $\xi_1(\omega)$, $\xi_2(\omega)$,... – последовательность независимых СВ. В случае, когда k – точный квадрат, $\xi_k(\omega)$ принимает значения $-\sqrt{k}$, $\sqrt{k}$ с вероятностью $\frac{1}{2}$ каждое; при остальных k $\xi_k(\omega)$ принимает значения -2^{-k}, 2^{-k} с вероятностью $\frac{1}{2}$ каждое. Применим ли к этой последовательности закон больших чисел?

10.10. Среднее значение скорости ветра у земли в данном пункте равна 10 км/ч. Оценить вероятность того, что в этом пункте скорость ветра при одном наблюдении не превысит 80 км/ч.

10.11. Среднее квадратичное отклонение ошибки измерения курса самолета $\sigma = 3^\circ$. Среднее значение ошибки измерения равно нулю. Оценить вероятность того, что ошибка при данном измерении курса самолета будет более 4°.

10.12. ЭВМ вырабатывает случайные двоичные числа так, что знаки 0 и 1 на каждой позиции появляются с одинаковой вероятностью и независимо от других позиций. Последовательность знаков делится на группы, состоящие из одинаковых знаков, например, 001101001110. Подсчитывается число знаков в каждой группе и делится на число групп. Как будет себя вести эта средняя величина при неограниченном увеличении числа групп n?

10.13. Среднее квадратичное отклонение ошибки измерения азимута равна 30', а математическое ожидание равно нулю. Оценить вероятность того, что ошибка среднего арифметического трех независимых измерений не превзойдет .

10.14. Нужно сделать 10 измерений x_1, $x_2, \ldots, x_{10}$ неизвестной величины a. Будем считать их независимыми и . Подобрать Δ так, чтобы

.

10.15. Проводятся n независимых измерений некоторой неизвестной величины . Ошибки измерений , $\delta_2, \ldots, \delta_n$ – СВ, $M\delta_k = 0$, $D\delta_k = \sigma^2$, $k = \overline{1, n}$. За значение величины a примем среднее арифметическое результатов измерений, тогда ошибка для будет равна

.

Оценить количество измерений n, при которых ошибка будет не больше Δ с достаточно большой вероятностью . Рассмотрим случай, когда $P = 0{,}99$, $\Delta = 0{,}1$, . Может ли быть меньше, чем ?

10.16. Можно ли принять величину

$$s_n^2 = \frac{1}{n} \sum_{k=1}^{n} \left(\xi_k(\omega) - a\right)^2$$

в качестве приближенного значения дисперсии ошибок прибора, если $\xi_1(\omega)$, $\xi_2(\omega), \ldots, \xi_n(\omega)$ – независимые измерения постоянной величины a, имеющие одинаковые ф.р.?

10.17. Пусть , $\xi_2(\omega)$,... – последовательность независимых одинаково распределенных невырожденных СВ, $D\xi_k(\omega)<\infty$, $k=1,\ 2,\ \ldots,$. Доказать, что для любых конечных чисел a и

$$\lim_{n\to\infty} P\{\omega : a \le \eta_n(\omega) \le b\} = 0 .$$

10.18. Пусть $\xi_1(\omega)$, $\xi_2(\omega)$,... – последовательность независимых одинаково распределенных СВ, $M\xi_k(\omega)=0$, $D\xi_k(\omega)<\infty$, $k=1,\ 2,\ \ldots,$. Найти $D\xi_k(\omega)$, если

$$\lim_{n\to\infty} P\left\{\omega : \frac{\eta_n(\omega)}{\sqrt{n}} > 1\right\} = \frac{1}{3} .$$

10.19. Будет ли выполняться центральная предельная теорема для последовательности независимых СВ $\xi_1(\omega)$, $\xi_2(\omega)$,... с распределениями, задаваемыми следующим образом:

$$\text{а)}\ P\{\omega : \xi_k(\omega) = \pm 2^k\} = \frac{1}{2} ;$$

$$\text{б)}\ P\{\omega : \xi_k(\omega) = \pm k\} = \frac{1}{2} k^{-\frac{1}{2}},\ P\{\omega : \xi_k(\omega) = 0\} = 1 - k^{-\frac{1}{2}} .$$

10.20. СВ $\eta(\omega)$ является средним арифметическим одинаково распределенных ошибок независимых измерений некоторой величины, дисперсия каждой из которых равна 5. Сколько нужно сделать измерений, чтобы СВ $\eta(\omega)$ с вероятностью, не меньшей $0{,}9973$, имела отклонение от своего среднего значения, не превосходящее $0{,}01$?

10.21. Для предыдущей задачи известно, что ошибки измерений являются независимыми одинаково распределенными случайными величинами с математическим ожиданием, равным 3, и дисперсией, равной 2. Произведено 3000 независимых измерений. Найти вероятность того, что $\eta(\omega)$ примет значение в промежутке $(2, 3)$.

10.22. В задаче 10.20 известно, что среднее квадратичное отклонение каждой из ошибок равно 2; произведено 10000 независимых измерений. Какое максимальное отклонение величины $\eta(\omega)$ от ее среднего значения можно ожидать с вероятностью, не меньшей $0,9544$?

10.23. Для измерения некоторой величины с помощью прибора, лишенного систематической ошибки, но имеющего случайные с дисперсией $\sigma^2 = 0,22$, сделано 100 независимых измерений. Найти вероятность того, что среднее арифметическое результатов отклонится от истинной величины больше, чем на $0,05$.

10.24. Производится выборочное обследование партии электролампочек для определения средней продолжительности их горения. Каков должен быть объем выборки, чтобы с вероятностью, не меньшей $0,99$, утверждать, что средняя продолжительность горения лампочки по всей партии отклонится от средней, полученной в выборке, не более чем на 100 часов, если среднее квадратичное отклонение продолжительности горения лампочки равно 200 часов?

10.25. В условиях предыдущей задачи найти наименьшее число ламп, которые нужно взять для обследования, чтобы с вероятностью $0,99$ утверждать, что средняя продолжительность горения лампы во всей партии отклонится от полученной в выборке не более чем на 50 часов.

10.26. Для определения средней продолжительности работы некоторого прибора из данной партии выбирают наугад 100 штук. Оценить снизу вероятность того, что средняя продолжительность работы 100 отобранных приборов отличается от средней продолжительности работы приборов всей партии по абсолютной величине меньше, чем на месяц, если известно, что среднее квадратичное отклонение продолжительности работы прибора не превышает двух месяцев.

10.27. В университет поступило 10 одинаковых ящиков с приборами. Среднее число приборов в каждом ящике, которые пришли в негодность за время транспортировки, равно трем, а среднее квадратичное отклонение – двум. Определить границы, в которых с вероятностью не менее $0,9$ будет заключено общее число приборов, пришедших в негодность за время транспортировки.

10.28. Стрелок попадает при выстреле по мишени в десятку с вероятностью $0{,}5$; в девятку – $0{,}3$; в восьмерку – $0{,}2$; в семерку – $0{,}1$. Сделано 100 выстрелов. Какова вероятность того, что выбито более 980 очков?

10.29. Напряжение на входе приемного устройства является случайной величиной, которую можно рассматривать как сумму достаточно большого числа независимых синусоидальных величин, амплитуды которых распределены по нормальному закону с параметрами $\left(0, \sigma^2\right)$, а фазы равномерно распределены в промежутке $\left(-\pi, \pi\right)$. Можно ли считать распределение этого напряжения почти нормальным?

10.30. В кольце, ограниченном концентрическими окружностями радиусов 1 и $R > 1$, расположено n излучателей равной мощности . Каждый из них может рассматриваться как случайная точка, равномерно распределенная в этом кольце независимо от всех остальных. Мощность сигнала, наводимая каждым излучателем в приемнике, находящемся в центре кольца, равна $M_{вх} = \frac{a}{\rho_i^4}$, где ρ_i – расстояние излучателя до центра. Какова ф.р. мощности сигнала, наведенного суммарным воздействием всех n излучателей, если они работают на одной и той же частоте и их число достаточно велико?

10.31. В предыдущей задаче оценить нижнюю границу мощности на входе приемника (как функцию от n), если число излучателей неограниченно возрастает.

10.32. При производстве деталей для автомагнитол рабочий выполняет с каждой деталью однотипные операции. На это уходит случайное время, распределенное по показательному закону. Найти вероятность того, что на выполнение 100 операций рабочему понадобится время от 6 до 7 часов, если среднее время, необходимое для выполнения одной операции равно 3 минуты.

10.33. В кассе в день зарплаты получают деньги человек. Размер выплаты каждому – СВ со средним значением \$100 и средним квадратичным отклонением \$50. Если выплаты отдельным клиентам независимы, то:

а) сколько должно быть денег в кассе, чтобы их с вероятностью $0{,}95$ хватило на выплату всем клиентам?

б) каков будет гарантированный с вероятностью $0{,}95$ остаток денег в кассе после выплаты всем клиентам, если в начале в кассе было \$11000?

10.34. Среднее суточное потребление электроэнергии в населенном пункте 12000 кВт • ч. Оцените вероятность того, что потребление в этом населенном пункте в течении данных суток превзойдет 50000 кВт • ч. Какого потребления энергии в том населенном пункте можно ожидать в ближайшие сутки с вероятностью, не меньшей $0{,}96$, если среднее квадратичное отклонение равно 200 кВт • ч. Сколько потребителей следует проверить, чтобы с вероятностью $0{,}95$ можно было утверждать, что отклонение среднего потребления отобранной группы от среднегородского потребления по модулю не превосходит 1 %?

10.35. Статистические данные свидетельствуют о вероятности 10 % краж из квартир граждан. Застрахована группа из 100 человек, сумма страхового взноса составила \$1000 в год. В случае ограбления клиента страховая фирма выплачивает потерпевшему $\$A$.

$P\{\xi=1\}=0{,}3,$ 1) какова должна быть страховая выплата A, чтобы с вероятностью $0{,}95$ фирма не оказалась в убытке?

2) какова вероятность фирмы получить доход, превосходящий \$10000?

10.36. Коммивояжер для выполнения своего задания должен посетить $n=60$ городов, затрачивая в каждом из них случайное время ξ (измеряемое в сутках) со следующим законом распределения:

$$P\{\xi=2\}=0{,}4,\ \ P\{\xi=3\}=0{,}3\,.$$

Оценить вероятность того, что коммивояжер затратит на работу от 110 до 130 суток.

10.37. Наудачу выбранный посетитель банка делает вклад, величина которого является СВ $\xi_k(\omega)$ с равномерным распределением на интервале $[-80{,}100]$. Если банк принимает в течении суток $n=200$ посетителей, то какова вероятность отрицательного суточного сальдо $P(s_n<0)$, где $s_n=\sum_{k=1}^{n}\xi_k(\omega)$.

Глава 2
ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика занимается изучением методов сбора и обработки опытных данных для получения научных и практических выводов и является разделом математики, очень близким к теории вероятностей. Лучше сказать, математическая статистика в решении своих специфических задач пользуется результатами теории вероятностей. Однако следует иметь в виду, что это другая, отдельная наука, решающая в каком-то смысле обратные задачи по сравнению с задачами теории вероятностей.

Типичная задача теории вероятностей. Задана вероятность p наступления случайного события в одном опыте. Какова вероятность того, что в 200 опытах событие A наступит 3 раза?

Типичная задача математической статистики. Произвели 200 опытов, случайное событие A при этом наступило 3 раза. Какова вероятность наступления события A в одном опыте?

В математической статистике объектом исследования являются данные эксперимента.

§11. Вариационные ряды и их графическое изображение

Установление статистических закономерностей, присущих случайным явлениям, основано на изучении статистических данных – сведений о том, какие значения принял в результате наблюдения интересующий нас признак (случайная величина ξ). Например, исследуется спрос на определенные размеры мужской обуви по имеющимся статистическим данным о размерах 350 пар обуви, проданных магазином за неделю; исследуется точность измерительного прибора по результатам 50 независимых измерений.

Генеральной совокупностью называется совокупность объектов или наблюдений, все элементы которой подлежат изучению при статистическом анализе. Генеральная совокупность может быть конечной или бесконечной.

На практике изучение всего набора элементов генеральной совокупности в большинстве случаев оказывается невозможным. Часть объектов генеральной совокупности, используемая для исследования, называется ***выборочной совокупностью*** или ***выборкой*** ( ). Число объектов в генеральной совокупности (N) или выборке (n) называют их объемами. В примере об исследовании спроса на размеры мужской обуви используется выборка объема , в примере об исследовании точности измерительного прибора $n = 50$. Сущность выборочного метода состоит в том, что выводы, сделанные на основе изучения части совокупности (выборки), статистика позволяет распространять на всю генеральную совокупность.

Полученные в результате выборки значения x_i признака ξ называются ***вариантами*** или ***элементами выборки,*** . Упорядоченная по возрастанию значений совокупность вариант носит название ***вариационного ряда***.

Дискретным вариационным (статистическим) рядом или статистическим распределением выборки называют таблицу, которая

$X = \{x_1, x_2, ..., x_n\}$ $n = 350$

в первой строке содержит значения $x_1, x_2, ..., x_n$ ($x_1 < x_2 < ... < x_n$), а во второй числа их повторений $n_1, n_2, ..., n_s$, $n_1 + n_2 + ... + n_s = n$:

X	x_1	x_2	$\cdots$	x_s
n	n_1	n_2	$\cdots$	n_s

Число n_i называют ***частотой***, а отношение $w_i = \dfrac{n_i}{n}$ ***относительной частотой*** элемента x_i , $i = \overline{1, s}$. Дискретный вариационный ряд, как правило, представляет собой выборку значений дискретной случайной величины.

При изучении непрерывных признаков, для которых x_i могут принимать как угодно близкие значения, пользуются ***интервальным вариационным (статистическим) рядом***. Также использование интервальных рядов целесообразно при изучении выборок большого объема. Интервал $J = [x_1, x_n]$ разбивают на s промежутков одинаковой длины . *При этом считают, что каждый промежуток содержит свой левый конец, но лишь последний промежуток содержит и свой правый конец.* Далее для каждого промежутка подсчитывают число элементов выборки, попавших в него – n_i, $i=\overline{1,s}$, в результате данные наблюдений представляют в виде таблицы:

X	$[x_1, x_2)$	$[x_2, x_3)$	...	$[x_s, x_{s+1}]$
n	n_1	n_2	...	n_s

Число промежутков s следует брать не очень большим, чтобы после группировки ряд не был громоздким, и не очень малым, чтобы не потерять особенности распределения признака. Согласно формуле Стерджеса, рекомендуемое число интервалов , а

величина частичного интервала .

Для графического изображения вариационных рядов наиболее часто используются полигон и гистограмма.

Полигон, как правило, служит для изображения дискретного вариационного ряда и представляет собой ломаную, отрезки которой последовательно соединяют точки с координатами (x_i, n_i), $i=\overline{1,s}$. При построении полигона для интервального вариационного ряда в качестве x_i используют середины интервалов.

Гистограмма (частот, относительных частот) служит только для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру из прямоугольников с основаниями равными интервалам значений признака $J_i = [x_i, x_{i+1})$ и высотами, равными плотности частоты $\dfrac{n_i}{n\Delta}$ (частоте n_i, относительной частоте $\dfrac{n_i}{n}$). Если соединить середины верхних оснований прямоугольников отрезками прямой, то можно получить полигон того же распределения. Суммарная площадь всех прямоугольников гистограммы равна 1:

$$\sum_{i=1}^{s} \frac{n_i}{n\Delta}\Delta = \sum_{i=1}^{s} \frac{n_i}{n} = \frac{1}{n}\sum_{i=1}^{s} n_i = \frac{n}{n} = 1 .$$

Кроме того, площадь каждого прямоугольника $\frac{n_i}{n}$ есть относительная частота попадания элементов выборки в соответствующий интервал J_i статистического ряда. Гистограмма является статистическим аналогом кривой распределения (кривой плотности распределения $p_\xi(x)$), наблюдаемой случайной величины ξ. При большом объеме выборки и достаточно малом с вероятностью, близкой к 1, можно считать кривую распределения и гистограмму приблизительно совпадающими.

Весьма важным является понятие эмпирической функции распределения.

Эмпирической функцией распределения называется функция , определяющая для каждого значения x относительную частоту события , то есть

,

где n_x – накопленная частота, равная числу вариант меньших, чем x, – объем выборки.

$\xi < x$ $F_n(x) = \frac{n_x}{n}$

Функция является статистическим аналогом функции распределения $F_\xi(x)$ генеральной совокупности. Функцию распределения генеральной совокупности $F_\xi(x)$ в математической статистике называют теоретической функцией распределения. Различие между эмпирической и теоретической функциями распределения состоит в том, что $F_\xi(x)$ определяет вероятность события $\xi < x$, а – относительную частоту этого события, в силу закона больших чисел $F_n(x) \xrightarrow[n\to\infty]{P} F_\xi(x)$.

Функция $F_n(x)$ обладает всеми свойствами функции распределения:

1. $0 \le F_n(x) \le 1$;
2. $F_n(x)$ – неубывающая функция;
3. $F_n(x) = 0$ при $x \le x_{\min}$, $F_n(x) = 1$ при $x > x_{\max}$.

Кумулятивная кривая (кумулята) – ломаная, соединяющая точки с координатами (x_i, n_{x_i}), где n_{x_i} – накопленные частоты; для интервального ряда n_{x_i} – число вариант меньших значений вариант интервала J_i.

Итак, эмпирическая функция распределения и кумулята служат для оценки вида теоретической функции распределения дискретного и непрерывного признаков, полигон и гистограмма – для оценки вида теоретической кривой распределения.

Пример 11.1. В течение суток измеряют напряжение тока в электросети в вольтах. В результате опыта получена выборка X объема $n = 30$:

107	108	110	109	110	111	109	110	111	107
108	109	110	108	107	111	109	111	111	110
109	112	113	110	106	110	109	110	108	112

Построить:

1) статистический ряд данной выборки,

2) полигон,

3) эмпирическую функцию распределения.

Решение.

1) Для построения статистического ряда различные значения признака располагаем в порядке их возрастания и под каждым из этих значений записываем его частоту

2) Построим полигон этого распределения.

