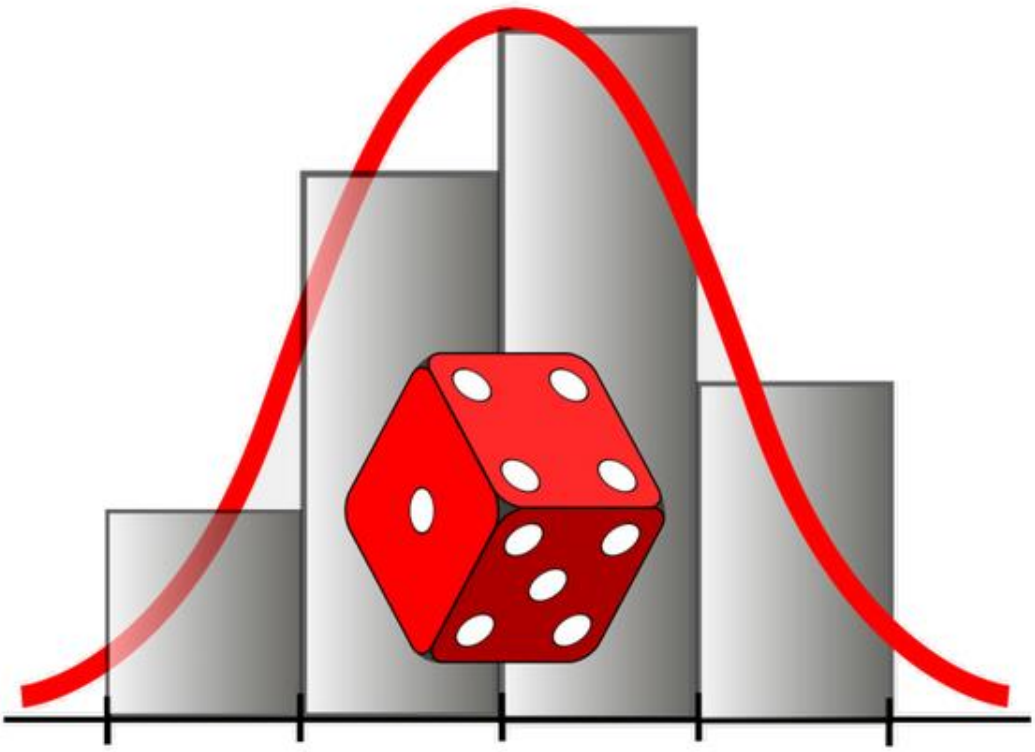
Федеральное агентство по образованию

Государственное образовательное учреждение
высшего профессионального образования

Воронежский государственный технический университет

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие



Воронеж 2022

УДК 517.8
ББК 22.17

Рецензенты:
кафедра высшей математики
МО РФ Воронежского государственного педагогического университета;
Задорожний В. Г., д.ф.-м.н., проф. кафедры системного анализа
ФГБОУ ВО «Воронежский государственный университет»

Акчурина Л.В.
Теория вероятностей и математическая статистика. Учебное пособие / Воронеж. гос. технического ун-та; сост.: Л.В. Акчурина, А.Б. Кущев, С.С. Сумера – Воронеж, 2022. – 117 с.

Учебное пособие содержит основные сведения по теории вероятностей и по математической статистике, иллюстрируемые большим набором задач. Содержит индивидуальные задания. Приведены 25 вариантов заданий. Предназначены для студентов тех специальностей, для которых этот раздел математики входит в учебный план.

Предназначено для студентов направления 08.03.01 «Строительство»

Ил. 31. Табл. 10. Библиогр.: 7 назв.

УДК 517.8
ББК 22.17

Печатается по решению редакционно-издательского совета Воронежского государственного технического университета.

ISBN

ВВЕДЕНИЕ

Без преувеличения можно сказать, что нас окружает во многом случайный мир. Мы часто пытаемся предугадать – случится ли то или иное событие, проработает ли заявленный срок приобретенная техника, совпадут ли номера в спортлото – иначе говоря, произойдет ли некоторое событие или нет. В то же время со школьной скамьи складывается убеждение, что всё в мире происходит по однозначным, строгим законам. Поэтому кажется, что и там, где имеет место случай, нам просто пока не хватает данных и знаний, чтобы объяснить, почему произошло именно это событие, а не другое, но оказалось, что дело не только в нехватке знаний. Например, в микромире, как это следует из законов квантовой механики, все устроено случайным образом. Оказалось, что нельзя достоверно утверждать, расположен электрон в данной области пространства или нет, а можно говорить только о его шансах нахождения там. Раздел математики, разработанный для случайных событий, которые не могут быть строго или точно предсказаны, называется **теорией вероятностей**. На ее основе получила развитие наука **математическая статистика**, она применяется для математической обработки информации – это контроль качества выпускаемой продукции, анализ различных процессов, анализ больших массивов данных и т. д. В конце учебного пособия приведены 25 вариантов индивидуальных заданий. Каждый вариант содержит 12 задач, нумерация примеров сквозная, поэтому номера каждого варианта идут с интервалом в 25 единиц. Например, для варианта 5 следует выполнять задания под №№ 5, 30, 55, 80, 105 и т. д.

Первое задание на непосредственный подсчет вероятности или задача, решаемая с использованием формул комбинаторики. Вторая и третья задачи на формулы вероятностей суммы, произведения вероятностей и формулу полной вероятности. Четвертая задача на схему Бернулли. Пятое задание на нахождение закона распределения дискретной случайной величины и вычисление ее основных числовых характеристик. В шестом задании непрерывная случайная величина задана с помощью функции плотности распределения и с помощью функции распределения, соответственно. Седьмое и восьмое задания на нормальный закон и другие основные законы распределения, девятое задание – на корреляционный анализ. Наконец, десятое и одиннадцатое задания на изучение математической статистики.

Для большинства направлений это последнее расчетно-графическое задание (РГЗ) по курсу математики. В отличие от предыдущих РГЗ здесь много интересных задач из окружающего нас мира, строительного производства. Применяемый для их решения математический аппарат доступен даже школьнику, однако существенной трудностью является то, что желательно проникнуться качественно новым, вероятностным взглядом на окружающий мир. Поэтому, если сразу не удается справиться с задачей, посмотрите в соответствующем разделе данного пособия (а также в другой литературе) образцы решения задач по этой теме, подумайте сами, проконсультируйтесь – и вероятность достигнуть правильного ответа устремится к единице. Успеха вам!

1. СЛУЧАЙНЫЕ СОБЫТИЯ. ВЕРОЯТНОСТЬ СЛУЧАЙНОГО СОБЫТИЯ

1.1. Основные понятия

Изложение теоретического материала в любой науке начинается с введения основных (базовых) определений, на которых далее будет строится вся теория. Исходным пунктом науки теория вероятностей, а тем более математическая статистика, являются экспериментальные данные, полученные в результате наступления некоторых условий. Если перейти на более простой уровень понимания, то изучается некоторый факт, который произошел в итоге наступления некоторых предшествующих ему событий, которые называются опытом. Далее необходимо более четко ввести основные базовые определения.

Опытом (испытанием) будем называть комплекс наступающих одновременно условий, которые при повторении одного и того же опыта должны строго повторяться. Опыт – это игра в шахматы, условия – это наличие игральной доски со всеми фигурами и двух игроков, а факт – результат партии.

Качественная характеристика опыта состоит в регистрации какого-нибудь факта, чаще его называют событием. При этом говорят, что «событие появилось (произошло)» или «событие не появилось (не произошло)» в результате опыта. Далее обозначать события будем прописными латинскими буквами A, B, C.

Событие называется ***случайным***, если оно может произойти или не произойти в результате определенного опыта.

Событие называется ***достоверным***, если оно обязательно появится в результате проведения опыта.

Событие называется ***невозможным***, если оно никогда не появится в результате проведения опыта.

В качестве примера рассмотрим опыт – бросание игральной кости (кубика с указанными числами первых шести цифр на его гранях). Само бросание кубика – это опыт. Выпадение в результате опыта любого числа от “1” до “6” – это достоверное событие; выпадение “10” есть событие невозможное; а выпадение цифры “6” есть случайное событие, т. к. оно может и не произойти.

Событие $\bar{A}$ называется ***противоположным*** **(*дополнительным*)** событию A, если оно означает не осуществление события A.

Для примера опять “бросим” кубик, где событием A будем считать выпадение четного числа, соответственно, противоположное $\bar{A}$ – это выпадение нечетного числа.

События называются ***несовместными***, если появление одного из них в единичном испытании исключает появление другого. Например, при бросании монеты появление герба и решки в одном испытании − события несовместные.

События называются ***совместными,*** если их совместное появление в одном испытании не исключено.

Снова в опыте с игральным кубиком обозначим A – выпадение четного числа, B – выпадение нечетного числа, C – выпадение “5”, тогда A и C события несовместные, а B и C – совместные.

Основной идеей теории вероятности является то, что каждому событию A можно поставить в соответствие число $p=p(A)$ $(0 \le p(A) \le 1)$, характеризующее частоту наступления события A (см. стр. 7). Вторым важнейшим понятием является понятие случайной величины (см. стр. 25).

1.2. Понятие вероятности. Классическое, геометрическое и статистическое определения вероятности

Анализируя комплекс условий, повторяющийся при проведении испытания, иногда можно выделить так называемое ***множество элементарных исходов* Ω**. Это множество равновероятных исходов $\Omega = \{\omega_1,\ \omega_2, \ldots,\ \omega_n\}$ включает в себя единственно возможные и попарно несовместные в одном испытании исходы.

Пусть опытом будет бросание игрального кубика. Испытанием в этом опыте является фиксирование числа очков на верхней грани кубика после его бросания. Множество исходов Ω состоит из 6 единственно возможных, несовместных между собой элементарных исходов – 1, 2, 3, 4, 5, 6.

Если испытание состоит в определении оценки, полученной студентом на экзамене, то здесь множество Ω состоит из 4 единственно возможных и несовместных элементарных исходов, соответствующих возможным оценкам $\Omega = \{2, 3, 4, 5\}$, т. к. на экзамене возможно получить только одну из этого набора оценок.

Благоприятным исходом событию A называется такой исход опыта, при котором событие A произошло. Например, считается, что студент сдал экзамен, если полученная им на экзамене оценка не ниже оценки “3”, тогда благоприятным исходом событию A – экзамен сдан, является получение любой оценки из набора – “3”, “4” или “5”, т. е. $\Omega_A = \{3, 4, 5\}$.

По ***классическому определению*** вероятностью появления события A, называется отношение числа $\boldsymbol{m}$ благоприятных этому событию элементарных исходов к общему числу $\boldsymbol{n}$ всех равновозможных несовместных элементарных исходов $\Omega = \{\omega_1,\ \omega_2, \ldots,\ \omega_n\}$:

$$p(A) = \frac{m}{n}. \tag{1}$$

При сдаче студентом экзамена можно подумать, исходя из формулы (1), что вероятность сдать экзамен всегда равна 3/4, т. к. из возможных четырех оценок три положительные, и это действительно было бы так, если бы испытуемый вместо билета тянул карточки с оценками – одну из четырех. Однако это не так, оценка зависит от объема выученного к экзамену теоретического материала и умения его применения на практике.

В опыте с игральной костью вероятностью события A – выпадения четного числа, где благоприятных исходов m только три $\Omega_A = \{2, 4, 6\}$, а общее число исходов n равно шести, имеет вероятность $p(A) = 3/6 = 0{,}5$. Полученная вероятность подсказывает вывод, что при бросании кубика каждый второй раз должно выпадать четное число, однако это совсем не обязательный итог каждого второго испытания. Полученная вероятность обозначает то, что при большом количестве бросаний кубика в среднем половина испытаний завершится выпадением четного числа.

Пример 1. Два раза бросается монета. Какова вероятность того, что хотя бы один раз выпадет “орел” (событие A)?

Решение. Равновозможными несовместными элементарными исходами являются четыре пары исходов (о,о), (о,р), (р,о) и (р,р), где буквами “о” и “р” обозначены выпадение, соответственно, орла или решки; общее число элементарных исходов $n = 4$. Событию A соответствуют первые три пары исходов, поэтому число благоприятных исходов $m = 3$. Следовательно, $p(A) = \frac{m}{n} = \frac{3}{4}$.

Одним из недостатков классического определения вероятности является то, что в нем предполагается конечное множество числа исходов испытания. Чтобы преодолеть этот недостаток А. Н. Колмогоров ввел геометрическое определение вероятности. Иногда очевидным является то, что множество исходов представляется как интервал (отрезок) или как некоторое плоское множество и т. д. Классический случай – вероятность попасть “в 10” при стрельбе из оружия. Пронумеровать и пересчитать множество исходов такого опыта не представляется возможным, но геометрически нарисовать просто, где полезными исходами естественно считать площадь круга с числом “10”.

Геометрической вероятностью события A называется отношение меры области, благоприятствующей данному событию к мере всей области:

$$p(A) = \frac{mes\Omega_A}{mes\Omega}, \tag{2}$$

где $mes\Omega_A$ – мера области исходов, при которых событие A произошло, а $mes\Omega$ - это мера всей области исходов опыта.

Пример 2. На отрезок длины 5 см произвольным образом поставлена точка. Какова вероятность того, что расстояние от нее до краев отрезка будет не меньше 2 см?

Решение. От каждого из концов отрезка откладываем по 2 см – сюда попадать нельзя, тогда благоприятным является только средняя часть отрезка длиной в 1 см, а значит, вероятность выполнения условия задачи

$$p(A) = \frac{mes\Omega_A}{mes\Omega} = \frac{1}{5} = 0{,}2.$$

Замечание. В некоторых задачах и в теории удобно все множество исходов обозначать простейшим множеством меры равной единице. На плоскости –

это квадрат со стороной, равной 1: $mes\Omega$ =S□=1, тогда $mes\Omega_A$ равна площади фигуры, соответствующей событию A.

Пример 3. Два друга договорились о встрече в определенном месте между 12 и 13 часами дня. Пришедший первым ждет второго 15 минут, а не дождавшись, уходит. Найти вероятность встречи, считая, что время прихода каждого в заданном часовом интервале в любой момент времени равновероятно.

Решение. Пусть время прихода первого в назначенное место есть x, а второго y, тогда множество всех исходов будет квадрат со стороной равной 1 (на рис 1 это квадрат OACB). Переведя 15 минут в часы, получим условие встречи

$$|y-x| \le \frac{1}{4},$$

или, что тоже самое,

$$x-\frac{1}{4} \le y \le x+\frac{1}{4}.$$

Считая встречу интересующую нас событием A, двойное неравенство задает заштрихованное множество на рисунке 1 – это шестиугольник OEFCHG.

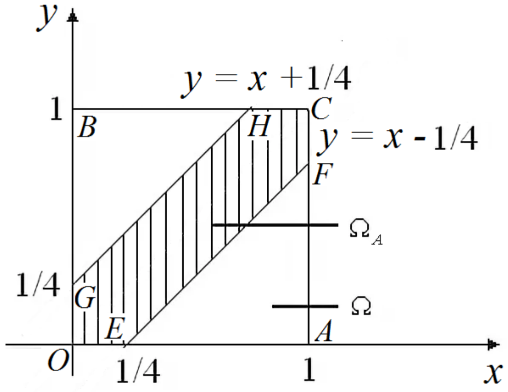


Рис. 1

Очевидно, вероятность события A

$$p(A)=S_{OEFCHG}=S_{OACB}-S_{AEF}-S_{BHG}=1-\frac{1}{2}\left(\frac{3}{4}\right)^2-\frac{1}{2}\left(\frac{3}{4}\right)^2=\frac{7}{16}.$$

На основе рассмотренных определений выделим основные свойства вероятности.

Свойства вероятности

1. Из самого определения вероятности, задаваемого формулой (1), следует, что вероятность ***случайного*** события $0 < p(A) < 1$, т. к. представляет собой отношение двух натуральных чисел, где количество всех исходов всегда больше, чем число благоприятных исходов.

   Отметим, что вероятность любого события $0 \le p(A) \le 1$.

2. Вероятность ***невозможного*** события A равна нулю:

$$p(A) = 0. \quad (3)$$

3. Вероятность ***достоверного*** события D равна единице:

$$p(D) = 1. \quad (4)$$

4. Вероятность случайного события $\overline{A}$, ***противоположного*** событию A, равна дополнению $p(A)$ до единицы, т. е.

$$p(\overline{A}) = 1 - p(A), \quad (5)$$

где $A + \overline{A} = D$ – достоверное событие.

Смысл формулы (5) хорошо проиллюстрирован в примере 2, где “полезную площадь” определили вычитанием (см. так же формулу (15) на стр. 14).

Однако чаще всего на практике ни классический, ни геометрический подход применить невозможно. В этом случае находят вероятность опытным путем. Проводится достаточно большое количество испытаний n в одинаковых условиях, в результате каждого из которых интересующее событие A может появиться или не появиться. Подсчитывают количество n_A появления событий A, статистической вероятностью называют

$$p^*(A) = \frac{n_A}{n}. \quad (6)$$

В математической статистике формула (6) задает относительную частоту, она может отличаться по значению для различных серий опытов. Например, пусть стрелок сделал 100 выстрелов и попал 89 раз, тогда для события A – попадание в цель конкретным стрелком при одном выстреле $p^*(A)=0.89$, но это не значит, что в каждой серии из 100 выстрелов всегда будет 89 попаданий. Это может быть или 87, или 90 попаданий, или другое близкое число. Число, около которого будет колебаться $p^*(A)$ в различных сериях опытов при увеличении количества серий опытов и будет интересующей нас вероятностью $p(A)$.

1.3. Элементы комбинаторики

Изучение данной темы традиционно начинают с классической игры в карты. Например, наилучшей комбинацией является возможность иметь при сдаче шести карт на руках четырех тузов. А какова вероятность данного события, зная формулу классической вероятности, это можно подсчитать, но проблема в том, как суметь пересчитать все благоприятные исходы и всевозможные исходы этого опыта, ведь простой перебор может потребовать значительные временные ресурсы.

Еще один пример – это игра в спортлото 5 из 36, где успешным будем считать выпадение единственной комбинации набора шаров, а прямое перечисление комбинаций шаров очень трудоемко. В этих случаях на помощь приходит раздел математики под названием комбинаторика.

Комбинаторика – это раздел математики, занимающийся изучением количества комбинаций, которые можно составить из элементов заданного конечного множества (или его подмножества).

Числом размещений A_n^k (читается “число размещений из n по k”) называется число различных способов упорядочить k различных предметов из всего набора, содержащего n предметов ($k \le n$, т. е. выбирается подмножество), где комбинации отличаются набором предметов и порядком их следования:

$$A_n^k = n(n-1)\times\ldots\times(n-k+1) = \frac{n!}{(n-k)!}. \qquad (7)$$

Смысл данной формулы будет более понятен, если представить, что мы рассаживаем на k пронумерованных стульев n человек и пересчитываем количество способов их размещения ($k \le n$, т. е. не всем хватит места на стульях). Очевидно, что на первый стул может сесть один из n человек (это первый множитель в произведении формулы (7)). На второй стул может сесть один из (n-1) человек, один уже сидит на первом стуле – это второй множитель в произведении. Соответственно, на третий стул может сесть уже один из (n-2) претендентов. И т. д., на последний k стул, при условии, что (k-1) стульев уже заняты, имеются $(n-(k-1))$ претендентов – это последний множитель в произведении. Далее формулу записывают в более компактном виде, что возможно с применением факториала $n! = 1\cdot 2\cdot\ldots\cdot n$.

В частном случае, когда $k = n$, т. е. работаем со всеми элементами множества, получается ***число перестановок***

$$P_n = A_n^n = \frac{n!}{(n-n)!} = n!, \text{ (заметим, что } 0! = 1). \qquad (8)$$

Числом сочетаний $\boldsymbol{C_n^k}$ называется число различных способов упорядочивания k различных предметов из всего набора, содержащего n предметов ($k \le n$, т. е. тоже выбирается подмножество из k элементов), комбинации отличаются только набором предметов, порядок их следования не важен:

$$C_n^k = \frac{A_n^k}{P_k} = \frac{n!}{(n-k)!k!}. \qquad (9)$$

Отметим, что число сочетаний из n по k дает количество его подмножеств, содержащих k элементов.

