Министерство образования Республики Беларусь
Учреждение образования «Витебский государственный
университет имени П.М. Машерова»
Кафедра геометрии и математического анализа

С.М. Бородич, Т.В. Кавитова

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации

Витебск
ВГУ имени П.М. Машерова
2017

УДК 519.2(075.8)
ББК 22.17я73
Б83

Печатается по решению научно-методического совета учреждения образования «Витебский государственный университет имени П.М. Машерова». Протокол № 3 от 28.02.2017 г.

Авторы: старшие преподаватели кафедры геометрии и математического анализа ВГУ имени П.М. Машерова **С.М. Бородич, Т.В. Кавитова**

Научный редактор:
доцент кафедры геометрии и математического анализа
ВГУ имени П.М. Машерова, кандидат физико-математических наук
Т.Л. Сурин

Рецензент:
доцент кафедры математики и информационных технологий УО «ВГТУ»,
кандидат физико-математических наук *Т.В. Никонова*

Бородич, С.М.

Б83 Теория вероятностей и математическая статистика : методические рекомендации / С.М. Бородич, Т.В. Кавитова. – Витебск : ВГУ имени П.М. Машерова, 2017. – 52 с.

Настоящее учебное издание может использоваться для организации самостоятельной работы студентов, а также при проведении практических занятий. По каждой теме приводится краткий теоретический материал и типовые задачи.

Предназначается для студентов, обучающихся по специальности «Программное обеспечение информационных технологий» на дневной и заочной формах обучения.

УДК 519.2(075.8)
ББК 22.17я73

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Курс «Теория вероятностей и математическая статистика» является составной частью цикла математических дисциплин, из которых складывается фундамент математического образования современного специалиста.

Издание призвано помочь студентам овладеть основами теории вероятностей и математической статистики в такой степени, чтобы они могли не только осознанно применять полученные знания в процессе обучения и работы, но и, по мере необходимости, углублять и расширять их путём дальнейшего самообразования.

Учебное издание состоит из четырех разделов: «Случайные события и их вероятности», «Случайные величины», «Предельные теоремы теории вероятностей», «Элементы математической статистики». В каждом пункте по всем изучаемым разделам приводится краткий теоретический материал и типовые задачи, предназначенные для самостоятельной работы.

В конце издания помещён список литературы, изучение которой поможет студентам подробнее разобраться в отдельных вопросах курса.

Настоящее учебное издание адресуется прежде всего студентам факультета математики и информационных технологий, обучающимся по специальности «Программное обеспечение информационных технологий» на дневной и заочной формах обучения.

Часть I. СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

1.1. Случайные события. Пространство элементарных событий. Алгебраические операции над событиями

Случайным событием (или просто: *событием*) называется любой факт, который может произойти или не произойти в результате некоторого испытания (опыта, эксперимента, наблюдения).

Событие называется *достоверным*, если оно обязательно наступит в данном испытании, и *невозможным*, если оно заведомо в данном испытании не произойдёт.

Достоверное событие будем обозначать буквой Ω, а невозможное – символом $\varnothing$.

Простейшие события, являющиеся непосредственными (теоретически возможными) исходами данного эксперимента и обладающие тем свойством, что в результате эксперимента происходит одно и только одно из них, называются *элементарными событиями (элементарными исходами)*.

Для любого события A в данном эксперименте можно выделить совокупность элементарных событий, при наступлении которых происходит это событие. Такие элементарные события называются *благоприятствующими событию A*. Таким образом, любое событие, связанное с данным экспериментом, можно представить как совокупность всех благоприятствующих этому событию элементарных исходов.

Очевидно, что достоверное событие представляется совокупностью всех элементарных событий.

Множество всех элементарных событий данного эксперимента называется *пространством элементарных событий* и обозначается, как и достоверное событие, буквой Ω.

Два события называются *несовместными*, если появление одного из них исключает появление другого в одном и том же испытании. В противном случае события называются *совместными*.

События $A_1, A_2, A_3, \ldots, A_n$ называются *попарно несовместными* (по-другому: *взаимоисключающими*), если любые два из них несовместны.

Говорят, что событие A *влечёт* событие B, и пишут $A \subset B$ или $B \supset A$, если при наступлении события A обязательно происходит и событие B.

Если $A \subset B$ и $B \subset A$, то события A и B называют *эквивалентными (равносильными, равными)*; записывают: $A = B$.

Суммой событий $A_1, A_2, \ldots, A_n$ называется событие A, которое состоит в появлении хотя бы одного из этих событий. Записывают:

$$A = A_1 + A_2 + \ldots + A_n \text{ или } A = \sum_{i=1}^{n} A_i .$$

Произведением событий $A_1, A_2, \ldots, A_n$ называется событие A, состоящее в появлении каждого из этих событий. Пишут:

$$A = A_1 A_2 \ldots A_n \quad \text{или} \quad A = \prod_{i=1}^{n} A_i .$$

Разностью двух событий A и B называется событие C, состоящее в том, что событие A произошло, а событие B не произошло. Записывают:

$$C = A - B.$$

Событие $\overline{A} = \Omega - A$ называется *дополнительным (противоположным)* к событию A. Оно заключается в непоявлении события A. Очевидно, что $\overline{\Omega} = \varnothing$, $\overline{\varnothing} = \Omega$, $\overline{\overline{A}} = A$.

Задачи

1. Игральная кость (кубик) подбрасывается два раза. Рассматриваются события:

A – оба раза выпало число очков, кратное 3;

B – сумма выпавших очков кратна 5;

C – оба раза выпало число очков, меньшее 3;

D – выпало одинаковое число очков.

Описать пространство элементарных событий Ω. Описать события A, B, C, D. Среди этих событий найти пары несовместных.

2. Бросают три монеты. События: A – только на одной монете из трёх появился герб, B – хотя бы на одной монете появился герб. Что означают события: а) $A+B$; б) AB; в) $\overline{A}$; г) $\overline{B}$; д) $B-A$?

3. Два игрока играют в шахматы. Событие A – выиграл первый игрок, событие B – выиграл второй игрок. Что означают события

а) $\overline{A}\,\overline{B}$; б) $\overline{A}B$; в) $\overline{B}-A$; г) $\overline{A}-\overline{B}$?

4. Из урны, в которой находится 2 белых и 3 чёрных шара, наудачу извлекают 3 шара. Рассматриваются события:

A – хотя бы один из вынутых шаров – чёрный;

B – вынули не менее двух чёрных шаров;

C – вынули не более одного белого шара;

D – хотя бы один из вынутых шаров – белый.

Описать пространство элементарных событий Ω. Описать события A, B, C, D, DC, $D-C$, $\overline{C}$, $\overline{B}C$.

5. Прибор состоит из двух блоков первого типа и трёх блоков второго типа. Событие A_i – исправлен i-й блок первого типа $(i = 1, 2)$, событие B_k – исправлен k-й блок второго типа $(k = 1, 2, 3)$. Прибор работает, если исправлен хотя бы один блок первого типа и не менее двух блоков второго типа. Событие C – прибор работает. Выразить событие C через события A_i и B_k.

6. Электрическая цепь составлена по схеме, приведённой на рисунке 1. Событие A_i – элемент с номером i вышел из строя $(i = 1, 2, 3, 4)$. Написать выражения для событий A и $\overline{A}$, если A означает разрыв цепи.

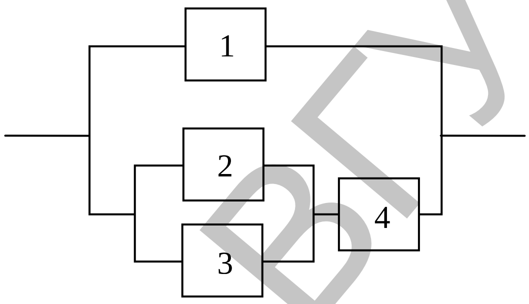

Рис. 1

1.2. Вероятность события. Классическое и геометрическое определения вероятности

Вероятность события – это числовая характеристика степени возможности его появления в рассматриваемом опыте. Вероятность события A обозначается символом $P(A)$.

Рассмотрим так называемую *классическую вероятностную модель*, которая используется для описания опытов с конечным числом равновозможных элементарных исходов.

Несколько событий в данном опыте называются *равновозможными*, если ни одно из них не является объективно более возможным, чем другие.

Пусть пространство элементарных событий эксперимента состоит из конечного числа элементарных событий, причём все они равновозможны. Пусть A – произвольное событие, связанное с данным экспериментом. *Вероятность (классическая вероятность) события A* определяется формулой

$$P(A) = \frac{|A|}{|\Omega|}, \tag{1.1}$$

где $|\Omega|$ – число всех элементарных событий, $|A|$ – число элементарных событий, благоприятствующих событию A.

Из определения (1.1) вытекают следующие свойства вероятности:

1) $0 \le P(A) \le 1$; 2) $P(\Omega) = 1$; 3) $P(\varnothing) = 0$;

4) Если A и B – несовместные события (т.е. $AB = \varnothing$), то

$$P(A+B)=P(A)+P(B);$$

5) $P(A)+P(\overline{A})=1$.

Геометрическое определение вероятности является обобщением классического определения на случай, когда пространство элементарных событий Ω не является конечным множеством.

Рассматривается некоторый случайный эксперимент с бесконечным числом равновозможных элементарных исходов. Допустим, что каждому элементарному исходу ставится в соответствие некоторая точка на прямой, плоскости или в пространстве, причём разным исходам соответствуют разные точки. Пусть Ω – множество всех таких точек. Таким образом, рассматриваемый эксперимент можно интерпретировать как случайный выбор точки $X\in\Omega$, а множество Ω – как пространство его элементарных исходов. Случайным событиям данного эксперимента соответствуют различные подмножества множества Ω, при этом подмножество A интерпретируется как случайное событие, заключающееся в том, что $X\in A$.

Будем предполагать, что и для самого множества Ω, и для всех рассматриваемых его подмножеств определена мера μ (длина, площадь, объём), причём предполагаем, что мера $\mu(\Omega)$ конечна.

Геометрической вероятностью события A называется число

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

Геометрическая вероятность обладает всеми приведёнными выше свойствами классической вероятности.

Пример 1.1. На линии связи длиной 10 км произошёл разрыв. Найти вероятность того, что разрыв произошёл не далее, чем в 2-х км от начала линии. Предполагается равная возможность разрыва в любых точках линии связи.

Решение. $\Omega=[0,10]$, $A=[0,2]$. По формуле геометрической вероятности находим $P(A)=\dfrac{2}{10}=0{,}2$. ●

Задачи

1. В цехе работают шесть мужчин и четыре женщины. По табельным номерам наудачу отобраны семь человек. Найти вероятность того, что среди отобранных лиц окажутся три женщины.

2. В лифт семиэтажного дома на первом этаже вошли 3 человека. Каждый из них с одинаковой вероятностью выходит на любом из этажей, начиная со второго. Найти вероятности следующих событий:

A – все пассажиры выйдут на четвертом этаже;

B – все пассажиры выйдут на одном и том же этаже;

C – все пассажиры выйдут на разных этажах.

3. Восемь человек садятся за круглый стол в произвольном порядке. Какова вероятность того, что два определённых лица будут сидеть рядом?

4. Пять мужчин и десять женщин случайным образом по трое рассаживаются за 5 столиков. Какова вероятность того, что за каждым столиком окажется мужчина?

5. Из промежутка $[0,2]$ наугад выбирается два числа. Какова вероятность, что их произведение больше 2?

6. Два парохода должны подойти к одному и тому же причалу. Время прихода обоих пароходов независимо и равновозможно в течение данных суток. Определить вероятность того, что одному из пароходов придётся ожидать освобождения причала, если время стоянки одного парохода – 1 час, а другого – 2 часа.

1.3. Условная вероятность. Формула умножения вероятностей. Независимые события

Пусть A и B – два события, причём $P(B) \neq 0$.

Условной вероятностью события A при условии, что произошло событие B, называется величина

$$P(A \mid B) = \frac{P(AB)}{P(B)}. \qquad (1.2)$$

Из формулы (1.2) следует равенство

$$P(AB) = P(B)P(A \mid B), \qquad (1.3)$$

называемое *формулой (теоремой) умножения вероятностей*.

Формула (1.3) обобщается на случай n событий:

Теорема 1.1. Пусть события $A_1, A_2, \ldots, A_n$ таковы, что $P(A_1 A_2 \ldots A_{n-1}) \neq 0$. Тогда

$$P(A_1 A_2 \ldots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 A_2)\ldots P(A_n \mid A_1 A_2 \ldots A_{n-1}).$$

События A и B, вероятности которых $P(A) \neq 0$ и $P(B) \neq 0$, называются *независимыми*, если

$$P(A \mid B) = P(A) \ \text{ и } \ P(B \mid A) = P(B). \qquad (1.4)$$

Замечание 1.1. Легко доказать, что если выполняется одно из равенств (1.4), то выполняется и другое. ●

Теорема 1.2. События A и B, имеющие ненулевую вероятность, являются независимыми тогда и только тогда, когда

$$P(AB) = P(A)P(B). \qquad (1.5)$$

С учётом теоремы 1.2 можно дать также следующее определение независимости событий, распространяющееся и на события с нулевой вероятностью.