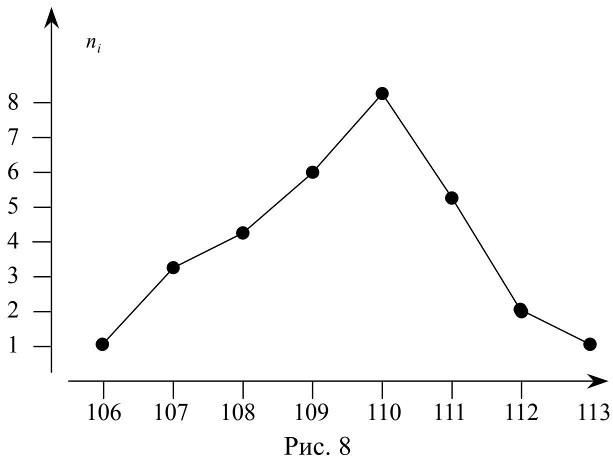

Рис. 8

3) Построим эмпирическую функцию распределения. Наименьшая варианта $x_{\min} = 106$, значит, $F_n(x) = 0$ при $x \le 106$. Пусть $106 < x \le 107$, меньше таких x только 106, это значение наблюдалось 1 раз, следовательно, $F_n(x) = \frac{1}{30}$ при $106 < x \le 107$. Пусть $107 < x \le 108$, меньше значения x, удовлетворяющего этому интервалу и $x_2 = 107$, $n_1 + n_2 = 1 + 3 = 4$, следовательно, $F_n(x) = \frac{4}{30} = \frac{2}{15}$ при $107 < x \le 108$. Аналогично находим $F_n(x)$ на остальных интервалах. Убеждаемся, что действительно $F_n(x) = 1$ при $x > x_{\max} = 113$.

$x_1 = 106$

$$F_n(x) = \begin{cases} 0 \text{ при } x \le 106, \\ \frac{1}{30} \text{ при } 106 < x \le 107, \\ \frac{1}{30}(1+3) = \frac{4}{30} = \frac{2}{15} \text{ при } 107 < x \le 108, \\ \frac{1}{30}(4+4) = \frac{8}{30} = \frac{4}{15} \text{ при } 108 < x \le 109, \\ \frac{1}{30}(8+6) = \frac{14}{30} = \frac{7}{15} \text{ при } 109 < x \le 110, \\ \frac{1}{30}(14+8) = \frac{22}{30} = \frac{11}{15} \text{ при } 110 < x \le 111, \\ \frac{1}{30}(22+5) = \frac{27}{30} \text{ при } 111 < x \le 112, \\ \frac{1}{30}(27+2) = \frac{29}{30} \text{ при } 112 < x \le 113, \\ \frac{1}{30}(29+1) = 1 \text{ при } x > 113; \end{cases} = \begin{cases} 0 \text{ при } x \le 106, \\ \frac{1}{30} \text{ при } 106 < x \le 107, \\ \frac{2}{15} \text{ при } 107 < x \le 108, \\ \frac{4}{15} \text{ при } 108 < x \le 109, \\ \frac{7}{15} \text{ при } 109 < x \le 110, \\ \frac{11}{15} \text{ при } 110 < x \le 111, \\ \frac{27}{30} \text{ при } 111 < x \le 112, \\ \frac{29}{30} \text{ при } 112 < x \le 113, \\ 1 \text{ при } x > 113. \end{cases}$$

Построим график эмпирической функции распределения.

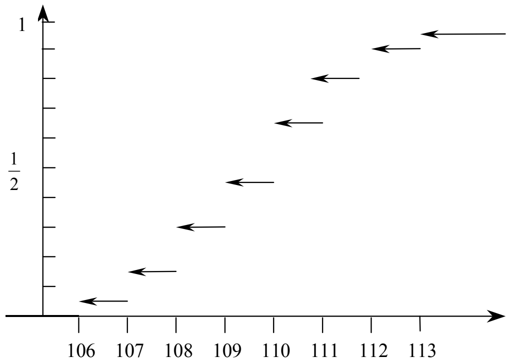

Рис. 9

Пример 11.2. В таблице приведены значения промежутков времени τ (в минутах) между вызовами такси в городе Гродно.

Построить по этим данным:

1) интервальный вариационный ряд;
2) полигон;
3) гистограмму;
4) эмпирическую функцию распределения.

Решение.

1) По условию объем выборки $n = 50$. Определим оптимальную длину частичного интервала с помощью формулы Стерджеса:

$$\Delta = \frac{x_{\max} - x_{\min}}{s} = \frac{x_{\max} - x_{\min}}{1 + 3{,}322\lg n} = \frac{0{,}702 - 0}{1 + 3{,}322\lg 50} \approx 0{,}106 .$$

За начало первого интервала можно выбрать величину $x_{нач} = x_{\min} - \Delta/3$ (или $x_{нач} = x_{\min} - \Delta/2$), при этом первый интервал должен покрывать $x_{\min}$, а последний – $x_{\max}$. В данном случае $x_{нач} = 0 - 0{,}106/3 = -0{,}035$. Сгруппированный ряд представим в виде таблицы 2, где n_i – число вариант, принадлежащих i-му интервалу.

Таблица 2

i	Интервалы $x_i - x_{i+1}$	Частота n_i	С[illegible] интервалов x_i^*	Относительная частота $w_i = \frac{n_i}{n}$	$\frac{n_i}{n\Delta}$	Накопленная частота n_{x_i}	Накопленная частость $w_i^{нак} = \frac{n_{x_i}}{n}$
1	-0,035 – 0,071	13	0,018	0,26	2,45	13	0,26
2	0,071 – 0,177	12	0,124	0,24	2,26	25	0,50
3	0,177 – 0,283	9	0,230	0,18	1,70	34	0,68
4	0,283 – 0,389	6	0,336	0,12	1,13	40	0,80
5	0,389 – 0,495	6	0,442	0,12	1,13	46	0,92
6	0,495 – 0,601	2	0,548	0,04	0,38	48	0,96
7	0,601– 0,707	2	0,654	0,04	0,38	50	1
$\sum$		50					

2) Согласно определению полигон имеет следующий вид:

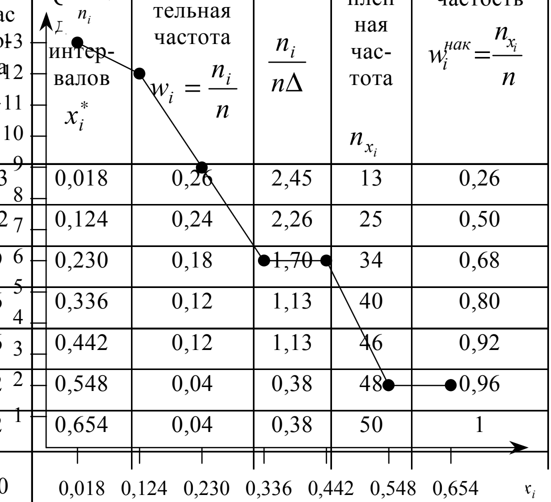

Рис. 10

3) Построим гистограмму.

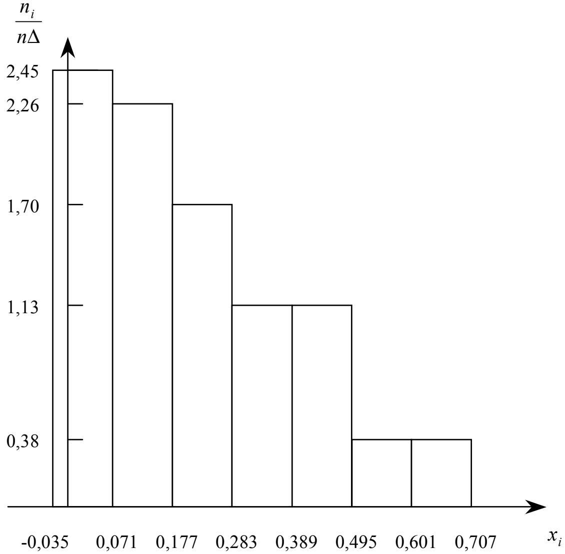

Рис. 11

4) В данном случае исследуется интервальный вариационный ряд, для непрерывно распределенного случайного признака τ. Эмпирическую функцию распределения находим таким же способом, как и в примере 11.1. Однако, учитывая, что теоретическая функция распределения является функцией распределения непрерывной случайной величины, ее эмпирическое приближение можем определить лишь на концах интервалов. Для графического изображения этой функции целесообразно ее доопределить, соединив точки графика, соответствующие концам интервалов, отрезками прямой. В результате полученная ломанная совпадет с кумулятой. В последней графе таблицы 2 приведены значения функции распределения на концах интервалов:

$$, \; F_n(0{,}071) = 0{,}26 \, , \; F_n(0{,}177) = 0{,}5 \, , \; F_n(0{,}283) = 0{,}68 \, ,$$
$$F_n(0{,}389) = 0{,}8 \, , \; F_n(0{,}495) = 0{,}92 \, , \; F_n(0{,}601) = 0{,}96 \, , \; F_n(0{,}707) = 1 \, .$$

Построим график $F_n(x)$.

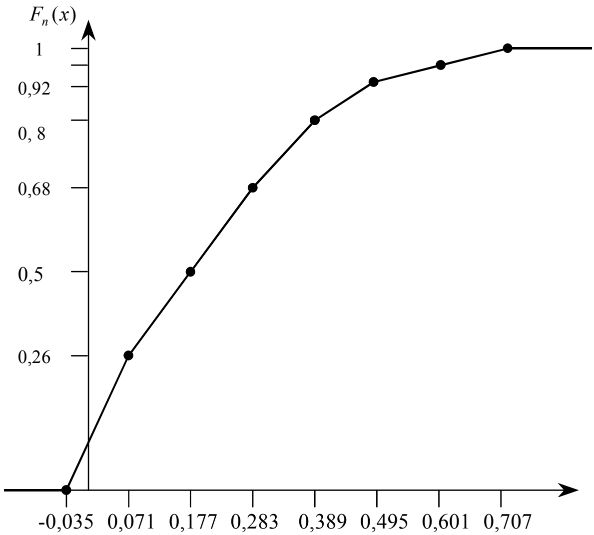

Рис. 12

Анализируя полученные результаты, можем предположить, интервалы времени между поступлениями вызовов такси распределены по показательному закону, так как полученная гистограмма схожа с кривой показательного распределения, график $F_n(x)$ также схож с функцией показательного распределения.

Задачи

11.1. На телефонной станции производились наблюдения за числом неправильных соединений в минуту. Наблюдения в течение часа дали следующие результаты:

3	1	3	4	2	1	1	3	2	7	2	0	1	2	1
4	0	3	0	2	0	2	1	3	3	1	2	0	3	2
0	2	0	4	3	4	2	0	2	0	1	1	2	2	3
4	1	4	2	0	1	0	1	1	0	1	1	1	5	5

Построить дискретный вариационный ряд, полигон, эмпирическую функцию распределения.

11.2. Пятьдесят наблюдений за жирностью молока дали следующие результаты в %:

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.3. Данные ежедневных измерений температуры в течение месяца представлены в таблице:

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.4. Ошибки 40 измерений приведены в таблице

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.5. При сверлении отверстий одним и тем же сверлом и последующем измерении диаметров получены следующие данные (в мм.):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.6. В ОТК завода была измерена глубина паза ста плашек и результаты приведены в следующей таблице (в мм.):

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Построить по этим данным [illegible] дискретный вариационный ряд. Построить полигон и эмпирическую функцию распределения.

11.7. Ошибки при округлении чисел заданы в таблице:

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.8. В таблице приведены данные исследований срока работы электрических лампочек (в годах):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.9. Время ожидания водителем зеленого света на перекрестке представлено случайной выборкой (в мин):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.10. Наблюдении за месячным доходом 50 жителей региона дали следующие результаты (в тыс. руб.):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.11. Имеются результаты наблюдений за числом сделок на фондовой бирже за квартал 40 инвесторов:

Построить по этим данным вариационный ряд, дискретный вариационный ряд. Построить полигон и эмпирическую функцию распределения.

11.12. Имеются статистические данные об удое 50 коров за лактационный период (в л):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.13. В таблице приведено распределение 50 рабочих по производительности труда (единиц за смену):

75	80	79	82	110	120	74	92	65	84
81	72	99	92	105	101	84	79	76	105
80	81	91	92	79	90	72	78	107	87
84	80	90	99	106	115	106	111	117	100
99	84	82	78	76	100	67	96	73	104

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.14. Имеются статистические данные об урожайности ржи на 40 участках колхозного поля (ц/га):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.15. В таблице приведены статистические данные о числе поврежденных стеклянных изделий в одном контейнере:

Построить по этим данным вариационный ряд, дискретный вариационный ряд. Построить полигон и эмпирическую функцию распределения.

11.16. Проведены исследования ошибки измерения дальности радиодальнометром, результаты экспериментов представлены в таблице:

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.17. Для определения надежности металлорежущих станков на заводе фиксировалось время их непрерывной работы до первого отказа. Полученные данные (в месяцах) приведены в таблице:

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.18. Интервал движения поездов метро составляет 2 минуты. В таблице приведены значения времени ожидания пассажиром поезда (в мин).

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.19. В таблице приведены численные значения времени обслуживания клиентов кассиром некоторого банка (в мин.):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.20. В таблице приведены значения прибыли 50 малых предприятий города Минска (в тыс. у.е.) за один из кварталов:

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.21. В фирме, производящей микропроцессоры, было исследовано их качество. Для этого испытаны 50 образцов и зафиксированы моменты выхода их из строя.

0,14	0,09	0,25	0,17	0,14	0,25	0,16	0,12	0,18	0,19
0,17	0,05	0,14	0,21	0,32	0,44	0,14	0,16	0,18	0,17
0,10	0,21	0,22	0,17	0,15	0,13	0,05	0,07	0,20	0,04
0,17	0,18	0,23	0,09	0,09	0,07	0,08	0,11	0,15	0,09
0,20	0,16	0,15	0,18	0,15	0,09	0,16	0,18	0,15	0,17

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.22. В одном из магазинов города исследовался спрос на прохладительные напитки в течение 50 летних дней. Имеются данные о количестве проданных бутылок:

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.23. В одном из магазинов города исследовался спрос на мороженное в течение 50 летних дней. Имеются данные о количестве проданных порций за день:

20541	20256	15264	10248	15248	20145	19254	18254	17458	14578
15487	12458	12425	21478	17145	16547	14258	14689	12658	15847
16478	16874	20458	19845	17844	17458	17452	16458	14532	14789
13578	14578	14795	12458	18547	18754	18264	17548	16548	16958
14875	12548	14587	15478	16145	15124	16015	17002	18048	16023

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.24. По данным выборочного обследования получено распределение семей по среднедушевому доходу (в у. е.):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

11.25. В страховой компании имеются данные об объемах страховых премий, получаемых компанией ежедневно (в у.е.):

Построить по этим данным интервальный вариационный ряд с равными интервалами. Построить полигон, гистограмму и эмпирическую функцию распределения.

§12. Средние величины

Средние величины характеризуют значения признака, вокруг которого концентрируются наблюдения. Наиболее распространенной из средних величин является средняя арифметическая.

Средней арифметической (выборочной средней) вариационного ряда называется величина:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{s} x_i n_i ,$$

где x_i – варианты дискретного ряда или середины интервалов интервального вариационного ряда, n_i – соответствующие им частоты, $n = \sum_{i=1}^{s} n_i$.

Для несгруппированного ряда все частоты $n_i = 1$, а $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ есть «невзвешенная» средняя арифметическая.

Пример 12.1. Найти среднее напряжение тока в электросети для примера 11.1.

Решение.

$$\bar{x} = \frac{1}{30}(106\cdot 1 + 107\cdot 3 + 108\cdot 4 + 109\cdot 6 + 110\cdot 8 + $$
$$+ 111\cdot 5 + 112\cdot 2 + 113\cdot 1) = 109{,}5 .$$

Отметим основные свойства выборочной средней, аналогичные свойствам математического ожидания случайной величины:

1. $\overline{C} = C$, если $C = const$;

2. , $C = const$;

3. , $C = const$;

4. .

Кроме рассмотренной средней арифметической, в статистическом анализе применяются структурные средние – медиана и мода.

Медианой Me вариационного ряда называется значение признака, приходящееся на середину ранжированного ряда наблюдений.

Для дискретного ряда с нечетным числом членов медиана равна серединному варианту, а для ряда с четным числом членов – полусумме двух серединных вариантов. Например, для примера 11.1

$Me = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(110 + 110) = 110$.

Для интервального ряда сначала находят медианный интервал $J_l = [x_l, x_{l+1})$, на который приходится середина ряда. Номер его будет соответствовать интервалу, кумулятивная частота которого равна или превышает половину суммы частот:

$$\sum_{i=1}^{l} n_i \le \frac{n}{2} < \sum_{i=1}^{l+1} n_i \ .$$

В случае выполнения равенства в предыдущей формуле номер медианного интервала равен l, в противном случае – $l + 1$. Медиану вычисляют по формуле

$$Me = x_l + \Delta \frac{\frac{n}{2} - \sum_{i=1}^{l-1} n_i}{n_{Me}} .$$

Здесь l – порядковый номер интервала, где находится медиана, Δ – величина медианного интервала, – накопленная частота домедианного интервала, n_{Me} – частота медианного интервала.

При получении медианы ряд разбивается на 2 равные части. Если ряд разбить на 4 части, то получатся ***квартили*** (q_1, q_2, q_3), на 10 частей – ***децили***. Второй квартиль q_2 равен медиане, а $q_{1,}$ q_3 вычисляются аналогично медиане с учетом разбиения.

Модой Mo вариационного ряда называется варианта, которой соответствует наибольшая частота. Например, в примере 11.1 $Mo = 110$.

Если распределение интервальное, то определяется модальный интервал $J_l = \lfloor x_{l,} x_{l+1} \big)$, которому соответствует наибольшая частота n_l, мода вычисляется по формуле:

$$Mo = x_l + \Delta \frac{n_l - n_{l-1}}{(n_l - n_{l-1}) + (n_l - n_{l+1})},$$

где n_{l-1}, n_{l+1} – частоты предмодального и послемодального интервалов.

Пример 12.2. Обследование качества пряжи дало следующие результаты, представленные в таблице. Найти моду и медиану этого распределения.

$x_l =$ □А.

Прочность нити, г	Частота	Накопленная частота
120 – 140	1	1
140 – 160	6	7
160 – 180	19	26
180 – 200	58	84
200 – 220	53	137
220 – 240	24	161
240 – 260	16	177
260 – 280	3	180
$\sum$	180	

Решение. Так как наибольшая частота $m_{Mo} = 58$ отвечает интервалу $180 - 200$, то , $m_{l-1} = 19$, $m_{l+1} = 53$, $\Delta = 20$. Мода равна:

.

Определим номер медианного интервала:

$$\sum_{i=1}^{l} n_i \le \frac{180}{2} < \sum_{i=1}^{l+1} n_i \ , \ \sum_{i=1}^{4} n_i \le 90 < \sum_{i=1}^{5} n_i \ , \ 84 \le \frac{180}{2} < 137 \ .$$

Следовательно, номер медианного интервала 5, а сам интервал $200-220$. Тогда получаем

.

Задачи

12.1. – 12.25. Используя данные задачи 11.1 – 11.25, вычислить:

а) выборочную среднюю,

б) моду,

в) медиану распределения.

Согласно условию задачи указать смысл полученных характеристик.