Пример 4. Пусть на окружность нанесено десять точек. Вычислим, сколько возможно по этим точкам построить: а) хорд; б) векторов отличной от нуля длины; в) треугольников.

Решение.

а) Хорда – это отрезок, соединяющий две точки окружности. Мы выбираем две точки из десяти. Порядок следования точек в выбранных двух неважен, тогда применяем формулу (9) имеем $C_{10}^2 = \frac{10!}{(10-2)!2!} = \frac{8!\cdot 9\cdot 10}{8!2!} = \frac{9\cdot 10}{2!} = 45$ хорд.

б) Вектор – это направленный отрезок, на одной хорде есть два вектора, которые отличаются направлением, т. е. ответ уже очевиден – 90 векторов.

Но и здесь применим комбинаторику. Чтобы составить вектор, нужно выбирать 2 точки из 10, а т. к. порядок следования элементов важен, то возьмем формулу размещений (7) и получим $A_{10}^2 = \frac{10\,!}{(10-2)\,!} = \frac{10\,!}{8\,!} = \frac{8\,!\cdot 9\cdot 10}{8\,!} = 90$ векторов**.**

в) Чтобы вычислить число треугольников, нужно выбирать 3 точки из 10 с применением формулы сочетаний, т. к. нумеровать вершины в этой геометрической фигуре не нужно, то получим $C_{10}^3 = \frac{10!}{(10-3)!3!} = \frac{7!\cdot 8\cdot 9\cdot 10}{7!3!} = \frac{8\cdot 9\cdot 10}{1\cdot 2\cdot 3} = 120$ треугольников.

Ответ: а) 45 хорд; б) 90 векторов; в) 120 треугольников.

Пример 5. Возвратимся к задаче о спортлото 5 из 36. Поскольку выигрышная комбинация одна и перемешивание шаров в ней не меняет сути, поэтому для вычисления всех исходов можно применять формулу сочетаний, считая m = 1, а $n = C_{36}^5 = \frac{36\,!}{(36-5)\,!5!} = \frac{36\,!}{31!5!} = \frac{31!\cdot 32\cdot 33\cdot 34\cdot 35\cdot 36}{31!\cdot 1\cdot 2\cdot 3\cdot 4\cdot 5} = 376992$.

Вероятность выигрыша равна $p(A) = \frac{1}{C_{36}^5} = \frac{1}{\frac{36\,!}{31!5!}} = 0{,}0000027$.

Ответ будет аналогичный, если считать полезными комбинациями перестановки из 5 элементов, тогда число всех комбинации исходов считаем через размещения $p(A) = \frac{P_5}{A_{36}^5} = \frac{5!}{\frac{36!}{31!}} = \frac{5!\cdot 31!}{36!} = \frac{P_6}{376992} = 0{,}0000027$.

Пример 6. Студент идет на экзамен, зная 16 вопросов из 25. Студент точно сдаст экзамен, если из трех вопросов в билете два будут ему известны. Какова вероятность студенту сдать экзамен?

Решение. Множество всех элементарных исходов равно количеству способов, сколькими можно выбрать 3 вопроса из 25, а так как порядок следования вопросов в билете не важен, то это будет число сочетаний $n = C_{25}^3$ $= \frac{25!}{(25-3)!3!} = \frac{23\cdot 24\cdot 25}{3!} = 2300$.

Определим число исходов, благоприятных событию A - среди трех вопросов не менее двух ему известны:

• два известных вопроса можно выбрать C_{16}^2 способами, а один неизвестный из оставшихся 9 невыученных (25–16=9) можно выбрать C_9^1 способами; число таких благоприятных событию A исходов равно произведению полученных вариантов: $m_1 = C_{16}^2 C_9^1 = \frac{16!\cdot 9}{(16-2)!2!} = \frac{15\cdot 16}{2!}\cdot 9 = 1080$;

• благоприятным исходом так же будет билет с тремя выученными вопросами, а значит, $m_2 = C_{16}^3 C_9^0 = \dfrac{16!}{(16-3)!3!} = \dfrac{14 \cdot 15 \cdot 16}{3!} = 560$.

Очевидно, что количество комбинаций вытянуть билет с двумя или с тремя выученными вопросами равно 1080+560=1640, тогда вероятность сдать экзамен для конкретного студента $p(A) = \dfrac{1640}{2300} = 0{,}71$.

Пример 7. Сборная конструкция состоит из четырех объемных элементов, которые завозятся на стройку по одному случайным образом. Какова вероятность того, что конструкция будет смонтирована без простоев, связанных с завозом, если 1) все элементы различны; 2) среди них два одинаковых?

Решение.

Общее число исходов одинаково в обеих задачах и равно числу перестановок $n = P_4 = 4! = 24$.

1) В первом случае благоприятным исходом m_1 будет один, т.к. задержек не будет только в том случае, если элементы конструкции завозятся в том же порядке, в котором они должны быть смонтированы. Следовательно, $p_1 = \dfrac{m_1}{n} = \dfrac{1}{24} \approx 0{,}04$.

2) Во втором случае число благоприятных исходов m_2 равно двум, так как одинаковые элементы можно поменять между собой. Следовательно, $p_2 = \dfrac{m_2}{n} = \dfrac{2}{24} \approx 0{,}08$.

В теории вероятностей над событиями, как над объектами изучения, можно производить различные операции, аналогичные действиям над множествами:

- умножение (пересечение);
- сложение (объединение);
- вычитание (дополнение).

1.4. Вероятность произведения событий

Произведением **(*пересечением*)** событий АВ (или $A \cap B$) называется событие, состоящее в совместном наступлении обеих этих событий – заштрихованная область на рисунке 2.

Случайные события *A* и *B* называются ***независимыми***, если наступление одного из них не влияет на вероятность появление второго; в противном случае они будут ***зависимыми***.

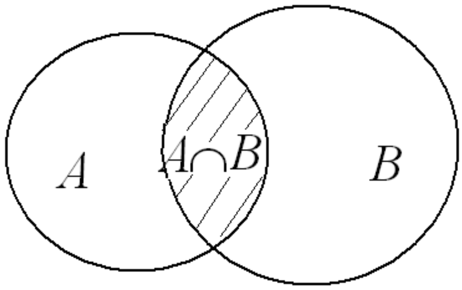


Рис. 2. Пересечение множеств

Например, два события – выпадение цифры “5” на верхней грани каждого из двух бросаемых игральных кубиках есть события независимые. А опыт, состоящий из двух последовательных событий, – извлечение из коробки с карандашами последовательно двух карандашей красного цвета, если их изначально было всего 2 из 10, есть события зависимые.

Пусть $P(A)$ и $P(B)$ – вероятности наступления случайных событий A и B, подсчитанные до испытаний. Пусть событие A произошло, что изменило вероятность наступления события B, тогда вероятность наступления события B обозначают $P(B/A)$ или $P_A(B)$ и называют ***условной вероятностью***.

Итак, ***условная вероятность*** $P(B/A)$ – это вероятность наступления события B, рассчитанная при условии, что событие A уже произошло.

Теорема 1. (*Вероятность произведения зависимых событий*).

Вероятность произведения двух зависимых событий равна произведению вероятности первого из наступивших событий на условную вероятность наступления второго события:

$$p(AB) = p(A)p_A(B) \qquad (\text{или } p(AB) = p(B)p_B(A)) \tag{10}$$

Доказательство. Воспользуемся геометрическим определением вероятности. На рисунке 3 множество Ω – есть все множество исходов, а $\Omega_{A\cap B}$ – исходы, соответствующие выпадению и события A и события B.

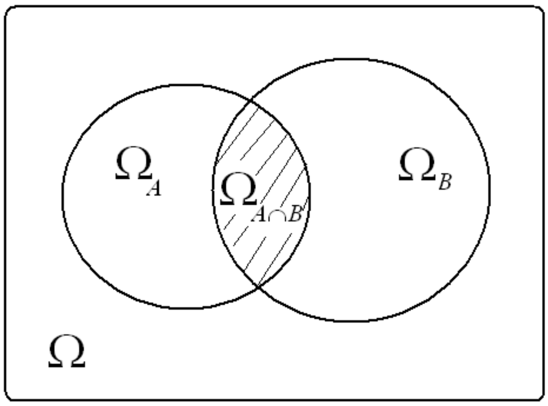


Рис. 3. Произведение событий $\Omega_{A\cap B}$ на всем множестве исходов Ω

Тогда $p(AB) = \dfrac{\Omega_{A\cap B}}{\Omega} = \dfrac{\Omega_{A\cap B}}{\Omega}\dfrac{\Omega_A}{\Omega_A} = \dfrac{\Omega_A}{\Omega}\cdot\dfrac{\Omega_{A\cap B}}{\Omega_A} = p(A)\cdot p_A(B)$, что и требовалось доказать.

Теорема 2. (Вероятность произведения независимых событий).

Вероятность произведения двух независимых событий равна произведению их вероятностей:

$$p(AB) = p(A)p(B) \quad . \tag{11}$$

Доказательство. Независимость событий A и B обозначает, что выпадение любого первого из них не изменяет вероятности появления последующего, т. е. справедливы равенства $p(B) = p_A(B)$ и $p(A) = p_B(A)$. Из равенства (10) получим $p(AB) = p(A)p_A(B) = \left| \begin{matrix} \text{т.к.} \\ p_A(B) = p(B) \end{matrix} \right| = p(A)p(B)$, из чего следует (11).

Замечание. Теоремы можно обобщить на большее число событий:

- для зависимых событий

$$p(ABC) = p(A)p_A(B)p_{AB}(C); \tag{12}$$

- для независимых событий

$$p(ABC) = p(A)p(B)p(C); \tag{13}$$

Пример 8. Перегорела одна из трех лампочек, включенных в сеть последовательно. Произвольно выбранную лампочку заменяют хорошей, после чего проверяют работу линии. Если линия не
исправна, то заменяется другая лампочка. Какова вероятность того, что повреждение будет устранено после второй замены лампочки.

Решение. Пусть событие A – линия исправна после второй замены лампочки. Событие A появляется в результате последовательного наступления (произведения) событий: A_1 – линия неисправна после первой замены лампочки, A_2 – линия исправна после второй замены лампочки, т. е. $A = A_1A_2$. Поскольку это зависимые события, то по формуле (10) $p(A_1A_2) = p(A_1)p_{A_1}(A_2)$.

Вероятность заменить первый раз годную лампочку, которых две из трех, $p(A_1) = 2/3$. Вероятность во второй раз заменить перегоревшую лампочку – одну из двух оставшихся непроверенных, $p_{A_1}(A_2) = 1/2$, тогда получим

$$p(A) = p(A_1A_2) = p(A_1)p_{A_1}(A_2) = \frac{2}{3}\cdot\frac{1}{2} \approx 0{,}33.$$

1.5. Вероятность суммы событий

Суммой (***объединением***) событий называется событие, состоящее в наступлении хотя бы одного из них – заштрихованная область на рисунке 4.

Если события ***совместны***е, то их сумма $A + B$ (или $A \cup B$) означает наступление или события A или B, или их обоих.

Если A и B несовместные, то их сумма означает появление или события A или события B.

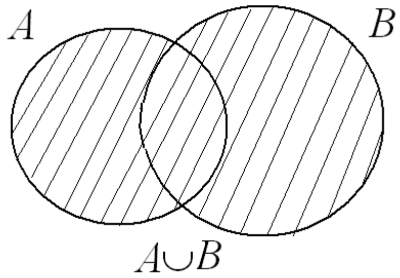


Рис. 4. Объединение множеств

Теорема 3. (*Вероятность суммы совместных событий*).

Вероятность суммы двух совместных событий равна сумме их вероятностей без вероятности их произведения:

$$p(A+B) = p(A) + p(B) - p(AB). \tag{14}$$

Доказательство. Воспользуемся геометрическим определением вероятности. На рис. 5 множество Ω – есть все множество исходов, а $\Omega_{A\cup B}$ – исходы, соответствующие выпадению или событию *A* или событию *B*.

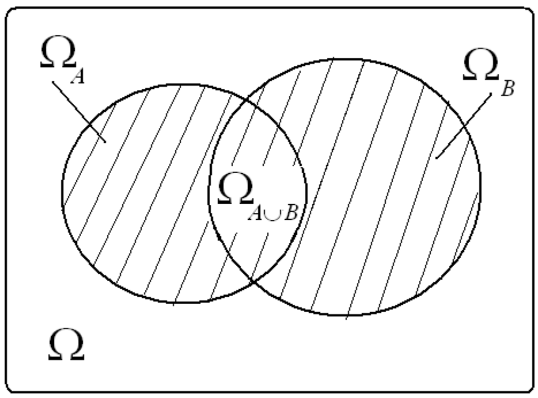


Рис. 5. Сумма событий $\Omega_{A\cup B}$ на всем множестве исходов Ω

$p(A+B) = \dfrac{\Omega_{A\cup B}}{\Omega} = \dfrac{\Omega_A + \Omega_B - \Omega_{A\cap B}}{\Omega} = \dfrac{\Omega_A}{\Omega} + \dfrac{\Omega_B}{\Omega} - \dfrac{\Omega_{A\cap B}}{\Omega} = p(A) + p(B) - p(AB)$, что и требовалось доказать.

Теорема 4. (*Вероятность суммы несовместных событий*).

Вероятность суммы двух несовместных событий равна сумме их вероятностей:

$$p(A+B) = p(A) + p(B)\ . \tag{15}$$

Доказательство. Несовместность событий *A* и *B* обозначает, что в одном опыте их совместное появление исключается, т. е. верно $p(AB) = 0$.

После подстановки $p(AB) = 0$ в формулу (14) получим частный случай для несовместных событий:

$$p(A+B)=p(A)+p(B)-p(AB)=\left|\begin{matrix}\text{т.к.}\\ p(AB)=0\end{matrix}\right|=p(A)+p(B),$$ из чего имеем (15).

Замечание. Теоремы можно обобщить на большее число слагаемых:

- для несовместных событий
$$p(A+B+C)=p(A)+p(B)+p(C); \qquad (16)$$
- для совместных событий
$$p(A+B+C)=p(A)+p(B)+p(C)-p(AB)-p(AC)-p(BC)+p(ABC). \qquad (17)$$

В случае *n* событий будут складываться вероятности комбинации каждого из *n* событий, вычитаться вероятности всевозможных комбинации из двух событий, складываться вероятности всевозможных комбинации из трех событий, вычитаться вероятности всевозможных комбинации из четырех событий и т. д.

Пример 9. По мишени стреляют три стрелка. Вероятность попадания в мишень первым стрелком – 0,9, вторым – 0,8 и третьим – 0,5. Найти вероятность следующих событий:

B_0 - никто не попал в мишень;

B_1 - только один стрелок попал в мишень;

B_2 - только два стрелка попали в мишень;

B_3 - все три стрелка попали в мишень;

B_4 - хотя бы один стрелок попал в мишень.

Решение. Заметим, что каждое из событий B_k состоит из нескольких событий – попадания или не попадания в цель каждого из трех стрелков.

Введем обозначения:

A_1 - первый стрелок попал, $p(A_1)=0{,}9$, $p(\overline{A}_1)=1-p(A_1)=1-0{,}9=0{,}1$;

A_2 - второй стрелок попал, $p(A_2)=0{,}8$, $p(\overline{A}_2)=1-p(A_2)=1-0{,}8=0{,}2$;

A_3 - третий стрелок попал, $p(A_3)=0{,}5$, $p(\overline{A}_3)=1-p(A_3)=1-0{,}5=0{,}5$.

С целью упрощения решения начнем с события B_3 – все три стрелка попали в мишень. Событие B_3 состоит в том, что и первый стрелок попал в цель, и второй попал, и третий попал, следовательно, $B_3=A_1A_2A_3$. Так как события A_1, A_2 и A_3 независимы, то по формуле (13)

$$p(B_3)=p(A_1A_2A_3)=p(A_1)p(A_2)p(A_3)=0{,}9\cdot 0{,}8\cdot 0{,}5=0{,}36.$$

Событие B_2 – только два попадания складывается из следующих возможных вариантов: первый стрелок не попал, а второй и третий попали, или второй стрелок не попал, а первый и третий попали, или третий стрелок не попал, а первый и второй попали. Перебор трех вариантов, когда один из стрелков не попал в мишень, а остальные двое попали представим формулой $B_2 = \overline{A}_1 A_2 A_3 + A_1 \overline{A}_2 A_3 + A_1 A_2 \overline{A}_3$, тогда

$$p(B_2) = p(\overline{A}_1 A_2 A_3 + A_1 \overline{A}_2 A_3 + A_1 A_2 \overline{A}_3),$$

а т. к. события в скобках $\overline{A}_1 A_2 A_3$, $A_1 \overline{A}_2 A_3$ и $A_1 A_2 \overline{A}_3$ попарно несовместны, то

$$p(B_2) = p(\overline{A}_1 A_2 A_3) + p(A_1 \overline{A}_2 A_3) + p(A_1 A_2 \overline{A}_3).$$

Далее из независимости событий A_1, A_2 и A_3 следует

$$p(B_2) = p(\overline{A}_1)p(A_2)p(A_3) + p(A_1)p(\overline{A}_2)p(A_3) + p(A_1)p(A_2)p(\overline{A}_3) =$$
$$= 0{,}1 \cdot 0{,}8 \cdot 0{,}5 + 0{,}9 \cdot 0{,}2 \cdot 0{,}5 + 0{,}9 \cdot 0{,}8 \cdot 0{,}5 = 0{,}49.$$

Событие B_1 – только одно попадание – складывается из следующих возможных вариантов: первый стрелок попал, а второй и третий не попали, или второй стрелок попал, а первый и третий не попали, или третий стрелок попал, а первый и второй не попали. Следовательно, $B_1 = A_1 \overline{A}_2 \overline{A}_3 + \overline{A}_1 A_2 \overline{A}_3 + \overline{A}_1 \overline{A}_2 A_3$. Отметим, что одновременно события $A_1 \overline{A}_2 \overline{A}_3$, $\overline{A}_1 A_2 \overline{A}_3$ и $\overline{A}_1 \overline{A}_2 A_3$ произойти не могут, т. е. являются несовместными и в каждом произведении события независимы, поэтому

$$p(B_1) = p(A_1 \overline{A}_2 \overline{A}_3 + \overline{A}_1 A_2 \overline{A}_3 + \overline{A}_1 \overline{A}_2 A_3) =$$
$$= p(A_1 \overline{A}_2 \overline{A}_3) + p(\overline{A}_1 A_2 \overline{A}_3) + p(\overline{A}_1 \overline{A}_2 A_3) =$$
$$= p(A_1)p(\overline{A}_2)p(\overline{A}_3) + p(\overline{A}_1)p(A_2)p(\overline{A}_3) + p(\overline{A}_1)p(\overline{A}_2)p(A_3) =$$
$$= 0{,}9 \cdot 0{,}2 \cdot 0{,}5 + 0{,}1 \cdot 0{,}8 \cdot 0{,}5 + 0{,}1 \cdot 0{,}2 \cdot 0{,}5 = 0{,}14.$$

Событие B_0 состоит в том, что ни один из стрелков не попал. Другими словами, это событие состоит в пересечении событий – первый стрелок не попал в цель, и второй не попал, и третий не попал, следовательно, $B_0 = \overline{A}_1 \overline{A}_2 \overline{A}_3$. Так как события $\overline{A}_1$, $\overline{A}_2$ и $\overline{A}_3$ независимы, то по формуле (13)

$$p(B_0) = p(\overline{A}_1 \overline{A}_2 \overline{A}_3) = p(\overline{A}_1)p(\overline{A}_2)p(\overline{A}_3) = 0{,}1 \cdot 0{,}2 \cdot 0{,}5 = 0{,}01.$$

Событие B_4 складывается из большого количества вариантов – или все попали, или попал только один, или попали только двое. Поэтому выгодно перейти к противоположному событию $\overline{B}_4$ – никто не попал, такое событие рассмотрено выше $\overline{B}_4 = B_0$, тогда $p(B_4) = 1 - p(\overline{B}_4) = 1 - p(B_0) = 1 - 0{,}01 = 0{,}99$.