События A и B называются *независимыми*, если для них выполнено равенство (1.5).

События $A_1, \ldots, A_n$ называют *независимыми в совокупности* (или просто: *независимыми*), если независимы каждые два из них и независимы каждое событие и всевозможные произведения остальных.

Формула (1.5) распространяется на n событий, независимых в совокупности:

$$P(A_1 A_2 \ldots A_n) = P(A_1) P(A_2) \ldots P(A_n).$$

Несколько событий называют *попарно независимыми*, если каждые два из них независимы.

Из независимости событий в совокупности следует их попарная независимость. Обратное, вообще говоря, не верно.

Задачи

1. Одновременно бросают две игральных кости – белую и чёрную. Рассматриваются события: A – на белой кости выпало более двух очков; B – в сумме выпало более двух очков; C – в сумме выпало менее десяти очков. Вычислить условные вероятности $P(A|B)$, $P(B|A)$, $P(A|C)$, $P(C|A)$, $P(B|C)$, $P(C|B)$. Являются ли независимыми события A и B, A и C, B и C? Являются ли независимыми в совокупности события A, B и C?

2. Из колоды в 36 карт наугад берут одну. Рассматриваются события: A – вынут туз; B – вынута пиковая карта. Являются ли события A и B независимыми?

3. В первом ящике 2 белых и 10 чёрных шаров, во втором – 3 белых и 9 чёрных шаров, в третьем – 6 белых и 6 чёрных шаров. Из каждого ящика вынули по шару. Найти вероятность того, что все вынутые шары белые.

4. Каждая буква слова «МАТЕМАТИКА» написана на отдельной карточке. Карточки перемешали и из них наугад последовательно извлекли и выложили слева направо четыре карточки. Найти вероятность того, что получилось слово «ТЕМА».

5. Имеется коробка с девятью новыми теннисными мячами. Для игры берут три мяча; после игры их кладут обратно. При выборе мячей играные от неиграных не отличают. Какова вероятность того, что после трёх игр в коробке не останется неиграных мячей?

6. Из урны, содержащей 6 белых и 4 чёрных шара, наудачу и последовательно извлекают по одному шару до появления чёрного шара. Найти вероятность того, что придётся производить четвёртое извлечение, если выборка производится: а) с возвращением; б) без возвращения.

1.4. Формулы сложения вероятностей

Как известно (см. пункт 1.2), вероятность суммы двух несовместных событий *A* и *B* равна сумме вероятностей этих событий:

$$P(A+B)=P(A)+P(B). \tag{1.6}$$

Эта формула легко обобщается на случай *n* попарно несовместных событий $A_1,\ \ldots,\ A_n$:

$$P(A_1+\ldots+A_n)=P(A_1)+\ldots+P(A_n). \tag{1.7}$$

Теорема 1.3. Для любых событий *A* и *B* имеет место формула

$$P(A+B)=P(A)+P(B)-P(AB). \tag{1.8}$$

Теорема 1.4. Для любых событий $A_1, A_2, \ldots,\ A_n$ имеет место формула

$$P(A_1+A_2+\ldots+A_n)=\sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_iA_j)+$$
$$+\sum_{i<j<k}P(A_iA_jA_k)-\ldots+(-1)^{n-1}P(A_1A_2\ldots A_n). \tag{1.9}$$

Формулы (1.6)-(1.9) называют *формулами сложения вероятностей*.

Задачи

1. В первой урне находится 2 белых и 3 чёрных шара, а во второй – 4 белых и 2 чёрных. Из каждой урны наугад извлекается по одному шару. Найти вероятность того, что вынутые шары будут одного и того же цвета.

2. Три орудия ведут огонь по цели, вероятность попадания в которую при одном выстреле из первого орудия равна $0{,}5$, из второго – $0{,}6$ и из третьего – $0{,}7$. Зная, что каждое орудие стреляет один раз, найти вероятность поражения цели, если для этого достаточно двух попаданий.

3. Два игрока поочерёдно бросают игральную кость. Выигрывает тот, у кого первым выпадет 6 очков. Определить вероятность выигрыша для каждого из игроков.

4. Два стрелка делают по одному выстрелу по одной и той же мишени, а затем каждый из стрелков стреляет ещё раз, если при первом сделанном им выстреле он промахнулся. Вероятность попадания в мишень при каждом выстреле для первого стрелка равна $0{,}7$, а для второго – $0{,}8$. Найти вероятность того, что мишень будет поражена дважды.

5. Рабочий обслуживает четыре станка, работающих независимо друг от друга. Вероятность того, что в течение часа первый станок не потребует внимания рабочего, равна $0{,}92$, для второго – такая вероятность равна $0{,}9$, для третьего – $0{,}85$ и для четвёртого – $0{,}8$. Какова вероятность того, что в течение часа хотя бы один станок не потребует внимания рабочего?

1.5. Формула полной вероятности. Формула Байеса

Говорят, что события $H_1, H_2, \ldots, H_n$ образуют *полную группу*, если они попарно несовместны и $\sum\limits_{i=1}^{n} H_i = \Omega$.

События $H_1, H_2, \ldots, H_n$, образующие полную группу, называются *гипотезами*.

Теорема 1.5. Пусть события $H_1, H_2, \ldots, H_n$ образуют полную группу. Тогда для любого события A имеет место формула

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

называемая *формулой полной вероятности*.

Теорема 1.6. Пусть события $H_1, H_2, \ldots, H_n$ образуют полную группу. Тогда условная вероятность события H_k $(k = 1, 2, \ldots, n)$ при условии, что в результате эксперимента произошло событие A, может быть вычислена по формуле

$$P(H_k \mid A) = \frac{P(H_k) P(A \mid H_k)}{\sum\limits_{i=1}^{n} P(H_i) P(A \mid H_i)}.$$

Эта формула называется *формулой Байеса* (*формулой гипотез*).

Пример 1.2. В первой коробке содержится 20 ламп, из них 18 стандартных; во второй коробке – 10 ламп, из них 9 стандартных. Из второй коробки наудачу взята лампа и переложена в первую. Затем из первой коробки наудачу извлекли одну лампу. Она оказалась стандартной. Какова вероятность того, что из второй коробки в первую переложили нестандартную лампу?

Решение. Введём обозначения событий: A – из первой коробки извлечена стандартная лампа, H_1 – лампа, переложенная из второй коробки в первую, была стандартной, H_2 – эта лампа была нестандартной. Тогда

$$P(H_1)=\frac{9}{10},\ \ P(H_2)=\frac{1}{10},\ \ P(A\,|\,H_1)=\frac{19}{21},\ \ P(A\,|\,H_2)=\frac{18}{21}.$$

По формуле Байеса находим

$$P(H_2\,|\,A)=\frac{P(H_2)P(A\,|\,H_2)}{P(H_1)P(A\,|\,H_1)+P(H_2)P(A\,|\,H_2)}=$$

$$=\frac{\frac{1}{10}\cdot\frac{18}{21}}{\frac{9}{10}\cdot\frac{19}{21}+\frac{1}{10}\cdot\frac{18}{21}}=\frac{2}{21}.\ \bullet$$

Задачи

1. В первой урне находится 3 белых шара и 2 чёрных, во второй – 4 белых и 4 чёрных. Из первой урны во вторую перекладывают, не глядя, два шара. После этого из второй урны берут наугад один шар. Найти вероятность того, что этот шар будет белым.

2. В ящике лежат 20 теннисных мячей, в том числе 15 новых и 5 играных. Для игры наудачу выбираются два мяча и после игры возвращаются обратно. Затем для второй игры также наудачу извлекаются ещё два мяча. Какова вероятность того, что вторая игра будет проводиться новыми мячами?

3. В торговую фирму поступили телевизоры от трёх поставщиков в отношении $1:4:5$. Практика показала, что телевизоры, поступающие от 1-го, 2-го и 3-го поставщиков не потребуют ремонта в течение гарантийного срока соответственно в 98, 88 и 92% случаев.

1) Найти вероятность того, что поступивший в торговую фирму телевизор, не потребует ремонта в течение гарантийного срока.

2) Проданный телевизор потребовал ремонта в течение гарантийного срока. От какого поставщика вероятнее всего поступил этот телевизор?

4. В урне лежит шар неизвестного цвета – с равной вероятностью белый или чёрный. В урну опускается один белый шар и после тщательного перемешивания наудачу извлекается один шар. Он оказался белым. Какова вероятность того, что в урне остался белый шар?

5. Три стрелка производят по одному выстрелу по одной и той же мишени. Вероятность попадания для первого стрелка равна 0,6, для второго – 0,5, для третьего – 0,4. В результате произведённых выстрелов в мишени оказались две пробоины. Найти вероятность того, что в мишень попали второй и третий стрелки.

1.6. Схема Бернулли. Формула Бернулли

Пусть проводится серия n испытаний, в каждом из которых может произойти или не произойти событие A. При этом предполагается, что вероятность наступления события A в каждом отдельном испытании не зависит от исходов других испытаний и равна одному и тому же числу p $(0 < p < 1)$. Такая последовательность испытаний называется *схемой Бернулли*.

Вероятность того, что в n испытаниях, проведённых по схеме Бернулли, событие A произойдёт ровно k раз $(0 \le k \le n)$, находится по *формуле Бернулли*:

$$P_n(k) = C_n^k p^k q^{n-k} \qquad (q = 1 - p). \tag{1.10}$$

Наивероятнейшее число k_0 наступлений события A в серии из n независимых испытаний, удовлетворяющих схеме Бернулли, определяется из двойного неравенства

$$np - q \le k_0 \le np + p\,.$$

Задачи

1. Известно, что 5% изделий, изготовляемых заводом, являются нестандартными. Проверено семь изделий. Какова вероятность того, что нестандартными окажутся: а) два изделия? б) хотя бы два изделия?

2. Вероятность поражения движущейся цели при каждом выстреле равна 0,4. Цель будет уничтожена при попадании в неё не менее двух раз. Найти вероятность того, что цель уничтожена, если в неё произведено пять независимых выстрелов.

3. В каждой из восьми урн имеется 10 белых и 5 чёрных шаров. Из каждой урны наудачу извлекли по одному шару. Что вероятнее: появление двух чёрных шаров и шести белых или трёх чёрных и пяти белых?

4. Вероятность попадания в десятку при одном выстреле равна 0,3. Сколько должно быть произведено независимых выстрелов, чтобы вероятность по меньшей мере одного попадания в десятку была больше 0,9?

5. В результате каждого визита страхового агента договор заключается с вероятностью 0,1. Найти наивероятнейшее число заключённых договоров после 25 визитов.

1.7. Предельные теоремы в схеме Бернулли

Использование формулы Бернулли (1.10) при больших значениях n затруднительно в вычислительном плане. Возникает необходимость в отыскании приближённых формул для вычисления вероятности $P_n(k)$. Такие формулы дают нам предельные теоремы.

Теорема 1.7 (теорема Пуассона). Пусть при неограниченном увеличении числа n испытаний $(n \to \infty)$ вероятность p наступления события A в каждом испытании стремится к нулю $(p \to 0)$, причём $np = \lambda$ – постоянная величина. Тогда

$$\lim_{n \to \infty} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \qquad (1.11)$$

Из равенства (1.11) при больших n и малых p вытекает приближённая *формула Пуассона*

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda = np, \quad k = 0, 1, 2, \ldots, n.$$

Эту формулу обычно используют, когда $p < 0{,}1$, а $np < 10$.

Пример 1.3. Завод отправил на базу 500 доброкачественных изделий. Вероятность того, что в пути изделие повредится, равна 0,002. Найти вероятность того, что на базу прибудут 3 негодных изделия.

Решение. Так как $n = 500$, $p = 0{,}002$, то $\lambda = 500 \cdot 0{,}002 = 1$. По формуле Пуассона находим

$$P_{500}(3) = \frac{1}{3!} e^{-1} \approx 0{,}06. \bullet$$

Теорема 1.8 (локальная теорема Муавра – Лапласа). Если вероятность p появления события A в каждом из n испытаний постоянна и отлична от 0 и 1, то

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi\left(\frac{k - np}{\sqrt{npq}}\right),$$

где

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ (функция Гаусса).}$$

Приближение тем точнее, чем больше n.

Для вычисления значений функции $\varphi(x)$ используется специальная таблица.