§13. Показатели вариации, моменты

Средние величины не отражают изменчивости (вариации) значений признака.

Простейшим показателем вариации является ***вариационный размах*** $R = x_{\max} - x_{\min}$.

Наибольший интерес представляет мера рассеяния наблюдений вокруг средней арифметической – дисперсия.

Дисперсией (выборочной дисперсий) вариационного ряда называется величина

$$D = \overline{(x-\overline{x})^2} = \frac{1}{n}\sum_{i=1}^{s}(x_i-\overline{x})^2 n_i \ .$$

При расчете дисперсии и других числовых характеристик интервальных рядов в качестве x_i также используют середины интервалов. Часто для вычисления дисперсии используют упрощенную формулу:

$$D = \overline{x^2} - \overline{x}^2 ,$$

где $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{s} x_i^{\,2} n_i$.

Если признак ξ измеряется в метрах, то, очевидно, его дисперсия – в метрах квадратных. Желательно в качестве меры вариации иметь характеристику, выраженную в тех же единицах, что и значения признака. Такой характеристикой является ***среднее квадратическое отклонение***:

.

Отношение среднего квадратического отклонения к средней величине признака, выраженное в процентах, называют ***коэффициентом вариации***:

$$v = \frac{\sigma}{\overline{x}} \cdot 100\ \% .$$

Если коэффициент вариации признака, принимающего только положительные значения, высок (например, более 100 %), то, как правило, это свидетельствует о неоднородности значений признака.

Дисперсия обладает следующими свойствами, аналогичными свойствам дисперсии случайной величины:

1. $D(C) = 0$, $C = const$;
2. , $C = const$;
3. , .

Пример 13.1. Вычислить дисперсию, среднее квадратическое отклонение и коэффициент вариации напряжения тока в электросети для примера 11.1.

Решение.

Вычислим дисперсию по упрощенной формуле .

$$\overline{x^2} = \frac{1}{30}(106^2 + 107^2 \cdot 3 + 108^2 \cdot 4 + 109^2 \cdot 6 + 110^2 \cdot 8 +$$
$$+ 111^2 \cdot 5 + 112^2 \cdot 2 + 113^2) = 11992{,}9 .$$

$$D = 11992{,}9 - 109{,}5^2 = 2{,}65 .$$

Среднее квадратическое отклонение $\sigma = \sqrt{2{,}65} = 1{,}63$. Вариация $v = \dfrac{1{,}63}{109{,}5} \cdot 100\ \% = 1{,}49\ \%$.

Средняя арифметическая и дисперсия вариационного ряда являются частными случаями более общего понятия – моментов.

Начальный момент $\tilde{\nu}_k$ -го порядка вариационного ряда определяется по формуле:

$$\tilde{\nu}_k = \frac{1}{n}\sum_{i=1}^{s} x_i^k n_i .$$

Центральный момент $\tilde{\mu}_k$ -го порядка вариационного ряда определяется по формуле:

$$\tilde{\mu}_k = \frac{1}{n}\sum_{i=1}^{s} (x_i - \bar{x})^k n_i .$$

Очевидно, что $\tilde{\nu}_1 = \bar{x}$, , $\tilde{\mu}_2 = D$.

Коэффициентом асимметрии вариационного ряда называется число

.

Если $\tilde{A}s = 0$, то распределение имеет симметричную форму, то есть варианты, равноудаленные от $\bar{x}$, имеют одинаковую частоту. При $\tilde{A}s > 0$ ($\tilde{A}s < 0$) говорят о положительной (отрицательной) или правосторонней (левосторонней) асимметрии.

Эксцессом вариационного ряда называется число

$$\tilde{E}x = \frac{\tilde{\mu}_4}{\sigma^4} - 3 = \frac{1}{n\sigma^4}\sum_{i=1}^{s} (x_i - \bar{x})^4 n_i - 3 .$$

Эксцесс является показателем крутости кривой распределения вариационного ряда по сравнению с нормальным распределением, дисперсия которого равна σ^2. При $\widetilde{E}x = 0$ распределение нормальное. Если $\widetilde{E}x > 0$, то кривая распределения имеет более острую вершину, чем при нормальном распределении, если $\widetilde{E}x < 0$ – более плоскую.

Пример 13.2. Вычислить коэффициент асимметрии и эксцесс распределения напряжения тока в электросети для примера 11.1.

Решение. Сначала находим центральные моменты третьего и четвертого порядков:

$$\widetilde{\mu}_3 = \frac{1}{30}\sum_{i=1}^{8}(x_i - 109{,}5)^3 n_i = \text{-}0{,}4;\ \widetilde{\mu}_4 = \frac{1}{30}\sum_{i=1}^{8}(x_i - 109{,}5)^4 n_i = 18{,}06.$$

Тогда $\widetilde{A}s = \dfrac{\widetilde{\mu}_3}{\sigma^3} = \dfrac{-0.4}{1{,}63^3} = -0{,}09$; $\widetilde{E}x = \dfrac{\widetilde{\mu}_4}{\sigma^4} - 3 = \dfrac{18{,}06}{1{,}63^4} - 3 = -0{,}44$. Поскольку найденные показатели близки к нулю, то можно сделать вывод, что рассматриваемое в примере 11.1 распределение по асимметрии и крутости приближается к нормальной кривой.

C

Вычисление выборочной средней и дисперсии можно упростить, если использовать не первоначальные варианты x_i, а новые варианты

$$u_i = \frac{x_i - C}{k},$$

где C и k – специально подобранные постоянные. Тогда согласно свойствам средней арифметической и дисперсии

$$\bar{u} = \overline{\left(\frac{x - C}{k}\right)} = \frac{\bar{x} - C}{k}, \text{ следовательно } \bar{x} = k\bar{u} + C,$$

$$D(u) = D\left(\frac{x - C}{k}\right) = \frac{D(x)}{k^2}, \text{ следовательно } D(x) = k^2 D(u).$$

Данный метод дает существенное упрощение в случае больших значений x_i. В качестве постоянной k рекомендуется брать величину интервала по x, а в качестве   – варианту, имеющую наибольшую частоту (середину интервала, имеющего наибольшую частоту).

Задачи

13.1. – 13.25. Используя данные задачи 11.1 – 11.25, вычислить:
а) выборочную дисперсию,
б) среднее квадратическое отклонение,
в) вариацию,
г) коэффициент асимметрии,
д) эксцесс.
На основе полученных результатов сделать выводы.

§14. Статистические оценки параметров распределения. Методы нахождения оценок

Основная задача теории оценок выглядит следующим образом. Имеется случайная величина ξ, для которой известен вид ее плотности распределения вероятностей с точностью до неизвестных параметров. Например, известно, что величина ξ нормальная, то есть ее плотность вероятностей имеет вид

,

но параметры a и , характеризующие эту плотность вероятностей, нам неизвестны. Нашей задачей является оценка этих неизвестных параметров. Будем пока считать, для простоты, что у нас имеется всего лишь один неизвестный параметр , подлежащий оценке.

Разумеется, оценить неизвестный параметр можно только на основе опытных данных . Имея выборку , мы должны указать число $\hat{\theta}$, которое близко к истинному значению неизвестного параметра θ, то есть мы должны указать ***оценку*** неизвестного параметра θ. Значит, мы должны каждой выборке поставить в соответствии некоторое число $\hat{\theta}$, которое будет называться оценкой неизвестного параметра θ. Другими словами, оценка есть функция от опытных данных:

$$\hat{\theta} = T(x_1, x_2, ..., x_n) = T(X).$$

Задачей теории оценки как раз и является указание вида функции $T(X)$.

Ясно, что функцию $T(X)$ следует выбирать таким образом, чтобы ее значения как можно точнее оценивали значения неизвестного параметра θ. К оценкам предъявляются требования, ограничивающие выбор функции . Рассмотрим эти требования.

1. ***Несмещенность*** – требование отсутствия систематических ошибок, или требование того, чтобы оценка в среднем совпадала с истинным значением неизвестного параметра:

$$M(\hat{\theta}) = \theta.$$

2. ***Эффективность***. Оценкой качества оценка является ее ***вариация***:

$$V(\hat{\theta}) = M(\hat{\theta} - \theta)^2.$$

Для несмещенной оценки она совпадает, очевидно, с дисперси-

$$\hat{\theta} \quad T(X) \quad \hat{\theta} \xrightarrow[n\to\infty]{P} \theta \qquad s^2 = \frac{n}{n-1} D = \frac{\sum_{i=1}^{s} (x_i - \bar{x})^2 n_i}{n-1}$$

ей. Оценка называется эффективной, если ее вариация является минимальной среди вариаций всех возможных оценок параметра θ, вычисленных по одному и тому же объему выборки .

3. ***Состоятельность***. Данное требование состоит в том, чтобы

.

Желательно использовать оценки, удовлетворяющие одновременно трем перечисленным требованиям.

Несмещенной оценкой математического ожидания случайной величины ξ служит выборочная средняя . Смещенной оценкой дисперсии случайной величины ξ служит выборочная дисперсия . Несмещенной оценкой дисперсии случайной величины ξ служит «***исправленная» выборочная дисперсия*** .

Точечной называют оценку, которая определяется одним числом. Рассмотрим основные методы нахождения точечных оценок.

Метод моментов

Метод моментов точечной оценки неизвестных параметров заданного распределения состоит в приравнивании теоретических моментов (начальных или центральных) соответствующим эмпирическим моментам того же порядка.

Напомним, что теоретические моменты для дискретной величины определяются по формулам:

$$\nu_k = \sum_{i=1}^{n} x_i^k p_i \,, \; \mu_k = \sum_{i=1}^{n} (x_i - M_\xi)^k p_i \,,$$

для непрерывных:

$$\nu_k = \int_{-\infty}^{+\infty} x^k p_\xi(x)dx \,, \; \mu_k = \int_{-\infty}^{+\infty} (x - M_\xi)^k p_\xi(x)dx \,.$$

Если распределение определяется одним параметром, то для его отыскания достаточно одного уравнения, чаще всего используют уравнение

$$\nu_1 = \tilde{\nu}_1, \text{ то есть } M_\xi = \overline{x}\,.$$

Если распределение определяется двумя параметрами, то чаще всего используют систему:

$$\begin{cases} \nu_1 = \tilde{\nu}_1, \\ \mu_2 = \tilde{\mu}_2; \end{cases} \text{, то есть } \begin{cases} M_\xi = \overline{x}, \\ D_\xi = D. \end{cases}$$

Разумеется, что для вычисления выборочных характеристик надо располагать выборкой.

Пример 14.1. Найти оценку параметра λ распределения Пуассона с помощью метода моментов.

Решение. Распределение Пуассона задается вероятностями

, $k = 0,1,2,\ldots$. В данном случае для нахождения единственного параметра достаточно приравнять $\nu_1 = \tilde{\nu}_1$ или $M_\xi = \overline{x}$. Математическое ожидание распределения Пуассона равно λ. Следовательно, оценка параметра закона Пуассона есть выборочная средняя :

$$\hat{\lambda} = \bar{x}.$$

Пример 14.2. Случайная величина ξ – время безотказной работы прибора, имеет показательное распределение:

$$, \ x \geq 0.$$

Ниже приведено эмпирическое распределение среднего времени работы $n = 200$ элементов:

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

Найти методом моментов точечную оценку неизвестного параметра показательного распределения.

Решение. Будем использовать уравнение , то есть $M_\xi = \bar{x}$.

λ $\bar{x}$ $\nu_1 = \tilde{\nu}_1$ $f_\xi(x) = \lambda e^{-\lambda x}$ $\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{5} = 0{,}2$ Математическое ожидание показательного распределения равно $\frac{1}{\lambda}$.

Значит $\hat{\lambda} = \frac{1}{\bar{x}}$.

Определим выборочную среднюю:

$$\bar{x} = \frac{1}{200}(2{,}5 \cdot 133 + 7{,}5 \cdot 45 + 12{,}5 \cdot 15 + 17{,}5 \cdot 4 + 22{,}5 \cdot 2 + 27{,}5 \cdot 1) = 5.$$

Тогда получим оценку параметра λ:

.

Оценки метода моментов состоятельны, однако по эффективности они не являются наилучшими. Тем не менее метод моментов часто используется на практике, так как приводит к сравнительно простым вычислениям.

Метод максимального правдоподобия

Основным методом получения оценок параметров генеральной совокупности по данным выборки является метод максимального правдоподобия, предложенный Р.Фишером.

Основу метода составляет ***функция правдоподобия***, выражающая плотность вероятности (вероятность) совместного появления результатов выборки $x_1, x_2, ..., x_n$:

$$L(x_1, x_2, ..., x_n; \theta) = L(X; \theta) = p(x_1, \theta) p(x_2, \theta) ... p(x_n, \theta),$$

где <u>в случае непрерывного распределения</u> $p(x, \theta)$ – плотность распределения вероятностей исследуемой случайной величины, <u>в случае дискретного распределения</u> $p(x, \theta)$ – вероятность того, что в результате испытания случайная величина примет значение x.

Согласно методу максимального правдоподобия в качестве оценки неизвестного параметра принимается такое значение , которое максимизирует функцию L**.** То есть оценка $\hat{\theta}$ является точкой максимума функции правдоподобия.

Нахождение оценки $\hat{\theta}$ упрощается, если максимизировать не саму функцию L, а $\ln L$, поскольку максимум обеих функций достигается при одном и том же значении θ. Функция называется ***логарифмической функцией правдоподобия***.

Точку максимума функции $\ln L$ по аргументу θ можно искать, например, так:

1. Найти производную .

2. Найти критическую точку θ^* из уравнения $\dfrac{d \ln L}{d\theta} = 0$.

3. Найти вторую производную $\dfrac{d^2 \ln L}{d\theta^2}$, если вторая прозводная при $\theta = \theta^*$ отрицательна, то θ^* – точка максимума.

Найденную точку максимума θ^* принимают в качестве оценки максимального правдоподобия параметра θ, .

В случае, когда надо оценить не один параметр θ, а несколько $(\theta_1, \theta_2, ..., \theta_k)$, оценки максимального правдоподобия для этих параметров находят из системы уравнений:

$$\frac{\partial L(X; \theta_1, \theta_2, ..., \theta_k)}{\partial \theta_i} = 0, \; i = \overline{1, k}.$$

Пример 14.3. Для случайной величины ξ, распределенной по закону Пуассона, по результатам выборки $(x_1, x_2, ..., x_n)$ оценить неизвестный параметр λ.

Решение. Для распределения Пуассона

$$p(x_i; \lambda) = P(\xi = x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}, \; i = \overline{1, n}.$$

Построим функцию правдоподобия:

$$L = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} ... \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! ... x_n!} e^{-n\lambda}.$$

Перейдем к логарифмической функции правдоподобия:

$$\ln L = -n\lambda + \ln \lambda \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!).$$

Дифференцируя это равенство, имеем

$$\frac{d \ln L}{d\lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i.$$

Приравнивая нулю, имеем

$$-n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0, \; \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}.$$

Находя вторую производную, получим

$\frac{d^2 \ln L}{d\lambda^2} = -\frac{1}{\lambda^2} \sum_{i=1}^{n} x_i < 0$, при $\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$, для распределения Пуассона $x_i \geq 0$.

Следовательно $\hat{\lambda} = \overline{x}$.

Пример 14.4. Пусть случайная величина ξ имеет показательное распределение с плотностью вероятностей , $x \geq 0$. Требуется по результатам выборки $x_1, x_2, ..., x_n$ найти оценку неизвестного параметра λ.

Решение. Для показательного распределения функция правдоподобия имеет вид:

.

Логарифмируя эту функцию, получим:

$$\ln L = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i \ .$$

Тогда

$$\frac{d \ln L}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \ .$$

Отсюда $\lambda = \dfrac{n}{\sum\limits_{i=1}^{n} x_i} = \dfrac{1}{x}$.

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{1}{\lambda^2} < 0 \, .$$

Значит, оценка неизвестного параметра имеет вид $\hat{\lambda} = \dfrac{1}{x}$.

Пример 14.5. Для нормально распределенной случайной величины ξ по результатам выборки найти оценку неизвестных параметров a и .

Решение. Плотность распределения вероятностей нормального закона имеет вид:

.

В этом случае функция правдоподобия имеет вид:

$$L = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2}.$$

Прологарифмировав L, получим:

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2 .$$

Дифференцируя по a и σ, будем иметь:

$$\begin{cases} \dfrac{\partial \ln L}{\partial a} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = 0, \\ \dfrac{\partial \ln L}{\partial \sigma} = -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - a)^2 = 0. \end{cases}$$

Отсюда находим оценки неизвестных параметров a и σ:

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \quad \hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sigma^2 .$$

Важность метода максимального правдоподобия связана с оптимальными свойствами его оценок. Основной недостаток метода максимального правдоподобия – трудность вычисления оценок, связанных с решением уравнений правдоподобия, чаще всего нелинейных. Существенно и то, что для построения оценок максимального правдоподобия и обеспечения их «хороших» свойств необходимо знание типа анализируемого закона распределения $p(x, \theta)$, что во многих случаях оказывается практически нереальным.

Широкое распространение в практике статистических исследований получил **метод наименьших квадратов**, так как он, во-первых, не требует знания закона распределения выборочных данных, во-вторых, достаточно хорошо разработан в плане вычислительной реализации. Суть его заключается в том, что ***оценка определяется из условия минимизации суммы квадратов отклонений выборочных данных от определяемой оценки***.

Применение метода наименьших квадратов будет рассмотрено при решении задач корреляционного и регрессионного анализа.

Задачи

14.1. Случайная величина ξ – число семян сорняков в пробе зерна, распределена по закону Пуассона, ниже приведено распределение семян сорняков в 1000 пробах зерна:

Найти методом моментов точечную оценку неизвестного параметра распределения Пуассона.

14.2. По данным задачи 14.1 найти методом максимального правдоподобия точечную оценку неизвестного параметра распределения Пуассона.

14.3. Найти методом моментов по выборке:

Точечную оценку параметра биномиального распределения, –

число появления события в i -м опыте, , n – количество испы-

таний в одном опыте, .

14.4. Найти методом моментов точечную оценку параметра p

геометрического распределения , где x_i – число испытаний, проведенных до появления события A, p – вероятность появления события в одном испытании.

14.5. Случайная величина (уровень воды в реке по сравнению с номиналом) подчинена гамма-распределению, плотность которого определяется параметрами и ( ):

, $x \geq 0$.

Ниже приведено распределение среднего уровня воды по данным 45 паводков:

x_i	37,5	62,5	87,5	112,5	137,5	162,5	187,5	250	350
n_i	1	3	6	7	7	5	4	8	4

Найти методом моментов точечные оценки неизвестных параметров α и рассматриваемого гамма-распределения.

14.6. Случайная величина (отклонение контролируемого размера изделия от номинала) подчинена нормальному закону распределения с неизвестными параметрами и . Ниже приведено эмпирическое распределение отклонения от номинала 200 изделий:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,2	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Найти методом моментов точечные оценки неизвестных параметров a и .