Замечание. Отметим, что сумма событий $B_0 + B_1 + B_2 + B_3 = D$ образует достоверное событие, т. к. B_0, B_1, B_2 и B_3 попарно несовместные события, образующие полную группу исходов опыта:

$$p(D) = p(B_0 + B_1 + B_2 + B_3) = p(B_0) + p(B_1) + p(B_2) + p(B_3) = 1.$$

Это условие можно использовать как критерий проверки правильности вычислений:

$$p(B_0) + p(B_1) + p(B_2) + p(B_3) = 0{,}01 + 0{,}14 + 0{,}49 + 0{,}36 = 1.$$

1.6. Формула полной вероятности. Формула Бейеса

Пусть событие B может наступить при условии появления одного из следующего набора событий $A_1, A_2, ..., A_n$, которые иногда называют ***гипотезами***.

События $A_1, A_2, ..., A_n$ образуют ***полную группу событий***, если в результате опыта происходит только одно и только из этого набора событие, что эквивалентно тому, что они

- попарно несовместны
$$p(A_k \cap A_p) = 0,\ k \neq p,\ k = 1,...,n,\ p = 1,...,n; \tag{18}$$

- в сумме дают достоверное событие $A_1 + A_2 + ... + A_n = D$, откуда в силу (16)
$$\sum_{k=1}^{n} p(A_k) = 1. \tag{19}$$

Простейшим примером полной группы событий является бросание игральной кости, где полная группа событий содержит 6 исходов: $\{A_k$ – выпадение на верхней грани цифры $k\}_{k=1}^{6}$.

В примере 9 события B_0, B_1, B_2 и B_3 тоже образуют полную группу событий (исходов).

Теорема 5. (*Формула полной вероятности*).

Вероятность наступления события B, которое может наступить при условии появления одного из полной группы событий $A_1, A_2, ..., A_n$, вычисляется по формуле
$$p(B) = p(A_1)p_{A_1}(B) + p(A_2)p_{A_2}(B) + ... + p(A_n)p_{A_n}(B). \tag{20}$$

Доказательство. Умножим обе части равенства $D = A_1 + A_2 + ... + A_n$ на событие B: $DB = (A_1 + A_2 + ... + A_n)B \Rightarrow B = A_1B + A_2B + ... + A_nB$. Вычислим вероятность событий левой и правой части последнего равенства, а т. к. справа события попарно несовместные, то, применив (15), получим:
$$p(B) = p(A_1B + A_2B + ... + A_nB) = p(A_1B) + p(A_2B) + ... + p(A_nB).$$

Поскольку событие B может наступить в результате каждого из событий $A_1, A_2, ..., A_n$, причем с разной вероятностью, то в произведениях $A_1B, A_2B, ..., A_nB$ события зависимые, тогда по формуле (10)
$$p(A_1B) + p(A_2B) + ... + p(A_nB) = p(A_1)p_{A_1}(B) + p(A_2)p_{A_2}(B) + ... + p(A_n)p_{A_n}(B),$$
что доказывает равенство (20).

Теорема 6. (*Формула Бейеса*).

Пусть событие B произошло в результате наступления события A_k из полной группы событий $A_1, A_2, ..., A_n$, тогда

$$p_B(A_k)=\frac{p(A_k)p_{A_k}(B)}{p(B)}. \qquad (21)$$

Доказательство. Т. к. $A_k B = BA_k$, то $p(A_k B)=p(BA_k)$. Дважды применив формулу (10), получим $p(A_k)p_{A_k}(B)=p(B)p_B(A_k)$, откуда выразив $p_B(A_k)$, получим (21).

Пример 10. В магазине поступают электрические лампочки, изготовленные на трех заводах. На данный момент имеется 3 ящика с первого завода, 5 – со второго завода и 2 – с третьего завода. Процент бракованных лампочек (то есть тех, что перегорают раньше положенного срока) составляет на первом заводе – 10 %, на втором – 20 %, на третьем – 25 %. Известно, что ящики с лампочками заносятся в зал случайным образом. Найти вероятность того, что:

1) купленная в этом магазине лампочка окажется бракованной;
2) если лампочка оказалась бракованной, то она изготовлена на 1-м заводе.

Решение. Обозначим, через событие В – лампочка с браком. Поскольку ящик с лампочками может быть ящиком с любого из трех заводов, то гипотезы всего три: A_1 – лампочка сделана на первом заводе, A_2 – лампочка сделана на втором заводе, A_3 – на третьем заводе. Так как в магазине всего $3+5+2=10$ ящиков с лампочками, то вероятность, что лампочка изготовлена первым заводом, $p(A_1)=3/10$, соответственно, $p(A_2)=0{,}5$, $p(A_3)=0{,}2$.

1) Если лампочка произведена на первом заводе, то вероятность события B– деталь бракованная $p_{A_1}(B)=0{,}1$, так как на первом заводе брак составляет 10 %. Аналогично, $p_{A_2}(B)=0{,}2$, $p_{A_3}(B)=0{,}25$.

По формуле полной вероятности (20) при $n=3$:

$$\begin{array}{ccccccc} p(B)= & p(A_1) & p_{A_1}(B) & + \; p(A_2) & p_{A_2}(B) & + \; p(A_3) & p_{A_3}(B) = \\ & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow & \Downarrow \\ = & 0{,}3 & \cdot\, 0{,}1 & + \; 0{,}5 & \cdot\, 0{,}2 & + \; 0{,}2 & \cdot\, 0{,}25 \; = 0{,}18. \end{array}$$

2) Если лампочка бракованная, то вероятность что она произведена на первом заводе определяется формулой (21): $p_B(A_1)=\dfrac{p(A_1)p_{A_1}(B)}{p(B)}=\dfrac{0{,}3\cdot 0{,}1}{0{,}18}=\dfrac{1}{6}$.

<u>Замечание.</u> Заметим, что $p(A_1)=3/10$ больше, чем $p_B(A_1)=1/6$, что, очевидно, связано с тем, что брака на первом заводе меньше, чем на остальных заводах. Формула Бейеса позволяет произвести переоценку ценностей после поступления дополнительной информации.

2. ПОВТОРЕНИЕ ИСПЫТАНИЙ. СХЕМА БЕРНУЛЛИ

Пусть производится серия из n независимых между собой одинаковых испытаний, где в результате каждого из них с одинаковой вероятностью $p(A) = p$ может произойти случайное событие A. Соответственно, событие A не наступает с вероятностью $q = p(\overline{A}) = 1 - p(A) = 1 - p$.

Определить нужно вероятность $p_n(k)$ того, что из n независимых испытаний событие A происходит ровно k раз или вероятность $p_n(k_1, k_2)$, что событие происходит от k_1 до k_2 раз включительно.

Построенная математическая модель главой знаменитого математического семейства Якобом Бернулли (1654 – 1705 г.г.), описывала серии независимых испытаний, была в последствии названа схемой Бернулли.

Например, определить вероятность для равносильных игроков выиграть 3 партии в шахматы из 5. Эту вероятность в принципе можно посчитать, используя теоремы сложения и умножения вероятностей, как это было рассмотрено в пунктах 1.4 и 1.5.:

$$p(AAA\overline{A}\overline{A} + AA\overline{A}A\overline{A} + A\overline{A}AA\overline{A} + \overline{A}AAA\overline{A} + AA\overline{A}\overline{A}A + A\overline{A}A\overline{A}A + \overline{A}AA\overline{A}A +$$
$$+ A\overline{A}\overline{A}AA + \overline{A}A\overline{A}AA + \overline{A}\overline{A}AAA) = \ldots = 10\left(p(A)\right)^3\left(p(\overline{A})\right)^2 = 10\left(\frac{1}{2}\right)^3\left(1 - \frac{1}{2}\right)^2 = \frac{5}{16}.$$

Уже вычисления трудоемки, а при достаточно большом количестве испытаний это будет практически невозможно.

2.1. Формулы Бернулли, Лапласа и Пуассона

Пусть в результате n опытов событие A наступает первые k раз, тогда остальные $(n-k)$ раз это событие не наступает, а вероятность этого события равна $p^k(1-p)^{n-k}$.

Вообще говоря, событие A может появиться k раз в n испытаниях в различных комбинациях – их число равно количеству сочетаний из n элементов по k, что находится по формуле: $C_n^k = \dfrac{n!}{k!(n-k)!}$. Применяя теорему сложения вероятностей несовместных событий, получим

$$C_n^k p^k (1-p)^{n-k}.$$

Формула Бернулли, вероятности того, что при n испытаниях событие A наступит ровно k раз, имеет вид:

$$p_n(k) = C_n^k p^k q^{n-k}. \tag{22}$$

Наивероятнейшим числом k_0 появления события A при n испытаниях называется такое число, которому соответствует наибольшая вероятность $p_n(k_0)$, такое значение может быть не одно.

Наивероятнейшее число k_0 определяется из двойного неравенства

$$(n+1)\,p-1 \le k_0 \le (n+1)\,p.$$

Ниже в п. 3.3.1 будет показано, что среднее число появления события A в n испытаниях $\lambda = np$.

Пользоваться *формулой Бернулли* при больших n затруднительно из-за имеющегося в ней факториала.

Если число испытаний велико, то применяются либо формула Лапласа, либо формула Пуассона. У каждой из них есть свои ограничения.

Если число испытаний n велико, а вероятность события p мала ($p \le 0.1$), то в таком случае может быть использована приближенная ***формула Пуассона***

$$p_n(k) \approx \frac{\lambda^k e^{-\lambda}}{k!}, \text{ где } \lambda = np, \qquad (23)$$

обычно $\lambda \le 10$.

Замечание. Формула Пуассона обобщается на непрерывный случай простейшего (пуассоновского) потока событий. *Потоком событий* называется последовательность аналогичных событий, наступающих в случайные моменты времени. Это могут быть вызовы в справочную службу или вызовы поступающие в скорую помощь и т. д. Простейший поток событий обладает свойствами:

- стационарности (вероятность появления k событий за период T, т. е. за промежуток $(t; t+T)$ не зависит от t, а зависит только от k и T);
- отсутствие последействия (неважно, что происходит до данного момента времени t, что сравнимо с независимостью испытаний в схеме Бернулли);
- ординарностью (все события разделены между собой некоторыми, может быть очень малыми, промежутками времени).

Обозначим через λ интенсивность потока – среднее число событий, появляющихся в единицу времени, тогда вероятность появления k событий простого потока за время длительности t определяется формулой Пуассона, обобщающей формулу (22)

$$p_t(k) \approx \frac{(\lambda t)^k e^{-\lambda t}}{k!}. \qquad (23')$$

Если же число n велико и не мала вероятность p (обычно $p > 0{,}1$), то можно воспользоваться другим приближением – ***локальной теоремой Лапласа***

$$p_n(k) \approx \frac{1}{\sqrt{npq}}\varphi(x), \qquad \text{где } x = \frac{k-np}{\sqrt{npq}}, \qquad (24)$$

значения функции Лапласа

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{25}$$

затабулированы и помещены в специальные таблицы, которые всегда приведены в любой учебной литературе по теории вероятностей (например, в [1]). Значения функции Лапласа приведены с малым шагом на отрезке $[0;5]$, это связано с тем, что на отрезке $[-5;5]$ находится ее наиболее выпуклая часть, а вне этого отрезка она принимает малые значения и поэтому в вычислениях просто считается равной нулю.

Поскольку функция (25) четная $\varphi(-x) = \varphi(x)$, то затабулирован только отрезок $[0;5]$. График локальной функции Лапласа указан ниже на рис. 6 и называется кривой Гаусса.

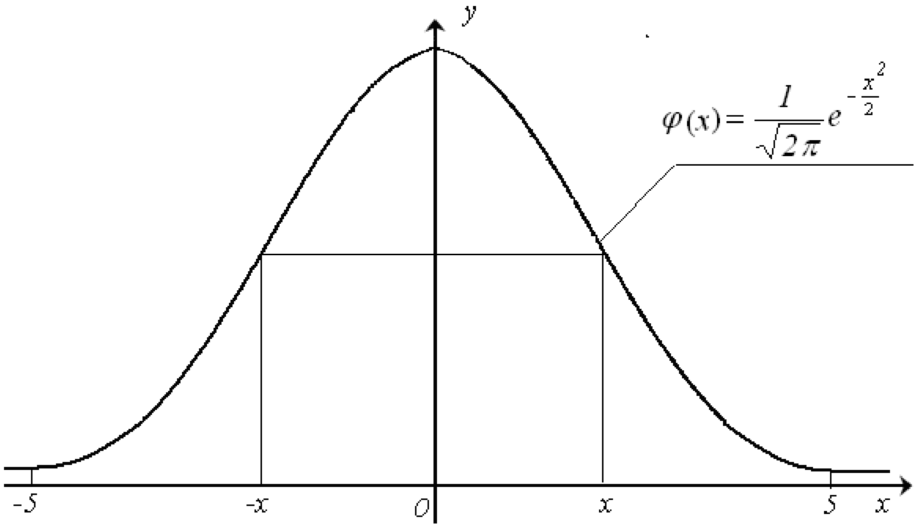


Рис. 6. Локальная функция Лапласа

Вероятность того, что событие A в n испытаниях произойдет от k_1 до k_2 раз (включительно), обозначают $p_n(k_1,k_2)$, что численно равно сумме

$$p_n(k_1,k_2) = p_n(k_1) + p_n(k_1+1) + \ldots + p_n(k_2),$$

можно применять формулы (22), (23) или (24), если попадаем в условия, описанные перед этими формулами, и число слагаемых в сумме справа невелико, что позволяет произвести подсчет быстро.

Например, в задаче с игроками в шахматы, определить вероятность выигрыша не менее 3 партий из 5 сыгранных, легко вычислить

$$p_5(3,5) = p_5(3) + p_5(4) + p_5(5),$$

трижды применив формулу Бернулли.

Однако, если число испытаний велико, а так же и разность $(k_2 - k_1)$, то применяют ***интегральную формулу Лапласа***

$$p_n(k_1,k_2)=p_n(k_1)+p_n(k_1+1)+\ldots+p_n(k_2)=\int_{x_1}^{x_2}\varphi(x)dx\approx\Phi(x_2)-\Phi(x_1)\text{, т. е.}$$

$$p_n(k_1,k_2)\approx\Phi(x_2)-\Phi(x_1), \tag{26}$$

где $x_1=\dfrac{k_1-np}{\sqrt{npq}}$ и $x_2=\dfrac{k_2-np}{\sqrt{npq}}$.

Значения функции Лапласа

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz \tag{27}$$

затабулированы и помещены в специальные таблицы, которые всегда приведены в любой учебной литературе по теории вероятностей ([1] в Приложении 2).

Отметим, что неопределенный интеграл из (27) является “неберущимся”. Значения функции Лапласа приведены с малым шагом на отрезке $[0;5]$, что связано с тем, что на отрезке $[-5;5]$ находится главная часть функции и функция является нечетной $\Phi(-x)=-\Phi(x)$, поэтому затабулирован только отрезок $[0;5]$. Вне этого отрезка она просто считается равной $\Phi(x)\big|_{x<-5}=-0.5$ и $\Phi(x)\big|_{x>5}=0.5$.

График функции Лапласа приведен на рисунке 7.

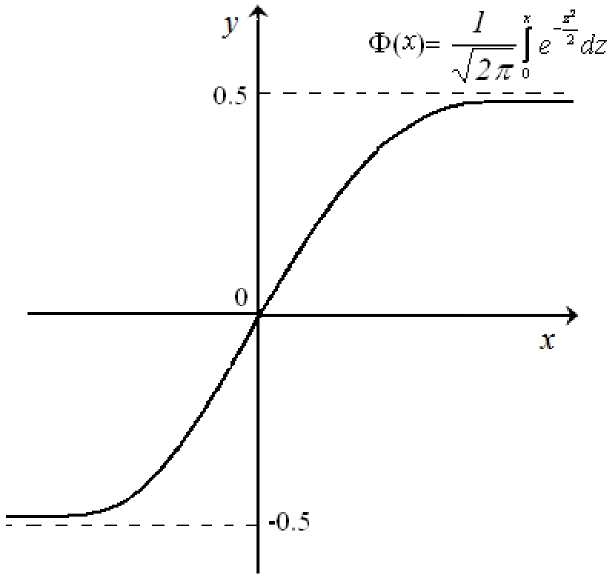


Рис. 7. Функция Лапласа

Как уже отмечалось в первой главе, очень часто найти вероятность p по классическому или геометрическому определению вероятности невозможно. В этом случае на практике используют схему Бернулли при достаточно большом числе независимых испытаний.

Пример 1. Прибор состоит из пяти узлов, каждый из которых работает безотказно в течение времени t с вероятностью 0,9. Работа узлов независима друг от друга. Найти вероятность отказа за время t двух узлов.

Решение. Пусть событие A есть выход одного узла из строя за время t. Число узлов $n=5$, а число отказавших узлов из условия задачи $k=2$.

Вероятность безотказной работы узла $p(A)=p=0{,}9$, тогда $q=1-p=1-0{,}9=0{,}1$. Вычислим искомую вероятность по формуле Бернулли (22): $p_5(2)=C_5^2\,0{,}9^3 0{,}1^{5-3}=\dfrac{5!}{3!2!}0{,}729\cdot 0{,}01=0{,}073$.

Ответ: $p_5(2)=0{,}073$.

Пример 2. Два шахматиста, первый из которых выигрывает в два раза чаще другого, играют матч из нескольких партий. Считая все партии результативными, найти:

1) вероятность выигрыша матча из трех партий первым игроком;
2) вероятность выигрыша первым игроком 13 партий из 18.

Решение.

1) Пусть событие A – матч выигран первым игроком, т. е. первый игрок выиграет две или три партии из трех. Вероятность выиграть одну партию для первого игрока, т. к. он выигрывает в два раза чаще, $p=2/3$, тогда вероятность проиграть $q=1-p=1/3$.

Матч состоит из трех повторяющихся событий. Так как $n=3$, то считать надо с помощью формулы Бернулли (22). Вероятность выиграть две партии из трех $p_3(2)=C_3^2\left(\dfrac{2}{3}\right)^2\left(\dfrac{1}{3}\right)^{3-2}=\dfrac{4}{9}$, а три партии из трех $p_3(3)=C_3^3\left(\dfrac{2}{3}\right)^3\left(\dfrac{1}{3}\right)^{3-3}=\dfrac{8}{27}$.

Итак, $p(A)=p_3(2)+p_3(3)=\dfrac{4}{9}+\dfrac{8}{27}=\dfrac{20}{27}\approx 0{,}74$.

2) В нашем случае $n=18$, $k=13$, $p=2/3$, $q=1/3$. Вычислим $p_{18}(13)$, т. к. $n=18$ большое число, то воспользуемся локальной формулой Лапласа (24)

$$p_{18}(13)=\frac{1}{\sqrt{18\cdot {}^2\!/_3\cdot {}^1\!/_3}}\cdot\varphi(x)=\frac{1}{2}\cdot\varphi(x),\text{ где }\ x=\frac{k-np}{\sqrt{npq}}=\frac{13-18\cdot 2/3}{\sqrt{18\cdot {}^2\!/_3\cdot {}^1\!/_3}}=\frac{1}{2}$$

и найдем из таблицы [1, прил. 2] $\varphi(0{,}5)\approx 0{,}3521$, тогда $p_{18}(13)=\dfrac{0{,}3521}{2}\approx 0{,}176$.

Ответ: вероятность выиграть матч из трех партий для первого игрока 0,74; вероятность выиграть 13 партий из 18 для первого игрока 0,176.