Обозначим через $P_n(k_1, k_2)$ вероятность того, что в n испытаниях, проводимых по схеме Бернулли, событие A наступит не менее k_1 и не более k_2 раз.

Теорема 1.9 (интегральная теорема Муавра – Лапласа). Если вероятность p наступления события A в каждом испытании постоянна и отлична от 0 и 1, то

$$P_n(k_1, k_2) \approx \Phi_0\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{k_1 - np}{\sqrt{npq}}\right),$$

где

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \text{ (нормированная функция Лапласа).}$$

Значения функции $\Phi_0(x)$ находятся по таблице. Отметим некоторые важные свойства этой функции:

1. $\Phi_0(x)$ – нечётная функция, т.е. $\Phi_0(-x) = -\Phi_0(x)$;
2. $\Phi_0(x)$ монотонно возрастает;
3. $\lim\limits_{x \to +\infty} \Phi_0(x) = 0{,}5$.

Задачи

1. На телефонной станции неправильное соединение происходит с вероятностью $\frac{1}{200}$. Найти вероятность того, что среди 200 соединений произойдёт:

а) ровно одно неправильное соединение;

б) не более двух неправильных соединений;

в) не менее трёх неправильных соединений.

2. Вероятность выигрыша по одному лотерейному билету равна 0,01. Сколько нужно купить билетов, чтобы выиграть хотя бы по одному из них с вероятностью, не меньшей чем 0,95?

3. Вероятность изделию некоторого производства оказаться бракованным равна 0,005. Чему равна вероятность того, что из 10000 наудачу взятых изделий бракованных окажется ровно 40?

4. Страховая компания заключила 40000 договоров. Вероятность страхового случая по каждому из них в течение года равна 0,02. Найти вероятность того, что таких случаев будет не более 870.

5. Отдел технического контроля проверяет 900 изделий на стандартность. Вероятность того, что изделие стандартно, равна 0,9. Найти вероятность того, что из проверенных изделий: а) 800 стандартных; б) не менее 800 стандартных.

Часть II. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

2.1. Понятие случайной величины. Функция распределения случайной величины. Дискретные случайные величины

Переменная величина, которая в зависимости от исхода опыта, т.е. в зависимости от случая, принимает различные действительные значения, называется *случайной величиной* (сокращённая запись: с.в.).

Случайные величины будем обозначать прописными латинскими буквами (при необходимости с индексами): X, Y_1, Z_i и т.д., а их возможные значения – соответствующими строчными буквами с индексами: $x_1, x_2, \ldots, y_{11}, y_{12}, \ldots, z_{i1}, z_{i2}, \ldots$.

Функцией распределения с.в. X называется функция $F(x)$, которая для любого $x \in \mathbf{R}$ равна вероятности события $X < x$:

$$F(x) = P(X < x).$$

Теорема 2.1. Функция распределения $F(x)$ обладает следующими свойствами:

1. $0 \le F(x) \le 1$ для любого $x \in \mathbf{R}$;
2. $F(x)$ – неубывающая функция на $\mathbf{R}$;
3. $\lim\limits_{x\to-\infty} F(x) = 0$, $\lim\limits_{x\to+\infty} F(x) = 1$;
4. $P(x_1 \le X < x_2) = F(x_2) - F(x_1)$, если $x_2 > x_1$;
5. $F(x)$ непрерывна слева в любой точке $x_0 \in \mathbf{R}$.

Дискретной случайной величиной называют с.в., которая принимает отдельные, изолированные друг от друга возможные значения. Очевидно, что множество возможных значений дискретной с.в. является конечным или счётным (т.е. его элементы могут быть занумерованы натуральными числами).

Пусть X — дискретная с.в., которая принимает значения x_1, x_2, ... *Законом распределения с.в.* X называется соответствие между возможными значениями x_i $(i = 1, 2, \ldots)$ и их вероятностями $p_i = P(X = x_i)$; его можно задать таблично, аналитически (в виде формулы) или графически. При табличном задании закона распределения дискретной с.в. первая строка таблицы содержит возможные значения с.в., а вторая – их вероятности:

X	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

(очевидно, $\sum_{i=1}^{n} p_i = 1$). Такую таблицу обычно называют *рядом распределения с.в. X*.

Рассмотрим некоторые наиболее часто встречающиеся на практике законы распределений дискретных случайных величин.

Биномиальное распределение. Дискретная с.в. X имеет *биномиальное распределение* (или *распределена по биномиальному закону*), если она принимает значения $k = 0, 1, 2, \ldots, n$ с вероятностями

$$p_k = P(X = k) = C_n^k p^k q^{n-k},$$

где $0 < p < 1,\ q = 1 - p.$

Случайную величину X, распределённую по биномиальному закону, можно рассматривать как число появлений события A в n испытаниях, проводимых по схеме Бернулли; при этом p – вероятность наступления события A в каждом из испытаний.

Распределение Пуассона. Дискретная с.в. X имеет *распределение Пуассона* (или *распределена по закону Пуассона*), если она принимает значения $k = 0, 1, 2, \ldots$ с вероятностями

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

где $\lambda > 0$ – параметр распределения.

Отметим, что распределение Пуассона возникло в теореме Пуассона (п. 1.7) как предельное распределение для числа появлений события A в n испытаниях схемы Бернулли, когда $n \to \infty$, а $p \to 0$ так, что $np = \lambda$ – постоянная величина.

Задачи

1. Стрелок, имея 5 патронов, стреляет до первого попадания в цель. Вероятность попадания при каждом выстреле равна 0,7. Пусть случайная величина X – число использованных патронов. Построить ряд распределения и функцию распределения $F(x)$ случайной величины X. Найти вероятность $P(2 < X < 5)$.

2. Дана функция распределения дискретной случайной величины X:

$$F(x) = \begin{cases} 0 & \text{при } x \le 0, \\ 0{,}4 & \text{при } 0 < x \le 1, \\ 0{,}7 & \text{при } 1 < x \le 2, \\ 0{,}8 & \text{при } 2 < x \le 3, \\ 1 & \text{при } x > 3. \end{cases}$$

Построить ряд распределения этой случайной величины.

3. Функция распределения дискретной случайной величины X имеет следующий вид:

$$F(x)=\begin{cases}0 & \text{при } x\le 2,\\ 0{,}3 & \text{при } 2<x\le 3,\\ 0{,}5 & \text{при } 3<x\le 4,\\ 1 & \text{при } x>4.\end{cases}$$

Вычислить вероятности $P(X\ge 3{,}5)$ и $P(|X|<2{,}5)$.

4. В одной студенческой группе обучается 25 студентов, среди которых 6 отличников. По жребию отобрано три студента. Пусть случайная величина X – число отличников среди отобранных студентов. Построить ряд распределения и функцию распределения случайной величины X.

5. В урне 9 шаров, из них 4 белых. Наудачу последовательно, без возвращения извлекаются шары до первого появления белого шара. Случайная величина X – число извлечённых шаров. Построить ряд распределения и функцию распределения случайной величины X.

6. Рабочий обслуживает четыре станка. Вероятность того, что в течение часа станок не потребует внимания рабочего, равна для первого станка 0,7, для второго – 0,75, для третьего – 0,8, для четвёртого – 0,9. Пусть случайная величина X – число станков, которые не потребуют внимания рабочего в течение часа. Построить ряд распределения случайной величины X.

2.2. Непрерывные случайные величины

Случайная величина X называется *непрерывной* с.в., если существует такая неотрицательная, интегрируемая по Риману на интервале $(-\infty,\ +\infty)$ функция $p(x)$, что при всех $x\in\mathbf{R}$

$$F(x)=\int_{-\infty}^{x}p(t)dt,$$

где $F(x)$ – функция распределения с.в. X. Функция $p(x)$ называется *плотностью распределения вероятностей* (*плотностью распределения* или просто *плотностью*) случайной величины X.

Функция распределения непрерывной с.в. непрерывна на всей числовой оси.

Теорема 2.2. Плотность распределения вероятностей непрерывной с.в. обладает следующими свойствами:

1. $p(x)\ge 0$ при всех $x\in\mathbf{R}$;

2. $P(a \le X < b) = \int_a^b p(x)dx$ при $a < b$;

3. $\int_{-\infty}^{+\infty} p(x)dx = 1$ (условие нормировки);

4. $F'(x) = p(x)$ в точках непрерывности функции $p(x)$;

5. $P(x \le X < x + \Delta x) = p(x)\Delta x + o(\Delta x)$ в точках непрерывности $p(x)$.

Замечание 2.1. Плотность распределения непрерывной с.в. называют также *дифференциальным законом распределения* (или просто *законом распределения*) этой с.в. ●

Теорема 2.3. Если X – непрерывная с.в., то для любого $x_0 \in \mathbf{R}$

$$P(X = x_0) = 0.$$

Следствие. Для непрерывной с.в. X справедливы равенства

$$P(a \le X < b) = P(a < X < b) = P(a \le X \le b) = P(a < X \le b).$$

Рассмотрим основные законы распределения непрерывных случайных величин.

Равномерное распределение. Непрерывная с.в. X имеет *равномерное распределение* на отрезке $[a, b]$, если её плотность распределения задана формулой

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

Показательное распределение. Непрерывная с.в. X имеет *показательное* (*экспоненциальное*) распределение, если её плотность имеет вид

$$p(x) = \begin{cases} 0 & \text{при } x < 0, \\ \lambda e^{-\lambda x} & \text{при } x \ge 0, \end{cases}$$

где $\lambda > 0$ – параметр распределения.

Нормальное распределение. Непрерывная с.в. X имеет *нормальное* распределение с параметрами $m \in \mathbf{R}$ и $\sigma > 0$, если её плотность имеет вид

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbf{R}.$$

Функция распределения с.в. X, распределённой по нормальному закону, выражается формулой

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt = \Phi\left(\frac{x-m}{\sigma}\right),$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \ \ (\textit{функция Лапласа}).$$

Замечание 2.2. Функция Лапласа $\Phi(x)$ связана с нормированной функцией Лапласа $\Phi_0(x)$ (см. п. 1.7) равенством

$$\Phi(x) = \Phi_0(x) + 0{,}5. \ \bullet$$

Графики плотности и функции распределения для нормального распределения изображены на рисунке 2.

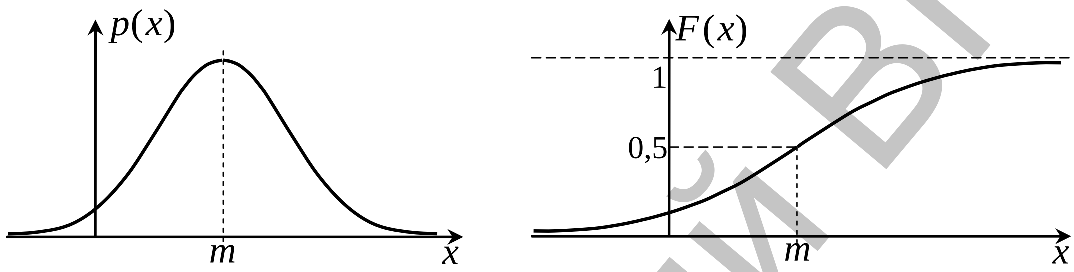

Рис. 2

Вероятность попадания с.в. X, распределённой по нормальному закону, в интервал $(a,\ b)$, находится по формуле

$$P(a < X < b) = \Phi_0\left(\frac{b-m}{\sigma}\right) - \Phi_0\left(\frac{a-m}{\sigma}\right).$$

Задачи

1. Функция распределения непрерывной случайной величины X задана выражением

$$F(x) = \begin{cases} 0 & \text{при } \ x \le 0, \\ ax^3 & \text{при } \ 0 < x \le 2, \\ 1 & \text{при } \ x > 2. \end{cases}$$

Найти коэффициент a и плотность распределения случайной величины X.

2. Задана плотность распределения непрерывной случайной величины X:

$$p(x) = \begin{cases} x & \text{при } \ 0 \le x \le 1, \\ 2 - x & \text{при } \ 1 < x \le 2, \\ 0 & \text{при } \ x < 0 \text{ или } x > 2. \end{cases}$$

Найти функцию распределения $F(x)$ этой случайной величины и вероятность события $\{0{,}5 \le X \le 1\}$.

3. Плотность распределения $p(x)$ случайной величины X определяется формулой $p(x)=ae^{-|x|}$. Найти коэффициент a и функцию распределения случайной величины X.

4. Случайная величина X имеет плотность распределения

$$p(x)=\frac{a}{1+x^2}.$$

Требуется: а) найти коэффициент a и функцию распределения $F(x)$; б) найти вероятность события $\{X>\sqrt{3}\}$.