14.7. Найти методом моментов по результатам выборки $x_1, x_2, \ldots, x_n$

точечные оценки параметров a и равномерного расределения, плотность которого $p_\xi(x) = \frac{1}{b-a}$, $b > a$,

14.8. Найти методом максимального правдоподобия по выборке $x_1, x_2, \ldots, x_n$ точечную оценку параметра β гамма-распределения (параметр известен), плотность которого

$$p_\xi(x) = \frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)} x^{\alpha} e^{-\frac{x}{\beta}}, \; x \geq 0, \; \alpha > -1, \; \beta > 0.$$

14.9. Найти методом максимально правдоподобия по выборке $x_1, x_2, \ldots, x_n$ точечную оценку параметра p геометрического распределения $P(\xi = x_i) = p(1-p)^{x_i-1}$, где x_i – число испытаний, проведенных до появления события A, p – вероятность появления события в одном испытании.

14.10. Найти методом максимального правдоподобия по выборке точечную оценку параметра a (параметр известен) распределения Кэптейна, плотность которого

,

где $g(x)$ – дифференцируемая функция.

14.11. По методу наибольшего правдоподобия найти оценку неизвестного параметра α случайной величины , распределенной по закону Максвелла с плотностью вероятностей

.

14.12. Из генеральной совокупности, распределенной по закону χ^2 с неизвестным параметром α, сделана выборка . Найти методом максимального правдоподобия оценку параметра α, если плотность вероятности

, $x > 0$.

14.13. Продолжительность безотказной работы датчика является случайной величиной с плотностью распределения вероятностей

$$p_\xi(t) = \frac{1}{\alpha} e^{-\frac{t}{\alpha}},\ t \ge 0,\ \alpha \ge 0.$$

По фиксированным значениям методом максимального правдоподобия оценить неизвестный параметр α.

14.14. Случайная величина представляет собой количество срывов поставок потребителям фирмами, производящими однородную продукцию. За определенный период обследовано 8 фирм, у которых количество срывов поставок соответственно равно: 6, 3, 1, 1, 3, 4, 0, 2. Полагая, что случайная величина распределена по закону Пуассона, методом максимального правдоподобия найти оценку параметра .

§15. Интервальные оценки

Интервальной называют оценку, которая определяется двумя числами – концами интервала, покрывающего оцениваемый параметр.

Доверительным называют интервал , который покрывает (содержит) неизвестный параметр θ с заданной ***надежностью*** (***доверительной вероятностью***) , то есть .

При этом δ называют ***точностью оценки***.

Следует обратить внимание на то, что границы интервала и его величина находятся по выборочным данным и поэтому являются случайными величинами в отличие от оцениваемого параметра – величины неслучайной, поэтому говорят, что интервал «покрывает» («накрывает»), а не «содержит» истинное значение θ.

Величина доверительного интервала существенно зависит от объема выборки (уменьшается с ростом n) и значения доверительной вероятности (увеличивается с приближением γ к единице).

Интервальной оценкой с надежностью γ *математического ожидания нормально распределенного количественного признака* по выборочной средней при известном среднем квадратическом отклонении σ генеральной совокупности служит доверительный интервал

,

где $\delta = t\dfrac{\sigma}{\sqrt{n}}$ – точность оценки, n – объем выборки, – значение аргумента функции Лапласа , при котором .

При неизвестном среднем квадратическом отклонении σ используют интервал:

,

где $s = \sqrt{s^2}$ – «исправленное» выборочное среднее квадратическое отклонение, t_γ находят по таблице приложения 4 по заданным n и .

Интервальной оценкой с надежностью γ *среднего квадратического отклонения нормально распределенного количественного признака* по «исправленному» выборочному среднему квадратическому отклонению служит доверительный интервал

где q находят по таблице приложения 5 по заданным и .

Первый из рассмотренных интервалов строится на основе следствия из теоремы Ляпунова (центральная предельная теорема):

.

Для определения необходимого объема выборки, при котором с вероятностью γ можно утверждать, что для нормально распределенной величины выборочное среднее отличается от генеральной средней M_ξ по абсолютной величине меньше чем на δ, пользуются формулой:

.

Пример 15.1. Из генеральной совокупности извлечена выборка объема $n = 25$:

x_i	4	6	7	8	9	10	11
n_i	1	6	3	3	7	3	2

Найти: 1) несмещенную оценку генеральной средней;

2) несмещенную оценку генеральной дисперсии.

Решение.

1) Как было рассмотрено выше, несмещенной оценкой генеральной средней является выборочная средняя:

$$\bar{x} = \frac{1}{25}(4 \cdot 1 + 6 \cdot 6 + 7 \cdot 3 + 8 \cdot 3 + 9 \cdot 7 + 10 \cdot 3 + 11 \cdot 2) = 8 .$$

2) Несмещенной оценкой генеральной дисперсии является «исправленная» выборочная дисперсия:

$$s^2 = \frac{n}{n-1} D = \frac{\sum_{i=1}^{s} (x_i - \bar{x})^2 n_i}{n-1} =$$

$$= \frac{1}{24}((4-8)^2 \cdot 1 + (6-8)^2 \cdot 6 + (7-8)^2 \cdot 3 + (8-8)^2 \cdot 3 +$$

$$+ (9-8)^2 \cdot 7 + (10-8)^2 \cdot 3 + (11-8)^2 \cdot 2) = 3{,}33 .$$

Пример 15.2. Найти доверительный интервал для оценки математического ожидания a и среднего квадратического отклонения нормально распределенного признака с надежностью $\lambda = 0{,}99$, зная выборочную среднюю , объем выборки $n = 25$ и среднее квадратическое отклонение $\sigma = 3$.

$\bar{x} - \frac{t\sigma}{\sqrt{n}} < a < \bar{x} + \frac{t\sigma}{\sqrt{n}}$ $\Phi(t) = \frac{\gamma}{2}$

– *Решение.* Доверительный интервал для математического ожидания имеет вид

.

Все величины, кроме t, известны. Найдем t из соотношения , то есть $\Phi(t) = \frac{0{,}99}{2} = 0{,}495$. По таблице приложения 2 находим $t = 2{,}57$. Тогда

$$10{,}5 - 2{,}57 \frac{3}{\sqrt{25}} < a < 10{,}5 + 2{,}57 \frac{3}{\sqrt{25}} ,$$

окончательно имеем

$$8{,}958 < a < 12{,}042 .$$

Найдем доверительный интервал для среднего квадратического отклонения

$$s(1-q) < \sigma < s(1+q) \quad \text{при} \quad q < 1,$$
$$0 < \sigma < s(1+q) \quad \text{при} \quad q > 1,$$

Сначала определим параметр q по таблице приложения 4. При , $\lambda = 0{,}99$, параметр . Находим «исправленное» среднее квадратическое отклонение по формуле

$$s = \sqrt{\frac{n}{n-1} D} = \sigma\sqrt{\frac{n}{n-1}} = 3\sqrt{\frac{25}{24}} = 3{,}06.$$

Тогда получаем доверительный интервал

$$3{,}06(1-0{,}49) < \sigma < 3{,}06(1-0{,}49),$$

После вычислений

.

Пример 15.3. Найти минимальный объем выборки, при котором с надежностью 0,95 точность оценки математического ожидания генеральной совокупности по выборочной средней равна , если известно среднее квадратическое отклонение нормально распределенной генеральной совокупности.

<u>Решение.</u> Воспользуемся формулой, определяющей точность оценки математического ожидния генеральной совокупности по выборочной средней: . Отсюда

$$n = \frac{t^2\sigma^2}{\delta^2}.$$

Определим величину t из условия , то есть

$\Phi(t) = \dfrac{0{,}95}{2} = 0{,}475$. По таблице приложения 2 найдем $t = 1{,}96$. Тогда

$$n = \frac{1{,}96^2 1{,}2^2}{0{,}3^2} = 61.$$

Задачи

15.1. При паспортизации партии из 100 арматурных стержней производилось выборочное исследование на разрыв опытных образцов. По результатам выборки составлен следующий вариационный ряд

Границы прочности (МПа)	300 – 370	370 – 440	440 – 510	510 – 580	580 – 650	650 – 720	720 – 790
Число стержней	1	2	5	8	4	3	2

Определить несмещенные оценки генеральной средней и генеральной дисперсии.

15.2. В итоге десяти измерений некоторой физической величины одним прибором получены следующий результаты: 10, 11, 10, 12, 11, 9, 9, 10, 11, 10. Найти выборочную среднюю результатов измерений и выборочную «исправленную» дисперсию ошибок прибора.

15.3. При обследовании средней зарплаты рабочих предприятия выборочным путем отобраны 100 человек. Средняя выборочная зарплата оказалась 150 у.е., а дисперсия 100 у.е. Определить с надежностью 0,99 доверительные интервалы для генеральной средней и среднего квадратического отклонения.

$\sigma = 1{,}5$

15.4. По данным 25 измерений прочности асфальтобетона установлено, что $\bar{x} = 40$ кг/см, $s = 1{,}5$. Полагая, что данные выборки имеют нормальное распределение, определить с надежностью 0,95 доверительный интервал для генеральной средней.

15.5. Выборка из большой партии электроламп содержит 100 ламп. Средняя продолжительность горения лампы по выборке оказалась 1000 ч. Найти с надежностью 0,95 доверительный интервал для средней продолжительности a горения лампы по всей партии, если среднее квадратическое отклонение продолжительности горения лампы 40 ч. Предполагается, что продолжительность горения ламп распределена нормально.

15.6. Найти минимальный объем выборки, при котором с надежностью 0,925 точность оценки математического ожидания нормально распределенной генеральной совокупности по выборочной средней равна 0,2, если известно среднее квадратическое отклонение генеральной совокупности .

15.7. По данным шестнадцати независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений и «исправленное» среднее квадратическое отклонение $s = 5$. Оценить истинное значение измеряемой величины с помощью доверительного интервала с надежностью 0,99. Предполагается, что результаты измерений распределены нормально. *Указание*: истинное значение измеряемой величины равно ее математическому ожиданию.

15.8. Произведено 14 измерений одним прибором (без систематической ошибки) некоторой физической величины, причем «исправленное» среднее квадратическое отклонение случайных ошибок измерения оказалось равным 0,6. Найти точность прибора с надежностью 0,99. Предполагается, что результаты измерений распределены нормально. *Указание*: точность прибора характеризуется средним квадратическим отклонением случайных ошибок измерений.

15.9. С целью определения средней суммы a вкладов в банке, была исследована величина 100 случайно отобранных вкладов. Найти с вероятностью 0,96 доверительные границы для a.

15.10. Определить численность выборки при обследовании остатков на расчетных счетах у клиентов банка, чтобы с вероятностью 0,683 предельная ошибка равнялась 5 у.е., если $\sigma = 120$ у.е.

§16. Некоторые статистические распределения

Рассмотрим три основных закона распределения, составляющих необходимый аппарат для построения в дальнейшем статистических критериев. Плотности этих распределений можно найти в любом справочнике по теории вероятностей, здесь из-за громоздкости они не приводятся.

Пусть – независимые случайные величины, распределенные по стандартному нормальному закону:

$$\xi_1, \xi_2, ..., \xi_k \sim N(0;1).$$

Говорят, что сумма квадратов этих случайных величин распределена по ***закону*** χ^2 ***(Хи-квадрат) с*** k ***степенями свободы***. Эту случайную величину обозначают $\chi^2(k)$:

$$\chi^2(k) = {\xi_1}^2 + {\xi_2}^2 + ... + {\xi_k}^2.$$

Графики плотности распределения $\chi^2(k)$ при различных k изображены на рис. 13.

$\xi_{\square}, \xi_{\square}, ..., \xi_j$

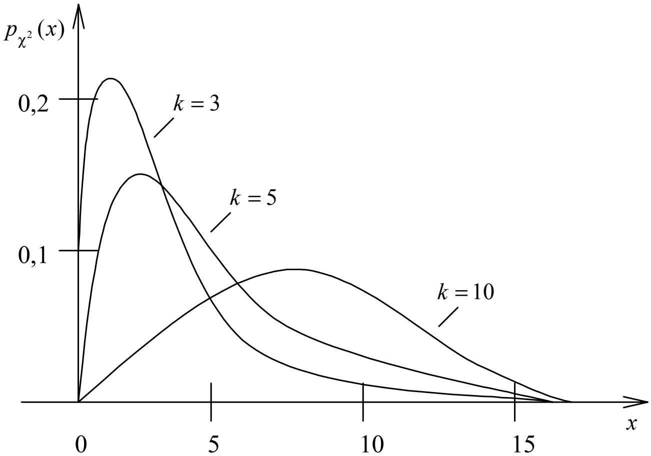

Рис. 13

1. Случайная величина $\chi^2(k)$ принимает неотрицательные значения, то есть нулевую плотность распределения при $x \le 0$. Это следует из опредления.

2. При большом числе степеней свободы k распредление $\chi^2(k)$ близко к нормальному. Этот факт иллюстрируется графиком на рис. 13.

3. Математическое ожидание случайной величины, распределенной по закону $\chi^2(k)$, равно k.

Пусть случайная величина ξ распределена по стандартному нормальному закону: $\xi \sim$. ***Распределением Стьюдента*** (или t ***-распределением***) с степенями свободы называется распределение случайной величины

$$t(k) = \frac{\xi}{\sqrt{\dfrac{\chi^2(k)}{k}}} .$$

Стьюдент – псевдоним английского статистика В.Госсета.

Графики плотности распределения Стьюдента при различном числе степеней свободы приведены на рис. 14. Из вида графиков и определения можно сделать некоторые наблюдения о свойствах распределения Стьюдента.

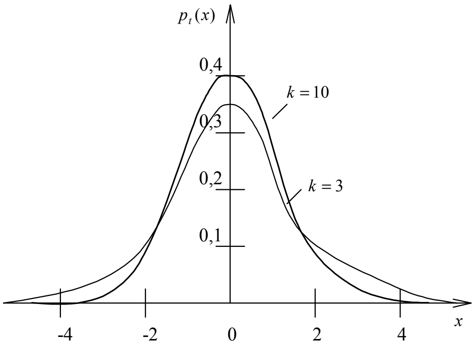

Рис. 14

1. Распределение Стьюдента симметрично, причем .

2. При больших k распределение Стьюдента близко к стандартному нормальному распределению $N(0;1)$.

На основе распределения χ^2 вводится еще одна случайная величина. Распределением ***Фишера*** (***Фишера – Снедекора*** или F ***-распределением***) с k_1 и k_2 степенями свободы называется распределение случайной величины

$$F(k_1, k_2) = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}.$$

1. Из определения видно, что данная случайная величина не может принимать орицательные значения, то есть имеет нулевую плотность распределения вероятностей при $x \le 0$.

2. Графики плотности распределения при различном числе степеней свободы изображены на рис. 15. При некоторых значениях числа степеней свободы k_1 и k_2 F-распределение приближается к нормальному.

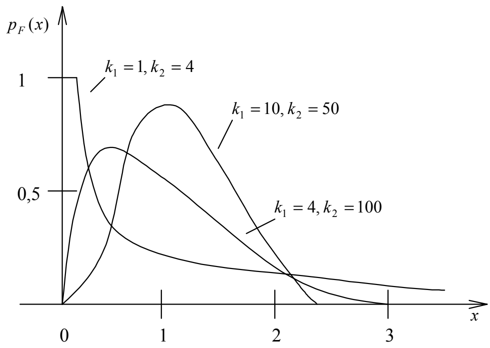

Рис. 15

§17. Проверка статистических гипотез

Наряду с задачами оценивания параметров большую группу задач математической статистики составляют так называемые задачи проверки статистических гипотез. ***Статистической гипотезой*** называется предположение относительно параметров или вида распределения изучаемой случайной величины. Например, после анализа выборки, то есть после того, как мы построили статистический ряд, определили выборочные характеристики, построили полигон, гистограмму, функцию распределения, мы делаем предположение, что данная случайная величина распределена по нормальному закону с определенными параметрами. Выдвинутое предположение является статистической гипотезой. После этого нужно принять решение:

противоречат экспериментальные данные высказанной гипотезе или нет. Процесс принятия решения называется ***проверкой статистической гипотезы***, а алгоритм проверки – ***решающим правилом***. Поскольку мы выдвигали гипотезу, опираясь только на случайные выборочные значения, наши выводы будут носить вероятностный характер. Мы не дадим точного ответа: да или нет. Можно будет лишь с некоторой долей уверенности (с некоторой вероятностью) утверждать, что данные не противоречат или противоречат предположению.

Статистические гипотезы можно разделить на следующие основные группы:

1) гипотезы о параметрах распределения,

2) гипотезы о виде распределения.

Выдвинутую гипотезу называют ***нулевой*** и обозначают ее через H_0. Наряду с H_0 рассматривают ***конкурирующую*** (или ***альтернативную***) гипотезу H_1. Например:

а) если H_0: «генеральная совокупность распределена нормально», то H_1: «генеральная совокупность не распределена нормально»;

б) если H_0: «математическое ожидание ξ равно 5», то : «математическое ожидание ξ не равно 5».

Таким образом, ставится задача проверки гипотезы относительно конкурирующей гипотезы H_1 на основе выборки X объема n. Правило, по которому принимается или отвергается гипотеза, называется ***статистическим критерием***. Принципы проверки статистических гипотез впервые были сформулированы в работах известных математиков Е.Неймана и Э.Пирсона. Они исходили из того что, принимая или отвергая гипотезу , можно допустить ошибки двух видов.

Ошибка первого рода: H_0 отвергается (принимается H_1) в то время, как в действительности верна гипотеза H_0. Вероятность ошибки первого рода называют ***уровнем значимости*** и обозначают α:

.

Величину $1-\alpha$, то есть вероятность принять верную гипотезу, называют ***уровнем доверия*** (***доверительным уровнем***):

$$P(H_0 / H_0) = 1-\alpha .$$

Ошибка второго рода: H_0 принимается, в то время как верна гипотеза H_1. Вероятность ошибки второго рода обозначается β:

.

Вероятность принять гипотезу H_1, если она верна, называют ***мощностью критерия***:

$$P(H_1 / H_1) = 1-\beta .$$

Возможные ситуации наглядно иллюстрируются следующей таблицей.

Гипотеза H_0	Принимается	Отвергается
Верна	Правильное решение	Ошибка 1-го рода
Неверна	Ошибка 2-го рода	Правильное решение

Применяя юридическую терминологию, α – вероятность вынесения судом обвинительного приговора, когда на самом деле обвиняемый невиновен, – вероятность вынесения судом оправдательного приговора, когда на самом деле обвиняемый виновен в совершенном преступлении.