Пример 3. Игральная кость бросается 100 раз, найти вероятность того, что
1) шесть очков выпадут не менее 20 и не более 30 раз;
2) шесть очков выпадут не более 20 раз.

Решение.

1) Так как $n >> 1$ и $p > 0{,}1$, воспользуемся интегральной формулой Лапласа (26). В нашем случае $n = 100$, $k_1 = 20$, $k_2 = 30$, $p = 1/6$, $q = 1 - p = 1 - 1/6 = 5/6$.

$$p_{100}(20, 30) \approx \Phi(x_2) - \Phi(x_1),$$

где $x_1 = \dfrac{k_1 - np}{\sqrt{npq}} = \dfrac{20 - 100 \cdot 1/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}} = 0{,}89$, $x_2 = \dfrac{k_2 - np}{\sqrt{npq}} = \dfrac{30 - 100 \cdot 1/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}} = 3{,}58$,

тогда

$$p_{100}(20, 30) \approx \Phi(3{,}58) - \Phi(0{,}89) = 0{,}4998\text{-}0{,}3133 \approx 0{,}19.$$

Значения функции Лапласа $\Phi(3{,}58)$ и $\Phi(0{,}89)$ найдены из таблицы [1, прил. 2].

2) Воспользуемся интегральной теоремой Лапласа.

В нашем случае $n = 100$, $k_1 = 0$, $k_2 = 20$, $p = 1/6$, $q = 1 - p = 1 - 1/6 = 5/6$.

$x_1 = \dfrac{k_1 - np}{\sqrt{npq}} = \dfrac{0 - 100 \cdot 1/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}} = -4{,}47$, $x_2 = \dfrac{k_2 - np}{\sqrt{npq}} = \dfrac{20 - 100 \cdot 1/6}{\sqrt{100 \cdot 1/6 \cdot 5/6}} = 0{,}89$, тогда

$p_{100}(0, 20) \approx \Phi(0{,}89) - \Phi(-4{,}47) = \Phi(0{,}89) + \Phi(4{,}47) = 0{,}3133 + 0{,}4999 \approx 0{,}81$.

Ответ: $p_{100}(20, 30) \approx 0{,}19$, $p_{100}(0, 20) \approx 0{,}81$.

Пример 4. Каменщик за смену может уложить 1000 кирпичей. Вероятность того, что он уронит кирпич, равна 0,004. Найти вероятность того, что в течение смены произойдет падение

1) 5 кирпичей;
2) не более пяти кирпичей.

Решение. Так как $n >> 1$ и $p < 0{,}1$, то искомую вероятность найдем по формуле Пуассона (23). По условию задачи $k = 5$, $n = 1000$, $\lambda = np = 1000 \cdot 0{,}004 = 4$.

1) Подставив найденные значения в формулу (23) получим

$$p_{1000}(5) \approx \frac{4^5 e^{-4}}{5!} \approx 0{,}156.$$

2) Найдем $p_{1000}(0;5) =$

$p_{1000}(0) + p_{1000}(1) + p_{1000}(2) + p_{1000}(3) + p_{1000}(4) + p_{1000}(5) =$

$$= \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} + \frac{4^3 e^{-4}}{3!} + \frac{4^4 e^{-4}}{4!} + \frac{4^5 e^{-4}}{5!} = \frac{1 + 4 + 8 + \dfrac{64}{6} + \dfrac{256}{24} + \dfrac{1024}{120}}{e^4} \approx 0.785.$$

Ответ: $p_{1000}(5) \approx 0{,}156$; $p_{1000}(0;5) \approx 0.785$.

Пример 5. За час на телефонную станцию поступает 120 вызовов. Найти вероятность того, что за 3 минуты поступят 5 вызовов.

Решение. Так как поступающие случайным образом вызовы представляют собой простой поток событий, то воспользуемся формулой ($23'$). Найдем среднее количество вызовов за одну минуту $\lambda = \frac{120}{60} = 2$ – что соответствует 2 вызовам в минуту, $\lambda \cdot t = 2 \cdot 3 = 6$, $k = 5$, тогда из $p_t(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$ получим $p_3(5) = \frac{(6)^5 e^{-6}}{5!} \approx 0{,}16$.

Ответ: вероятность того, что за 3 минуты поступит 5 вызовов приблизительно равна 0,16.

3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

3.1. Дискретные и непрерывные случайные величины, способы их задания

Вторым основным понятием в теории вероятностей, на которое будет ориентироваться вся последующая теория, является ***случайная величина (СВ)***.

Случайная величина (в отличие от случайного события, которое может произойти либо не произойти) – это переменная величина, которая в результате испытания может принимать различные значения. Множество ее значений перед началом испытания известно, но только одно из своих возможных значений ей принимается, а какое именно, зависит от случайных причин.

Например, при бросании игральной кости случайная величина $\boldsymbol{X}$ – число выпавших очков может произвольно принимать только одно из шести значений.

При ответе студента на экзамене случайная величина $\boldsymbol{X}$ – это экзаменационная оценка, она может принимать одно из четырех значений от “2” до “5”, что зависит от степени подготовленности к экзамену и часто от везения.

В примерах, представленных выше, очевидным является то, что множество исходов можно пересчитать, а так бывает не всегда.

При стрельбе из орудия по мишени случайная величина $\boldsymbol{X}$ – это расстояние, на которое отклонился снаряд от поставленной цели. Измеряется в сантиметрах, метрах, километрах в зависимости от вида орудия. Пронумеровать в данном случае множество исходов не представляется возможным.

Исходя из приведенных примеров, можно случайные величины разделить на два класса – дискретные и непрерывные.

Случайная величина называется ***дискретной*** (ДСВ), если множество ее значений образует счетное множество (множество, элементы которого могут быть пронумерованы). В результате опыта принимается только одно из набора значений, причем с определенной вероятностью для каждого из них. Это множество может быть как конечным, так и бесконечным.

Например, X – количество выстрелов до первого попадания в цель – это дискретная случайная величина. Если стреляющий вообще не умеет стрелять, то эта величина может принимать и бесконечное, хотя и счетное количество значений.

Случайная величина называется ***непрерывной*** (НСВ), если множество ее значений может принимать любые значения из некоторого конечного или бесконечного промежутка (интервала).

Например, при стрельбе по мишени X – величина отклонения от центра есть непрерывная случайная величина.

Способы задания ДСВ и НСВ различны, но можно провести определенные параллели.

3.1.1. Дискретная случайная величина, способы ее задания

1. Табличный способ

Определение. Соотношение между всеми возможными значениями случайной величины и вероятностями их выпадения называется ***законом распределения дискретной*** случайной величины, сама таблица соответствия называется ***законом распределения*** (табл. 1). При этом сумма все вероятностей значений случайной величины равна единице.

Закон распределения ДСВ Таблица 1

X	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Можно сказать, что события A_i, где $i=1,\ldots,n$, заключаются в том, что случайная величина X принимает значения x_i, тогда $\{A_i\}_{i=1}^{n}$ – образуют полную группу исходов, для которой верно $\sum_{i=1}^{n} p(A_i) = \sum_{i=1}^{n} p(X = x_i) = 1$. Обозначив $p_i =$ $p(X = x_i)$, для $i=1,\ldots,n$ получим

$$\sum_{i=1}^{n} p_i = 1. \quad (28)$$

2. Графический способ

Закон распределения дискретной случайной величины можно изобразить графически в виде ломаной.

Многоугольником распределения ДСВ называется ломаная, соединяющая на плоскости xOp последовательно соседние точки $A_i(x_i; p_i)$, где $i=1, 2,\ldots,n$.

3. Аналитический способ

Определение. Функция распределения $F(x)$ – это функция, принимающая в точке $\boldsymbol{x}$ значение вероятности того, что случайная величина X принимает значение строго меньшее заданного значения аргумента:

$$F(x) = p(X < x), \tag{29}$$

где аргумент может принимать любое действительное значение.

Очевидно, что с увеличением аргумента x увеличивается и значение функции, т. е. функция представляет собой “накопительную вероятность”. Такой способ задания случайной величины является универсальным, поэтому применим как для дискретных, так и для непрерывных случайных величин.

Функция распределения случайной величины, заданной на интервале $(a;b)$, обладает следующими ***свойствами***:

1. Функция распределения принимает значения, которые меняются от нуля до единицы $0 \le F(x) \le 1$, так как она принимает значения вероятности.
2. $F(x) = 0$ при $x \le a$, т. к. левее точки $x = a$ значений ДСВ не имеет;
   $F(x) = 1$ при $b < x$, т. к. все значения уже пройдены (накоплены).

   Если случайная величина задана на всей числовой оси $(-\infty;+\infty)$, то данное свойство имеет вид
   $$\lim_{x\to-\infty} F(x) = 0,\ \lim_{x\to+\infty} F(x) = 1.$$
3. Функция распределения есть неубывающая ступенчатая функция, значения которой меняются от нуля до единицы и скачкообразно увеличивается при переходе через x_i на соответствующую вероятность p_i:
   $$F(\alpha) \le F(\beta),\ \text{если}\ \alpha < \beta.$$
4. В силу определения функции распределения вероятность попадания ее значений в некоторый интервал (α, β) равна разности значений этой функции:
   $$p(\alpha < x < \beta) = F(\beta) - F(\alpha). \tag{30}$$

Пример 1. По мишени стреляют три стрелка. Вероятность попадания в мишень первым стрелком равна 0,9; вторым – 0,8; третьим – 0,5. Для случайной величины X, равной числу промахов, требуется:

1) составить закон распределения;
2) построить многоугольник распределения;
3) построить функцию распределения $F(x)$.

Решение. Случайная величина X – число промахов, может принимать четыре значения: $X = 0$, если все попали в цель; $X = 1$, если кто-то один из трех промахнулся и т.д.

Вычисление вероятностей для каждого значения случайной величины было проведено в примере 9 на стр. 15:

$$p(X=0)=p(B_3)=0{,}36;$$
$$p(X=1)=p(B_2)=0{,}49;$$
$$p(X=2)=p(B_1)=0{,}14;$$
$$p(X=3)=p(B_0)=0{,}01.$$

1) Составим закон распределения (табл. 2).

Таблица 2

X	0	1	2	3
p	0,36	0,49	0,14	0,01

Критерием проверки правильности вычислении является условие (28), которое в рассматриваемом примере выполняется 0,36+0,49+0,14+0,01=1.

2) Чтобы построить многоугольник распределения, построим на плоскости XOP четыре точки: $A_1(0;0{,}36)$, $A_2(1;0{,}49)$, $A_3(2;0{,}14)$, $A_4(3;0{,}01)$, координаты которых, очевидно, взяты из таблицы и соединим их ломаной.

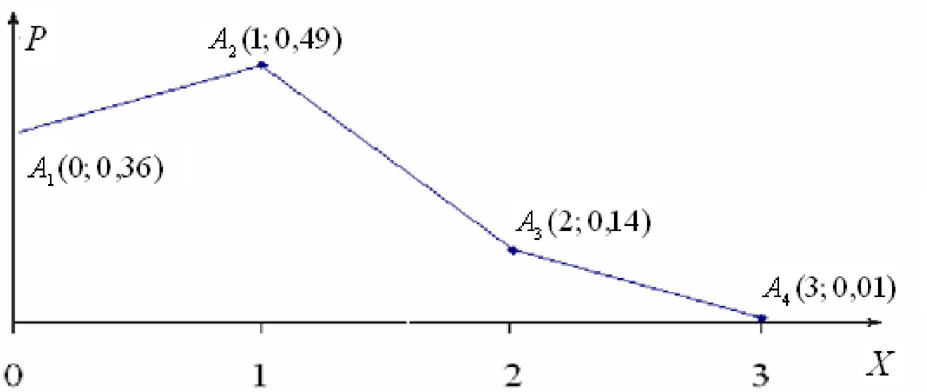


Рис. 8. Многоугольник распределения ДСВ X

3) Для построения функции распределения $F(x)$ еще раз отметим, что ее значения могут меняться только в точках $X=0,1,2,3$, точнее, должны увеличиваться на соответствующие вероятности.

Если $X\le 0$, то для любого x из этого бесконечного интервала
$F(x)=P(X<x)=0$, т. к. левее нет значений случайной величины.

Если $0<X\le 1$, то для любого x из этого полуинтервала
$F(x)=p(X<x)=p(X=0)$=0,36, т. к. левее только одно значение $X=0$.

Если $1<X\le 2$, то для любого x из этого полуинтервала
$F(x)=p(X<x)=p(X=0)+p(X=1)$=0,36+0,49=0,85.

Если $2<X\le 3$, то для любого x из этого полуинтервала
$F(x)=p(X<x)=p(X=0)+p(X=1)+p(X=2)$=0,36+0,49+0,14=0,99.

Если $3<X$, то для любого x из этого бесконечного промежутка
$F(x)=p(X=0)+p(X=1)+p(X=2)+p(X=3)$=0,36+0,49+0,14+0,01=1.

Итак, функция $F(x)$ является неубывающей, ступенчатой функцией

$$F(x)=\begin{cases} 0, & x\le 0; \\ 0.36, & 0<x\le 1; \\ 0.85, & 1<x\le 2; \\ 0.99, & 2<x\le 3; \\ 1, & 3<x. \end{cases},$$

а ее график имеет вид:

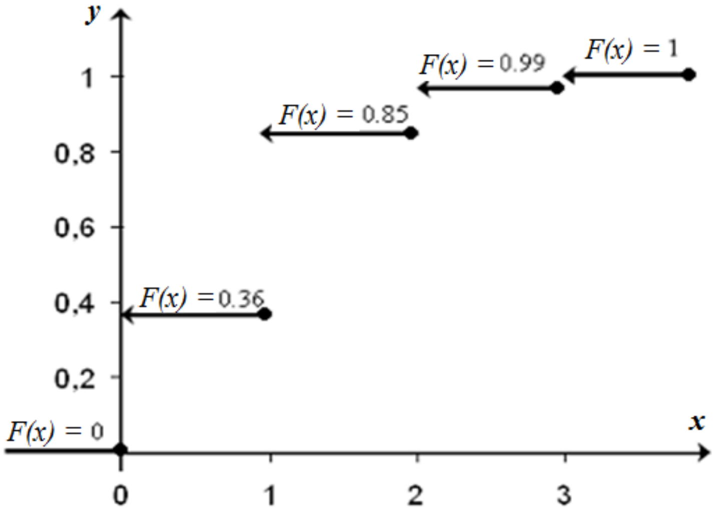


Рис. 9. График функции распределения

3.1.2. Непрерывная случайная величина, способы ее задания

Одним из двух возможных способов описания непрерывной случайной величины является ***функция распределения*** $F(x)$.

Определение ее по формуле (29) дано в предыдущем пункте 3.1.1., но для непрерывной случайной величины требуется, чтобы сама функция $F(x)$ была непрерывной, а ее производная – кусочно-непрерывной. Функция распределения непрерывной случайной величины, заданной на интервале $(a;b)$, обладает следующими свойствами:

1. Функция распределения принимает значения, которые меняются от нуля до единицы $0 \le F(x) \le 1$, так как она принимает значения вероятности.
2. Если случайная величина задана на интервале $(a;b)$, то при $x\le a$ $F(x)=0$ и $F(x)=1$ при $b < x$.

Если случайная величина задана на всей числовой оси $(-\infty;+\infty)$, то данное свойство имеет вид

$$\lim_{x\to-\infty} F(x)=0,\ \lim_{x\to+\infty} F(x)=1.$$

3. $F(x)$ является *непрерывной неубывающей* функцией

$$F(\alpha) \le F(\beta), \text{ если } \alpha < \beta.$$

$$F(\beta) = p(X < \beta) = p(X < \alpha) + p(\alpha \le X < \beta) = F(\alpha) + p(\alpha \le X < \beta),$$

т. е. $F(\beta) - F(\alpha) = p(\alpha \le X < \beta)$, где справа вероятность, а она всегда неотрицательна.

4. Вероятность попадания ее значений в некоторый интервал (α, β) можно получить из предыдущего свойства:

$$p(\alpha < x < \beta) = F(\beta) - F(\alpha),$$

т. е. формула (30) справедлива и для дискретной и для непрерывной случайной величины.

В заключение отметим основные свойства функции распределения для непрерывной случайной величины: она является непрерывной, монотонной, неубывающей функцией, значения которой меняются от нуля до единицы.

Функцией плотности $f(x)$ распределения вероятностей случайной величины X называется производная от функции распределения $F(x)$

$$f(x) = F'(x). \tag{31}$$

Она также используется для задания закона распределения вероятностей непрерывной случайной величины.

Укажем вероятностный смысл функции плотности

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{\Delta F}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x},$$

тогда, исходя из 4 свойства функции $F(x)$, имеем в числителе $F(x + \Delta x) - F(x)$ вероятность попадания случайной величины в интервал $(x; x + \Delta x)$, а дробь при уменьшении интервала $\Delta x \to 0$ указывает плотность значений СВ на данном интервале. Предельный переход указывает *плотность* значений СВ в фиксированной точке: $\dfrac{F(x + \Delta x) - F(x)}{\Delta x} \underset{\Delta x \to 0}{\to} p(X = x)$.

Свойства функции *плотности распределения случайной величины*:

1. Так как функция плотности $f(x)$ является производной монотонной, не убывающей функции распределения $F(x)$ (см. (31)), то

$$f(x) \ge 0.$$

2. Вероятность попадания значений НСВ в заданный интервал (α, β) равна

$$p(\alpha < x < \beta) = \int_{\alpha}^{\beta} f(t)dt, \tag{32}$$

что следует из свойства 4 функции распределения, то по формуле Ньютона-Лейбница имеем $p(\alpha < x < \beta) = F(\beta) - F(\alpha) = \int\limits_{\alpha}^{\beta} F'(t)dt = \int\limits_{\alpha}^{\beta} f(t)dt$.

3. Несобственный интеграл по всей числовой оси от функции распределения дает вероятность достоверного события $X \in (-\infty, +\infty)$, равную единице

$$\int\limits_{-\infty}^{+\infty} f(x)dx = 1. \tag{33}$$

В случае, когда непрерывная случайная величина принимает значения на интервале $X \in (a,b)$, то имеем обычный интеграл:

$$\int\limits_{a}^{b} f(x)dx = 1. \tag{34}$$

4. Чтобы восстановить по функции плотности $f(x)$ функцию распределения $F(x)$, которая задает по определению вероятность попадания в интервал $(-\infty, x)$, надо взять несобственный интеграл от функции плотности в этих пределах

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt. \tag{35}$$

Замечание. Из геометрического смысла определенного интеграла от неотрицательной функции $f(x)$, вероятность попадания случайной величины в соответствующий интервал равна площади фигуры, расположенной под кривой $y = f(x)$ на данном промежутке, т. е. $S_{[\alpha;\beta]} = p(\alpha < x < \beta) = \int\limits_{\alpha}^{\beta} f(x)dx$.

Пример 2. Задана функция $F(x) = \begin{cases} 0, & x \le 0; \\ Ax^4, & 0 < x \le 2; \\ 1, & 2 < x. \end{cases}$.

Требуется:

1) найти параметр A, при котором эта функция будет функцией распределения непрерывной случайной величины;
2) построить функцию распределения $F(x)$;
3) найти функцию плотности $f(x)$ и построить ее график;
4) найти вероятность попадания случайной величины в интервал $(-2;1)$.

Решение.

1) В силу непрерывности функции распределения $F(x)$ (слева она, очевидно, непрерывна) она должна быть непрерывна и справа при $x = 2$: $Ax^4\big|_{x=2} = 1$, тогда $16A = 1$ и $A = 1/16$.