5. Функция распределения случайного времени безотказной работы радиоаппаратуры имеет вид

$$F(t)=1-e^{-\frac{t}{T}} \qquad (0\le t<\infty).$$

Найти вероятность безотказной работы аппаратуры в течение промежутка времени $(0,T)$.

6. Случайная величина X распределена нормально с параметрами $m=6$, $\sigma=2$. Найти вероятности событий:

а) $\{4<X<7\}$; б) $\{|X-m|<0{,}3\}$.

2.3. Понятия многомерной случайной величины и её функции распределения. Дискретные двумерные случайные величины

Пусть в некотором эксперименте одновременно наблюдается n случайных величин $X_1,\ X_2,\ \ldots,\ X_n$.

Упорядоченный набор случайных величин $(X_1,\ X_2,\ldots,\ X_n)$ называется *n-мерной (многомерной) случайной величиной* или *n-мерным случайным вектором*.

Одномерные случайные величины X_i $(i=1,\ 2,\ \ldots,\ n)$ называются *компонентами (составляющими)* n-мерной с.в. $(X_1,\ X_2,\ldots,\ X_n)$.

Функцией распределения n-мерной с.в. $(X_1,\ X_2,\ldots,\ X_n)$ называется функция $F(x_1,\ x_2,\ \ldots,\ x_n)$, которая для любых действительных $x_1,\ x_2,\ldots,\ x_n$ равна вероятности произведения событий $X_1<x_1$, $X_2<x_2,\ \ldots,\ X_n<x_n$:

$$F(x_1,\ x_2,\ \ldots,\ x_n)=P(X_1<x_1,\ X_2<x_2,\ \ldots,\ X_n<x_n).$$

В частности, для двумерной с.в. $(X,\ Y)$ функция распределения определяется равенством

$$F(x,y)=P(X<x,\ Y<y).$$

Геометрически функция распределения $F(x,y)$ означает вероятность попадания случайной точки $(X,\ Y)$ в бесконечный квадрант с вершиной в точке $M(x, y)$, лежащий левее и ниже её (заштрихованная область на рисунке 3).

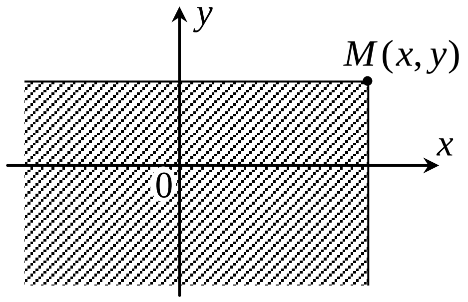

Рис. 3

Теорема 2.4. Функция распределения $F(x, y)$ обладает следующими свойствами:

1. $0 \le F(x, y) \le 1$ в любой точке $(x, y) \in \mathbf{R}^2$;
2. $F(x, y)$ – неубывающая функция по каждому из аргументов;
3. $\lim\limits_{x \to -\infty} F(x, y) = \lim\limits_{y \to -\infty} F(x, y) = \lim\limits_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y) = 0$;
4. $\lim\limits_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$;
5. $\lim\limits_{y \to +\infty} F(x, y) = F_X(x)$, $\lim\limits_{x \to +\infty} F(x, y) = F_Y(y)$,

где $F_X(x)$ и $F_Y(y)$ – функции распределения случайных величин X и Y, т.е. $F_X(x) = P(X < x)$, $F_Y(y) = P(Y < y)$;

6. $F(x, y)$ непрерывна слева в любой точке $(x_0, y_0) \in \mathbf{R}^2$ по каждому из аргументов x и y.

Рассмотрим двумерную с.в. (X, Y), компонентами которой являются дискретные случайные величины X и Y. Такая случайная величина называется *дискретной двумерной случайной величиной*. Пусть с.в. X может принимать только значения $x_1, x_2, \ldots$, а с.в. Y — только значения $y_1, y_2, \ldots$. Тогда множество возможных значений двумерной с.в. (X, Y) составляют пары значений (x_i, y_j), $i = 1, 2, \ldots$, $j = 1, 2, \ldots$.

Перечень всех возможных пар значений (x_i, y_j) и соответствующих этим парам вероятностей

$$p_{ij} = P(X = x_i, Y = y_j)$$

называется *законом распределения дискретной двумерной с.в.* (X, Y).

Если множества возможных значений случайных величин X и Y конечны, то распределение дискретной двумерной с.в. (X, Y) можно задать в виде *таблицы совместного распределения*:

$X \backslash Y$	y_1	…	y_j	…	$\boldsymbol{y_m}$
x_1	p_{11}	…	p_{1j}	…	$\boldsymbol{p_{1m}}$
…	…	…	…	…	**…**
x_i	p_{i1}		p_{ij}	…	$\boldsymbol{p_{im}}$
…	…	…	…	…	**…**
$\boldsymbol{x_n}$	$\boldsymbol{p_{n1}}$		$\boldsymbol{p_{nj}}$	**…**	$\boldsymbol{p_{nm}}$

Очевидно, что $\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1$.

Одномерные законы распределения отдельных компонент выражаются через вероятности совместных значений по формулам

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1}^{m} p_{ij}, \qquad p_{\bullet j} = P(Y = y_j) = \sum_{i=1}^{n} p_{ij}.$$

Через $p(x_i \mid y_j)$ будем обозначать условную вероятность того, что с.в. X примет значение x_i при условии, что с.в. Y приняла значение y_j:

$$p(x_i \mid y_j) = P(X = x_i \mid Y = y_j).$$

Согласно определению условной вероятности (п. 1.3),

$$p(x_i \mid y_j) = \frac{P(X = x_i,\, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\bullet j}}.$$

Совокупность условных вероятностей

$$p(x_1 \mid y_j),\;\; p(x_2 \mid y_j),\;\; \ldots,\;\; p(x_n \mid y_j)$$

представляет собой *условный закон распределения с.в. X при условии, что $Y = y_j$.*

Аналогично определяется *условный закон распределения с.в. Y при условии, что $X = x_i$*:

$$p(y_j \mid x_i) = P(Y = y_j \mid X = x_i) = \frac{P(X = x_i,\, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i\bullet}},$$

где $j = 1, 2, \ldots, m$.

Замечание 2.3. Сумма вероятностей условного закона распределения равна единице:

$$\sum_{i=1}^{n} p(x_i \mid y_j) = 1, \qquad \sum_{j=1}^{m} p(y_j \mid x_i) = 1. \bullet$$

Задачи

1. Бросаются две одинаковые игральные кости. Рассматриваются случайные величины: X – индикатор чётности суммы выпавших очков (т.е. $X=1$, если эта сумма чётна, и $X=0$ в противном случае), Y – индикатор чётности произведения выпавших очков ($Y=1$, если это произведение чётно, и $Y=0$ в противном случае). Найти закон распределения двумерной случайной величины (X,Y) и одномерные законы распределения компонент X и Y.

2. Распределение вероятностей дискретной двумерной случайной величины (X,Y) задано таблицей

x_i \ y_j	2	4	6
–1	0,08	0,12	0,20
1	0,12	0,18	0,30

Найти одномерные законы распределения компонент X и Y, а также условный закон распределения случайной величины X при условии, что $Y=4$. Построить функцию распределения двумерной случайной величины (X,Y).

3. Распределение вероятностей дискретной двумерной случайной величины (X,Y) задано таблицей

x_i \ y_j	10	20	30	40
0,5	0,05	0,12	0,08	0,04
2,5	0,09	0,30	0,11	0,21

Требуется:

а) построить ряды распределения случайных величин X и Y;

б) найти значения функции распределения $F(x,y)$ двумерной случайной величины (X,Y) в точках $(2{,}5;25)$ и $(9;11)$, а также вероятность события $\{2\le X<9,\ 10\le Y<30\}$.

4. Функция распределения дискретной двумерной случайной величины (X,Y) имеет вид, определяемый таблицей

x \ y	$y\le 1$	$1<y\le 3$	$3<y\le 5$	$y>5$
$x\le 2$	0	0	0	0
$2<x\le 4$	0	0,25	0,3	0,4
$x>4$	0	0,4	0,75	1

Найти закон распределения двумерной случайной величины (X,Y).

2.4. Непрерывные двумерные случайные величины

Двумерная случайная величина (X,Y) называется *непрерывной*, если её функция распределения $F(x,y)$ непрерывна в $\mathbf{R}^2$ и существует такая неотрицательная интегрируемая по Риману в бесконечных пределах по каждой из переменных функция $p(x,y)$, что

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

Функция $p(x,y)$ называется *плотностью распределения* (или *совместной плотностью*) двумерной с.в. (X,Y).

Теорема 2.5. Пусть (X,Y) – непрерывная двумерная с.в. с плотностью распределения $p(x,y)$. Тогда:

1. $p(x,y)\geq 0$ при всех $(x,y)\in\mathbf{R}^2$;
2. $P((X,Y)\in G)=\iint\limits_G p(x,y)dxdy$, G – область в $\mathbf{R}^2$;
3. $\int_{-\infty}^{+\infty} ds\int_{-\infty}^{+\infty} p(s,t)dt=1$;
4. $p(x,y)=\dfrac{\partial^2 F(x,y)}{\partial x\partial y}$, если $(x,\ y)$ – точка непрерывности функции $p(x,y)$;
5. $P(x\leq X<x+\Delta x,\ y\leq Y<y+\Delta y)\approx p(x,y)\Delta x\Delta y$;
6. X и Y – непрерывные случайные величины, причём их плотности $p_X(x)$ и $p_Y(y)$ выражаются через совместную плотность $p(x,y)$ следующим образом:

$$p_X(x)=\int_{-\infty}^{+\infty} p(x,y)dy,\quad p_Y(y)=\int_{-\infty}^{+\infty} p(x,y)dx. \tag{2.1}$$

Равномерное распределение. Непрерывная двумерная с.в. $(X,\ Y)$ имеет *равномерное распределение* в области $D\subset\mathbf{R}^2$, если её плотность распределения имеет вид

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

где $S(D)$ – площадь области D.

Замечание 2.4. По аналогии с двумерным случаем определяется непрерывная n-мерная с.в. $(X_1,\ X_2,\ \ldots,\ X_n)$; её функция распределения имеет вид

$$F(x_1, \ldots, x_n) = \int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_n} p(t_1, \ldots, t_n) dt_1 \ldots dt_n .$$

Свойства плотности $p(x_1, \ldots, x_n)$ аналогичны свойствам плотности $p(x, y)$. ●

Пусть (X, Y) – непрерывная двумерная с.в. с совместной плотностью распределения $p(x, y)$. Через $p_X(x)$ и $p_Y(y)$ обозначаем, как и выше, плотности распределений одномерных случайных величин X и Y.

Условной плотностью с.в. X при условии $Y = y$ называется функция $p_X(x \mid y)$, определяемая соотношением

$$p_X(x \mid y) = \frac{p(x, y)}{p_Y(y)} \qquad (p_Y(y) \neq 0).$$

Аналогично определяется *условная плотность с.в. Y при условии $X = x$*:

$$p_Y(y \mid x) = \frac{p(x, y)}{p_X(x)} \qquad (p_X(x) \neq 0).$$

В силу равенств (2.1) условные плотности можно выразить через совместную плотность следующим образом:

$$p_X(x \mid y) = \frac{p(x, y)}{\int\limits_{-\infty}^{+\infty} p(x, y) dx}, \qquad p_Y(y \mid x) = \frac{p(x, y)}{\int\limits_{-\infty}^{+\infty} p(x, y) dy}.$$

Как и любая плотность распределения, условные плотности обладают следующими свойствами:

$$p_X(x \mid y) \geq 0, \quad p_Y(y \mid x) \geq 0, \quad \int\limits_{-\infty}^{+\infty} p_X(x \mid y) dx = 1, \quad \int\limits_{-\infty}^{+\infty} p_Y(y \mid x) dy = 1.$$

Задачи

1. Плотность распределения непрерывной двумерной случайной величины (X, Y) имеет вид

$$p(x, y) = \begin{cases} axy & \text{при } (x, y) \in D, \\ 0 & \text{при } (x, y) \notin D, \end{cases}$$

где область $D = \{(x, y) : x \geq 0,\ y \geq 0,\ x + y \leq 2\}$. Найти:

а) постоянную a;

б) вероятность попадания случайной точки (X, Y) в область $G = \{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$;

в) одномерные плотности распределения $p_X(x)$ и $p_Y(y)$ случайных величин X и Y соответственно.

2. Плотность распределения непрерывной двумерной случайной величины (X, Y) имеет следующий вид:

$$p(x, y) = \begin{cases} c(x+y) & \text{при } 0 \le x \le 1,\ 0 \le y \le 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Найти постоянную c и вычислить вероятность события $\{X + Y < 1\}$.