β $P(H_0/H_1)=\beta$ $K=K(x_1,\ldots,x_n)$

Нам хотелось бы, конечно, сделать вероятности ошибок первого и второго рода нулевыми. Однако это оказывается невозможным. Более того, как правило, уменьшая вероятность ошибки первого рода, мы увеличиваем вероятность ошибки второго рода и наоборот.

Суть проверки статистической гипотезы заключается в том, что используется специально составленная выборочная характеристика (***статистика***) , полученная по выборке X, так, чтобы в случае, если гипотеза H_0 верна, точное или приближенное распределение K было бы известным, например χ^2-распределением. Построение критерия, в зависимости от вида гипотезы H_0, заключатся в выборе таких значений $K_{кр}^1$ и $K_{кр}^2$, что если $K_{кр}^1 < K < K_{кр}^2$, то гипотеза H_0 принимается. При этом возможно $K_{кр}^1 = -\infty$ или $K_{кр}^2 = +\infty$. Значения $K_{кр}^1$ и $K_{кр}^2$ называются критическими, а область

$$K_{Д} = \left\{K : K_{кр}^1 < K < K_{кр}^2\right\}$$

называется областью допустимых значений.

Таким образом, множество возможных значений статистики K разбивается на два непересекающихся подмножества: ***критическую область*** – множество значений K, при которых H_0 отвергается – $\overline{K_{Д}}$, и ***область допустимых значений*** (область принятия решений) – множество значений K, при которых H_0 принимается – $K_{Д}$. При этом точки $K_{кр}^1$ и $K_{кр}^2$, отделяющие эти два множества, называют ***критическими точками***. Если фактически наблюдаемое (полученное по выборке) значение статистики критерия K попадает в критическую область, то гипотезу H_0 отвергают, в противном случае принимают.

В зависимости от вида конкурирующей гипотезы H_1 выбирают ***правостороннюю***, ***левостороннюю*** или ***двухстороннюю*** критическую область. При конкурирующей гипотезе $H_1 : \theta > \theta_0$ следует использовать правостороннюю критическую область ($K_{кр}^1 = -\infty$), в случае $H_1 : \theta < \theta_0$ – левостороннюю ($K_{кр}^2 = +\infty$), а при гипотезе $H_1 : \theta \neq \theta_0$ – двухстороннюю критическую область. Границы критических областей $K_{кр}^1$ и $K_{кр}^2$ при заданном уровне значимости α определяются соответственно из соотношений:

для правосторонней критической области

,

для левосторонней критической области

$$P(K < K_{кр}^1) = \alpha ,$$

для двухсторонней критической области

$$P(K > K_{кр}^2) = P(K < K_{кр}^1) = \frac{\alpha}{2} .$$

Нейман и Пирсон предложили следующий принцип построения критической области. *Критическую область следует выбирать так, чтобы вероятность попадания в нее статистического критерия* K *была минимальной и равной* α*, если верна нулевая гипотеза* , *и максимальной в противоположном случае.* Другими словами, критическая область должна быть такой, чтобы при заданном уровне значимости α мощность критерия была максимальной. Задача построения такой области решается с помощью теоремы Неймана – Пирсона, излагаемой в более полных курсах математической статистики. Уровень значимости α обычно задают значениями 0,05 и 0,01.

Наиболее распространена правосторонняя критическая область, рассмотрим подробнее принцип ее построения. По выборочному распределению статистики определяется критическое значение $K_{кр}^2$ – такое что, если гипотеза H_0 верна, то вероятность $P(K > K_{кр}^2) = \alpha$ мала, то есть событие $K > K_{кр}^2$ можно считать практически невозможным. Поэтому, если окажется $K > K_{кр}^2$, то гипотеза H_0 отвергается, в то время как обнаружение того, что $K < K_{кр}^2$, подтверждает справедливость H_0.

Предположим, что если верна гипотеза H_0, то статистика K имеет распределение с плотностью $p = p_1(K)$, а если верна гипотеза H_1, то $p = p_2(K)$. Тогда описанные выше построения критической области можно изобразить графически.

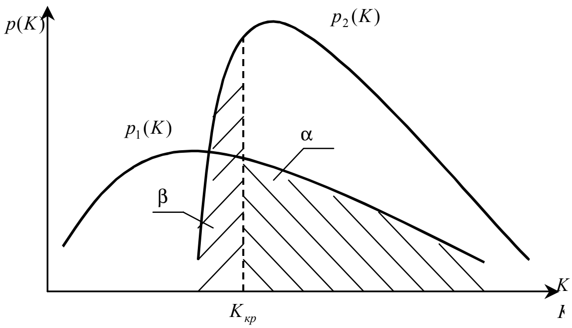

Рис. 16

На рис. 16 или $K_{кр}^{2} = K_{кр}$. Критической правосторонней областью является множество $\{K : K > K_{кр}\}$. Гипотеза H_0 принимается, если $K < K_{кр}$. Площади заштрихованных областей представляют собой вероятности ошибок 1-го и 2-го рода, α и соответственно.

Рис. 17

На рис. 17 представлена правосторонняя критическая область, принцип построения которой описан выше.

Итак, процедуру проверки статистической гипотезы можно разбить на следующие основные шаги.

1. Сформулировать нулевую и конкурирующую H_1 гипотезы.

2. Задать уровень значимости α.

3. Выбрать статистику для проверки гипотезы H_0.

4. Найти плотность распределения $p = p_1(K)$ статистики в предположении, что гипотеза H_0 верна.

5. Определить критическую область $\overline{K_Д}$. Для наиболее часто используемых распределений (Стьюдента, Фишера, χ^2) составлены таблицы критических точек, которые можно найти в учебниках по математической статистике.

6. По выборке вычислить выборочное значение K статистики критерия.

7. Принять решение: если $K \in \overline{K_Д}$, то H_0 отклоняется (то есть принимается H_1), если $K \in K_Д$, то H_0 принимается.

Принятое решение, разумеется, носит вероятностный характер. Поэтому обычно применяют более осторожные формулировки. Вместо того чтобы сказать «гипотеза H_0 отклоняется», говорят «данные эксперимента не подтверждают гипотезу H_0 «, «гипотеза не согласуется с экспериментом» и т.д.

Критерий согласия χ^2 Пирсона

Одной из главных задач математической статистик является установление истинного закона распределения случайной величины на основании экспериментальных данных. На практике о законе распределения можно судить, например, по виду полигона и гистограммы. Однако полной уверенности в сделанном предположении о законе распределения нет, поэтому вопрос может стоять лишь о проверке гипотезы о предполагаемом законе распределения. Критерии, устанавливающие закон распределения, называются критериями согласия. Для проверки гипотезы о нормальном распределении генеральной совокупности используется критерий Пирсона.

Алгоритм применения критерия Пирсона заключается в следующем.

h, $n_i' = \frac{nh}{\sigma}\varphi(u_i)$

1. Из генеральной совокупности образовывается случайная выборка, и на ее основе делается предположение о нормальном законе распределения. Выдвигается гипотеза H_0: «генеральная совокупность распределена нормально».

2. Вычисляются выборочные числовые характеристики $\overline{x}$, σ.

3. Вычисляются теоретически частоты:

3.1. <u>Для дискретного ряда</u>

,

где n – объем выборки, – шаг (разность между двумя соседними вариантами),

$$u_i = \frac{x_i - \overline{x}}{\sigma},\ \varphi(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}},$$

значения $\varphi(u)$ определяются о таблице приложения 1;

3.2. Для интервального ряда

,

где – объем выборки, – теоретические вероятности попадания в интервалы $x_i - x_{i+1}$, $z_i = \frac{x_i - \overline{x}}{\sigma}$, $z_{i+1} = \frac{x_{i+1} - \overline{x}}{\sigma}$,

$\Phi(z)$ – функция Лапласа, значения которой определяются по таблице приложения 2.

4. Находится наблюдаемое значение критерия Пирсона по формуле

.

5. По таблице критических точек распределения χ^2, по заданному уровню значимости α и числу степеней свободы (s – число групп для дискретно ряда или число интервалов для интервального ряда) находят критическую точку правосторонней критической области.

6. Если $\chi^2 < \chi^2_{кр}$ – нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо. Если $\chi^2 > \chi^2_{кр}$ – гипотезу отвергают.

Замечание. Малочисленные варианты и интервалы (содержащие малочисленные частоты) $n_i < 5$ следует объединить, а соответствующие им частоты сложить. Если производилось объединение частот, то в формуле $k = s - 3$ следует в качестве s принять число групп или интервалов выборки, оставшихся после объединения частот.

Пример 17.1. Для интервального статистического ряда, полученного в результате наблюдения случайной величины, требуется:

1) вычислить числовые характеристики данного эмпирического распределения: выборочную среднюю и выборочную дисперсию;

2) вычислить теоретические частоты предполагаемого нормального распределения;

3) при заданном уровне значимости $\alpha = 0{,}05$ проверить гипотезу о нормальном распределении генеральной совокупности, пользуясь критерием Пирсона.

Даны результаты наблюдения за распределением 60 валиков по диаметру:

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Решение.

1). Найдем середины интервалов и примем их в качестве вариант для расчета числовых характеристик

x_i	13,99	14,09	14,19	14,29	14,39	14,49	14,59	14,69
n_i	1	1	4	10	15	13	10	6

$$n = \sum n_i = 60\,.$$

Выборочную среднюю определим по формуле

$$\bar{x} = \frac{1}{n}\sum x_i n_i = 14{,}4333.$$

Определим $\overline{x^2}$

$$\overline{x^2} = \frac{1}{n}\sum x_i^2 n_i = 208{,}3456.$$

Дисперсия $D = \overline{x^2} - (\bar{x})^2 = 208{,}3456 - 14{,}4333 \cdot 14{,}4333 = 0{,}0255$. Среднее квадратическое отклонение $\sigma = \sqrt{D} = 0{,}1597$.

2). Найдем теоретические частоты. Для этого пронормируем данную случайную величину X и перейдем к величине Z, $Z = \dfrac{X - \bar{x}}{\sigma}$. Затем найдем теоретические вероятности, пользуясь формулой $P_i = \Phi(z_{i+1}) - \Phi(z_i)$, $\Phi(z)$ – функция Лапласа. И, наконец, по формуле $n_i' = 60 \cdot P_i$ определим теоретические частоты n_i'. Расчеты приведем в следующей таблице:

3). Сравним эмпирические и теоретические частоты, используя критерий Пирсона. Вычислим наблюдаемое значение критерия Пирсона по формуле

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'} .$$

Необходимым условием применения критерия Пирсона является наличие в каждом из интервалов не менее 5 наблюдений. Однако, учитывая, что первые три интервала содержат малочисленные частоты, объединим их, а соответствующие частоты и теоретические частоты сложим. Данные расчетов приведем в таблице.

№	x_i	x_{i+1}	n_i	n_i'	$\frac{(n_i - n_i')^2}{n_i'}$	$\frac{n_i^2}{n_i'}$
1	13,94	14,24	6	6,786	0,0910	5,3050
2	14,24	14,34	10	9,984	2,56E-05	10,0160
3	14,34	14,44	15	14,232	0,0414	15,8094
4	14,44	14,54	13	13,878	0,0556	12,1776
5	14,54	14,64	10	9,252	0,0605	10,8085
6	14,64	14,74	6	5,868	0,0030	6,1350
$\sum$			60	60	0,2515	60,2515

Значит, $\chi^2 = 0{,}2515$. Столбец $\frac{n_i^2}{n_i'}$ последней таблицы нужен для контроля, так как, если вычисления произведены правильно, то должно выполняться равенство $\sum \frac{n_i^2}{n_i'} - n = \chi^2$. *Контроль*:

$$\sum \frac{n_i^2}{n_i'} - n = 60{,}2515 - 60 = 0{,}2515 = \chi^2$$ выполняется.

По таблице критических точек распределения χ^2 (приложение 6), по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = s - 3$ (s – число интервалов) находим критическую точку правосторонней критической области $\chi^2_{кр}(0{,}05;3) = 7{,}8$. Так как $\chi^2 < \chi^2_{кр}$, то гипотезу о нормальном распределении генеральной совокупности принимаем.

Пример 17.2. Используя критерий Пирсона, при уровне значимости $0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с эмпирическим распределением выборки объема $n = 200$:

x_i	6,68	6,70	6,72	6,74	6,76	6,78	6,80	6,82
n_i	5	17	24	54	52	23	18	7

Решение. Во-первых, найдем выборочную среднюю $\overline{x} = 6{,}7507$ и выборочное среднее квадратическое отклонение $\sigma = 0{,}0316$.

Вычислим теоретические частоты по формуле

$$n_i' = \frac{nh}{\sigma}\varphi(u_i) = \ldots\varphi(u_i),$$

для этого составим расчетную таблицу.

i	x_i	$u_i = \dfrac{x_i - \overline{x}}{\sigma}$	$\varphi(u_i)$	$n_i' = 126{,}58\varphi(u_i)$
1	6,68	-2,24	0,0325	4,12
2	6,70	-1,60	0,1109	14,04
3	6,72	-0,97	0,2492	31,54
4	6,74	-0,34	0,3765	47,66
5	6,76	0,29	0,3825	48,42
6	6,78	0,93	0,2589	32,78
7	6,80	1,56	0,1182	14,96
8	6,82	2,19	0,0363	4,60

Составим расчетную таблицу, из которой найдем наблюдаемое значение критерия Пирсона

i	n_i	n_i'	$\dfrac{(n_i - n_i')^2}{n_i'}$	n_i''
1	5	4,12	0,1880	4
2	17	14,04	0,6240	14
3	24	31,54	1,8025	32
4	54	47,66	0,8434	48
5	52	48,42	0,2647	49
6	23	32,78	2,9179	33
7	18	14,96	0,6178	15
8	7	4,60	1,2522	5
$\sum$	200		8,5105	200

Значит, $\chi^2 = 8{,}5105$. По таблице критических точек распределения χ^2 находим критическую точку правосторонней критической области $\chi^2_{кр}(0{,}05; 5) = 11{,}1$.

Так как $\chi^2 < \chi^2_{кр}$ – гипотезу о нормальном распределении генеральной совокупности принимаем. Другими словами, эмпирические и теоретические частоты различаются незначимо.

По смыслу частота есть целое число, поэтому иногда целесообразно округлить n_i' до целых, следя при этом за тем, чтобы сумма полученных таким образом теоретических частот была равна объему выборки. Получим частоты , сумма которых равна .

Замечание. Так как нормальное распределение является непрерывным, то, проверяя гипотезу о нормальном распределении на основе дискретного вариационного ряда данных, можно осуществить переход к интервальному вариационному ряду, считая варианты дискретного ряда серединами интервалов. Например, в примере 17.2 перейти к интервальному ряду

Далее осуществлять проверку гипотезы аналогично примеру 17.1.

Задачи

17.1. – 17.25. Используя данные задачи 11.1 – 11.25, при заданном уровне значимости $\alpha = 0{,}05$ проверить гипотезу о нормальном распределении генеральной совокупности, пользуясь критерием Пирсона.

§18. Элементы теории корреляционного и регрессионного анализа

Методы теории корреляции позволяют определить зависимость между различными факторами или случайными величинами. Термин корреляция произошел от латинского «correlatio» – соотношение, взаимосвязь.

В естественных науках часто речь идет о ***функциональной зависимости***, когда каждому значению одной величины соответствует вполне определенное значение другой. Случайные величины обычно не связаны функциональной зависимостью. В экономике в большинстве случаев между переменными величинами существуют зависимости, когда каждому значению одной переменной соответствует не какое-то определенное, а множество возможных значений другой переменной. Иначе говоря, каждому значению одной переменной соответствует определенное условное распределение другой переменной. Например, значению соответствует распределение величины η

$\eta/\xi = x_1$	y_1'	$\dots$	y_k'
n	n_1'	$\dots$	n_k'

,

значению соответствует распределение

$\eta/\xi = x_2$	y_1''	$\dots$	y_k''
n	n_1''	$\dots$	n_k''

и т.д. Такая зависимость получила название ***статистической*** (или ***стохастической***, ***вероятностной***). Примером статистической связи является зависимость урожайности от количества внесенных удобрений, производительности труда на предприятии от его энерговооруженности и т.п.

В силу неоднозначности статистической зависимости между и для исследователя представляет интерес усредненная схема зависимости – зависимость условного математического ожидания или его статистического аналога $\overline{y}_x$ от значений x случайной величины , то есть или $\overline{y}_x = f(x)$. Здесь $\overline{y}_x$ – условная средняя, кото-

рая определяется как среднее арифметическое значений η (   ), соответствующих значению $\xi = x$. Такая зависимость получила название корреляционной. ***Корреляционной зависимостью*** между двумя величинами называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой. Уравнение
называют ***уравнением регрессии*** η на , уравнение   называют ***выборочным уравнением регрессии*** η на . Функцию   называют ***функцией регрессии***, а ее график – ***линией регрессии***.

Статистические связи между переменными можно изучать методами корреляционного и регрессионного анализа. Основной задачей корреляционного анализа является выявление связи между случайными величинами и оценка ее тесноты. Основной задачей регрессионного анализа – установление и изучение формы зависимости между переменными.

Данные о статистической зависимости удобно представлять в виде корреляционной таблицы.

	η	$y_1^* - y_2^*$	$y_2^* - y_3^*$	$y_3^* - y_4^*$	…	$y_m^* - y_{m+1}^*$	
ξ	Серед. интерв	y_1	y_2	y_3	…	y_m	n_{x_i}
$x_1^* - x_2^*$	x_1	n_{11}	n_{12}	n_{13}	…	–	n_{x_1}
$x_2^* - x_3^*$	x_2	–	n_{22}	n_{23}	…	–	n_{x_2}
$x_3^* - x_4^*$	x_3	–	–	n_{33}	…	–	$n_{x_{i3}}$
…	…	…	…	…	…	…	…
$x_s^* - x_{s+1}^*$	x_s	–	–	–	…	n_{sm}	n_{x_s}
	n_{y_j}	n_{y_1}	n_{y_2}	n_{y_3}	…	n_{y_m}	$n = \sum$

Здесь n_{ij} – частоты появления пар (x_i, y_j), прочерк говорит о том, что соответствующая пара (x_i, y_j) не встречалась, $n_{y_j} = \sum_{i=1}^{s} n_{ij}$, $n_{x_i} = \sum_{j=1}^{m} n_{ij}$, $n = \sum_{i=1}^{s} n_{x_i} = \sum_{j=1}^{m} n_{y_j}$.

Наличие корреляции приближенно может быть определено с помощью ***корреляционного поля***. Его получим, если нанесем на график в определенном масштабе точки, соответствующие наблюдаемым одновременным значениям двух величин (x_i, y_j).

Пример 18.1. В таблице приведены данные, отражающие зависимость урожайности зерновой культуры η (ц) от расстояния до реки (км). Построить поле корреляции, сделать вывод.