2) Поскольку $A = 1/16$, то функция распределения имеет вид

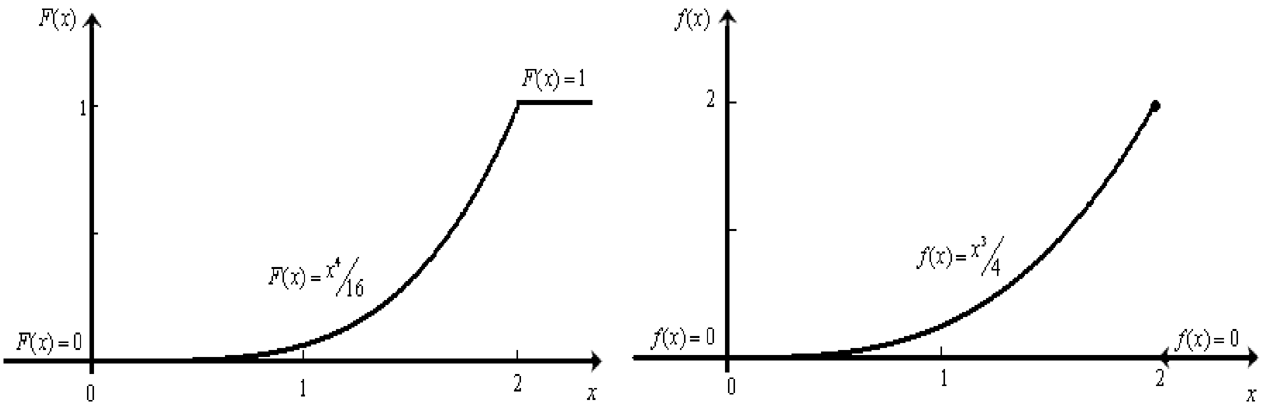
$$F(x) = \begin{cases} 0, & x \le 0, \\ x^4/16, & 0 < x \le 2, \\ 1, & 2 < x. \end{cases}.$$

График этой функции укажем ниже.

3) Дифференцируя на каждом из интервалов $F(x)$, получим функцию плотности распределения $f(x)$:

$$f(x) = \begin{cases} 0, & x \le 0, \\ x^3/4, & 0 < x \le 2, \\ 0, & 2 < x. \end{cases}$$

Построим графики функций $F(x)$ и $f(x)$:



а) Функция распределения б) Функция плотности

Рис. 10. Графики функций распределения и плотности НСВ

4) Вероятность попадания случайной величины в интервал $(-2;1)$ определяется формулой (30) $p(-2 < X < 1) = F(1) - F(-2) = \left.\frac{x^4}{16}\right|_0^1 - 0 = \frac{1}{16}$ или формулой (32) $p(-2 < X < 1) = \int_{-2}^{1} f(x)dx = \int_{-2}^{0} f(x)dx + \int_{0}^{1} f(x)dx = \int_{-2}^{0} 0dx + \int_{0}^{1} \frac{x^3}{4} dx = \left.\frac{x^4}{16}\right|_0^1 = \frac{1}{16}$.

Ориентироваться нужно на то, какая из функций задана в условии: или $F(x)$, или $f(x)$.

Ответ: $A = 1/16$, $p(0 < X < 1) = 1/16$.

Для определения функции распределения $F(x)$ при условии задания функции плотности $f(x)$ нужно применять интегрирование.

Пример 3. Задана функция $f(x) = \begin{cases} 0, & x \le 0; \\ Ax^3, & 0 < x \le 4; \\ 0, & 4 < x. \end{cases}.$

Требуется:

1) найти параметр A, при котором эта функция будет функцией плотности непрерывной случайной величины;

2) найти функцию распределения $F(x)$.

Решение.

1) Т. к. функция $f(x)$ непрерывна на конечном интервале $X \in (0,4)$, то по формуле (34): $\int_0^4 Ax^3dx = 1 \Rightarrow A\int_0^4 x^3dx = 1 \Rightarrow A\left(\frac{x^4}{4}\right)\Big|_0^4 = 1 \Rightarrow A\left(\frac{4^4}{4}\right) = 1 \Rightarrow A = \frac{1}{64}$.

Итак, функция плотности имеет вид

$$f(x) = \begin{cases} 0, & x \le 0; \\ \frac{x^3}{64}, & 0 < x \le 4; \\ 0, & 4 < x. \end{cases}.$$

2) Для нахождения функции распределения будем поинтервально интегрировать функцию плотности, увеличивая аргумент.

Если случайная величина $X \le 0$, то для любого $x_1 \in (-\infty;0]$:

$$F(x_1) = \int_{-\infty}^{x_1} 0 \cdot dt = 0.$$

Если случайная величина $0 < X \le 4$, то для любого $x_2 \in (0;4]$:

$$F(x_2) = \int_{-\infty}^{x_2} f(t)dt = \int_{-\infty}^{0} 0 \cdot dt + \int_0^{x_2} \frac{t^3}{64} \cdot dt = 0 + \frac{1}{64}\left(\frac{t^4}{4}\right)\Big|_0^{x_2} = \frac{1}{64}\left(\frac{{x_2}^4}{4}\right) = \frac{{x_2}^4}{256}.$$

Если случайная величина $4 < X$, то для любого $x_3 \in (4;+\infty)$:

$$F(x_3) = \int_{-\infty}^{x_3} f(t)dt = \int_{-\infty}^{0} 0 \cdot dt + \int_0^4 \frac{t^3}{64} \cdot dt + \int_4^{x_3} 0 \cdot dt = 0 + \frac{1}{64}\left(\frac{t^4}{4}\right)\Big|_0^4 + 0 = 1.$$

Итак, функция распределения имеет вид

$$F(x) = \begin{cases} 0, & x \le 0; \\ x^4/64, & 0 < x \le 4; \\ 1, & 4 < x. \end{cases}.$$

Критерием проверки верности вычислений является удовлетворение полученной функции всем свойствам функции распределения. Нарисовав ее, можно убедиться, что она действительно принимает значения $0 \le F(x) \le 1$ и является непрерывной, монотонной, неубывающей функцией.

Пример 4. Дана функция плотности распределения $f(x)$ случайной величины X: $f(x) = \begin{cases} 0, & x \le 1; \\ (x-1)/2, & 1 < x \le 3; \\ 0, & 3 < x. \end{cases}$

Требуется:

1) найти функцию распределения $F(x)$;

2) построить графики функций плотности $f(x)$ и распределения $F(x)$;

3) найти вероятность того, что случайная величина X примет значение, принадлежащее интервалу $(2;4)$.

Решение.

1) Найдем функцию распределения $F(x)$:

если $x \le 1$, то $F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{x} 0dt = 0$;

если $1 < x \le 3$, то

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{1} 0dt + \int\limits_{1}^{x} \frac{t-1}{2}dt = 0 + \int\limits_{1}^{x}\left(\frac{t}{2} - \frac{1}{2}\right)dt = \left.\left(\frac{t^2}{4} - \frac{t}{2}\right)\right|_1^x = \frac{t^2}{4} - \frac{t}{2} + \frac{1}{4};$$

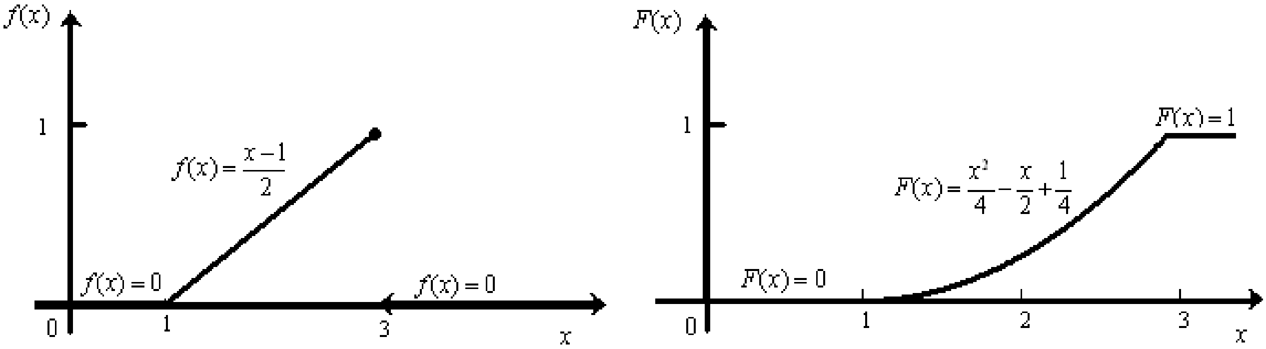
если $3 < x$, то

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{1} 0dt + \int\limits_{1}^{3} \frac{t-1}{2}dt + \int\limits_{3}^{+\infty} 0dt = 0 + \int\limits_{1}^{3}\left(\frac{t}{2} - \frac{1}{2}\right)dt = \left.\left(\frac{t^2}{4} - \frac{t}{2}\right)\right|_1^3 = 1.$$

Итак, искомая функция распределения имеет вид

$$F(x) = \begin{cases} 0, & x \le 1, \\ \dfrac{x^2}{4} - \dfrac{x}{2} + \dfrac{1}{4}, & 1 < x \le 3, \\ 1, & 3 < x. \end{cases}.$$

2) Графики функций $f(x)$ и $F(x)$ приведены на рис. 11.



а) Функция распределения б) Функция плотности

Рис. 11. Графики функций распределения и плотности НСВ

3) Вероятность того, что случайная величина X примет значение, принадлежащее интервалу $(2;4)$, по формуле (32):

$$p(2 < X < 4) = \int\limits_{2}^{4} f(x)dx = \int\limits_{2}^{3} \frac{x-1}{2}dx + \int\limits_{3}^{4} 0dx = \left.\frac{(x-1)^2}{4}\right|_2^3 = \frac{(3-1)^2}{4} - \frac{(2-1)^2}{4} = \frac{3}{4}$$

или по формуле (30):

$$p(2 < X < 4) = F(4) - F(2) = 1\Big|_{x=4} - \left(\frac{x^2}{4} - \frac{x}{2} + \frac{1}{4}\right)\Bigg|_{x=2} = 1 - \left(\frac{4}{4} - \frac{2}{2} + \frac{1}{4}\right) = \frac{3}{4}.$$

Ответ: $F(x) = \begin{cases} 0, & x \le 1 \\ \dfrac{x^2}{4} - \dfrac{x}{2} + \dfrac{1}{4}, & 1 < x \le 3, \\ 1, & 3 < x \end{cases}$ $p(2 < X < 4) = \dfrac{3}{4}$.

Пример 5. Дана функция плотности распределения $f(x)$ непрерывной случайной величины X: $f(x) = \dfrac{1}{\pi}\dfrac{1}{x^2+1}$.

Требуется:

1) проверить, что $f(x)$ является функцией плотности;

2) найти функцию распределения $F(x)$;

3) найти вероятность того, что случайная величина X примет значение, принадлежащее интервалу $(0;\pi/4)$.

Решение.

1) Проверим выполнение условия (33):

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{1}{x^2+1}dx = \frac{2}{\pi}\int_{0}^{+\infty}\frac{1}{x^2+1}dx = \frac{2}{\pi}\lim_{N\to+\infty}\int_{0}^{N}\frac{1}{x^2+1}dx = \frac{2}{\pi}\lim_{N\to+\infty} arctg\, x\Big|_0^N =$$

$$= \frac{2}{\pi}\lim_{N\to+\infty}(arctgN - arctg\,0) = \frac{2}{\pi}\left(\frac{\pi}{2} - 0\right) = 1.$$

Следовательно, данная функция $f(x)$ является функцией плотности.

2) Случайная величина непрерывна и задана на всей числовой оси, тогда функция распределения определяется несобственным интегралом

$$F(x) = \frac{1}{\pi}\int_{-\infty}^{x}\frac{1}{t^2+1}dt = \frac{1}{\pi}\lim_{N\to-\infty}\int_{N}^{x}\frac{1}{t^2+1}dt = \frac{1}{\pi}\lim_{N\to-\infty} arctg\, x\Big|_N^x =$$

$$= \frac{1}{\pi}\lim_{N\to-\infty}(arctgx - arctgN) = \frac{1}{\pi}\left(arctgx + \frac{\pi}{2}\right).$$

3) Вероятность того, что случайная величина X примет значение, принадлежащее интервалу $(0;\pi/4)$, вычислим по формуле (30):

$$p\left(0 < X < \frac{\pi}{4}\right) = F\left(\frac{\pi}{4}\right) - F(0) = \frac{1}{\pi}\left(arctg\frac{\pi}{4} + \frac{\pi}{2}\right) - \frac{1}{\pi}\left(arctg\,0 + \frac{\pi}{2}\right) =$$

$$= \frac{1}{\pi}\left(\left(1 + \frac{\pi}{2}\right) - \left(0 + \frac{\pi}{2}\right)\right) = \frac{1}{\pi}.$$

Ответ: $F(x) = \dfrac{1}{\pi}\left(arctgx + \dfrac{\pi}{2}\right)$; $p\left(0 < X < \dfrac{\pi}{4}\right) = \dfrac{1}{\pi}$.

3.2. Числовые характеристики случайных величин и их свойства

3.2.1. Математическое ожидание

Определение. Математическим ожиданием $M(x)$ случайной величины X называется ее среднее значение.

Для дискретной случайной величины оно равно сумме произведений всех ее значений на соответствующие вероятности $M(x) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$, т. е.

$$M(x) = \sum_{k=1}^{n} x_i p_i . \qquad (36)$$

Для непрерывной случайной величины

$$M(x) = \int_{-\infty}^{\infty} x f(x) dx . \qquad (37)$$

Отметим, что, несмотря на различие формул, математическое ожидание случайной величины всегда определяет "среднее" значение, около которого группируются возможные ее значения.

Свойства математического ожидания

1. Математическое ожидание константы равно самой константе
$$M(C) = C .$$
2. Постоянный множитель можно выносить за знак математического ожидания
$$M(Cx) = CM(x) .$$
3. Математическое ожидание ограничено наибольшим и наименьшим значениями случайной величины:
$$x_{\min} \le M(x) \le x_{\max} .$$
4. Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий слагаемых:
$$M(X+Y) = M(X) + M(Y) .$$
   Свойство справедливо для произвольного конечного числа с. в.
5. Математическое ожидание произведения двух независимых случайных величин (для которых закон или функция распределения одной случайной величины не зависит от того, какое значение приняла другая случайная величина) равно произведению их математических ожиданий:
$$M(XY) = M(X)M(Y).$$
   Это свойство также справедливо для произвольного конечного числа случайных величин.

Попробуйте самостоятельно вывести свойства 1-3 для дискретной случайной величины с помощью формулы (36), а для непрерывной с помощью формулы (37).

Например, если $M(x)=\sum_{k=1}^{n} x_i p_i$, то $M(Cx)=\sum_{k=1}^{n}(Cx_i)p_i = C\sum_{k=1}^{n}(x_i)p_i = CM(x)$ что доказывает свойство 2 для дискретной случайной величины.

3.2.2. Дисперсия случайной величины

Определение. Дисперсией случайной величины $D(x)$ называется математическое ожидание квадрата отклонения случайной величины от ее среднего значения: $D(x)=M\left(X-M(X)\right)^2$.

Другими словами, дисперсия есть величина, характеризующая среднее значение квадратов разброса значений случайной величины вокруг её среднего значения.

Вычисление дисперсии по определению может быть громоздко, проще с помощью свойств математического ожидания упростить вычисления по формуле

$$D(x)=M(X^2)-\left(M(X)\right)^2. \tag{38}$$

Действительно, $D(x)=M\left(X-M(X)\right)^2=M\left(X^2-2XM(X)+\left(M(X)\right)^2\right)=$ $=M\left(X^2\right)+M\left(-2XM(X)\right)+M\left(\left(M(X)\right)^2\right)=M\left(X^2\right)-2M\left(XM(X)\right)+\left(M(X)\right)^2=$ $=M\left(X^2\right)-2\left(M(X)\right)^2+\left(M(X)\right)^2=M\left(X^2\right)-\left(M(X)\right)^2$.

Для дискретной случайной величины $D(x)=\sum_{k=1}^{n}\left(x_i-M(x)\right)^2 p_i$ проще считать с помощью формулы

$$D(x)=\sum_{k=1}^{n} x_i^{\,2} p_i-\left(\sum_{k=1}^{n} x_i p_i\right)^2. \tag{39}$$

Для непрерывной случайной величины $D(x)=\int_{-\infty}^{\infty}(x-M(X))^2 f(x)dx$ после упрощения имеем

$$D(x)=\int_{-\infty}^{\infty} x^2 f(x)dx-\left(\int_{-\infty}^{\infty} xf(x)dx\right)^2. \tag{40}$$

Свойства дисперсии случайной величины

1. Дисперсия постоянной величины равна нулю

$$D(C)=0.$$

2. Постоянный множитель можно выносить за знак дисперсии по правилу

$$D(CX)=C^2D(X).$$

3. Дисперсия непостоянной случайной величины положительна

$$D(X)>0.$$

4. Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин

$$D(X+Y)=D(X)+D(Y).$$

5. Дисперсия разности двух независимых случайных величин равна сумме дисперсий этих величин

$$D(X-Y)=D(X)+D(Y).$$

Попробуйте самостоятельно вывести свойства 1-3 с помощью свойств математического ожидания и определения дисперсии.

Например, если $D(x)=M\left(X-M(X)\right)^2$, где $\left(X-M(X)\right)^2\geq 0$, то среднее значение положительных чисел (квадратов отклонений) положительно $M\left(X-M(X)\right)^2>0$, что доказывает свойство 3.

3.2.3. Среднее квадратическое отклонение случайной величины

Дисперсия характеризует среднее значение *квадратов* разброса значений случайной величины вокруг среднего значения $M(x)$, то $\sigma(x)$ и есть средняя величина разбросов. Следуя сказанному можно привести определение.

Средним квадратическим отклонением $\sigma(x)$ случайной величины X называется квадратный корень из дисперсии $D(x)$

$$\sigma(x)=\sqrt{D(x)}. \qquad (41)$$

Свойства среднего квадратического отклонения

1. Среднее квадратическое отклонение постоянной величины равно нулю

$$\sigma(C)=0.$$

2. Постоянный множитель можно выносить за знак среднего квадратичного отклонения

$$\sigma(CX)=C\sigma(X).$$

3. Среднее квадратическое отклонение непостоянной случайной величины положительно

$$\sigma(X)>0.$$

Пример 6. Для дискретной случайной величины, заданной законом распределения,

X	0	1	2	3
p	0,36	0,49	0,14	0,01

вычислить основные числовые характеристики: математическое ожидание, дисперсию и среднее квадратическое отклонение.

Решение.

Математическое ожидание дискретной случайной величины вычисляется по формуле (36) $M(x)=\sum_{k=1}^{n} x_i p_i =0\cdot 0{,}36+1\cdot 0{,}49+2\cdot 0{,}14+3\cdot 0{,}01=0{,}8$.

Дисперсия дискретной случайной величины вычисляется по формуле (39)

$$D(X)=\sum_{k=1}^{n} x_i^{\ 2} p_i -\left(M(x)\right)^2 =0^2\cdot 0{,}36+1^2\cdot 0{,}49+2^2\cdot 0{,}14+3^2\cdot 0{,}01-(0.8)^2 =0{,}5.$$

Среднее квадратическое отклонение случайной величины вычисляется по формуле (41)

$$\sigma(x)=\sqrt{D(x)}=\sqrt{0{,}5}\approx 0{,}71.$$

Пример 7. Функция распределения непрерывной случайной величины имеет вид

$$F(x)=\begin{cases} 0, & x\le 0,\\ x^4/16, & 0<x\le 2,\\ 1, & 2<x. \end{cases}$$

Вычислить основные числовые характеристики: математическое ожидание, дисперсию и среднее квадратическое отклонение.

Решение.

Для вычисления числовых характеристик нужна функция плотности $f(x)$, а т. к. $f(x)=F'(x)$, то, поинтервально продифференцировав заданную в условии функцию $F(x)$, получим $f(x)=\dfrac{x^3}{4}$ при $x\in(0;2]$ и $f(x)=0$ вне этого промежутка (см. пример 2 на стр. 32).