3. Известна функция распределения непрерывной двумерной случайной величины (X, Y):

$$F(x, y) = \begin{cases} 1 - 2^{-x} - 2^{-y} + 2^{-x-y} & \text{при } x \ge 0, y \ge 0, \\ 0 & \text{при } x < 0 \text{ или } y < 0. \end{cases}$$

Найти совместную плотность распределения двумерной случайной величины (X, Y) и вероятность её попадания в прямоугольник, ограниченный прямыми $x = 1,\ x = 2,\ y = 3,\ y = 5$.

4. Плотность распределения непрерывной двумерной случайной величины (X, Y) имеет вид

$$p(x, y) = \begin{cases} cxy^2 & \text{при } (x, y) \in D, \\ 0 & \text{при } (x, y) \notin D, \end{cases}$$

где D – прямоугольник, ограниченный прямыми $x = 0,\ x = 1,\ y = 0,\ y = 2$. Найти: а) постоянную c; б) функцию распределения двумерной случайной величины (X, Y).

2.5. Независимые случайные величины

Случайные величины $X_1, \ldots, X_n$ называются *независимыми* (по-другому: *независимыми в совокупности, взаимно независимыми*), если для любых $x_1, \ldots, x_n \in \mathbf{R}$ являются независимыми в совокупности события $X_1 < x_1, \ldots, X_n < x_n$ (см. п. 1.3).

Теорема 2.6. Для независимости случайных величин $X_1, \ldots, X_n$ необходимо и достаточно, чтобы для любых $x_1, \ldots, x_n \in \mathbf{R}$ выполнялось равенство

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n),$$

где $F(x_1, \ldots, x_n)$ – функция распределения n-мерной с.в. $(X_1, \ldots, X_n)$, $F_{X_i}(x_i)$ – функция распределения с.в. X_i, $i = 1, \ldots, n$.

Теорема 2.7. Непрерывные случайные величины $X_1, \ldots, X_n$ независимы тогда и только тогда, когда n-мерная с.в. $(X_1, \ldots, X_n)$ непрерывна и её плотность $p(x_1, \ldots, x_n)$ представляется в виде

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n),$$

где $p_{X_i}(x_i)$ – плотность с.в. X_i, $i = 1, \ldots, n$.

Для дискретных случайных величин справедлива следующая теорема.

Теорема 2.8. Дискретные случайные величины $X_1, \ldots, X_n$ независимы тогда и только тогда, когда для любых $x_1, \ldots, x_n \in \mathbf{R}$ выполняется равенство

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n).$$

Замечание 2.5. Из теорем 2.7 и 2.8, в частности, следует:

1. Если компоненты непрерывной с.в. (X, Y) являются независимыми случайными величинами, то их условные плотности равны безусловным:

$$p_X(x \mid y) = p_X(x), \quad p_Y(y \mid x) = p_Y(y).$$

2. Если компоненты дискретной с.в. (X, Y) независимы, то их условные законы распределения совпадают с безусловными:

$$p(x_i \mid y_j) = p_{i\bullet}, \quad p(y_j \mid x_i) = p_{\bullet j}. \; \bullet$$

Задачи

1. Распределение вероятностей дискретной двумерной случайной величины (X, Y) задано таблицей

x_i \ y_j	1	2	3
1	0,05	0,1	0,05
2	0,15	0,2	0,1
3	0,1	0,2	0,05

Являются ли случайные величины X и Y независимыми?

2. Непрерывная двумерная случайная величина (X, Y) распределена равномерно в области $G = \{(x, y) : -1 \le x \le 2, 1 \le y \le 2\}$. Установить, зависимы или нет компоненты X и Y.

3. Непрерывная двумерная случайная величина (X, Y) распределена равномерно в квадрате со стороной a и диагоналями, лежащими на осях координат. Установить, зависимы или нет компоненты X и Y.

4. Непрерывная двумерная случайная величина (X, Y) распределена равномерно в треугольнике с вершинами в точках $(0;0)$, $(2;0)$, $(0;1)$. Найти плотности и условные плотности распределения компонент X и Y. Установить, зависимы или нет эти случайные величины.

2.6. Функции случайных величин. Закон распределения функции одной случайной величины

Пусть $X_1, \ldots, X_n$ – случайные величины, связанные с некоторым опытом, и $f(x_1, \ldots, x_n)$ – действительнозначная функция n действительных переменных, область определения которой содержит все возможные значения n-мерной с.в. $(X_1, \ldots, X_n)$. Тогда можно определить с.в. Y, которая принимает свои значения в зависимости от того, какие значения принимают случайные величины $X_1, \ldots, X_n$, а именно: если в результате опыта случайные величины $X_1, \ldots, X_n$ приняли значения $x_1, \ldots, x_n$, то с.в. Y принимает значение $y = f(x_1, \ldots, x_n)$. При этом Y называют *функцией случайных величин* $X_1, \ldots, X_n$ и записывают: $Y = f(X_1, \ldots, X_n)$.

Рассмотрим вопрос о том, как найти закон распределения функции $Y = f(X)$, если известен закон распределения аргумента X.

1. Пусть X — дискретная с.в., принимающая значения $x_1, x_2, \ldots, x_n$ с соответствующими вероятностями $p_1, p_2, \ldots, p_n$. Множество возможных значений с.в. Y составляют все различные числа среди чисел $f(x_1), f(x_2), \ldots, f(x_n)$. Пусть y_j – одно из возможных значений с.в. Y. Событие $Y = y_j$ эквивалентно сумме тех событий $X = x_i$, для которых $f(x_i) = y_j$. Поскольку все слагаемые в этой сумме — события несовместные, то

$$P(Y = y_j) = \sum_{i: f(x_i) = y_j} P(X = x_i) = \sum_{i: f(x_i) = y_j} p_i$$

(суммирование ведётся по всем i, для которых $f(x_i) = y_j$).

2. Пусть X – непрерывная с.в. с плотностью распределения $p_X(x)$. Предположим, что все возможные значения с.в. X составляют интервал (a, b) (в частности, $a = -\infty$, $b = +\infty$), а функция $f(x)$ строго монотонна и дифференцируема на интервале (a, b). В этом случае множеством всех значений с.в. $Y = f(X)$ также является интервал; обозначим его (c, d). Можно доказать, что плотность с.в. Y имеет следующий вид:

$$p_Y(y) = \begin{cases} p_X(f^{-1}(y)) \cdot |(f^{-1}(y))'| & \text{при } y \in (c, d), \\ 0 & \text{при } y \notin (c, d), \end{cases} \qquad (2.2)$$

где $f^{-1}(y)$ – функция, обратная функции $f(x)$.

Пример 2.1. Пусть с.в. X имеет нормальное распределение с параметрами m и σ (см. п. 2.2). Найти плотность распределения с.в. $Y = aX + b$, где $a, b \in \mathbf{R}$, $a \neq 0$.

Решение. В данном случае имеем: $f(x) = ax + b$ – строго возрастающая (при $a > 0$) или строго убывающая (при $a < 0$) на всей числовой оси функция. Находим

$$f^{-1}(y) = \frac{y-b}{a}, \quad (f^{-1}(y))' = \frac{1}{a}.$$

Как легко видеть, множеством возможных значений с.в. Y является интервал $(-\infty, +\infty)$. Согласно формуле (2.2), получаем

$$p_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{y-b}{a}-m\right)^2}{2\sigma^2}} \cdot \frac{1}{|a|} = \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(y-(am+b))^2}{2(|a|\sigma)^2}}.$$

Таким образом, с.в. $Y = aX + b$ также распределена по нормальному закону с параметрами, равными $am + b$ и $|a|\sigma$. ●

Замечание 2.6. Из примера 2.1., в частности, следует: если с.в. X имеет нормальное распределение с параметрами m и σ, то с.в. $Y = \dfrac{X-m}{\sigma}$ также распределена нормально с параметрами 0 и 1. ●

Задачи

1. Шесть раз бросается правильная монета. Случайная величина X – модуль разности числа появлений герба и числа появлений цифры в данном эксперименте. Найти закон распределения этой случайной величины.

2. Распределение дискретной двумерной случайной величины (X, Y) задано таблицей

$x_i \backslash y_j$	10	12	14
1	0,08	0,02	0,10
2	0,32	0,08	0,40

Найти ряд распределения случайной величины $Z = (Y - 12)/X$.

3. Дискретная двумерная случайная величина (X, Y) распределена по закону, представленному таблицей

$x_i \backslash y_j$	–1	0	1
0	0,1	0,2	0,1
1	0,2	0,3	0,1

Найти функцию распределения случайной величины $Z = XY$.

4. Случайная величина X распределена равномерно на отрезке $[0;3]$. Найти функцию распределения случайной величины $Y = X^2 + 1$.

5. Случайная величина X распределена по показательному закону с параметром λ. Найти плотность распределения вероятностей случайной величины Y, если: а) $Y = \sqrt{X}$; б) $Y = X^2$.

6. На отрезке длиной a наудачу выбираются две точки; случайная величина Z – расстояние между ними. Найти математическое ожидание и дисперсию случайной величины Z.

2.7. Числовые характеристики случайных величин

Математическое ожидание. *Математическим ожиданием* (или *средним значением*) *дискретной с.в. X*, имеющей закон распределения $P(X = x_i) = p_i$, $i = 1, 2, \ldots$, называется число $M(X)$, определяемое формулой

$$M(X) = \sum_i x_i p_i .$$

Математическим ожиданием непрерывной с.в. X с плотностью распределения вероятностей $p(x)$ называется число

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

Теорема 2.9. Математическое ожидание обладает следующими свойствами:

1. $M(C) = C$, если C – константа (дискретная с.в., принимающая значение C с вероятностью 1);

2. $M(kX) = kM(X)$, если k – константа;

3. $M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n)$;

4. $M(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = M(X_1) \cdot M(X_2) \cdot \ldots \cdot M(X_n)$ для независимых (в совокупности) случайных величин $X_1, X_2, \ldots, X_n$;

5. Если $X \geq 0$ (т.е. с.в. X принимает только неотрицательные значения), то $M(X) \geq 0$.

Теорема 2.10. Пусть $f(X)$ – функция с.в. X. Тогда

$$M(f(X)) = \begin{cases} \sum\limits_i f(x_i) p_i, & \text{если с.в.} X \text{ дискретная}, \\ \int\limits_{-\infty}^{+\infty} f(x) p(x) dx, & \text{если с.в.} X \text{ непрерывная}, \end{cases}$$

где x_i $(i = 1, 2, \ldots)$ – возможные значения дискретной с.в. X, p_i – вероятности этих значений; $p(x)$ – плотность распределения вероятностей непрерывной с.в. X.

Дисперсия. *Дисперсией* с.в. X называется число $D(X)$, определяемое формулой

$$D(X) = M[(X - M(X))^2] =$$

$$= \begin{cases} \sum_i (x_i - M(X))^2 p_i, & \text{если с.в.} X \text{ дискретная}, \\ \int_{-\infty}^{+\infty} (x - M(X))^2 p(x) dx, & \text{если с.в.} X \text{ непрерывная}. \end{cases}$$

Теорема 2.11. Дисперсия обладает следующими свойствами:

1. $D(X) \geq 0$;
2. $D(C) = 0$, если C – константа;
3. $D(kX) = k^2 D(X)$, если k – константа;
4. $D(X_1 + X_2 + \ldots + X_n) = D(X_1) + D(X_2) + \ldots + D(X_n)$ для независимых (в совокупности) случайных величин $X_1, X_2, \ldots, X_n$;
5. $D(X - Y) = D(X) + D(Y)$ для независимых случайных величин X и Y;
6. $D(X) = M(X^2) - [M(X)]^2$.