ξ ξ η	100	120	140	160	180	n_{x_i}
5	–	–	–	3	2	5
10	–	–	–	4	1	5
15	–	–	8	3	–	11
20	–	–	10	–	1	11
25	–	–	5	–	–	5
30	6	4	–	–	–	10
35	1	2	–	–	–	3
n_{y_j}	7	6	23	10	4	50

Решение. Полученное корреляционное поле представлено на рис. 17. Так как точки поля корреляции концентрируются вдоль убывающей прямой, то можно сделать предположение об обратной линейной зависимости между урожайностью и расстоянием до реки. То есть чем больше расстояние до реки, тем меньше урожайность исследуемой зерновой культуры.

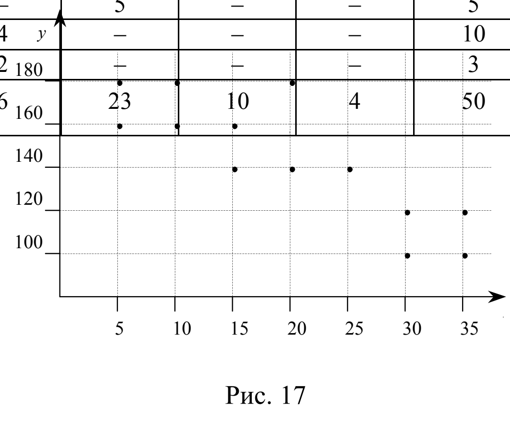

Рис. 17

Перейдем к оценке тесноты корреляционной зависимости. Рассмотрим наиболее важный для практики случай линейной зависимости. В теории вероятностей показателем тесноты линейной зависимости являлся коэффициент корреляции, в математической статистике таким показателем является выборочный коэффициент корреляции.

Выборочным коэффициентов корреляции называется величина, рассчитываемая по формуле

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y},$$

где $\overline{xy} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{m} x_i y_j n_{ij}$, $\overline{x}$, $\overline{y}$ – выборочные средние, σ_x, σ_y – выборочные средние квадратические отклонения.

Отметим основные свойства выборочного коэффициента корреляции, аналогичные свойствам коэффициента корреляции для случайных величин.

1. Коэффициент корреляции принимает значения на отрезке $[-1,1]$, то есть $-1 \le r \le 1$.

2. Чем ближе значение $|r|$ к единице, тем более тесная ***линейная*** зависимость между изучаемыми величинами. В зависимости от того, насколько $|r|$ приближается к единице, различают слабую, умеренную, заметную, достаточно тесную и весьма тесную линейную связь.

3. Если $r > 0$, то говорят о прямой зависимости, то есть с увеличением значений одной из величин значения другой также увеличиваются, при $r < 0$ – обратную зависимость.

4. Если все значения переменных увеличить (уменьшить) на одно и то же число или в одно и то же число раз, то величина коэффициента корреляции не изменится. Коэффициент корреляции есть безразмерная характеристика тесноты линейной связи.

5. При $r = \pm 1$ корреляционная связь представляет линейную функциональную зависимость, при этом все точки поля корреляции лежат на одной прямой.

6. При $r = 0$ или $|r|$ близком к нулю ***линейная*** корреляционная связь отсутствует. Но это не означает отсутствие другой зависимости, например, нелинейная связь может быть очень тесной.

Для ответа на вопрос о значимости коэффициента корреляции проверяют нулевую гипотезу $H_0 : r_г = 0$ о равенстве нулю генерального коэффициента корреляции. Если гипотеза принимается, то это означает, что между ξ и нет линейной корреляционной зависимости, в противном случае линейная зависимость признается значимой.

Для того чтобы при уровне значимости проверить нулевую гипотезу при конкурирующей , надо вычислить наблюдаемое значение критерия

$$t_{набл} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

и по таблице критических точек распределения Стьюдента (приложение 7), по заданному уровню значимости α и числу степеней свободы найти критическую точку $t_{кр}(\alpha; k)$ двухсторонней критической области. Если $|t_{набл}| < t_{кр}$ – нет оснований отвергнуть нулевую гипотезу. Если $|t_{набл}| > t_{кр}$ – нулевую гипотезу отвергаем.

Пример 18.2. По данным примера 18.1 рассчитать выборочный коэффициент корреляции. При уровне значимости 0,05 проверить нулевую гипотезу о равенстве генерального коэффициента корреляции нулю при конкурирующей гипотезе $H_1 : r_г \neq 0$. Сделать вывод.

Решение. Для удобства вычислений построим вспомогательную таблицу.

ξ η	100	120	140	160	180	n_{x_i}	$x_i n_{x_i}$	$x_i^2 n_{x_i}$
5	–	–	–	3	2	5	25	125
10	–	–	–	4	1	5	50	500
15	–	–	8	3	–	11	165	2475
20	–	–	10	–	1	11	220	4400
25	–	–	5	–	–	5	125	3125
30	6	4	–	–	–	10	300	9000
35	1	2	–	–	–	3	105	3675
n_{y_j}	7	6	23	10	4	$n=50$	990	23300
$y_i n_{y_j}$	700	720	3220	1600	720	6960	Σ	
$y_i^2 n_{y_j}$	70000	86400	450800	256000	129600	992800		

Находим средние значения:

$$\overline{x} = \frac{990}{50} = 19{,}8\,,\ \ \overline{y} = \frac{6960}{50} = 159{,}2\,,\ \ \overline{x^2} = \frac{23300}{50} = 466\,,$$

$$\overline{y^2} = \frac{992800}{50} 19856\,,\ \ \sigma_x = \sqrt{466 - 19{,}8^2} = 8{,}6\,,$$

$$\sigma_y = \sqrt{19856 - 139{,}2^2} = 21{,}89\,,$$

$$\overline{xy} = \frac{1}{50}(100 \cdot 30 \cdot 6 + 100 \cdot 1 \cdot 35 + 120 \cdot 30 \cdot 4 + 120 \cdot 2 \cdot 35 + 140 \cdot 8 \cdot 15 +$$
$$+ 140 \cdot 10 \cdot 20 + 140 \cdot 5 \cdot 25 + 160 \cdot 3 \cdot 5 + 160 \cdot 4 \cdot 10 + 160 \cdot 3 \cdot 15 + 180 \cdot 2 \cdot 5 +$$
$$+ 180 \cdot 1 \cdot 10 + 180 \cdot 1 \cdot 20) = 2596.$$

Находим коэффициент корреляции:

$$r = \frac{2596 - 19{,}8 \cdot 159{,}2}{8{,}6 \cdot 21{,}89} = -0{,}851\,.$$

Проверим гипотезу о равенстве генерального коэффициента корреляции нулю. Рассчитаем наблюдаемое значение критерия

$$t_{набл} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0{,}851\sqrt{50-2}}{\sqrt{1-(-0{,}851)^2}} = \frac{5{,}896}{0{,}525} = 11{,}23\,.$$

По таблице критических точек распределения Стьюдента определим $t_{кр}(0{,}05; 48) = 2{,}01$. Так как $|t_{набл}| > t_{кр}$, отвергаем нулевую гипотезу о равенстве генерального коэффициента корреляции нулю.

Таким образом, анализируя полученное значение выборочного коэффициента корреляции, делаем вывод о достаточно тесной обратной линейной зависимости между ξ и , что не противоречит выводам примера 18.1.

Рассмотрим ***уравнение парной линейной регрессии*** .

Найдем формулы расчета неизвестных параметров a и по имеющимся статистическим данным (x_i, y_i), $i = \overline{1, n}$.

Согласно ***методу наименьших квадратов*** неизвестные параметры выбираются таким образом, чтобы сумма квадратов отклонений выборочных значений y_i от значений $\overline{y}_{x_i} = a + bx_i$, полученных по уравнению регрессии, была минимальна:

$$S = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \to \min .$$

На основании необходимого условия экстремума приравниваем нулю частные производные, получим

$$\begin{cases} 2\sum_{i=1}^{n} (y_i - a - bx_i) = 0, \\ 2\sum_{i=1}^{n} (y_i - a - bx_i)x_i = 0; \end{cases}$$

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i , \\ a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^{2} = \sum_{i=1}^{n} x_i y_i . \end{cases}$$

После преобразований получим систему нормальных уравнений для определения параметров линейной регрессии:

ξ $\overline{y}_x - \overline{y} = b(x - \overline{x})$ $b = \dfrac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \overline{x}^2}$

$$\begin{cases} a + b\overline{x} = \overline{y}, \\ a\overline{x} + b\overline{x^2} = \overline{xy}. \end{cases}$$

Из последней системы следуют формулы для определения параметров уравнения парной линейной регрессии η на :

$$\text{или } b = r\frac{\sigma_y}{\sigma_x} ,$$

$$a = \overline{y} - b\overline{x} .$$

Уравнение регрессии $\overline{y}_x = a + bx$ можно с учетом формулы вычисления параметра a записать в виде

.

Коэффициент b показывает, на сколько единиц в среднем изменится переменная η при увеличении переменной на одну единицу.

Уравнение регрессии может быть использовано для прогнозирования значений при значениях не указанных в корреляционной таблице.

Замечание. Если значения переменных ξ и (то есть и y_j) достаточно велики, то при расчете параметров a и удобно перейти к условным вариантам $u_i = \dfrac{x_i - c}{k}$ и $v_j = \dfrac{y_j - c'}{k'}$, где k и k' – величины интервалов, а и – варианты (середины интервалов), имеющих наибольшую частоту. Тогда , $\overline{y} = k'\overline{v} + c'$, $\sigma_x^2 = k^2\sigma_u^2$, $\sigma_y^2 = k'^2\sigma_v^2$, $r(u,v) = r(x,y)$.

Пример 18.3. По данным примера 18.1 определить параметры уравнения парной линейной регрессии, построить линию регрессии на корреляционном поле. Спрогнозировать значение урожайности при км.

Решение.

Определим параметры уравнения регрессии

,

$a = \overline{y} - b\overline{x} = 139{,}2 + 2{,}17 \cdot 19{,}8 = 182{,}17$.

Запишем полученное уравнение регрессии $\overline{y}_x = 182{,}17 - 2{,}17x$ и нанесем полученную прямую на корреляционное поле.

5 10 15 20 25 30 35 x

Рис. 18

Найдем прогнозное значение урожайности η при км:

.

Пример 18.4. Найти коэффициент линейной корреляции между признаками   и , записать уравнение прямой регрессии η на , если распределение признаков приводится в таблице.

x_i	y_i
2,3	29
4,1	38
3,8	39
3,9	38
1,8	28
1,5	30
4,1	38
1,8	17
1,7	15
3	37

Решение. Составим следующую расчетную таблицу

№	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	2	23	4	529	46
2	4,1	31	16,81	961	127,1
3	3,8	35	14,44	1225	133
4	3,9	36	15,21	1296	140,4
5	2,1	23	4,41	529	48,3
6	4	34	16	1156	136
7	4,1	38	16,81	1444	155,8
8	1,8	17	3,24	289	30,6
9	1,7	13	2,89	169	22,1
10	3	37	9	1369	111
11	2,3	19	5,29	361	43,7
12	2,1	18	4,41	324	37,8
13	2,9	29	8,41	841	84,1
14	3	38	9	1444	114
15	1,8	18	3,24	324	32,4
16	1,5	20	2,25	400	30
17	2,1	29	4,41	841	60,9
18	3,2	36	10,24	1296	115,2
19	2,2	25	4,84	625	55
20	3	33	9	1089	99
Сумма	54,6	552	163,9	16512	1622,4
Среднее	2,73	27,6	8,195	825,6	81,12

Тогда $\overline{x} = 2{,}73$, $\overline{y} = 27{,}6$, $\overline{x^2} = 8{,}195$, $\overline{y^2} = 825{,}6$, $\overline{xy} = 81{,}12$,

$$\sigma_x = \sqrt{8{,}195 - 2{,}73^2} = 0{,}86, \ \sigma_y = \sqrt{825{,}6 - 27{,}6^2} = 7{,}99.$$

Выборочный коэффициент корреляции

$$r = \frac{81{,}12 - 2{,}73 \cdot 27{,}6}{0{,}86 \cdot 7{,}99} = 0{,}84,$$

параметры уравнения $b = 0{,}84 \dfrac{7{,}99}{0{,}86} = 7{,}8$, $a = 27{,}6 - 7{,}8 \cdot 2{,}73 = 6{,}31$.

Уравнение регрессии $\overline{y}_x = 6{,}31 + 7{,}8x$.

Задачи

18. Для исследования зависимости случайных величин η и получены статистические данные, представленные в корреляционной таблице (– наблюдаемые значения ξ , – значения η). Требуется:

а) построить корреляционное поле,

б) определить выборочный коэффициент корреляции,

в) при уровне значимости 0,05 проверить нулевую гипотезу о равенстве генерального коэффициента корреляции нулю при конкурирующей гипотезе ,

г) найти уравнение прямой регрессии η на ,

д) построить линию регрессии на корреляционном поле.

18.1.

18.2.

$x \backslash y$	11	17	23	29	35	41	n_x
15	5	1	–	–	–	–	6
25	–	6	2	–	–	–	8
35	–	–	5	26	5	–	36
45	–	–	7	12	10	–	29
55	–	–	–	6	7	8	21
n_y	5	7	14	44	22	8	100

18.3.

$x \backslash y$	15	25	35	45	55	65	n_x
14	2	4	–	–	–	–	6
22	–	3	7	–	–	–	10
30	–	3	30	15	4	–	52
38	–	–	11	7	5	–	23
46	–	–	–	1	2	2	5
54	–	–	–	–	2	2	4
n_y	2	10	48	23	13	4	100

18.4.

$x \backslash y$	14	22	30	38	46	54	n_x
16	3	7	2	–	–	–	12
24	–	2	12	6	–	–	20
32	–	7	27	11	–	–	45
40	–	–	10	6	–	–	16
48	–	–	–	2	1	1	4
56	–	–	–	–	2	1	3
n_y	3	16	51	25	3	2	100

18.5.

$x \backslash y$	13	19	25	31	37	n_x
17	3	–	–	–	–	3
23	7	2	6	–	–	15
29	2	11	32	–	–	45
35	–	10	12	2	–	24
41	–	–	–	6	3	9
47	–	–	–	–	4	4
n_y	12	23	50	8	7	100

18.6.

x \ y	50	60	70	80	90	n_x
0,5	–	–	12	2	3	17
1	–	2	6	6	4	18
1,5	2	11	32	–	–	45
2	7	10	–	–	–	17
2,5	3	–	–	–	–	3
n_y	12	23	50	8	7	100

18.7.

x \ y	300	350	400	450	500	550	n_x
0,1	–	–	–	–	3	2	5
0,2	–	–	4	5	5	2	16
0,3	–	8	14	5	–	–	27
0,4	3	4	10	–	–	–	17
0,5	2	3	–	–	–	–	5
n_y	5	15	28	10	8	4	70

18.8.

x \ y	1	3	5	7	9	n_x
1000	2	2	–	–	–	4
1500	3	4	–	–	–	7
2000	–	13	6	4	–	23
2500	–	9	21	13	2	45
3000	–	–	19	26	10	55
3500	–	–	–	9	17	26
n_y	5	28	46	52	29	160

18.9.

x \ y	10000	11000	12000	13000	14000	n_x
5	1	2	–	–	–	3
15	2	3	–	–	–	5
20	–	4	8	2	–	14
25	–	8	8	7	2	25
30	–	–	13	8	2	23
n_y	3	17	29	17	4	70

18.10.

x \ y	-1	-2	-3	-4	-5	n_x
10	1	2	–	–	–	3
11	1	2	4	–	–	7
12	–	4	4	8	–	16
13	–	2	4	7	–	13
14	–	–	1	7	1	9
15	–	–	–	1	1	2
n_y	2	10	13	23	2	50

18.11.

x \ y	18	20	22	24	26	28	n_x
-5	–	–	–	–	7	1	8
-10	–	–	–	1	5	1	7
-15	–	–	2	7	1	–	10
-20	–	3	5	7	–	–	15
-25	2	3	3	–	–	–	8
-30	1	1	–	–	–	–	2
n_y	3	7	10	15	13	2	50

18.12.

x \ y	1 – 2	2–3	3–4	4–5	5–6	n_x
2 – 4	1	2	–	–	–	3
4 – 6	2	3	–	–	–	5
6 – 8	–	4	8	2	–	14
8 – 10	–	8	8	7	2	25
10 – 12	–	–	13	8	2	23
n_y	3	17	29	17	4	70

18.13.

x \ y	10–14	14–18	18–22	22–26	26–30	n_x
2 – 4	3	4	–	–	–	7
4 – 6	–	10	9	3	–	22
6 – 8	–	6	40	5	–	51
8 – 10	–	–	4	8	3	15
10 – 12	–	–	–	2	3	5
n_y	3	20	53	18	6	100

18.14.

x \ y	12-14	14-16	16-18	18-20	20-22	22-24	24-26	n_x
10-20	–	–	–	–	7	5	3	15
20-30	–	–	–	5	4	2	–	11
30-40	–	–	6	4	2	–	–	9
40-50	–	2	6	4	–	–	–	12
50-60	3	5	3	–	–	–	–	11
60-70	4	2	–	–	–	–	–	6
n_y	7	9	12	13	13	7	3	64

18.15.

x \ y	0.1	0.2	0.3	0.4	0.5	0.6	0.7	n_x
225	3	6	2	–	–	–	–	11
375	1	4	5	1	3	–	–	14
525	1	3	8	10	2	2	7	33
675	–	6	2	9	8	4	–	29
825	–	–	2	5	1	2	–	10
975	–	–	–	–	2	1	–	3
n_y	5	19	19	25	16	9	7	100

18.16.

x \ y	0,02	0,06	0,10	0,14	0,18	0,22	n_x
72-75	3	1	1	–	–	–	5
75-78	1	2	4	3	2		12
78-81	–	1	3	2	3	2	11
81-84	–	–	1	4	4	3	12
84-87	–	–	–	1	2	5	8
n_y	4	4	9	10	11	10	48

18.17.

x \ y	15-20	20-25	25-30	30-35	35-40	40-45	n_x
2,1-2,2	–	–	1	1	–	–	2
2,2-2,3	–	1	4	3	2	–	10
2,3-2,4	2	7	–	9	7	3	36
2,4-2,5	–	3	4	3	3	–	13
2,5-2,6	–	–	3	2	2	–	7
2,6-2,7	–	–	2	2	–	–	4
n_y	2	11	22	20	14	3	72

18.18.

x / y	0,5	1,0	1,5	2,0	2,5	n_x
20	–	–	–	2	3	5
50	–	–	9	3	3	15
80	–	6	40	5	–	51
110	–	10	4	8	–	22
140	3	4	–	–	–	7
n_y	3	20	53	18	6	100

18.19.

x / y	56	68	80	92	104	116	128	140	n_x
0,9	2	3	5	–	–	–	–	–	10
1,3	–	6	3	5	–	–	–	–	14
1,7	–	–	5	8	15	–	–	–	28
2,1	–	–	–	6	9	10		–	25
2,5	–	–	–	–	1	6	8	–	15
2,9	–	–	–	–	–	3	4	1	8
n_y	2	9	13	19	25	19	12	1	100

18.20.

x / y	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	n_x
250	3	4	5	–	–	–	–	–	12
450	–	6	2	8	–	–	–	–	16
650	–	–	–	5	14	9	–	–	28
850	–	–	–	6	8	6		–	20
1050	–	–	–	–	5	7	4	–	16
1250	–	–	–	–	–	–	5	3	8
n_y	3	10	7	19	27	22	9	3	100

18.21.

x / y	-7	-6	-5	-4	-3	-2	-1	0	n_x
10	–	–	–	–	–	3	2	4	9
20	–	–	–		4	5	7	–	16
30	–	–	–	7	8	6		–	21
40	–	–	–	9	14	6	–	–	29
50	–	–	7	5	2	–	–	–	14
60	4	2	5	–	–	–	–	–	11
n_y	4	2	12	21	28	20	9	4	100

18.22.