Математическое ожидание НСВ вычисляется по формуле (37)

$$M(x)=\int_{-\infty}^{\infty} xf(x)dx=\int_{-\infty}^{0} x\cdot 0dx+\int_{0}^{2} x\frac{x^3}{4}dx+\int_{2}^{\infty} x\cdot 0dx=\int_{0}^{2}\frac{x^4}{4}dx=\left.\frac{x^5}{20}\right|_0^2=\frac{8}{5}.$$

Дисперсия НСВ вычисляется по формуле (40)

$$D(X)=\int_{-\infty}^{\infty} x^2 f(x)dx-(M(x))^2=\int_{-\infty}^{0} x^2\cdot 0dx+\int_{0}^{2} x^2\frac{x^3}{4}dx+\int_{2}^{\infty} x^2\cdot 0dx-\left(\frac{8}{5}\right)^2=$$

$$=\int_{0}^{2}\frac{x^5}{4}dx-\left(\frac{8}{5}\right)^2=\left.\frac{x^6}{24}\right|_0^2-\frac{64}{25}=\frac{8}{3}-\frac{64}{25}=\frac{8}{75}.$$

Среднее квадратичное отклонение случайной величины вычисляется по формуле (41) $\sigma(x)=\sqrt{D(x)}=\sqrt{\dfrac{8}{75}}\approx 0{,}33$.

Мы рассмотрели основные числовые характеристики случайной величины $M(x)$, $D(x)$, $\sigma(x)$. Отметим так же следующие числовые характеристики.

Определение. *Модой M_0* дискретной случайной величины называется такое ее значение, которому соответствует наибольшая вероятность. Для непрерывной случайной величины мода – это значение, при котором плотность распределения будет максимальна.

$$f(M_0) = \max .$$

В общем случае значение M_0 может быть не единственным.

Определение. *Медианой M_D* случайной величины X называется такое ее значение, относительно которого равновероятны события

$$p(X < M_D) = p(X > M_D) . \quad (42)$$

Определение. *Начальным моментом α_k порядка k* случайной величины X называется математическое ожидание величины X^k

$$\alpha_k = M[X^k], \quad (43)$$

где для дискретной случайной величины

$$\alpha_k = \sum_{i=1}^{n} x_i^k p_i , \quad (43.1)$$

для непрерывной случайной величины

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x) dx . \quad (43.2)$$

Очевидно, что $\alpha_0 = 1$, а математическое ожидание $\alpha_1 = M[X]$ есть начальный момент первого порядка.

Определение. *Центральным моментом порядка k* случайной величины X называется математическое ожидание величины $(X - M(x))^k$

$$\mu_k = M[(X - M(x))^k], \quad (44)$$

где для дискретной случайной величины

$$\mu_k = \sum_{i=1}^{n} (x_i - M(x))^k p_i , \quad (44.1)$$

для непрерывной случайной величины:

$$\mu_k = \int_{-\infty}^{\infty} (x - M(x))^k f(x) dx . \quad (44.2)$$

Очевидно, что $\mu_0 = 0$, центральный момент первого порядка равен нулю $\mu_1 = M[(X - M(x))^1] = 0$, а дисперсия есть центральный момент второго порядка $\mu_2 = M[(X - M(x))^2] = D(x)$.

Центральный момент третьего порядка характеризует асимметрию распределения.

Определение. Отношение центрального момента третьего порядка к среднему квадратическому отклонению в третьей степени называется *коэффициентом асимметрии*:

$$a_x = \frac{\mu_3}{\sigma_x^3} .$$

Определение. Для характеристики островершинности и плосковершинности распределения используется величина, называемая *эксцессом*:

$$C_x = \frac{\mu_4}{\sigma_x^4} - 3.$$

Кроме рассмотренных величин используются также так называемые абсолютные моменты:

- абсолютный начальный момент: $\beta_k = M[|X|^k]$;

- абсолютный центральный момент: $\nu_k = M[|X - m_x|^k]$. Абсолютный центральный момент первого порядка называется *средним арифметическим отклонением*.

3.3. Основные законы распределения случайных величин

3.3.1. Биномиальный закон распределения

Биномиальным законом распределения дискретной случайной величины называют случайную величину X, принимающую значения числа появлений события A в n независимых испытаниях.

В каждом из n испытаний вероятность появления события A постоянна и равна p, соответственно, вероятность противоположного события $\bar{A}$ равна $1 - p = q$.

Вероятность возможного значения $X = k$ (событие A выпадет k раз в n испытаниях) вычисляется по формуле Бернулли:

$$p(X = k) = p_n(k) = C_n^k \cdot p^k \cdot q^{n-k},\ k = 0, 1, \ldots, n.$$

Биномиальный закон распределения случайной величины X имеет вид (табл. 3).

Биномиальный закон распределения Таблица 3

X	0	1	…	k	…	n
p	q^n	npq^{n-1}	…	$C_n^k p^k q^{n-k}$	…	p^n

Основные числовые характеристики *биномиального* закона распределения равны

$$M(X) = np,\ D(X) = npq,\ \sigma(X) = \sqrt{npq}. \qquad (45)$$

Доказательство. Введем n случайных величин X_k – появление события A в k-ом испытании, где k=1, 2,…, n, тогда $X = \sum_{k=1}^{n} X_k$. В силу независимости испытаний у всех случайных величин X_k закон распределения одинаков (табл. 4).

Таблица 4

X_k	0	1
p_k	q	p

Очевидно, $M(X_k) = p$, тогда по свойству 4 для математического ожидания

$$M(X) = M\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} M(X_k) = \sum_{k=1}^{n} p = np.$$

Для дисперсии получим

$$D(X_k) = M(X_k^2) - (M(X_k))^2 = p - p^2 = p(1-p) = pq.$$

Так как случайные величины X_k независимы, то по свойству 4 для дисперсии – дисперсия суммы независимых случайных величин равна сумме дисперсий, получим

$$D(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} (DX_k) = npq.$$

3.3.2. Равномерный закон распределения

Равномерным называют распределение вероятностей непрерывной случайной величины X, если на промежутке $(a;b]$, на котором задана случайная величина X, плотность распределения случайной величины принимает постоянное значение

$$f(x) = \begin{cases} C, x \in (a;b]; \\ 0, \quad x \notin (a;b]. \end{cases} \tag{46}$$

По заданному промежутку значений можно определить константу C, т. к. есть условие равенства единице площади ограниченной кривой и осью Ox, откуда получим $C = \dfrac{1}{b-a}$. График функции плотности представлен на рисунке 12.

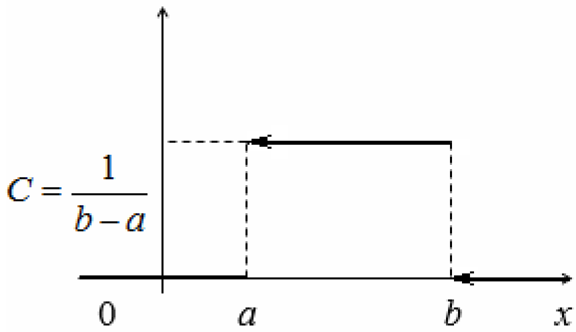


Рис. 12. Функция плотности равномерно распределенной случайной величины

Проинтегрируем функцию плотности $f(x)$ для определения функции распределения: на $[a;b]$ имеем $F(x) = \int_{-\infty}^{x} f(x)dx = \int_{a}^{x} \dfrac{1}{b-a}dx = \dfrac{x}{b-a}\Big|_a^x = \dfrac{x-a}{b-a}$.

Тогда для любого действительного x получим

$$F(x)=\begin{cases}0, & при \quad x\le a;\\ \dfrac{x-a}{b-a}, & при \quad a<x\le b;\\ 1, & при \quad b<x.\end{cases} \tag{47}$$

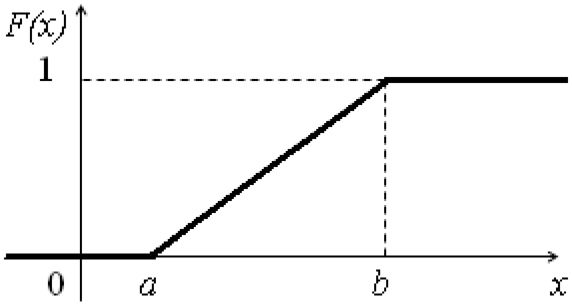


Рис. 13. Функция распределения равномерно распределенной СВ

Определим математическое ожидание и дисперсию случайной величины, подчиненной равномерному закону распределения.

$$M(x)=\int_a^b xf(x)dx=\int_a^b \frac{x}{b-a}dx=\frac{x^2}{2(b-a)}\Big|_a^b=\frac{a+b}{2}.$$

$$M(x^2)=\int_a^b x^2 f(x)dx=\int_a^b \frac{x^2}{b-a}dx=\frac{x^3}{3(b-a)}\Big|_a^b=\frac{b^3-a^3}{3(b-a)}=\frac{b^2+ab+a^2}{3}.$$

$$D(x)=M(x^2)-M^2(x)=\frac{b^2+ab+a^2}{3}-\frac{a^2+2ab+b^2}{4}=\frac{b^2-2ab+a^2}{12}=\frac{(b-a)^2}{12}.$$

$$\sigma(x)=\sqrt{D(x)}=\frac{b-a}{2\sqrt{3}}.$$

Вероятность попадания случайной величины в заданный интервал $(\alpha;\beta)$, где $(\alpha;\beta)\subset[a;b]$ определяется формулой:

$$P(\alpha<X<\beta)=\int_\alpha^\beta \frac{dx}{b-a}=\frac{\beta-\alpha}{b-a}. \tag{48}$$

Итак, основные числовые характеристики *равномерного* закона распределения

$$M(X)=\frac{a+b}{2},\ D(X)=\frac{(b-a)^2}{12},\ \sigma(X)=\frac{b-a}{2\sqrt{3}}. \tag{49}$$

Замечание 1. Формула (48) является частным случаем формулы (2) геометрического определения вероятности в одномерном случае.

Замечание 2. С помощью равномерного закона распределения на отрезке $[0;1]$ стало возможно моделирование вероятностных процессов на ЭВМ.

Программа датчик случайных величин выдает произвольное число из отрезка $[0;1]$, если $x<p=P(A)$, то считается, что событие A произошло, что часто применяется в математическом моделировании вероятностных событий.

Пример 8. Случайная величина X распределена равномерно на отрезке $[3;12]$.

Требуется:

1) найти функцию плотности $f(x)$, функцию распределения $F(x)$ и построить их графики;

2) вычислить числовые характеристики – математическое ожидание, дисперсию и среднее квадратичное отклонение;

3) найти вероятность попадания случайной величины в интервал $(1;8)$.

Решение.

1) Функция плотности распределения равномерной случайной величины X на отрезке $(3;12]$ равна $f(x)=1/(12-3)=1/9$, вне данного промежутка $f(x)=0$:

$$f(x)=\begin{cases}1/9, & x\in(3;12];\\ 0, & x\notin(3;12].\end{cases}.$$

Функция распределения $F(x)$ для равномерной случайной величины по формуле (47) имеет вид:

$$F(x)=\begin{cases}0, & x\le 3;\\ \dfrac{x}{9}-\dfrac{1}{3}, & 3<x\le 12;\\ 1, & 12<x.\end{cases}$$

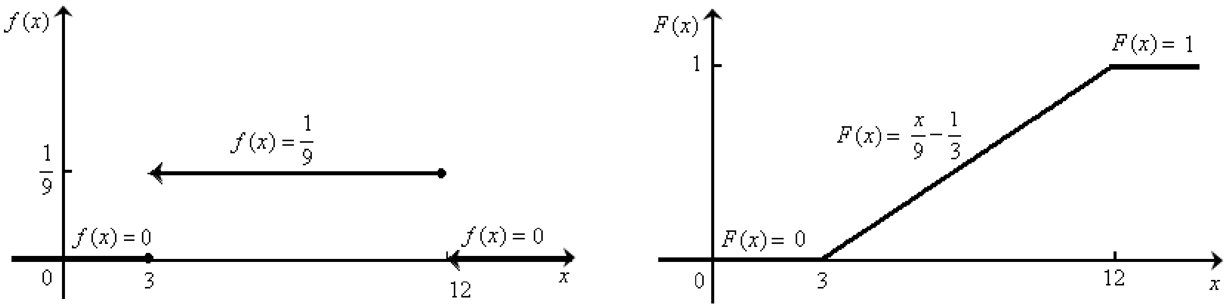


Рис. 14. Функция плотности и функция распределения равномерно распределенной случайной величины

2) Обратите внимание, что вычислить числовые характеристики в данном случае проще по формулам (49):

$$M(X)=\frac{a+b}{2}=\frac{3+12}{2}=7{,}5,\quad D(X)=\frac{(b-a)^2}{12}=\frac{(12-3)^2}{12}=6{,}75,$$

$$\sigma(X)=\frac{b-a}{2\sqrt{3}}=\frac{12-3}{2\sqrt{3}}\approx 2{,}6.$$

3) Вероятность того, что случайная величина X принимает значение из интервала $(1;8]$, по формуле (30):

$$p(1<X<8)=F(8)-F(1)=\left.\left(\frac{x}{9}-\frac{1}{3}\right)\right|_{x=8}-0=\frac{5}{9}.$$

Ответ: $M(x)=7{,}5$, $D(x)=6{,}75$, $\sigma(x)\approx 2{,}6$, $p(1<X<8)=5/9$.

3.3.3. Показательный закон распределения

Показательным называют распределение вероятностей непрерывной случайной величины, которое описывается функцией плотности распределения:

$$f(x)=\begin{cases}0, & x\le 0;\\ \lambda e^{-\lambda x}, & x>0;\end{cases}, \qquad (50)$$

где λ – положительная константа.

Найдем функцию распределения:

если $x\le 0$, то $F(x)=\int\limits_{-\infty}^{x} f(t)dt=\int\limits_{-\infty}^{x}0dt=0$;

если $x>0$, то $F(x)=\int\limits_{-\infty}^{x} f(t)dt=\int\limits_{-\infty}^{0}0dt+\lambda\int\limits_{0}^{x}e^{-\lambda t}dt=-\int\limits_{0}^{x}e^{-\lambda t}d(-\lambda t)=-e^{-\lambda t}\Big|_{0}^{x}=1-e^{-\lambda x}$.

Итак, функция распределения имеет вид

$$F(x)=\begin{cases}0, & x<0;\\ 1-e^{-\lambda x}, & x\ge 0.\end{cases} \qquad (51)$$

Графики функций распределения и плотности распределения (рис. 15):

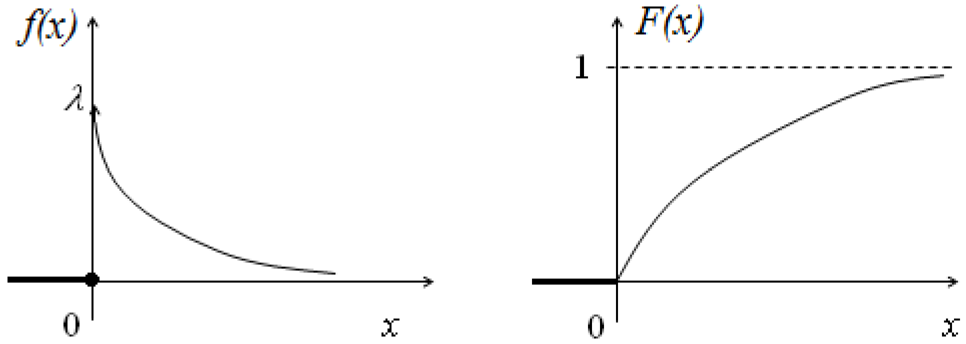


Рис. 15. Графики функции плотности и функции распределения показательно распределенной случайной величины

Найдем числовые характеристики случайной величины, подчиненной показательному распределению.

$$M(x)=\int\limits_{-\infty}^{+\infty}xf(x)dx=\int\limits_{0}^{+\infty}x\lambda e^{-\lambda x}dx=\lim_{N\to+\infty}\int\limits_{0}^{N}x\lambda e^{-\lambda x}dx=\left|\begin{array}{c}u=x\to du=dx\\ \lambda e^{-\lambda x}dx=dv\to -e^{-\lambda x}=v\end{array}\right|=$$

$$= \lim_{N\to+\infty}\left(-xe^{-\lambda x}\Big|_0^N + \int_0^N e^{-\lambda x}dx\right) = -\lim_{N\to+\infty}\left(xe^{-\lambda x} + \frac{e^{-\lambda x}}{\lambda}\right)\Bigg|_0^N = -\lim_{N\to+\infty}\left(\frac{N}{e^{\lambda N}} + \frac{1}{\lambda e^{\lambda N}} - 0\cdot e^{-\lambda\cdot 0} - \frac{e^0}{\lambda}\right) =$$

$$= -\lim_{N\to+\infty}\left(\frac{N}{e^{\lambda N}} + 0 - 0 - \frac{1}{\lambda}\right) = -\lim_{N\to+\infty}\left(\frac{(N)'}{(e^{\lambda N})'} - \frac{1}{\lambda}\right) = -\lim_{N\to+\infty}\left(\frac{1}{\lambda e^{\lambda N}} - \frac{1}{\lambda}\right) = -(0 - \frac{1}{\lambda}) = \frac{1}{\lambda}$$

Для нахождения дисперсии найдем величину $M(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^{\infty} \lambda x^2 e^{-\lambda x}dx$, дважды интегрируя по частям, получим $M(X^2) = \frac{2}{\lambda^2}$, тогда $D(X) = M(X^2) - [M(X)]^2 = \frac{1}{\lambda^2}$.

Среднее квадратичное отклонение равно $\sigma(x) = \sqrt{D(X)} = \frac{1}{\lambda}$.

Итак, основные числовые характеристики случайной величины с показательным распределением

$$M(X) = \frac{1}{\lambda},\ D(X) = \frac{1}{\lambda^2},\ \sigma(X) = \frac{1}{\lambda}. \qquad (52)$$

С помощью представленной выше функции распределения (51) вычисляется вероятность попадания случайной величины в заданный интервал, где x принимает неотрицательные значения

$$p(a < x < b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}. \qquad (53)$$

Показательное распределение имеет большое значение в теории надежности. ***Надежностью*** считается способность системы не отказывать в работе в заданный временной промежуток.

Предположим, устройство начинает работать в момент времени t_0=0, тогда правомерно возникает вопрос – в течение какого времени t отказа в работе устройства не будет, т. е. как долго будет светить лампочка или как долго безотказно будет работать стиральная машина, телевизор и т. д.

Введем непрерывную случайную величину T – длительность безотказной работы устройства. Оказывается, что она описывается показательным законом распределения. Сама функция распределения

$$F(t) = p\ (T < t) = 1 - e^{-\lambda t}$$

указывает вероятность отказа за время длительностью t, а вероятность противоположного события, которым является безотказность функционирования устройства в течение времени t , равна

$$R(t) = p\ (T > t) = 1 - p\ (T < t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

Функция $R(t)$ называется **функцией надежности**, с помощью ее значений вычисляют вероятность безотказной работы устройства за время t

$$R(t) = e^{-\lambda t}. \qquad (54)$$

Пример 9. Время безотказной работы башенного крана распределено по показательному закону. Кран ломается в среднем один раз в 50 суток.
Требуется:

1) найти функцию плотности $f(t)$, функцию распределения $F(t)$ и построить их графики (время t измеряется в сутках);

2) найти вероятность того, что кран проработает без отказа 100 дней.

Решение.

1) При показательном распределении математическое ожидание $M(x) = 1/\lambda$. Известно, что кран в среднем может безотказно проработать 50 суток, тогда $50 = 1/\lambda$, т. е. постоянная интенсивности отказов $\lambda = 0{,}02$. Соответственно, функция плотности (50) имеет вид:

$$f(t) = \begin{cases} 0, & t \le 0; \\ 0{,}02e^{-0{,}02t}, & t > 0. \end{cases}.$$

Соответственно, функция распределения (51) имеет вид:

$$F(t) = \begin{cases} 0, & t \le 0 \\ 1 - e^{-0{,}02t}, & t > 0 \end{cases}.$$

Укажем графики функций $f(t)$ и $F(t)$ (рис. 16).