Среднее квадратическое отклонение. *Средним квадратическим отклонением* с.в. X называется число

$$\sigma(X) = \sqrt{D(X)}.$$

Пример 2.2 (биномиальное распределение). Как известно (см. п. 2.1), с.в. X, имеющую биномиальное распределение, можно рассматривать как число появлений события A в n испытаниях, проводимых по схеме Бернулли. Пусть с.в. X_i – число появлений события A в i-м испытании. Тогда

$$X = X_1 + \ldots + X_n.$$

Каждая с.в. X_i $(i = 1, \ldots, n)$ может принимать только два значения: $x_1 = 1$ с вероятностью p и $x_2 = 0$ с вероятностью $q = 1 - p$. Поэтому

$$M(X_i) = 1 \cdot p + 0 \cdot q = p \qquad (i = 1, \ldots, n),$$

$$M(X) = M(X_1) + \ldots + M(X_n) = np.$$

В силу независимости исходов испытаний случайные величины $X_1, \ldots, X_n$ независимы. Поэтому, согласно свойству 4 дисперсии (теорема 2.11),

$$D(X) = D(X_1) + D(X_2) + \ldots + D(X_n).$$

Дисперсию каждой с.в. X_i найдём, используя свойство 6 дисперсии:

$$M(X_i^2) = 1^2 \cdot p + 0^2 \cdot q = p,$$

$$D(X_i) = M(X_i^2) - [M(X_i)]^2 = p - p^2 = p(1 - p) = pq.$$

Таким образом,

$$D(X) = npq\,. \bullet$$

Пример 2.3 (нормальное распределение). Пусть с.в. X имеет нормальный закон распределения с параметрами m и σ. Как известно (см. замечание 2.6), с.в. $Y = \dfrac{X - m}{\sigma}$ также распределена нормально с параметрами 0 и 1. Найдём сначала математическое ожидание и дисперсию с.в. Y:

$$M(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2}} dy = 0$$

(так как интеграл сходится, а подынтегральная функция нечётная);

$$M(Y^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} dy = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y d e^{-\frac{y^2}{2}} =$$

$$= -\frac{1}{\sqrt{2\pi}} y e^{-\frac{y^2}{2}} \Bigg|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = 0 + 1 = 1,$$

$$D(Y) = M(Y^2) - [M(Y)]^2 = 1 - 0 = 1.$$

Теперь с помощью соответствующих свойств математического ожидания и дисперсии находим математическое ожидание и дисперсию с.в. $X = \sigma Y + m$:

$$M(X) = \sigma M(Y) + m = m, \qquad D(X) = \sigma^2 D(Y) = \sigma^2\,. \bullet$$

Задачи

1. Дискретная случайная величина X имеет следующий закон распределения:

x_i	1	2	3	4	5
p_i	c	$4c$	$9c$	$16c$	$25c$

Найти: 1) постоянную c; 2) математическое ожидание $M(X)$; 3) дисперсию $D(X)$; 4) среднее квадратическое отклонение $\sigma(X)$.

2. В урне находится 4 белых и 6 чёрных шаров. Случайным образом из неё извлекается 3 шара. Пусть случайная величина X – количество белых шаров среди трёх вынутых. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение с.в. X.

3. Случайная величина X имеет плотность распределения

$$p(x)=\begin{cases}\frac{2}{9}x & \text{при } 0<x<3,\\ 0 & \text{при } x\le 0 \text{ или } x\ge 3.\end{cases}$$

Найти математическое ожидание $M(X)$, дисперсию $D(X)$ и среднее квадратическое отклонение $\sigma(X)$.

4. Плотность распределения $p(x)$ случайной величины X определяется формулой $p(x)=ae^{-|x|}$. Определить: а) коэффициент a; б) математическое ожидание $M(X)$ и дисперсию $D(X)$.

5. Функция распределения непрерывной случайной величины X задана выражением

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

Найти математическое ожидание и дисперсию с.в. X.

Часть III. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

3.1. Неравенства Маркова и Чебышева. Теоремы Чебышева и Бернулли

Теорема 3.1 (неравенство Маркова). Пусть X – неотрицательная с.в., имеющая математическое ожидание $M(X)$. Тогда для любого $\varepsilon>0$ справедливо неравенство

$$P(X\ge\varepsilon)\le\frac{M(X)}{\varepsilon}.$$

Неравенство Маркова можно записать в другой форме:

$$P(X<\varepsilon)\ge 1-\frac{M(X)}{\varepsilon}.$$

Теорема 3.2 (неравенство Чебышева). Пусть X – с.в., имеющая математическое ожидание $M(X)$ и дисперсию $D(X)$. Тогда для любого $\varepsilon>0$ справедливо неравенство

$$P(|X-M(X)|\ge\varepsilon)\le\frac{D(X)}{\varepsilon^2}.$$

Другая форма неравенства Чебышева:

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

Теорема 3.3 (теорема Чебышева). Пусть $X_1,\ X_2,\ \ldots$ – последовательность независимых случайных величин, обладающих математическими ожиданиями $M(X_i)$ и дисперсиями $D(X_i)$, причём дисперсии ограничены в совокупности (т.е. существует такая константа $C > 0$, что $D(X_i) \leq C$, $i = 1,\ 2,\ \ldots$). Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| \geq \varepsilon\right) = 0.$$

Следствие. Пусть $X_1,\ X_2,\ \ldots$ — последовательность независимых одинаково распределённых случайных величин, $M(X_i) = m$, $D(X_i) = \sigma^2$, $i = 1,\ 2,\ \ldots$. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| \geq \varepsilon\right) = 0. \qquad (3.1)$$

Последовательность случайных величин $Y_1,\ Y_2,\ \ldots$ называется *сходящейся по вероятности* к числу b, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|Y_n - b| \geq \varepsilon\right) = 0$$

(или $\lim\limits_{n\to\infty} P\left(|Y_n - b| < \varepsilon\right) = 1$); символическая запись:

$$Y_n \xrightarrow{P} b.$$

Замечание 3.1. Соотношение (3.1), выражающее частный случай теоремы Чебышева, можно записать в виде

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} m. \qquad \bullet$$

Теорема 3.4 (теорема Я. Бернулли). Пусть k – число наступлений события A в n испытаниях, проводимых по схеме Бернулли. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) = 0,$$

где p – вероятность появления события A в каждом из испытаний.

Таким образом, теорема Бернулли утверждает, что относительная частота $\frac{k}{n}$ сходится по вероятности к вероятности $P(A) = p$.

Задачи

1. По многолетним наблюдениям, средняя скорость ветра в некотором пункте равна 16 км/ч. Оценить вероятность того, что в случайный момент времени скорость ветра в этом пункте превысит 80 км/ч.

2. Оценить вероятность того, что случайная величина X отклонится от своего математического ожидания не менее чем на 3σ, где σ – среднее квадратическое отклонение этой случайной величины.

3. Выяснить, применима ли теорема Чебышева к последовательности независимых случайных величин $X_1, X_2, \ldots, X_n, \ldots$, если X_n имеет следующее распределение:

x_{ni}	$-5n$	0	$5n$
p_{ni}	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

4. Глубина моря измеряется прибором, не имеющим систематической ошибки. Среднее квадратическое отклонение измерений не превосходит 15 м. Сколько нужно сделать независимых измерений, чтобы с вероятностью, не меньшей 0,9, можно было утверждать, что среднее арифметическое этих измерений отличается от a (глубины моря) по модулю менее чем на 5 м?

5. Вероятность наличия опечатки на одной странице рукописи $p = 0{,}2$. Оценить вероятность того, что в рукописи, содержащей 400 страниц, частость (относительная частота) появления опечатки отличается от p по модулю менее чем на 0,05.

3.2. Центральная предельная теорема

Теорема 3.5. Пусть $X_1,\ X_2,\ \ldots$ – последовательность независимых одинаково распределённых случайных величин, имеющих математическое ожидание $M(X_i) = m$ и дисперсию $D(X_i) = \sigma^2 > 0$, $i = 1,\ 2,\ \ldots$. Тогда для любого $x \in \mathbf{R}$

$$P\left(\frac{X_1 + \ldots + X_n - nm}{\sigma\sqrt{n}} < x\right) \xrightarrow[n\to\infty]{} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt. \quad (3.2)$$

Замечание 3.2. Соотношение (3.2) означает, что при $n \to \infty$ закон распределения с.в.

$$Z_n = \frac{X_1 + \ldots + X_n - nm}{\sigma\sqrt{n}}$$

неограниченно приближается к нормальному закону распределения с параметрами 0 и 1. ●

Задачи

1. Случайная величина *X* является средним арифметическим 3200 независимых одинаково распределённых случайных величин: $X = \frac{1}{3200}\sum_{n=1}^{3200} X_i$, причём $M(X_i) = 3$, $D(X_i) = 2$, $i = \overline{1,3200}$. Найти вероятность того, что случайная величина *X* попадёт в интервал $(2{,}925;\, 3{,}075)$.

2. Складывается 10^3 чисел, каждое из которых округлено с точностью до 10^{-3}. Предполагая, что ошибки от округления независимы и равномерно распределены в интервале $(-0{,}5\cdot 10^{-3};\, 0{,}5\cdot 10^{-3})$, найти интервал, симметричный относительно математического ожидания, в котором с вероятностью 0,998 заключена суммарная ошибка.

3. Время ожидания автобуса пассажиром имеет показательное распределение со средним значением 9 минут. Найти число поездок, для которого суммарное время ожидания автобуса превысит 3 часа с вероятностью не более 0,2.

4. Независимые случайные величины X_i $(i = \overline{1,100})$ распределены равномерно на отрезке $[0;1]$. Написать приближенное выражение для плотности распределения случайной величины $Y = \sum_{i=1}^{100} X_i$, а также вероятность того, что $55 < Y < 70$.

5. Определить, сколько раз нужно бросить игральную кость, чтобы с вероятностью не менее 0,99 можно было утверждать, что относительная частота выпадения пятёрки будет находиться в пределах от $1/6 - 0{,}05$ до $1/6 + 0{,}05$?

Часть IV. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

4.1. Выборка. Статистическое распределение выборки. Эмпирическая функция распределения. Числовые характеристики выборки

Будем предполагать, что каждый исход некоторого случайного эксперимента характеризуется одним из возможных значений с.в. *X*. В таком случае говорят, что с.в. *X наблюдается* в данном эксперименте.

Повторив *n* раз эксперимент в одинаковых условиях, получим последовательность из *n* наблюдённых значений с.в. *X*: $x_1,\ x_2,\ \ldots,\ x_n$.

Набор чисел $x_1,\ x_2,\ \ldots,\ x_n$ называется *выборкой*.

Числа x_i $(i=1,\ 2,\ \ldots,\ n)$ называются *элементами выборки*, а их количество n — *объёмом выборки*.

Выборку можно упорядочить, расположив её элементы в неубывающем порядке:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}.$$

Полученная таким образом последовательность чисел

$$x_{(1)},\ x_{(2)},\ \ldots,\ x_{(n)}$$

называется *вариационным рядом*.

Различные значения с.в. X, представленные в выборке, называются *вариантами*.

Если в выборке объема n варианта x_i встретилась n_i раз, то число n_i называется *частотой*, а число $w_i = \dfrac{n_i}{n}$ – *относительной частотой (частостью)* этой варианты. Очевидно, что сумма всех частот равна объёму выборки n, а сумма всех относительных частот равна единице.

Статистическим рядом распределения выборки называют перечень вариант и соответствующих им частот или относительных частот. Обычно статистический ряд записывается в виде таблицы, первая строка которой содержит варианты x_i, а вторая — их частоты n_i (или относительные частоты w_i). При этом варианты располагаются в порядке возрастания.

Как правило, статистический ряд распределения составляется в случае, когда наблюдаемая с.в. X является дискретной.

Графически статистический ряд распределения выборки изображается следующим образом. В прямоугольной декартовой системе координат строят точки с координатами $(x_i,\ n_i)$. Последовательно соединив их отрезками, получают ломаную линию, называемую *полигоном частот*. Аналогично строится *полигон относительных частот*.

Пример 4.1. Имеется выборка значений с.в. X объёма $n=15$:

$$0,\ 3,\ -5,\ -3,\ 1,\ 0,\ 1,\ 3,\ 0,\ 0,\ -3,\ 1,\ -1,\ 0,\ -1.$$

Вариационным рядом для неё будет последовательность

$$-5,\ -3,\ -3,\ -1,\ -1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 3,\ 3.$$

Статистический ряд распределения данной выборки имеет вид

x_i	-5	-3	-1	0	1	3
n_i	1	2	2	5	3	2

Полигон частот изображён на рисунке 4.

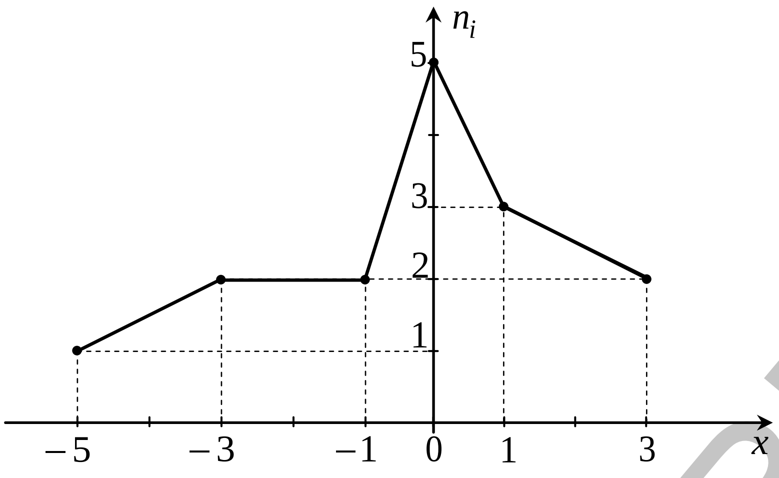

Рис. 4

В случае, когда X является непрерывной с.в. или же вариационный ряд имеет большое количество вариант, элементы выборки объединяют в группы (разряды), представляя результаты опытов в виде *интервального (группированного) статистического ряда распределения*. Для этого интервал, содержащий все значения выборки, разбивают на k непересекающихся интервалов

$$[a_0, a_1), \ [a_1, a_2), \ \ldots, \ [a_{k-1}, a_k].$$

Затем для каждого интервала разбиения определяют *частоту* – количество элементов выборки, попавших в этот интервал: n_i – частота, соответствующая i-му интервалу разбиения $(i = 1, 2, \ldots, k)$. Наряду с частотами находят также *относительные частоты* $w_i = \dfrac{n_i}{n}$ (n – объём выборки). Интервальный статистический ряд записывают в виде таблицы, в первой строке которой содержатся интервалы группировки, а во второй – соответствующие частоты n_i или относительные частоты w_i.