$x \backslash y$	14	16	18	20	22	24	26	28	n_x
19	3	2	3	–	–	–	–	–	8
29	–	1	4	5	–	–	–	–	10
39	–	–	7	13	8	–	–	–	28
49	–	–	–	–	9	6	6	–	21
59	–	–	–	–	–	7	8	3	18
69	–	–	–	–	–	4	6	5	15
n_y	3	3	14	18	17	17	20	8	100

18.23.

$x \backslash y$	18,5	19,7	20,9	22,1	23,3	24,5	25,7	26,9	n_x
100	3	3	4	6	–	–	–	–	16
200	–	5	8	9	–	–	–	–	22
300	–	–	–	–	8	9		–	30
400	–	–	–	–	9	2	4	–	15
500	–	–	–	–		1	3	5	9
600	–	–	–	–	–	–	5	3	8
n_y	3	8	12	28	17	12	12	8	100

18.24.

$x \backslash y$	36	56	76	96	116	136	156	176	n_x
5,4	–	–	–	–	–	2	1	2	5
7,0	–	–	–	–	–	4	3	1	8
8,6	–	–	–	–	–	8	3	2	13
10,2	–	–	–	–	17	5	4	–	26
11,8	–	–	6	14	5	–	–	–	25
13,4	2	4	3	10	4	–	–	–	23
n_y	2	4	9	24	26	19	11	5	100

18.25.

$x \backslash y$	4,5	6,0	7,5	9,0	10,5	12,0	13,5	15,0	n_x
60	–	–	–	–	–	–	4	4	8
90	–	–	–	–	–	6	3	2	11
120	–	–	–	–	1	6	5	–	12
150	–	–	–	–	17	9	6	–	32
180	–	4	8	9	4	–	–	–	25
210	5	3	2	2	–	–	–	–	12
n_y	5	7	10	11	22	21	18	6	100

§19. Однофакторный дисперсионный анализ

Дисперсионный анализ определяется как статистический метод, предназначенный для оценки влияния различных факторов на результат эксперимента, а также для последующего планирования аналогичных экспериментов. Например, необходимо выяснить, существенно ли различие между партиями некоторого изделия по определенному показателю качества, то есть проверить влияние на качество изделия одного фактора – партии изделия. По числу факторов, влияние которых исследуется, различают однофакторный и многофакторный дисперсионный анализ.

Пусть на количественный нормально распределенный признак ξ воздействует фактор , который имеет m постоянных уровней . Одновременно будем рассматривать пример об исследовании влияния технологии обработки почвы на урожайность. Задача, которую предстоит решить, ставится следующим образом: выяснить, влияет выбор технологии обработки почвы на урожайность культуры или нет. Выбор технологии естественно назвать фактором, если m – полное число применяемых технологий, то каждую отдельную технологию , $i=\overline{1,m}$, называют ***уровнем фактора***. Пусть на i-м уровне проведено наблюдений, в результате которых получено $n=\sum_{i=1}^{m} n_i$ значений x_{ij} признака ξ, – номер уровня фактора, , j – номер испытания на этом уровне, . В рассматриваемом примере x_{ij} – урожайность культуры, полученная в j-м году при использовании -й технологии, , где n_i – число лет, в течение которых производились наблюдения за применением технологии F_i. Сведем все данные в таблицу.

Уровни фактора	Номер испытания								Гр. средн.
	1	2	...	n_1	...	n_2	...	n_m	
F_1	x_{11}	x_{12}	...	x_{1n_1}					$\overline{x}_{гр1}$
F_2	x_{21}	x_{22}	...	x_{2n_1}	...	x_{2n_2}			$\overline{x}_{гр2}$
...	...	...	...	...	...	...	...		
F_m	x_{n1}	x_{n2}	...	x_{mn_1}	...	x_{mn_2}	...	x_{mn_m}	$\overline{x}_{грm}$

Рассмотрим математическую модель, в которой предполагается, что каждая случайная величина x_{ij} может быть представлена в виде $x_{ij} = x_{грi} + \varepsilon_{ij}$, где согласно условию примера $x_{грi}$ – урожайность, вызванная применением технологии F_i, а ε_{ij} – независимые случайные величины, которые описывают суммарный вклад всех случайных факторов, влияющих на итоговую урожайность. Чаще всего полагают, что все ε_{ij} распределены нормально с нулевым математическими ожиданиями и с одинаковыми неизвестными дисперсиями σ^2.

Задача об исследовании влияния технологии обработки почвы на урожайность культуры на математическом языке означает, что по результатам эксперимента необходимо проверить справедливость статистической гипотезы $H_0 : x_{гр1} = x_{гр2} = ... = x_{грm}$, против альтернативной гипотезы H_1 о том, что хотя бы одно равенство не выполнено. То есть на некотором уровне значимости α требуется проверить нулевую гипотезу о равенстве групповых средних при допущении, что групповые генеральные дисперсии неизвестны, но одинаковы.

Проверка гипотезы основана на сопоставлении двух оценок неизвестной дисперсии . Обозначим

$$\overline{x}_{грi} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}\ ,\ i = \overline{1,m}\ , \text{ – групповые средние и}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} \text{ – общую выборочную среднюю.}$$

Несмещенной оценкой для неизвестной дисперсии σ^2 является, как известно, сумма квадратов

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\overline{x})^2 ,$$

деленная на $n-1$, где $n=\sum_{i=1}^{m} n_i$ – количество всех наблюдений (если на каждом уровне проведено одинаковое количество наблюдений $n_1=n_2=...=n_m=n'$, то $n=n'\cdot m$). Основная идея дисперсионного анализа заключается в разбиении этой суммы квадратов отклонений на несколько компонент, каждая из которых соответствует предполагаемой причине изменения средних значений $x_{грi}$.

Справедливо тождество:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\overline{x})^2=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\overline{x}_{грi}-\overline{x})^2+\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{грi})^2 ,$$

или

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\overline{x})^2=\sum_{i=1}^{m}n_i(\overline{x}_{грi}-\overline{x})^2+\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{грi})^2 .$$

В результате получим следующее тождество:

$$S_{общ}=S_{факт}+S_{ост} ,$$

где $S_{общ}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\overline{x})^2$ – ***общая*** или ***полная*** сумма квадратов отклонений;

$S_{факт}=\sum_{i=1}^{m}n_i(\overline{x}_{грi}-\overline{x})^2$ – сумма квадратов отклонений групповых средних от общей средней или ***межгрупповая*** (***факторная***) сумма квадратов отклонений;

$S_{ост}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{грi})^2$ – сумма квадратов отклонений наблюдений от групповых средних или ***внутригрупповая*** (***остаточная***) сумма квадратов отклонений.

В разложении $S_{общ}=S_{факт}+S_{ост}$ заключена основная идея дисперсионного анализа. Применительно к рассматриваемому примеру данное равенство показывает, что общая вариация показателя урожайности культуры, измеренная суммой $S_{общ}$, складывается из двух компонент – $S_{факт}$ и $S_{ост}$, характеризующий изменчивость показателя урожайности между технологиями ($S_{факт}$) и изменчивость «внутри» технологи ($S_{ост}$).

В дисперсионном анализе анализируются не сами суммы квадратов отклонений, а так называемые *средние квадраты*, являющиеся несмещенными оценками соответствующих дисперсий, которые получаются делением сумм квадратов отклонений на соответствующее число степеней свободы (см. «исправленная» выборочная дисперсия). Число степеней свободы определяется как общее число наблюдений минус число связывающих их уравнений. Поэтому несмещенной оценкой ***межгрупповой дисперсии*** является

$$s^2_{факт} = \frac{S_{факт}}{m-1},$$

так как при расчете $S_{факт}$ используются m групповых средних, связанных между собой одним уравнением. Несмещенной оценкой ***внутригрупповой дисперсии*** является

,

ибо при расчете $S_{ост}$ используются все n наблюдений, связанных

между собой уравнениями . В случае однофакторного комплекса $s^2_{факт}$ и $s^2_{ост}$ являются несмещенными и независимыми оценками дисперсии σ^2.

Сравним обе оценки $s^2_{факт}$ и $s^2_{ост}$. Если гипотеза H_0 верна, то ***дисперсионное отношение*** (статистика)

$$F = \frac{s^2_{факт}}{s^2_{ост}}$$

имеет распределение Фишера с $m-1$ и $n-m$ степенями свободы. Следовательно, проверка нулевой гипотезы свелась к проверке существенности различия несмещенных выборочных оценок и $s^2_{ост}$ дисперсии σ^2.

Гипотеза H_0 отвергается, если фактически вычисленное значение статистики F больше критического $F_{кр}(\alpha; m-1; n-m)$ и принимается, если $F < F_{кр}(\alpha; m-1; n-m)$.

Применительно к рассматриваемому примеру опровержение гипотезы H_0 означает наличие существенных различий в урожайности культуры в зависимости от выбранной технологии.

Если установлено, что фактор F влияет на результативный признак, то возникает вопрос о степени этого влияния. Для измерения степени влияния фактора на результативный признак используют ***выборочный коэффициент детерминации***, равный

$$R^2 = \frac{S_{факт}}{S_{общ}}.$$

Коэффициент детерминации показывает, какая часть общей дисперсии результативного признака объясняется зависимостью признака ξ от фактора F. Тогда $1 - R^2$ – доля дисперсии результативного признака, обусловленная случайными факторами. Очевидны следующие свойства коэффициента детерминации:

1) $0 \le R^2 \le 1$;

2) чем ближе значение R^2 к единице, тем больше степень влияния фактора F на признак ξ.

Замечание. Для вычисления сумм квадратов в случае $n_1 = n_2 = \ldots = n_m = n'$ часто бывает удобно использовать следующие формулы:

$$S_{факт} = \frac{\sum_{i=1}^{m}\left(\sum_{j=1}^{n'} x_{ij}\right)^2}{n'} - \frac{\left(\sum_{i=1}^{m}\sum_{j=1}^{n'} x_{ij}\right)^2}{n'm},$$

$$S_{ост} = \sum_{i=1}^{m}\sum_{j=1}^{n'} x_{ij}{}^2 - \frac{\sum_{i=1}^{m}\left(\sum_{j=1}^{n'} x_{ij}\right)^2}{n'},$$

$$S_{общ} = \sum_{i=1}^{m}\sum_{j=1}^{n'} x_{ij}{}^2 - \frac{\left(\sum_{i=1}^{m}\sum_{j=1}^{n'} x_{ij}\right)^2}{n'm}.$$

Пример 19.1. В таблице приведены данные об урожайности сельскохозяйственной культуры за 6 лет при различных технологиях обработки почвы.

Номер технологии	Годы					
	1-й	2-й	3-й	4-й	5-й	6-й
1	140	141	140	141	142	145
2	150	149	150	147		
3	147	147	145	150	150	
4	144	147	142	146		

Выяснить на уровне значимости $\alpha = 0{,}05$, зависит ли урожайность сельскохозяйственной культуры от технологии обработки почвы. Установите степень влияния технологии обработки почвы на урожайность.

Решение. Рассчитаем групповые средние и общую среднюю.

$$\bar{x} = \frac{140 + 141 + 140 + \ldots + 147 + 142 + 146}{19} = 145{,}4211 .$$

Вычислим суммы квадратов отклонений:

$$S_{факт} = \sum_{i=1}^{m} n_i (\bar{x}_{грi} - \bar{x})^2 = 6(141{,}5 - 145{,}4211)^2 + 4(149 - 145{,}4211)^2 +$$

$$+ 5(147{,}8 - 145{,}4211)^2 + 4(144{,}75 - 145{,}4211)^2 = 173{,}582;$$

$$S_{ост} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{грi})^2 = (140 - 141{,}5)^2 + \ldots + (145 - 141{,}5)^2 +$$

$$+ (150 - 149)^2 + \ldots + (147 - 149)^2 +$$

$$+ (147 - 147{,}8)^2 + \ldots + (150 - 147{,}8)^2 +$$

$$+ (144 - 144{,}75)^2 + \ldots + (146 - 144{,}75)^2 = 57{,}050;$$

$$S_{общ} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 = (140 - 145{,}4211)^2 + ... + (145 - 145{,}4211)^2 +$$

$$+ (150 - 145{,}4211)^2 + ... + (147 - 145{,}4211)^2 +$$

$$+ (147 - 145{,}4211)^2 + ... + (150 - 145{,}4211)^2 +$$

$$+ (144 - 145{,}4211)^2 + ... + (146 - 145{,}4211)^2 = 230{,}632.$$

Результаты сведем в таблицу

Компоненты дисперсии	Суммы квадра-тов	Число степеней свободы	Средние квадраты	F	R^2
Межгрупповая	173,582	$m-1$ =3	$s^2_{факт}$ =57,861	15,215	0,753
Внутригрупповая	57,050	$n-m$ =15	$s^2_{ост}$ =3,803		
Общая	230,632	$n-1$ =18			

Сделаем выводы. По таблице критических точек распределения Фишера (приложение 8) определим $F_{кр}(0{,}05; 3; 15) = 3{,}29$. Так как $F = 15{,}215 > F_{кр}(0{,}05; 3; 15) = 3{,}29$, то нулевая гипотеза отвергается, то есть на уровне значимости $0{,}05$ (с надежностью $0{,}95$, или $95\ \%$) выбор технологии существенно влияет на урожайность. Анализируя значение коэффициента детерминации, можем утверждать, что $75{,}3\ \%$ общего изменения урожайности обусловлены технологией и лишь $24{,}7\ \%$ другими случайными составляющими.

	'					
r	44	44	44*	44)	44>	443
1	>	)			3	?
C	>		?	)	)	?
2		*?		**	4	*
s	)	)	*A	>.	*4	

З а д а ч и

19. Для заданного уровня значимости $\alpha = 0{,}05$ установите влияние типа используемой рекламы на объем продаж товара. Определите степень влияния типа используемой рекламы на объем продаж товара.

19.1.

19.2.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	207,33	221,83	241,34	223,13	221,83	221,60
B	239,27	216,28	220,53	207,15	209,76	235,13
C	217,80	222,71	225,32	193,85	216,37	
D	256,51	243,92	234,56	244,94		

19.3.

19.4.

	Годы					
Тип рекламы	1993	1994	1995	1996	1997	1998
A	224 85	238 64	222 86	221 81	231 95	215 19
B	215 31	205 01	223 35	189 37	224 08	224 87
C	232 26	240 63	226 25	215 57	233 51	
D	247 09	238 20	238 91	235 53		

19.5.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	215,19	224,97	235,58	211,15	222,19	217,20
B	213,61	236,47	238,55	212,78	224,66	229,00
C	212,35	214,91	221,44	227,28	230,34	
D	215,42	208,40	227,59	229,52		

19.6.

	Годы					
Тип рекламы	1991	1992	1993	1994	1995	1996
A	221,82	233,26	225,66	225,59	228,08	213,07
B	205,38	226,84	218,96	216,52	223,42	234,16
C	248,91	221,06	220,78	251,27	219,37	
D	244,76	239,35	233,18	243,94		

19.7.

	Годы					
Тип рекламы	1990	1991	1992	1993	1994	1995
A	228,08	213,07	201,07	232,94	235,75	229,09
B	223,42	234,16	205,96	234,04	231,84	240,47
C	219,37	217,77	218,48	234,35	232,74	
D	227,00	225,29	243,25	240,81		

19.8.

	Годы					
Тип рекламы	1991	1992	1993	1994	1995	1996
A	235,75	229,09	215,30	220,50	248,46	218,27
B	231,84	240,47	225,63	214,86	229,73	243,61
C	232,74	225,62	252,79	244,56	234,89	
D	241,19	225,51	248,01	236,42		

19.9.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	220 15	235 60	234 69	213 86	242 61	229 80
B	201 31	226 23	208 01	212 02	219 59	221 27
C	218 54	219 72	236 49	230 72	214 55	
D	242 49	230 79	252 28	258 23		

19.10.

	Годы					
Тип рекламы	1993	1994	1995	1996	1997	1998
A	242,61	229,80	218,36	223,89	236,38	235,37
B	203,74	242,10	220,57	229,62	224,19	210,96
C	220,63	210,11	223,19	224,36	234,87	
D	234,83	243,85	244,86	244,71		

19.11.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	224,19	210,96	233,70	220,25	215,96	205,11
B	212,19	229,79	235,49	220,67	212,68	226,58
C	234,87	242,22	237,09	223,85	224,41	
D	243,65	233,64	244,29	249,91		

19.12.

	Годы					
Тип рекламы	1991	1992	1993	1994	1995	1996
A	232,63	220,86	216,70	216,35	204,55	220,24
B	205,38	207,78	228,74	232,09	219,36	228,82
C	237,68	227,69	224,84	233,95	218,23	
D	243,08	241,23	243,86	219,11		

19.13.

	Годы					
Тип рекламы	1990	1991	1992	1993	1994	1995
A	215,96	205,11	232,63	220,86	216,70	216,35
B	219,11	230,07	224,70	213,32	227,08	228,29
C	227,69	224,84	233,95	218,23	221,92	
D	241,23	243,86	219,11	240,53		

19.14.

	Годы					
Тип рекламы	1991	1992	1993	1994	1995	1996
A	230 07	224 70	213 32	227 08	228 29	219 54
B	229 79	235 49	220 67	212 68	226 58	215 63
C	242 22	237 09	223 85	224 41	234 01	
D	233 64	244 29	249 91	244 29		

19.15.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	204,55	220,24	205,38	207,78	228,74	232,09
B	225,89	221,60	231,94	219,11	230,07	224,70
C	234,59	233,97	237,68	227,69	224,84	
D	250,22	246,63	243,08	241,23		

19.16.