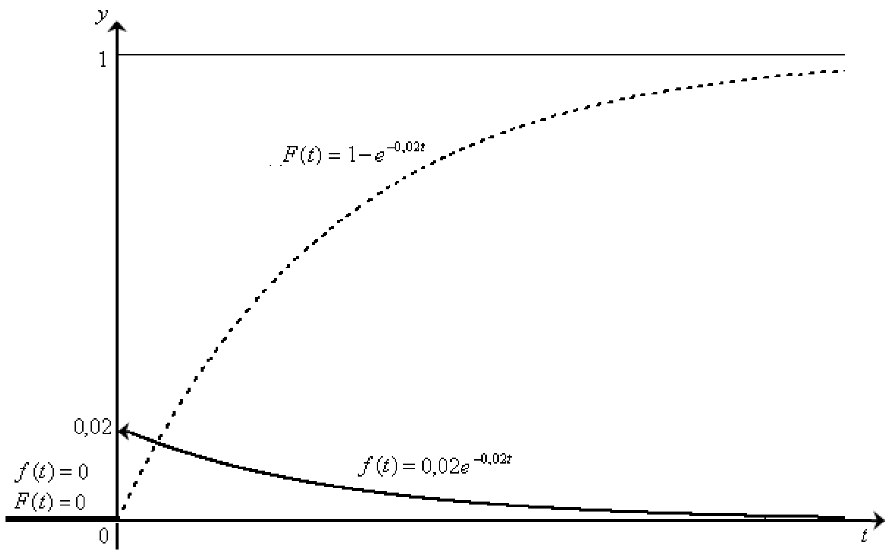


Рис. 16. Пример функции плотности и функции распределения показательно распределенной случайной величины

2) Вероятность того, что кран проработает без отказов 100 дней можно определить по формуле функции надежности (54): $R(100) = e^{-0.02\cdot 100} = e^{-2} \approx 0{,}14$, т. е. вероятность безотказной работы башенного крана не менее 100 суток достаточно мала и приближенно равна 0,14.

3.3.4. Нормальный закон распределения

Наиболее часто в окружающем мире встречается закон распределения, который называется *нормальным законом распределения*, он задается функцией плотности

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (55)$$

Особое внимание к этому закону распределения связано с тем, что именно он наиболее часто употребим для построения выводов и оценок массовых случайных событий.

Вычислим основные числовые характеристики *нормального* закона распределения.

$$M(X)=\int_{-\infty}^{+\infty}x\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\left|\begin{array}{l} t=\dfrac{x-a}{\sigma} \\ x=\sigma t+a \\ dx=\sigma dt \end{array}\right|=\left|\begin{array}{l} -\infty<x<\infty \\ \text{тогда} \\ -\infty<t<\infty \end{array}\right|=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma t+a)\cdot e^{-\frac{t^2}{2}}\sigma dt=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma t+a)\cdot e^{-\frac{t^2}{2}}dt=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t\cdot e^{-\frac{t^2}{2}}dt+\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt,$$

первый из полученных двух интегралов есть интеграл на симметричном промежутке от нечетной функции, он равен нулю $\int_{-\infty}^{+\infty}t\cdot e^{-\frac{t^2}{2}}dt=0$, а второй интеграл – есть интеграл Пуассона, он из “неберущихся”, но значение его известно

$\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$, тогда $M(x)=\frac{\sigma}{\sqrt{2\pi}}\cdot 0+\frac{a}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=a$.

$$D(X)=M(x^2)-\left(M(x)\right)^2=\int_{-\infty}^{+\infty}x^2\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}dx-a^2=\left|\begin{array}{l} t=\dfrac{x-a}{\sigma} \\ x=\sigma t+a \\ \sigma dt=dx \end{array}\right|=\left|\begin{array}{l} -\infty<x<\infty \\ \text{тогда} \\ -\infty<t<\infty \end{array}\right|=$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma t+a)^2\cdot e^{-\frac{t^2}{2}}\sigma dt-a^2=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma^2t^2+2\sigma ta+a^2)\cdot e^{-\frac{t^2}{2}}\sigma dt-a^2=$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^2\cdot e^{-\frac{t^2}{2}}dt+\frac{2\sigma a}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}t\cdot e^{-\frac{t^2}{2}}\sigma dt+\frac{a^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt-a^2=I_1+I_2+I_3-a^2.$$

Мы ввели обозначения для интегралов и отдельно каждый из них вычислим

$$I_1=\frac{\sigma^2}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}t^2\cdot e^{-\frac{t^2}{2}}dt=\frac{2\sigma^2}{\sqrt{2\pi}}\int\limits_{0}^{+\infty}t\cdot t\cdot e^{-\frac{t^2}{2}}dt=\left|\begin{array}{l}u=t\to du=dt\\ te^{-\frac{t^2}{2}}dt=dv\to\\ v=\int te^{-\frac{t^2}{2}}dt=\int e^{-\frac{t^2}{2}}d\frac{t^2}{2}=-e^{-\frac{t^2}{2}}\end{array}\right|=$$

$$=\frac{2\sigma^2}{\sqrt{2\pi}}\left(-\lim_{A\to+\infty}e^{-\frac{t^2}{2}}t\Big|_0^A+\int\limits_0^{+\infty}e^{-\frac{t^2}{2}}dt\right)=\frac{2\sigma^2}{\sqrt{2\pi}}\left(-\lim_{A\to+\infty}\left(e^{-\frac{A^2}{2}}A-e^{-\frac{0^2}{2}}0\right)+\frac{\sqrt{2\pi}}{2}\right)=\sigma^2.$$

$$I_2=\frac{2\sigma ta}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}t\cdot e^{-\frac{t^2}{2}}\sigma dt=\left|\text{интеграл от нечетной функции}\right|=0.$$

$$I_3=\frac{a^2}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\left|\text{содержит интеграл Пуассона}\right|=\frac{a^2}{\sqrt{2\pi}}\sqrt{2\pi}=a^2.$$

Итак, $D(X)=I_1+I_2+I_3-a^2=\sigma^2+0+a^2-a^2=\sigma^2$.

Случайная величина, описываемая нормальным законом распределения, имеет числовые характеристики

$$M(X)=a,\ D(X)=\sigma^2,\ \sigma(X)=\sigma. \qquad (56)$$

Графиком функции плотности нормального закона распределения (55) является кривая Гаусса (рис. 17) центром симметрии которой является прямая $x=a$, и очевидной является зависимость графика от характеристики σ.

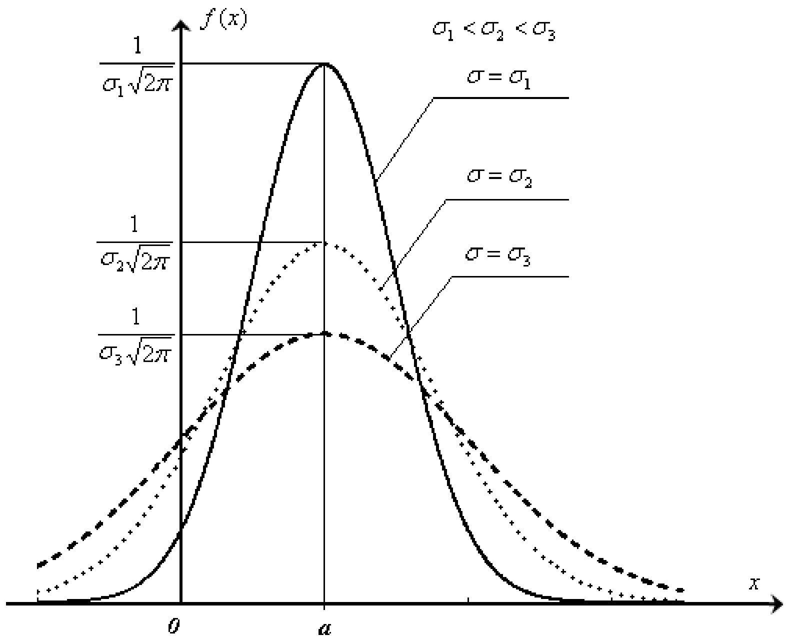


Рис. 17. Графики функции плотности нормально распределенной СВ

Точка максимума лежит на прямой $x = a$.

Чем меньше значение σ, тем выше будет точка максимума функции плотности, а график ближе прижат к прямой $x = a$ и тем быстрее приближается к оси Ox при $|x| \to \infty$ (т. к. площадь под любой кривой Гаусса равна единице), что наглядно характеризует вероятностный смысл σ как разброс значений случайной величины относительно среднего значения a.

Определим вероятность попадания случайной величины в интервал $(\alpha;\beta)$

$$p(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}}dx = \left| \begin{array}{l} t = \dfrac{x-a}{\sigma} \\ x = \sigma t + a \\ \sigma dt = dx \end{array} \right| = \left| \begin{array}{l} \alpha < x < \beta \\ \text{тогда} \\ \dfrac{\alpha - a}{\sigma} < t < \dfrac{\beta - a}{\sigma} \end{array} \right| =$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-\frac{t^2}{2}}\sigma dt = \frac{1}{\sqrt{2\pi}}\int_{0}^{\frac{\beta-a}{\sigma}} e^{-\frac{t^2}{2}}dt - \frac{1}{\sqrt{2\pi}}\int_{0}^{\frac{\alpha-a}{\sigma}} e^{-\frac{t^2}{2}}dt = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right),$$

где $\Phi(x)$ – функция Лапласа (27), т. е. справедлива формула

$$p(\alpha < X < \beta) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right). \qquad (57)$$

Из формулы (57) можно получить следующие важные следствия:

- вероятность того, что отклонение значения нормально распределенной случайной величины X от ее среднего значения (математического ожидания) по абсолютной величине меньше положительного числа δ

$$p(|X - a| < \delta) = p(a - \delta < X < a + \delta) = \Phi\left(\frac{a + \delta - a}{\sigma}\right) - \Phi\left(\frac{a - \delta - a}{\sigma}\right) =$$

$$= \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right) = \Phi\left(\frac{\delta}{\sigma}\right) + \Phi\left(\frac{\delta}{\sigma}\right) = 2\Phi\left(\frac{\delta}{\sigma}\right),$$

т. е.

$$p(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right). \qquad (58)$$

- правило “трех сигм” (“3σ”): для $\delta = 3\sigma$ по формуле (58)

$$p(a - 3\sigma < X < a + 3\sigma) = 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3) = 2 \cdot 0.49865 = 0.9973,$$

где значение $\Phi(3)$ нашли по таблице значений функции Лапласа [1, прил. 2], следовательно, почти достоверно, что нормально распределенная случайная величина отклонится от математического ожидания по абсолютной величине не больше чем на 3σ. На рис. 18 изображена кривая с учетом сказанного правила:

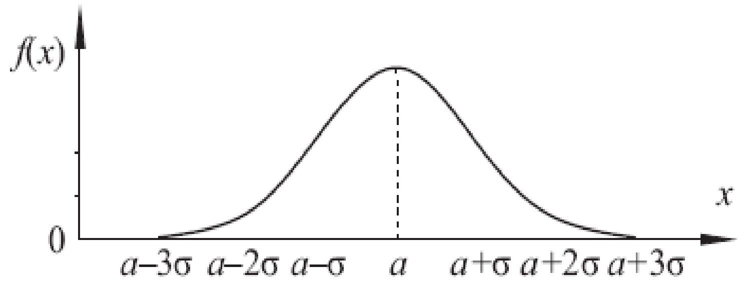


Рис. 18. Правило “трех сигм” в построении графика функции (55)

Пример 10. Диаметр деталей, вытачиваемых на станке, является случайной величиной, распределенной по нормальному закону с параметрами $a = 20$ *мм* (ожидаемое значение диаметра) и $\sigma = 2$ (допустимое отклонение от среднего значения).

Найти:

1) вероятность того, что диаметр детали d лежит в диапазоне $19 < d < 22$;

2) вероятность того, что отклонение диаметра от ожидаемого (от математического ожидания $a = 20$ *мм*) не превышает 1 *мм*;

3) указать функцию плотности распределения и нарисовать ее график.

Решение.

1) Вероятность того, что деталь будет нужного размера, определяется формулой (57), где $a = 20$, $\sigma = 2$, $\alpha = 19$ и $\beta = 22$:

$$p(19 < X < 22) = \Phi\left(\frac{22-20}{2}\right) - \Phi\left(\frac{19-20}{2}\right) = \Phi(1) - \Phi(-0{,}5) =$$
$$= \Phi(1) + \Phi(0{,}5) = 0{,}3413 + 0{,}1915 = 0{,}5328.$$

При вычислениях воспользовались свойствами интегральной функции Лапласа $\Phi(x)$, а именно тем, что она нечетная, а значения $\Phi(1)$ и $\Phi(0{,}5)$ нашли по таблице ее значений [1, прил. 2].

2) Вероятность того, что отклонение диаметра от ожидаемого не превышает 1 *мм*, можно вычислить по формуле (58), где в нашем случае $\delta = 1$:

$$p(|X - 20| < 1) = 2\Phi\left(\frac{\delta}{\sigma}\right) = 2\Phi\left(\frac{1}{2}\right) = 2\Phi(0{,}5) = 2 \cdot 0{,}1915 = 0{,}383.$$

3) Функция плотности распределения имеет вид $f(x) = \dfrac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{(x-20)^2}{8}}$, график которой указан ниже. Согласно правилу “3σ” наиболее выпуклая часть кривой находится на интервале $a - 3\sigma < X < a + 3\sigma \Rightarrow 20 - 3 \cdot 2 < X < 20 + 3 \cdot 2$, т.е.

это интервал $14 < X < 26$, вне этого интервала значение функции можно считать равной нулю.

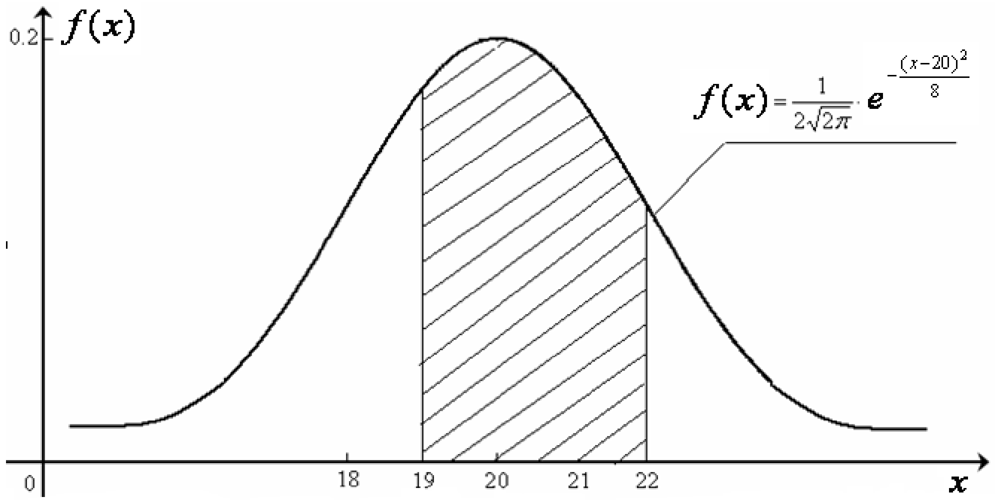


Рис. 19. График функции плотности
нормального распределения случайной величины

Замечание. Вероятности $p(19 < d < 22)$ соответствует заштрихованная площадь на рис. 19 под кривой.

Далее обоснуем широкое распространение данного закона распределения. Именно с важности нормального закона распределения начинался этот параграф. Приведенная ниже теорема раскрывает смысл этих высказываний

Теорема. (*Центральная предельная теорема Ляпунова*)

Если случайная величина X представляет собой сумму большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то X имеет распределение, близкое к нормальному.

Как применяется данная теорема на практике?

Рассматривая повторяющийся опыт, контролируется значение некоторой случайной величины, которая имеет возможность принимать различные значения.

Пусть опытом будем считать измерение показателей работы некоторого прибора. При многократном измерении они различны. Тогда можно считать, что влияние одного конкретного измерения на всю сумму измерений ничтожно мало. Между собой эти измерения независимы, а значит, в условия теоремы Ляпунова наши испытания попадают.

Что из полученного следует?

По набору измерений можно найти среднее значение и среднее квадратическое отклонение, по ним составляется функция плотности, по которой изучают величину и вероятность погрешностей.

На практике для большинства случайных величин выполняются условия теоремы Ляпунова, и поэтому именно нормальное распределение является основным в построении выводов.

4. Системы случайных величин

4.1. Дискретные и непрерывные случайные величины, их способы задания

Ранее рассматривались одномерные случайные величины, которые характеризовались одним числом, однако существуют случайные величины, которые описываются двумя, тремя и т. д. числами, они называются двумерными, трехмерными и т. д. Для них тоже есть деление на дискретные и непрерывные.

Например, дискретная трехмерная случайная величина $\left(X_1; X_2; X_3\right)$ - значения параметров выпускаемой детали (длина, ширина, высота) для проверки ее на качество и определение процента брака.

Законом распределения системы случайных величин называется соотношение между областями возможных значений системы случайных величин и вероятностями попадания ее значений в эти области.

Далее остановимся на системах случайных величин, заданных двумя значениями. Двумерная случайная величина $\left(X;Y\right)$ называется дискретной, если дискретны составляющие ее одномерные случайные величины $x_i \subset X$ и $y_j \subset Y$, где $i = 1,2,\ldots,n;\ j = 1,2,\ldots,m$. Закон распределения вероятностей такой случайной величины может быть задан таблицей с двумя переменными.

Двумерный закон распределения Таблица 5

Y	X			
	x_1	x_2	...	x_n
y_1	p_{11}	p_{21}	$\cdots$	p_{n1}
y_2	p_{12}	p_{22}	$\cdots$	p_{n2}
...	$\cdots$	$\cdots$	$\cdots$	$\cdots$
y_m	p_{1m}	p_{2m}	$\cdots$	p_{nm}

Это табличный способ задания двумерной случайной величины, где сумма вероятностей равна единице:

$$\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} = 1 \qquad (p_{ij} = p(x_i; y_j)). \tag{59}$$

Случайная величина может быть задана формулой – это, в первую очередь, функция распределения, которая подходит для описания и дискретных и непрерывных СВ.

Функцией распределения системы двух случайных величин называется функция двух аргументов $F(x;y)$, принимающая значение вероятности совместного выполнения двух условий $X < x$, $Y < y$:

$$F(x;y) = p(X < x; Y < y), \tag{60}$$

для дискретной СВ

$$F(x;y) = \sum_{x_i < x}\sum_{y_j < y} p_{ij}, \tag{60.1}$$

для непрерывной СВ

$$F(x;y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy, \tag{60.2}$$

где функция $f(x,y)$ – плотность совместного распределения (61).

Для дискретного случая область на плоскости функции указан на рис. 20.