Интервальный статистический ряд распределения изображают с помощью *гистограммы частот*. Для этого на оси абсцисс откладывают интервалы группировки и на них, как на основании, строят прямоугольники с высотами $h_i = \dfrac{n_i}{a_i - a_{i-1}}$ $(i = 1, 2, \ldots, k)$. Аналогично строится *гистограмма относительных частот*.

Пример 4.2. Выборка 100 значений наблюдаемой непрерывной с.в. X представлена в виде следующего интервального статистического ряда распределения:

Интервал	[22, 24)	[24, 26)	[26, 28)	[28, 30)	[30, 32)	[32, 34]
Частота	2	14	34	40	8	2

Построить гистограмму частот.

Решение. В данном случае $a_i - a_{i-1} = 2$ $(i = 1, 2, 3, 4, 5, 6)$. Находим высоты прямоугольников:

$$h_1 = \frac{2}{2} = 1, \quad h_2 = \frac{14}{2} = 7, \quad h_3 = \frac{34}{2} = 17,$$

$$h_4 = \frac{40}{2} = 20, \quad h_5 = \frac{8}{2} = 4, \quad h_6 = \frac{2}{2} = 1.$$

Гистограмма частот изображена на рисунке 5.

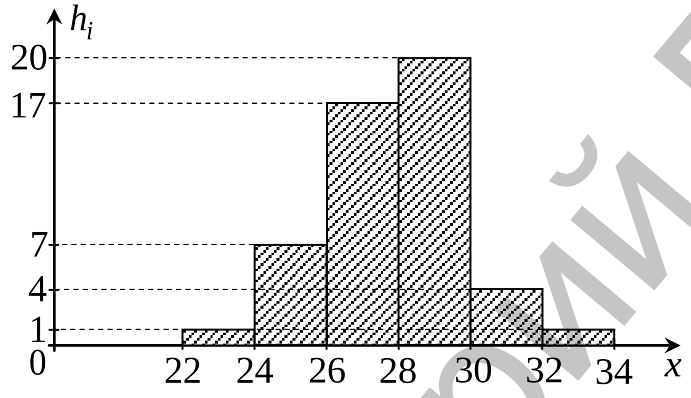

Рис. 5

●

Эмпирической (выборочной) функцией распределения с.в. X называют функцию $F_n^*(x)$, которая при каждом $x \in \mathbf{R}$ равна относительной частоте события $X < x$, т.е.

$$F_n^*(x) = \frac{n_x}{n},$$

где n_x – число элементов выборки, меньших x.

Эмпирическая функция распределения обладает следующими свойствами:

1. $0 \le F_n^*(x) \le 1$ для любого $x \in \mathbf{R}$;

2. $F_n^*(x)$ не убывает на $\mathbf{R}$;

3. $F_n^*(x) = 0$ при $x \le x_{\min}$, $F_n^*(x) = 1$ при $x > x_{\max}$, где $x_{\min}$ и $x_{\max}$ – соответственно наименьший и наибольший элементы выборки;

4. $F_n^*(x)$ непрерывна слева в любой точке $x_0 \in \mathbf{R}$.

График эмпирической функции распределения имеет ступенчатый вид. В промежутках между соседними вариантами выборки $F_n^*(x)$ сохраняет постоянное значение. В точках оси Ox, равных вариантам выборки, $F_n^*(x)$ претерпевает скачки; величина скачка в точке $x = x_i$ равна относительной частоте варианты x_i.

Выборочным средним $\bar{x}$ называется среднее арифметическое всех значений выборки:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонений значений выборки от выборочного среднего $\bar{x}$:

$$D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

Выборочную дисперсию можно вычислять также по формуле

$$D_в = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 .$$

Замечание 4.1. Если выборка представлена статистическим рядом распределения, то значения $\bar{x}$ и $D_в$ находят по формулам

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i , \qquad (4.1)$$

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - (\bar{x})^2, \qquad (4.2)$$

где x_i – варианты, n_i – соответствующие им частоты. ●

Замечание 4.2. Для интервального статистического ряда выборочное среднее и выборочную дисперсию также вычисляют по формулам (4.1) и (4.2). В качестве x_i берут середины интервалов ряда, а в качестве n_i – частоты соответствующих интервалов. ●

Выборочное среднее квадратическое отклонение $\sigma_в$ определяется формулой

$$\sigma_в = \sqrt{D_в} .$$

Задачи

1. Имеются результаты измерения длины (в мм) 30 случайно отобранных заготовок:

39, 41, 40, 40, 43, 41, 44, 42, 41, 41, 43, 42, 39, 40, 42,
43, 41, 42, 41, 39, 42, 41, 42, 40, 41, 43, 41, 39, 40, 41.

Составить вариационный и статистический ряды данной выборки. Построить полигон частот.

2. Найти эмпирическую функцию распределения и начертить ее график для выборки, представленной следующим статистическим рядом:

x_i	39	40	41	42	43	44
n_i	4	5	9	7	4	1

3. Построить гистограмму частот для выборки, представленной следующим интервальным статистическим рядом:

$x_{i-1} - x_i$	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
n_i	1	2	7	18	12	8	2

4. Найти среднее и дисперсию выборки, представленной следующим интервальным статистическим рядом:

$x_{i-1} - x_i$	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13
n_i	1	2	4	2	1	1

4.2. Точечные оценки неизвестных параметров распределения. Метод максимального правдоподобия

Пусть $F(x,\theta)$ – функция распределения с.в. X, θ – неизвестный параметр. Предполагаем, что общий вид функции $F(x,\theta)$ задан.

Пусть в результате n наблюдений за с.в. X получена выборка

$$x_1,\ x_2,\ \ldots,\ x_n. \tag{4.3}$$

По выборке требуется найти приближённое значение параметра θ.

Элементы выборки (4.3) можно рассматривать последовательно как частные значения n независимых случайных величин

$$X_1,\ X_2,\ \ldots,\ X_n,$$

каждая из которых имеет тот же закон распределения, что и с.в. X.

Любая функция случайных величин $X_1,\ X_2,\ \ldots,\ X_n$ называется *статистикой*.

Пусть $\tilde{\theta} = \tilde{\theta}(X_1,\ X_2,\ \ldots,\ X_n)$ – некоторая статистика. Значение $\tilde{\theta}(x_1,\ x_2,\ \ldots,\ x_n)$, принятое статистикой $\tilde{\theta}$ на выборке (4.3), называется её *выборочным значением*.

Важными примерами статистик являются *выборочное среднее*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и *выборочная дисперсия*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Определённые в пункте 4.1 числовые характеристики выборки $\bar{x}$ (выборочное среднее) и $D_в$ (выборочная дисперсия) являются выборочными значениями одноимённых статистик $\overline{X}$ и S^2.

Статистика $\tilde{\theta}$, выборочное значение которой принимается за приближённое значение неизвестного параметра θ, называется его *точечной оценкой* или просто *оценкой*.

Основными свойствами оценок, характеризующими их качество, являются *несмещённость*, *состоятельность* и *эффективность*.

Оценку $\tilde{\theta}$ параметра θ называют *несмещённой*, если

$$M(\tilde{\theta}) = \theta.$$

В противном случае оценка называется *смещённой*.

Оценку $\tilde{\theta}_n = \tilde{\theta}_n(X_1, X_2, \ldots, X_n)$ параметра θ называют *состоятельной*, если она сходится по вероятности к θ при $n \to \infty$, т.е. для любого $\varepsilon > 0$ выполняется соотношение

$$\lim_{n\to\infty} P(|\tilde{\theta}_n - \theta| \geq \varepsilon) = 0.$$

Несмещённая оценка $\tilde{\theta}$ параметра θ называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещённых оценок параметра θ.

Теорема 4.1. Выборочное среднее $\overline{X}$ является несмещённой и состоятельной оценкой математического ожидания $M(X)$.

Теорема 4.2. Выборочная дисперсия S^2 является смещённой и состоятельной оценкой дисперсии $D(X)$.

Легко доказать, что

$$M(S^2) = \frac{n-1}{n} D(X).$$

Замечание 4.3. Очевидно, статистика

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (4.4)$$

является несмещённой оценкой дисперсии $D(X)$. ●

Статистика S_0^2, определённая формулой (4.4), называется *исправленной выборочной дисперсией*.

Рассмотрим один из наиболее распространённых методов получения точечных оценок неизвестных параметров распределения – метод максимального правдоподобия.

Предположим, что известен вид закона распределения с.в. X, но неизвестен параметр θ, которым определяется этот закон (параметр θ может быть и векторным: $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$). Пусть $x_1, x_2, \ldots, x_n$ – выборка наблюдений с.в. X, по которой требуется оценить параметр θ.

Функцией правдоподобия для оценки параметра θ называется функция

$$L(\theta) = p(x_1,\ \theta)\, p(x_2,\ \theta)\ \ldots\ p(x_n,\ \theta),$$

где $p(x,\ \theta)$ – плотность распределения с.в. X, если X – непрерывная с.в., и $p(x,\ \theta) = P(X = x)$, если с.в. X дискретная.

Метод максимального правдоподобия состоит в том, что в качестве оценки параметра θ принимается статистика $\tilde{\theta}$ (векторная статистика $\tilde{\theta} = (\tilde{\theta}_1,\ \ldots,\ \tilde{\theta}_r)$, если $\theta = (\theta_1,\ \ldots,\ \theta_r)$), выборочное значение которой $\tilde{\theta}(x_1,\ x_2,\ \ldots,\ x_n)$ является точкой максимума функции правдоподобия. Такую оценку называют *оценкой максимального правдоподобия*.

Так как максимум функций $L(\theta)$ и $\ln L(\theta)$ достигается при одном и том же значении θ, то часто при нахождении оценки максимального правдоподобия используют функцию $\ln L(\theta)$.

Пример 4.3. Найти оценку максимального правдоподобия параметра λ распределения Пуассона.

Решение. В данном случае X – дискретная с.в.,

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots \quad (\lambda > 0).$$

Составляем функцию правдоподобия:

$$L(\lambda) = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda}\, \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} \ldots \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} = e^{-\lambda n} \cdot \lambda^{\sum\limits_{i=1}^{n} x_i} \cdot \frac{1}{x_1!\, x_2! \ldots x_n!}.$$

Следовательно,

$$\ln L(\lambda) = -\lambda n + \ln \lambda \cdot \left(\sum_{i=1}^{n} x_i \right) - \ln(x_1!\, x_2! \ldots x_n!).$$

Ищем точку максимума функции $\ln L(\lambda)$:

$$\frac{d \ln L(\lambda)}{d\lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0.$$

Отсюда получаем

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}.$$

Легко видеть, что $\lambda = \bar{x}$ является точкой максимума функции $\ln L(\lambda)$. Таким образом, $\tilde{\lambda} = \overline{X}$ – искомая оценка. ●

Задачи

1. Найти оценки максимального правдоподобия математического ожидания m и дисперсии σ^2 нормально распределенной генеральной совокупности.

2. Найти оценку максимального правдоподобия параметра θ показательного распределения с плотностью

$$p(x, \theta) = \begin{cases} (1/\theta)e^{-x/\theta} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0, \end{cases}$$

где $\theta > 0$.

3. Случайная величина X имеет плотность распределения

$$p(x) = \begin{cases} kx & \text{при } x \in [0, \sqrt{2/k}], \\ 0 & \text{при } x \notin [0, \sqrt{2/k}]. \end{cases}$$

Найти оценку максимального правдоподобия математического ожидания с. в. X по выборке объема n.

4. Найти оценку максимального правдоподобия параметра θ равномерного на отрезке $[0, \theta]$ распределения.

4.3. Интервальное оценивание. Доверительные интервалы для параметров нормального распределения

Пусть θ – неизвестный параметр распределения с.в. X.