	Годы					
Тип рекламы	1993	1994	1995	1996	1997	1998
A	219,36	228,82	241,84	225,77	232,40	211,66
B	213,32	227,08	228,29	219,54	220,41	216,58
C	218,23	221,92	229,96	235,59	231,93	229,57
D	240,53	234,71	241,64	223,11	249,54	230,98

19.17.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	214,64	214,67	222,74	227,60	212,36	242,42
B	234,41	216,47	224,85	215,05	217,21	233,43
C	221,09	229,87	238,87	225,70	228,47	
D	243,94	239,63	244,09	265,25	234,05	

19.18.

	Годы					
Тип рекламы	1991	1992	1993	1994	1995	1996
A	242,42	217,98	212,19	229,33	227,39	214,93
B	225,77	232,40	211,66	214,64	214,67	222,74
C	217,21	233,43	216,77	235,69	219,77	
D	213,89	219,45	241,32	225,38		

19.19.

	Годы					
Тип рекламы	1990	1991	1992	1993	1994	1995
A	239,52	231,84	225,21	233,43	240,23	222,63
B	231,64	229,17	224,81	221,88	231,19	214,54
C	234,14	236,16	234,24	236,58		
D	241,26	247,46	239,55	240,08		

19.20.

	Годы					
Тип рекламы	1991	1992	1993	1994	1995	1996
A	240,23	222,63	208,66	219,18	232,91	220,15
B	231,19	214,54	221,37	225,85	237,58	230,13
C	237,31	246,79	229,22	227,59	230,99	
D	261,66	241,93	232,66	229,58		

19.21.

	Годы					
Тип рекламы	1993	1994	1995	1996	1997	1998
A	22182	24133	22313	22183	22159	21339
B	22270	22532	19384	21637	22237	22122
C	22407	24170	23816	22347	23327	
D	24392	23456	24494	22861		

19.22.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	22159	21339	23927	21627	22052	20714
B	20975	23513	23235	23083	22514	21619
C	22638	23347	23207	24846	22951	
D	22638	23347	23207	24846	22951	

19.23.

	Годы					
Тип рекламы	1991	1992	1993	1994	1995	1996
A	22514 21	21619 72	24016 72	22233 65	22485 26	23863 57
B	22225 98	22039 69	23519 86	21861 24	21531 15	20501 24
C	22951 13	21423 50	21534 19	23554 48	23225 51	
D	23190 09	23882 36	24359 34	25609 16		

19.24.

	Годы					
Тип рекламы	1992	1993	1994	1995	1996	1997
A	21589,78	22462,79	22224,45	22139,56	20644,05	20732,94
B	21529,21	22217,05	20743,64	21484,79	22427,30	21780,19
C	22235,77	23741,13	22139,17	23331,27	22952,21	
D	24277,80	24276,62	23834,02	25067,19		

19.25.

	Годы					
Тип рекламы	1990	1991	1992	1993	1994	1996
A	10759	11248	11778	10557	11109	10860
B	11243	11283	10617	10889	10416	10363
C	11904	11315	11852	11268	11394	11080
D	11336	12584	12582	12241	11900	12930

П Р И Л О Ж Е Н И Я

ПРИЛОЖЕНИЕ 1

ТАБЛИЦА ЗНАЧЕНИЙ ФУНКЦИИ

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

ТАБЛИЦА ЗНАЧЕНИЙ ФУНКЦИИ $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz$

x	Φ*(x)*	*x*	Φ*(x)*	*x*	Φ*(x)*	*x*	Φ*(x)*
0,00	0,0000	0,32	0,1255	0,64	0,2389	0,96	0,3315
0,01	0,0040	0,33	0,1293	0,65	0,2422	0,97	0,3340
0,02	0,0080	0,34	0,1331	0,66	0,2454	0,98	0,3365
0,03	0,0120	0,35	0,1368	0,67	0,2486	0,99	0,3389
0,04	0,0160	0,36	0,1406	0,68	0,2517	1,00	0,3413
0,05	0,0199	0,37	0,1443	0,69	0,2549	1,01	0,3438
0,06	0,0239	0,38	0,1480	0,70	0,2580	1,02	0,3461
0,07	0,0279	0,39	0,1517	0,71	0,2611	1,03	0,3485
0,08	0,0319	0,40	0,1554	0,72	0,2642	1,04	0,3508
0,09	0,0359	0,41	0,1591	0,73	0,2673	1,05	0,3531
0,10	0,0398	0,42	0,1628	0,74	0,2703	1,06	0,3554
0,11	0,0438	0,43	0,1664	0,75	0,2734	1,07	0,3577
0,12	0,0478	0,44	0,1700	0,76	0,2764	1,08	0,3599
0,13	0,0517	0,45	0,1736	0,77	0,2794	1,09	0,3621
0,14	0,0557	0,46	0,1772	0,78	0,2823	1,10	0,3643
0,15	0,0596	0,47	,0,1808	0,79	0,2852	1,11	0,3665
0,16	0,0636	0,48	0,1844	0,80	0,2881	1,12	0,3686
0,17	0,0675	0,49	0,1879	0,81	0,2910	1,13	0,3708
0,18	0,0714	0,50	0,1915	0,82	0,2939	1,14	0,3729
0,19	0,0753	0,51	0,1950	0,83	0,2967	1,15	0,3749
0,20	0,0793	0,52	0,1985	0,84	0,2995	1,16	0,3770
0,21	0,0832	0,53	0,2019	0,85	0,3023	1,17	0,3790
0,22	0,0871	0,54	0,2054	0,86	0,3051	1,18	0,3810
0,23	0,0910	0,55	0,2088	0,87	0,3078	1,19	0,3830
0,24	0,0948	0,56	0,2123	0,88	0,3106	1,20	0,3849
0,25	0,0987	0,57	0,2157	0,89	0,3133	1,21	0,3869
0,26	0,1026	0,58	0,2190	0,90	0,3159	1,22	0,3883
0,27	0,1064	0,59	0,2224	0,91	0,3186	1,23	0,3907
0,28	0,1103	0,60	0,2257	0,92	0,3212	1,24	0,3925
0,29	0,1141	0,61	0,2291	0,93	0,3238	1,25	0,3944
0,30	0,1179	0,62	0,2324	0,94	0,3264		
0,31	0,1217	0,63	0,2357	0,95	0,3289		

	Φ(*x*)	*x*	Φ(*x*)	*x*	Φ(*x*)	*x*	Φ(*x*)
1,26	0,3962	1,59	0,4441	1,92	0,4726	2,50	0,4938
1,27	0,3980	1,60	0,4452	1,93	0,4732	2,52	0,4941
1,28	0,3997	1,61	0,4463	1,94	0,4738	2,54	0,4945
1,29	0,4015	1,62	0,4474	1,95	0,4744	2,56	0,4948
1,30	0,4032	1,63	0,4484	1,96	0,4750	2,58	0,4951
1,31	0,4049	1,64	0,4495	1,97	0,4756	2,60	0,4953
1,32	0,4066	1,65	0,4505	1,98	0,4761	2,62	0,4956
1,33	0,4082	1,66	0,4515	1,99	0,4767	2,64	0,4959
1,34	0,4099	1,67	0,4525	2,00	0,4772	2,66	0,4961
1,35	0,4115	1,68	0,4535	2,02	0,4783	2,68	0,4963
1,36	0,4131	1,69	0,4545	2,04	0,4793	2,70	0,4965
1,37	0,4147	1,70	0,4554	2,06	0,4803	2,72	0,4967
1,38	0,4162	1,71	0,4564	2,08	0,4812	2,74	0,4969
1,39	0,4177	1,72	0,4573	2,10	0,4821	2,76	0,4971
1,40	0,4192	1,73	0,4582	2,12	0,4830	2,78	0,4973
1,41	0,4207	1,74	0,4591	2,14	0,4838	2,80	0,4974
1,42	0,4222	1,75	0,4599	2,16	0,4846	2,82	0,4976
1,43	0,4236	1,76	0,4608	2,18	0,4854	2,84	0,4977
1,44	0,4251	1,77	0,4616	2,20	0,4861	2,86	0,4979
1,45	0,4265	1,78	0,4625	2,22	0,4868	2,88	0,4980
1,46	0,4279	1,79	0,4633	2,24	0,4875	2,90	0,4981
1,47	0,4292	1,80	0,4641	2,26	0,4881	2,92	0,4982
1,48	0,4306	1,81	0,4649	2,28	0,4887	2,94	0,4984
1,49	0,4319	1,82	0,4656	2,30	0,4893	2,96	0,4985
1,50	0,4332	1,83	0,4664	2,32	0,4898	2,98	0,4986
1,51	0,4345	1,84	0,4671	2,34	0,4904	3,00	0,49865
1,52	0,4357	1,85	0,4678	2,36,	0,4909	3,20	0,49931
1,53	0,4370	1,86	0,4686	2,38	0,4913	,3,40	0,49966
1,54	0,4382	1,87	0,4693	2,40	0,4918	3,60	0,499841
1,55	0,4394	1,88	0,4699	2,42	0,4922	3,80	0,499928
1,56	0,4406	1,89	0,4706	2,44	0,4927	4,00	0,499968
1,57	0,4418	1,90	0,4713	2,46	0,4931	4,50	0,499997
1,58	0,4429	1,91	0,4719	2,48	0,4934	5,00	0,499997

РАСПРЕДЕЛЕНИЕ ПУАССОНА $P(\xi = k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$

k	0	1	2	3	4	5	6	7	8	9	10	11
0,1	0,9048	0,0905	0,0045	0,0002	0,0000							
0,2	0,8187	0,1637	0,0164	0,0011	0,0001	0,0000						
0,3	0,7408	0,2222	0,0333	0,0033	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,4	0,6703	0,2681	0,0536	0,0072	0,0007	0,0001	0,0000	0,0000	0,0001	0,0002	0,0002	0,0002
0,5	0,6065	0,3033	0,0758	0,0126	0,0016	0,0002	0,0000	0,0000	0,0009	0,0009	0,0008	0,0007
0,6	0,5488	0,3293	0,0988	0,0198	0,0030	0,0004	0,0001	0,0001	0,0031	0,0027	0,0023	0,0019
0,7	0,4966	0,3476	0,1217	0,0284	0,0050	0,0007	0,0002	0,0008	0,0081	0,0066	0,0053	0,0043
0,8	0,4493	0,3595	0,1438	0,0383	0,0077	0,0012	0,0003	0,0034	0,0169	0,0132	0,0104	0,0082
0,9	0,4066	0,3659	0,1647	0,0494	0,0111	0,0020	0,0005	0,0099	0,0298	0,0232	0,0181	0,0225
1,0	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031	0,0035	0,0216	0,0463	0,0363	0,0413	0,0452
1,5	0,2231	0,3347	0,2510	0,1255	0,0471	0,0141	0,0120	0,0385	0,0653	0,0688	0,0710	0,0722
2,0	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0278	0,0595	0,1033	0,1014	0,0993	0,0970
2,5	0,0821	0,2052	0,2565	0,2138	0,1336	0,0668	0,0504	0,0824	0,1304	0,1241	0,1186	0,1137
3,0	0,0498	0,1494	0,2240	0,2240	0,1680	0,1008	0,0771	0,1044	0,1396	0,1318	0,1251	
3,5	0,0302	0,1057	0,1850	0,2158	0,1888	0,1322	0,1042	0,1377	0,1318	0,1251		
4,0	0,0183	0,0733	0,1465	0,1954	0,1954	0,1563	0,1281	0,1490	0,1126			
4,5	0,0111	0,0500	0,1125	0,1687	0,1898	0,1708	0,1462	0,1396				
5,0	0,0067	0,0337	0,0842	0,1404	0,1755	0,1755	0,1606	0,1171				
6,0	0,0025	0,0149	0,0446	0,0892	0,1339	0,1606	0,1490	0,0901				
7,0	0,0009	0,0064	0,0223	0,0521	0,0912	0,1277	0,1221					
8,0	0,0003	0,0027	0,0107	0,0286	0,0573	0,0916	0,0911					
9,0	0,0001	0,0011	0,0050	0,0150	0,0337	0,0607	0,0631					
10,0	0,0000	0,0005	0,0023	0,0076	0,0189	0,0378						

k	12	13	14	15	16	17	18	19	20	21	22	23	24
3,0	0,0001	0,0000											
3,5	0,0002	0,0001	0,0000										
4,0	0,0006	0,0002	0,0001	0,0000	0,0000	0,0001	0,0000	0,0001	0,0000	0,0001	0,0000	0,0000	0,0001
4,5	0,0016	0,0006	0,0002	0,0001	0,0000	0,0006	0,0002	0,0004	0,0002	0,0003	0,0001	0,0002	
5,0	0,0034	0,0013	0,0005	0,0002	0,0003	0,0021	0,0009	0,0014	0,0006	0,0009	0,0004		
6,0	0,0113	0,0052	0,0022	0,0009	0,0014	0,0058	0,0029	0,0037	0,0019				
7,0	0,0263	0,0142	0,0071	0,0033	0,0045	0,0128	0,0071						
8,0	0,0481	0,0296	0,0169	0,0090	0,0109								
9,0	0,0728	0,0504	0,0324	0,0194	0,0217								
10,0	0,0948	0,0729	0,0521	0,0347									

ПРИЛОЖЕНИЕ 4

ТАБЛИЦА ЗНАЧЕНИЙ

$t_{\gamma} = t(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

ПРИЛОЖЕНИЕ 5

ТАБЛИЦА ЗНАЧЕНИЙ $q_{\gamma} = q(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,95
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

ПРИЛОЖЕНИЕ 8

КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ ФИШЕРА – СНЕДЕКОРА

k_1 – число степеней свободы большей дисперсии
k_2 – число степеней свободы меньшей дисперсии

Уровень значимости α =0,01												
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,50	4,32	4,14	4,00	3,89	3,80	3,73	3,67
10	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости α =0,05												
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Рекомендуемая литература

Учебники и учебные пособия

1. Боровков, А.А. Теория вероятностей / А.А. Боровков. – М: Наука, 1986. – 428 с.
2. Боровков, А.А. Математическая статистика / А.А. Бровков. – М.: Наука, 1984. – 472 с.
3. Гихман, Н.Н. Теория вероятностей и математическая статистика / Н.Н. Гихман, А.В. Скороход , М.Н. Ядренко. – Киев: Высшая школа, 1979. – 408 с.
4. Гнеденко, Б.В. Курс теории вероятностей / Б.В. Гнеденко. – М: Наука, 1988. – 447 с.
5. Коваленко, И.Н. Теория вероятностей и математическая статистика / И.Н. Коваленко, А.А. Филиппова. – М: Высшая школа, 1982. – 256с.
6. Кремер, Н.Ш. Теория вероятностей и математическая статистика / Н.Ш. Кремер. – М.: Юнити-Дана, 2000. – 543 с.
7. Маталыцкий, М.А. Вероятность и случайные процессы: теория, примеры, задачи / М.А. Маталыцкий. – Гродно: ГрГУ, 2006. – 583 с.
8. Мацкевич, И.П. Высшая математика. Теория вероятностей и математическая статистика / И.П. Мацкевич, Г.П. Свирид. – Мн.: Высшая школа, 1993. – 272 с.
9. Пугачев, В.С. Теория вероятностей и математическая статистика / В.С. Пугачев. – М: Наука, 1973. – 496 с.
10. Севастьянов, Б.А. Курс теории вероятностей и математической статистики / Б.А. Севастьянов. – М: Наука, 1982. – 252 с.
11. Четыркин, Е.М. Вероятность и статистика / Е.М. Четыркин, И.Я. Калихман. – М.: Финансы и статистика, 1982. – 319 с.
12. Чистяков, В.П. Курс теории вероятностей / В.П. Чистяков. – М: Наука, 1987. – 240 с.

Сборники задач

1. Агапов, Г.И. Задачник по теории вероятностей / Г.И. Агапов. – М: Высшая школа, 1986. – 80 с.
2. Белько, И.В. Теория вероятностей и математическая статистика. Примеры и задачи / И.В. Белько, Г.П. Свирид. – Мн.: Новое знание. 2002. – 250 с.
3. Вентцель, Е.С. Прикладные задачи теории вероятностей / Е.С. Вентцель, Л.А. Овчаров. – М: Радио и связь, 1983. – 416 с.
4. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. – М.: Высш. шк., 2000. – 400 с.
5. Емельянов, Г.В. Задачник по теории вероятностей и математической статистике / Г.В. Емельянов, В.П. Скитович. – Ленинград: ЛГУ, 1967. – 331 с.
6. Зубков, А.М. Сборник задач по теории вероятностей / А.М. Зубков, Б.А. Севастьянов, В.П. Чистяков. – М: Наука, 1989. – 320 с.
7. Лаппо, П.М. Задачы па теорыі імавернасцей / П.М. Лаппо, М.А. Маталыцкі. – Мінск: Універсітэцкае, 1995. – 87 с.
8. Маталыцкий, М.А. Теория вероятностей в примерах и задачах / М.А. Маталыцкий, Т.В. Романюк. – Гродно: ГрГУ, 2002. – 247 с.
9. Мацкевич, И.П. Сборник зада и упражнений по высшей математике: Теории вероятностей и математическая статистика / И.П. Мацкевич, Г.П. Свирид, Г.М. Булдык. – Мн.: Высш. шк., 1996. – 318 с.
10. Сборник задач по теории вероятностей, математической статистике и теории случайных функций / под ред. А.А. Свещникова. – М: Наука, 1965. – 632 с.
11. Рябушко, А.П. Сборник индивидуальных заданий по теории вероятностей и математической статистике / А.П. Рябушко, В.В. Бархотов, В.В. Државец, И.Е. Юруть. – Мн.: Высш. шк., 1992. – 191 с.
12. Харин, Ю.С. Сборник задач по теории вероятностей, случайных процессов и математической статистике / Ю.С. Харин, Г.А. Хацкевич, В.И. Лобач. – Минск: БГУ, 1995. – 99 с.

Оглавление

Учебное издание

Маталыцкий Михаил Алексеевич
Русилко Татьяна Владимировна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебно-методическое пособие

Редактор *Н.Н. Красницкая*
Компьютерная верстка: *Р.Н. Баранчик*
Дизайн обложки: *О.В. Канчуга*

Сдано в набор 19.06.2007. Подписано в печать 20.11.2007.
Формат 60×84/16. Бумага офсетная.
Печать RISO. Гарнитура Таймс .
Усл. печ. л. 12,79. Уч.-изд. л. 11,66. Тираж экз. Заказ .

Учреждение образования «Гродненский государственный
университет имени Янки Купалы».
ЛИ № 02330/0133257 от 30.04.2004. Пер. Телеграфный, 15а, 230023, Гродно.

Отпечатано на технике издательского центра
Учреждения образования «Гродненский государственный
университет имени Янки Купалы».
ЛП № 02330/0056882 от 30.04.2004. Пер. Телеграфный, 15а, 230023, Гродно.