Пусть статистики $\tilde{\theta}_1 = \tilde{\theta}_1(X_1, \ldots, X_n)$ и $\tilde{\theta}_2 = \tilde{\theta}_2(X_1, \ldots, X_n)$ таковы, что интервал $(\tilde{\theta}_1, \tilde{\theta}_2)$ содержит (накрывает) истинное значение параметра θ с заданной вероятностью $p = 1 - \alpha$:

$$P(\tilde{\theta}_1 < \theta < \tilde{\theta}_2) = 1 - \alpha .$$

Тогда интервал $(\tilde{\theta}_1, \tilde{\theta}_2)$ называется *доверительным интервалом* для параметра θ, вероятность $p = 1 - \alpha$ – *доверительной вероятностью (надёжностью)*, а число α – *уровнем значимости*.

Пусть X – нормально распределённая с.в. Тогда:

1. Если среднее квадратическое отклонение σ известно, то доверительный интервал для математического ожидания m имеет вид

$$\left(\overline{X} - u_{\alpha/2} \frac{\sigma}{\sqrt{n}},\ \overline{X} + u_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right),$$

где n – объём выборки, $\overline{X}$ – выборочное среднее, $u_{\alpha/2}$ находится по таблице значений функции Лапласа $\Phi_0(x)$ (см. п. 1.7) из условия

$$\Phi_0(u_{\alpha/2}) = \frac{1-\alpha}{2}.$$

2. Если среднее квадратическое отклонение σ неизвестно, то доверительный интервал для математического ожидания m имеет вид

$$\left(\overline{X} - t_{\alpha/2,\,n-1}\frac{S_0}{\sqrt{n}},\ \overline{X} + t_{\alpha/2,\,n-1}\frac{S_0}{\sqrt{n}}\right),$$

где $S_0 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ (исправленное выборочное среднее квадратическое отклонение), $t_{\alpha/2,n-1}$ находится по таблице квантилей распределения Стьюдента из условия

$$P\left(T_{n-1} > t_{\alpha/2,\,n-1}\right) = \alpha/2$$

(здесь T_{n-1} – с.в., распределённая по закону Стьюдента с $n-1$ степенью свободы).

Доверительный интервал для среднего квадратического отклонения σ нормально распределённой с.в. X имеет вид

$$\left(S_0\sqrt{\frac{n-1}{\chi^2_{\alpha/2,\,n-1}}},\ S_0\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2,\,n-1}}}\right),$$

где $\chi^2_{1-\alpha/2,\,n-1}$ и $\chi^2_{\alpha/2,n-1}$ находятся по таблице квантилей χ^2-распределения из условий

$$P(\chi^2_{n-1} < \chi^2_{1-\alpha/2,\,n-1}) = \alpha/2, \quad P(\chi^2_{n-1} > \chi^2_{\alpha/2,\,n-1}) = \alpha/2$$

(здесь χ^2_{n-1} – с.в., имеющая χ^2-распределение с $n-1$ степенью свободы).

Задачи

1. По результатам 6 независимых наблюдений над нормально распределённой с.в. X найдены выборочное среднее $\bar{x} = 5{,}63$ и исправленная выборочная дисперсия $s_0^2 = 0{,}0625$. Требуется найти доверительный интервал для математического ожидания с.в. X, соответствующий доверительной вероятности $1-\alpha = 0{,}99$.

2. Из нормально распределенной генеральной совокупности извлечена выборка объема $n = 10$. Результаты приведены в таблице:

Наблюденное значение (x_i)	–2	1	2	3	4	5
Частота (n_i)	2	1	2	2	2	1

Найти доверительный интервал для математического ожидания m генеральной совокупности. Доверительную вероятность $1 - \alpha$ принять равной 0,95.

3. Получены следующие результаты 5 независимых измерений толщины металлической пластинки: $x_1 = 2{,}015$ мм, $x_2 = 2{,}020$ мм, $x_3 = 2{,}025$ мм, $x_4 = 2{,}020$ мм, $x_5 = 2{,}015$. Считая, что толщина пластинки X – нормально распределенная случайная величина, найти доверительный интервал, накрывающий среднее квадратическое отклонение этой с.в. с надежностью $1 - \alpha = 0{,}98$.

4. Построить доверительный интервал с доверительной вероятностью $1-\alpha = 0{,}96$ для дисперсии σ^2 нормально распределенной случайной величины X, если исправленная выборочная дисперсия $s_0^2 = 10$, а объем выборки $n = 20$.

4.4. Проверка гипотез о виде закона распределения

Пусть необходимо проверить гипотезу, состоящую в том, что наблюдаемая в эксперименте с.в. X распределена по некоторому известному закону (нормальному, биномиальному, Пуассона и т.д.). Проверяемая гипотеза называется *нулевой* и обозначается H_0.

Для проверки гипотезы H_0 производится выборка значений с.в. X. Требуется сделать заключение: согласуются ли данные выборки с высказанным предположением. Для этого используют специально подобранную статистику $Z(X_1, \ldots, X_n)$, по выборочному значению которой судят о справедливости гипотезы H_0.

Правило, согласно которому проверяемая гипотеза H_0 принимается или отвергается, называется *критерием согласия*, а статистика $Z(X_1, \ldots, X_n)$, с помощью которой это правило определяется, называется *статистикой критерия*.

Поскольку проверка статистической гипотезы осуществляется на основании выборочных данных, носящих случайный характер, то всегда присутствует возможность ошибочно отвергнуть гипотезу H_0, когда на самом деле она верна. Поэтому заранее задаётся малое число α – вероятность, с которой мы можем позволить себе отвергнуть верную гипотезу H_0. Это число называют *уровнем значимости критерия*. Обычно для α используются стандартные значения: $\alpha = 0{,}001$; $0{,}005$; $0{,}01$; $0{,}05$; $0{,}1$.

Критериев согласия существует много. Мы рассмотрим наиболее часто используемый на практике критерий согласия – *критерий* χ^2 *Пирсона*.

Предположим, что сформулирована гипотеза H_0, состоящая в том, что с.в. X имеет закон распределения известного вида, зависящий от параметров $\theta_1, \ldots, \theta_r$ (например, нормальный закон с параметрами m и σ или закон Пуассона с параметром λ).

Пусть в результате наблюдений за с.в. X получена выборка объёма n: $x_1, x_2, \ldots, x_n$.

Для проверки гипотезы H_0 с помощью критерия χ^2 Пирсона поступают следующим образом.

1. Выбирают уровень значимости α.

2. По выборке методом максимального правдоподобия (см. п. 4.2) оценивают параметры предполагаемого закона распределения.

3. Область возможных значений с.в. X разбивают на k непересекающихся множеств $S_1, S_2, \ldots S_k$. Эти множества представляют собой интервалы в случае, когда X – непрерывная с.в., либо группы отдельных значений, если с.в. X дискретная.

4. Для каждого множества S_i, $i = 1, 2, \ldots, k$, подсчитывают число n_i элементов выборки, попавших в это множество (т.е. находят *эмпирические частоты*).

5. Используя предполагаемый закон распределения с.в. X, вычисляют гипотетические вероятности p_i попадания с.в. X в множества S_i:

$$p_i = P(X \in S_i), \quad i = 1, 2, \ldots, k.$$

6. Находят *теоретические частоты* n_i' попадания значений с.в. X в множества S_i:

$$n_i' = np_i.$$

7. Вычисляют выборочное значение $\chi^2_в$ статистики критерия χ^2 Пирсона – случайной величины

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n_i')^2}{n_i'}. \qquad (4.5)$$

Применение статистики (4.5) для проверки гипотезы H_0 основано на следующей теореме.

Теорема 4.3. Если гипотеза H_0 верна, то статистика (4.5) критерия χ^2 асимптотически при $n \to \infty$ распределена по закону χ^2 с $\nu = k - r - 1$ степенями свободы, где k – число множеств S_i, r – число параметров предполагаемого распределения, оцененных по выборке.

8. По таблице квантилей χ^2-распределения по заданному уровню значимости α и числу степеней свободы $\nu = k - r - 1$ находят значение $\chi^2_{\alpha,\nu}$, удовлетворяющее условию

$$P(\chi^2_\nu > \chi^2_{\alpha,\nu}) = \alpha .$$

9. Сравнивают значения $\chi^2_в$ и $\chi^2_{\alpha,\nu}$. В соответствии с критерием χ^2 Пирсона гипотеза H_0 принимается, если $\chi^2_в \le \chi^2_{\alpha,\nu}$ (в этом случае говорят также, что гипотеза H_0 согласуется с результатами наблюдений). Если $\chi^2_в > \chi^2_{\alpha,\nu}$, то гипотеза H_0 отклоняется.

Замечание 4.4. Необходимым условием применения критерия χ^2 Пирсона является выполнение неравенства $n'_i \ge 5$ для всех множеств S_i. Если для некоторых множеств S_i это условие не выполняется, то их следует объединить с соседними. ●

Задачи

1. При 50 подбрасываниях монеты герб появился 20 раз. Можно ли считать монету симметричной? Принять $\alpha = 0{,}10$.

2. Приведены данные об отказах аппаратуры за 10 000 часов работы (всего обследовано n = 757 экземпляров):

Число отказов, k	0	1	2	3	4	5	≥ 6
Количество случаев, в которых наблюдалось k отказов, n_k	427	235	72	21	1	1	0

Проверить гипотезу H_0 о том, что число отказов имеет распределение Пуассона. Принять $\alpha = 0{,}10$.

3. Выборка представлена следующим интервальным статистическим рядом:

$x_{i-1} - x_i$	3,0–3,6	3,6–4,2	4,2–4,8	4,8–5,4	5,4–6,0	6,0–6,6	6,6–7,2
n_i	2	8	35	43	22	15	5

При уровне значимости $\alpha = 0{,}01$ проверить гипотезу о нормальном распределении генеральной совокупности.

4. Выборка представлена интервальным статистическим рядом:

$x_{i-1} - x_i$	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
n_i	133	45	15	4	2	1

При уровне значимости $\alpha = 0{,}05$ проверить гипотезу о показательном распределении генеральной совокупности.

ЛИТЕРАТУРА

1. Бородич, С.М. Теория вероятностей и математическая статистика: методические рекомендации / С.М. Бородич, Т.В. Кавитова. – Витебск : ВГУ имени П.М. Машерова, 2015. – 51 с.

2. Бородич, С.М. Теория вероятностей: задания для самостоятельной работы / С.М. Бородич, Т.В. Кавитова. – Витебск : ВГУ имени П.М. Машерова, 2016. – 52 с.

3. Бочаров, П.П. Теория вероятностей. Математическая статистика / П.П. Бочаров, А.В. Печинкин. – 2-е изд. – М.: ФИЗМАТЛИТ, 2005. – 296 с.

4. Герасимович, А.И. Математическая статистика / А.И. Герасимович, Я.И. Матвеева. – Минск: Выш. шк., 1978. – 200 с.

5. Гмурман, В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2003. – 479 с.

6. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для студентов вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2004. – 404 с.

7. Гнеденко, Б.В. Курс теории вероятностей: учебник / Б.В. Гнеденко. – 8-е изд., испр. и доп. – М.: Едиториал УРСС, 2005. – 448 с.

8. Емельянов, Г.В. Задачник по теории вероятностей и математической статистике: учеб. пособие / Г.В. Емельянов, В.П. Скитович. – 2-е изд., стер. – СПб. [и др.]: Лань, 2007. – 332 с.

9. Кремер, Н.Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н.Ш. Кремер. – 2-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2004. – 573 с.

10. Максимов, Ю.Д. Математика. Выпуск 8. Математическая статистика: опорный конспект / Ю.Д. Максимов. – СПб.: Изд-во СПбГПУ, 2002. – 96 с.

11. Теория вероятностей: учеб. для вузов / А.В. Печинкин [и др.]; под ред. В.С. Зарубина, А.П. Крищенко. – 3-е изд., испр. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2004. – 456 с.

12. Чистяков, В.П. Курс теории вероятностей / В.П. Чистяков. – М.: Агар, 2000. – 255 с.

Учебное издание

БОРОДИЧ Сергей Митрофанович

КАВИТОВА Татьяна Валерьевна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации

Технический редактор *Г.В. Разбоева*

Компьютерный дизайн *Т.Е. Сафранкова*

Подписано в печать .2017. Формат 60x84 $^{1}/_{16}$. Бумага офсетная.

Усл. печ. л. 3,02. Уч.-изд. л. 3,38. Тираж экз. Заказ .

Издатель и полиграфическое исполнение – учреждение образования «Витебский государственный университет имени П.М. Машерова».

Свидетельство о государственной регистрации в качестве издателя, изготовителя, распространителя печатных изданий № 1/255 от 31.03.2014 г.

Отпечатано на ризографе учреждения образования «Витебский государственный университет имени П.М. Машерова».

210038, г. Витебск, Московский проспект, 33.