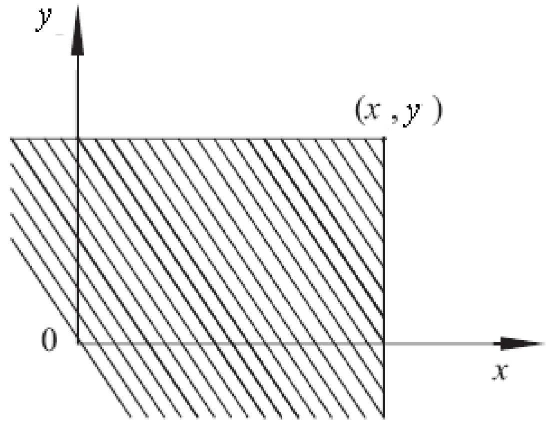


Рис. 20. Формула (60) определяет вероятность попадания системы случайных величин в заштрихованную область

Отметим ***свойства функции распределения*** системы двух СВ:

1. $0 \le F(x,y) \le 1$, т. к. является вероятностью;
2. Имеют место предельные соотношения
$$\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = \lim_{\substack{x\to-\infty \\ y\to-\infty}} F(x,y) = 0, \ \lim_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y) = 1;$$
3. По каждому из аргументов функция распределения является неубывающей функцией;
4. Если один из аргументов стремится к плюс бесконечности, то справедливы пределы
$$\lim_{y\to\infty} F(x,y) = F_1(x);$$
$$\lim_{x\to\infty} F(x,y) = F_2(y);$$
5. Вероятность попадания случайной точки $(X;Y)$ в произвольный прямоугольник со сторонами, параллельными координатным осям, вычисляется по формуле (см. рис. 21):

$$P(x' \le X \le x'', y' \le Y \le y') = F(x'', y'') - F(x', y'') - F(x'', y') + F(x', y').$$

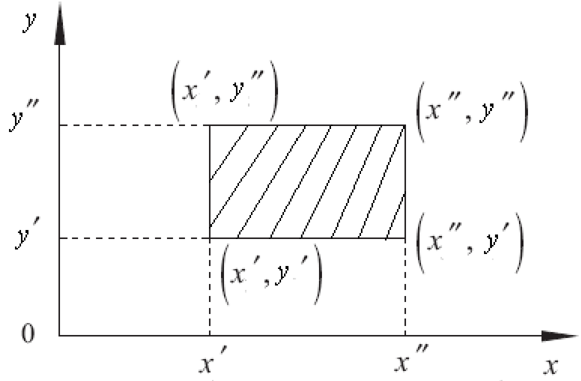


Рис. 21. Вероятность попадания случайной точки $(x; y)$ в произвольный прямоугольник со сторонами параллельными осям координат

Двумерная случайная величина $(X;Y)$ непрерывна, если непрерывными являются составляющие ее одномерные случайные величины X и Y. Для такой случайной величины тоже подходит ее описание с помощью функции распределения, имеющей такое же определение, что и для дискретной случайной величины. Однако к указанным ранее для нее свойствам добавляется свойство *непрерывности* $F(x; y)$.

Непрерывную функцию $F(x; y)$ можно дифференцировать, а значит получаем способ описания НСВ через функцию плотности $f(x; y)$, который неприменим для ДСВ.

Плотностью совместного распределения вероятностей двумерной случайной величины $(X;Y)$ называется вторая частная смешанная производная функции распределения:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \qquad (61)$$

По известной функции плотности распределения может быть найдена функция распределения:

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

Вероятность попадания значений случайной величины в область Ω определяется интегрированием:

$$p((x, y) \in \Omega) = \iint_{\Omega} f(x, y) dx dy.$$

Двойной интеграл по всей плоскости *xOy* от функции двумерной плотности равен единице:

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

Случайные величины называются ***независимыми***, если закон распределения одной из них не зависит от значений, принимаемой другой случайной величиной.

Теорема. *Для того чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы функция распределения системы (X, Y) была равна произведению функций распределения ее составляющих*

$$F(x,y) = F_1(x)F_2(y).$$

Аналогичную теорему можно сформулировать и для плотности распределения.

Теорема. *Для того чтобы случайные величины X и Y были независимы, необходимо и достаточно, чтобы плотность совместного распределения системы (X, Y) была равна произведению функций плотностей распределения ее составляющих*

$$f(x,y) = f_1(x)f_2(y).$$

4.2. Числовые характеристики системы случайных величин

Для системы нескольких случайных величин тоже можно ввести числовые характеристики. Их смысл аналогичен одномерным случайным величинам. Применять будем обозначения в соответствии с принятыми нормами.

Определение. *Начальным моментом порядка* k,s двумерной случайной величины (X,Y) называется математическое ожидание произведения X^k на Y^s:

$$a_{k,s} = M(X^kY^s), \tag{62}$$

для дискретных случайных величин формула имеет вид

$$a_{k,s} = \sum_i\sum_j x_i^k y_j^s p_{ij}, \tag{62.1}$$

а для непрерывных

$$a_{k,s} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^s f(x,y)dxdy. \tag{62.2}$$

Очевидно, $a_{1,0} = M(X)$, $a_{0,1} = M(Y)$ – математические ожидания случайных величин X и Y соответственно.

Определение. *Центральным моментом порядка* k,s двумерной случайной величины (X,Y) называется математическое ожидание произведения $(X - M(x))^k$ на $(Y - M(y))^s$:

$$\mu_{k,s} = M\left((X - M(x))^k (Y - M(Y))^s\right), \tag{63}$$

для дискретных случайных величин

$$\mu_{k,s} = \sum_i\sum_j \left(x_i - M(X)\right)^k \left(y_j - M(Y)\right)^s p_{ij}, \tag{63.1}$$

а для непрерывных

$$\mu_{k,s} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x - M(X)\right)^k \left(y - M(Y)\right)^s f(x,y)dxdy\,. \qquad (63.2)$$

Очевидно, $\mu_{2,0} = D(X)$, $\mu_{0,2} = D(Y)$ – дисперсии случайных величин X и Y соответственно.

При $k = s = 1$ в теории корреляции используется корреляционный момент $\mu = \mu_{1,1} = M\left((X - M(X)(Y - M(Y)\right)$. На практике для нахождения $M(X)$, $M(Y)$, а так же $D(X)$, $D(Y)$ переходят, суммируя в таблице p_{ij} по строкам или столбцам, к одномерным случайным величинам Y и X.

Пример 1. Вычислить числовые характеристики каждой из случайных величин.

Y	*X*			
	1	*3*	*4*	*10*
-2	0	0.04	0.03	0.07
1	0.1	0.15	0.15	0.25
4	0.05	0.06	0.07	0.03

Имея такое задание системы случайных величин, можно считать, что каждая из величин задана дискретно, просуммировав по строкам и столбцам выпишем их законы распределения

Y		
-2	*1*	*4*
0.14	0.65	0.21

X			
1	*3*	*4*	*10*
0.15	0.25	0.25	0.35

и далее вычислять характеристики для каждой из СВ:

$M(Y) = a_{1,0} = -2 \cdot 0.14 + 1 \cdot 0.65 + 4 \cdot 0.21 = 1.21;$

$M(X) = a_{0,1} = 1 \cdot 0.15 + 3 \cdot 0.25 + 4 \cdot 0.25 + 10 \cdot 0.35 = 5.4;$

$D(Y) = \mu_{2,0} = M(Y^2) - M^2(Y) = (-2)^2 \cdot 0.14 + 1^2 \cdot 0.65 + 4^2 \cdot 0.21 - (1.21)^2 = 3.11;$

$D(X) = \mu_{0,2} = M(X^2) - M^2(X) =$

$$= 1^2 \cdot 0.15 + 3^2 \cdot 0.25 + 4^2 \cdot 0.25 + 10^2 \cdot 0.35 - (5.4)^2 = 12.24.$$

В 1.4. мы рассматривали условную вероятность $p_A(B)$ наступления события B при условии, что уже произошло событие A. Аналогично для системы случайных величин рассматривается распределение одной СВ при условии, что другая приняла конкретные значения.

Условным законом распределения одной из случайных величин, входящих в систему, называется распределение, найденное из условия, что другая случайная величина приняла определенной значение.

Условная плотность распределения вычисляется по формулам:

$$\varphi(x/y)=\frac{f(x,y)}{f_2(y)}=\frac{f(x,y)}{\int\limits_{-\infty}^{\infty}f(x,y)dx};\qquad(64)$$

$$\psi(y/x)=\frac{f(x,y)}{f_1(x)}=\frac{f(x,y)}{\int\limits_{-\infty}^{\infty}f(x,y)dy}.\qquad(65)$$

Условная плотность распределения обладает всеми свойствами плотности распределения одной случайной величины.

Условным математическим ожиданием дискретной случайной величины Y при фиксированном $X=x$ называется произведение всех возможных значений Y на их условные вероятности

$$M(Y/X=x)=\sum_{j=1}^{m}y_j\,p(y_j/x),\qquad(66)$$

для непрерывных случайных величин имеем

$$M(Y/X=x)=\int\limits_{-\infty}^{\infty}y\psi(y/x)dy.\qquad(67)$$

где $\psi(y/x)$ – условная плотность случайной величины Y при $X=x$, определяемая формулой (65).

Условным математическим ожиданием дискретной случайной величины X при $Y=y$ (y – одно из возможных значений Y) называется произведение всех возможных значений X на их условные вероятности

$$M(X/Y=y)=\sum_{i=1}^{n}x_i\,p(x_i/y),\qquad(68)$$

для непрерывных случайных величин имеем

$$M(X/Y=y)=\int\limits_{-\infty}^{\infty}x\varphi(x/y)dx,\qquad(69)$$

где $\varphi(x/y)$ – условная плотность случайной величины X при $Y=y$, определяемая формулой (64).

$M(Y/X=x)=g(x)$ ($M(X/Y=y)=h(y)$) – условные математические ожидания являются функциями от x (от y) и называются ***функциями регрессии*** X *на* Y (Y *на* X).

Пример 2. Вычислить числовые условные характеристики каждой из случайных величин.

X	*Y*			
	1	*3*	*4*	*10*

-2	0	0.04	0.03	0.07
1	0.1	0.15	0.15	0.25
4	0.05	0.06	0.07	0.03

$$M(X/Y=1) = -2\cdot 0 + 1\cdot 0{,}1 + 4\cdot 0{,}05 = 0{,}3;$$

$$M(X/Y=3) = 0.31;\ M(X/Y=4) = 0{,}37;\ M(X/Y=10) = 0{,}23;$$

$$M(Y/X=-2) = 1\cdot 0 + 3\cdot 0{,}04 + 4\cdot 0{,}03 + 10\cdot 0{,}07 = 0{,}94;$$

$$M(Y/X=1) = 3{,}65;\ M(Y/X=4) = 0{,}81.$$

Замечание. Мы подробно расписали вычисления для первых столбца и строки. Для остальных строк и столбцов проверьте вычисления самостоятельно.

4.3. Статистическая и корреляционная зависимость. Коэффициент корреляции, уравнение прямой линии регрессии

При работе с системами случайных величин основным вопросом является их взаимодействие, их влияние друг на друга.

В естественных науках основной задачей является построение функциональной зависимости между переменными, т. е. когда одному конкретному значению одной переменной соответствует единственное значение другой переменной (например, зависимость расстояния от времени).

В прикладных задачах часто существуют другие зависимости, это связи, когда одному значению одной переменной может соответствовать множество значений другой переменной, что зависит от каких-то дополнительных условий, – такая зависимость называется *статистической*. Например, статистической зависимостью является зависимость успеваемости студента от его присутствия на занятиях, от его навыков, полученных в школе, от его целеустремленности.

В силу неоднозначности статистической зависимости рассматривается зависимость между случайными величинами, которая определяет зависимость среднего значения одной переменной от среднего значения другой переменной. Такая статистическая зависимость называется *корреляционной*. Она может описываться через линейные, квадратичные и другие функции, что и определяет вид связи.

На рис. 22 [6]. указаны задачи, решаемые для систем случайных величин.

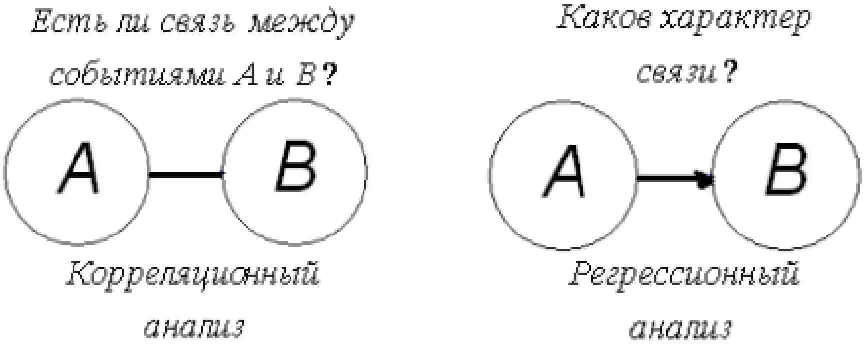


Рис. 22.

Подводя итог вышесказанному, можно сказать, что корреляционный анализ решает задачи об определении связи между двумя явлениями – скоростью в беге и ростом спортсмена или общей успеваемостью студента и его оценкой по математике и т. д., а регрессионный анализ позволяет понять характер этой зависимости. В нашем курсе рассматривается и устанавливается только линейная зависимость. При этом нужно понимать, что это далеко не единственный вид зависимости, который существует.

Корреляционная зависимость может быть представлена в виде:

$$M(Y/X=x)=\varphi(x),$$

$$M(X/Y=y)=\psi(y),$$

или в виде:

$$y_x=\varphi(x), \qquad (70)$$

$$x_y=\psi(y). \qquad (71)$$

Уравнения (70) и (71) называются *уравнениями регрессии* соответственно Y по X и X по Y, функции $\varphi(x)$, $\psi(y)$ – *функциями регрессии*, а их графики – *линиями регрессии*.

Пусть в результате n независимых опытов получены n пар чисел: $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$. По этим данным наблюдений надо составить уравнение прямой линии регрессии Y на X:

$$\overline{y_x}=kx+b.$$

Угловой коэффициент прямой линии регрессии Y на X называют *выборочным коэффициентом* регрессии Y на X и обозначают через ρ_{yx}.

Таким образом, будем искать выборочное уравнение прямой линии регрессии Y на X в виде уравнения

$$y_x=\rho_{yx}x+b. \qquad (72)$$

Подберём параметры ρ_{yx} и b так, чтобы точки (x_1,y_1), (x_2,y_2), (x_3,y_3),..., (x_n,y_n) построенные по данным наблюдения, на плоскости $x0y$ лежали как можно ближе к прямой (72), y_k – наблюдаемые ординаты, соответствующие x_k.

Применим метод наименьших квадратов, который заключается в том, чтобы сумма квадратов отклонений была наименьшей. Так как каждое отклонение зависит от отыскиваемых параметров, то и сумма квадратов отклонений есть функция F этих отклонений $F(\rho, b) = \sum_{k=1}^{n} (\rho x_k + b - y_k)^2$, где упрощаем запись $\rho = \rho_{yx}$.

Для вычисления минимума приравниваем нулю частные производные:

$$\begin{cases} \dfrac{\partial F}{\partial \rho} = 2\sum_{k=1}^{n} (\rho x_k + b - y_k) x_k = 0; \\ \dfrac{\partial F}{\partial b} = 2\sum_{k=1}^{n} (\rho x_k + b - y_k) = 0. \end{cases}$$

Преобразуя эти уравнения, получим систему двух линейных уравнений относительно ρ и b, далее не будем писать индексы у знаков сумм, считая k=1, 2,…, n:

$$\begin{cases} \rho \sum x_i^{\,2} + b \sum x_i = \sum x_i y_i; \\ \rho \sum x_i + nb = \sum y_i. \end{cases}$$

Решив эту систему, найдём искомые коэффициенты в уравнении (72)

$$\rho_{yx} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^{\,2} - (\sum x_i)^2},$$

$$b = \frac{\sum x_i^{\,2} \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^{\,2} - (\sum x_i)^2},$$

или, упростив запись,

$$\rho_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}, \; b = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2}. \qquad (73)$$

Аналогично можно найти выборочное уравнение прямой линии регрессии X на Y:

$$\overline{x}_y = \rho_{xy} y + c, \qquad (74)$$

где ρ_{xy} – выборочный коэффициент регрессии X на Y. Повторяя аналогичные рассуждения, получаем коэффициенты

$$\rho_{xy} = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}, \; c = \frac{\sum x \sum y^2 - \sum y \sum xy}{n\sum y^2 - (\sum y)^2}. \qquad (75)$$

Поскольку

$$\overline{x} = \frac{\sum x}{n}, \; \overline{y} = \frac{\sum y}{n}, \; \overline{xy} = \frac{\sum xy}{n}, \; \sigma_x^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n} \right)^2, \; \sigma_y^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n} \right)^2,$$

тогда

$$\rho_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{\frac{\sum xy}{n} - \frac{\sum x}{n}\frac{\sum y}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x^2} = \frac{\sigma_y}{\sigma_x}\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} = \frac{\sigma_y}{\sigma_x}\rho_r,$$

где коэффициент корреляции

$$\rho_r = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} \tag{76}$$

характеризует силу линейной связи между случайными величинами.

Преобразуем в (73) коэффициент

$$b = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = \frac{\frac{\sum x^2}{n}\frac{\sum y}{n} - \frac{\sum x}{n}\frac{\sum xy}{n}}{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - \left(\overline{x}\right)^2} =$$

$$= \frac{\overline{x^2}\,\overline{y} - \left(\overline{x}\right)^2\overline{y} + \left(\overline{x}\right)^2\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - \left(\overline{x}\right)^2} = \overline{y}\frac{\overline{x^2} - \left(\overline{x}\right)^2}{\overline{x^2} - \left(\overline{x}\right)^2} - \overline{x}\frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} = \overline{y} - \overline{x}\frac{\sigma_y}{\sigma_x}\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} = \overline{y} - \overline{x}\frac{\sigma_y}{\sigma_x}\rho_r$$

т. е.

$$b = \overline{y} - \overline{x}\frac{\sigma_y}{\sigma_x}\rho_r .$$

Подставив ρ_{yx} и b в уравнение (72), получим

$$y_x = \frac{\sigma_y}{\sigma_x}\rho_r x + \overline{y} - \overline{x}\frac{\sigma_y}{\sigma_x}\rho_r ,$$

которое задает уравнение прямой линии регрессии Y на X в виде прямой с угловым коэффициентом, проходящей через заданную точку $(\overline{x}; \overline{y})$:

$$y_x - \overline{y} = \rho_r \frac{\sigma_y}{\sigma_x}(x - \overline{x}) . \tag{77}$$

Аналогично уравнение прямой линии регрессии X на Y можно записать в виде:

$$x_y - \overline{x} = \rho_r \frac{\sigma_x}{\sigma_y}(y - \overline{y}) . \tag{78}$$

Коэффициент корреляции ρ_r (см. стр. 76), входящий в уравнения прямых линий регрессии (77) и (78), обладает важными свойствами (далее отметим их без доказательства), которые позволяют оценить тесноту связи между случайными величинами X и Y.

Свойства коэффициента корреляции

1. Коэффициент корреляции по модулю не превышает единицы, т. е. $|\rho_r| \le 1$.

2. Если случайные величины X и Y линейно независимы, то $\rho_r = 0$.
3. Если величины X и Y связаны линейной функциональной зависимостью, то $|\rho_r| = 1$.

Таким образом, по значению коэффициента корреляции оценивают, насколько тесная линейная взаимосвязь рассматриваемых случайных величин. Чем ближе $|\rho_r|$ к единице, тем теснее линейная корреляционная зависимость между X и Y, и наоборот, чем ближе $|\rho_r|$ к нулю, тем слабее линейная взаимосвязь.

величина коэффициента корреляции по абсолютной величине	название связи
$\lvert\rho_r\rvert = 0$	нет связи
$0 < \lvert\rho_r\rvert < \lvert 0,3\rvert$	слабая
$\lvert 0,3\rvert \le \lvert\rho_r\rvert < \lvert 0,7\rvert$	средняя
$\lvert 0,7\rvert \le \lvert\rho_r\rvert < \lvert 1\rvert$	сильная
$\lvert\rho_r\rvert = 1$	функциональная

Пример 3. На геодезическом пункте одновременно выполнялись измерения угла двумя теодолитами, из которых один был установлен на сигнале, а другой — на штативе под сигналом. При этом получены результаты наблюдений, которые представлены в следующей таблице [7].

№ наблюдений	1	2	3	4	5	6	7	8	9	10
Сигнал (X)	3,21	2,21	2,11	2,49	3,19	3,29	4,41	2,67	0,75	2,20
Штатив (Y)	5,10	4,02	3,22	2,57	3,62	4,93	6,53	4,59	2,84	3,46

Вычислить оценку коэффициента корреляции между наблюдениями X и Y. Построить уравнение регрессии значения угла, измеренного на штативе, в зависимости от значения угла, полученного на сигнале.

Решение. Составим вспомогательную таблицу.

№	X	X^2	Y	Y^2	XY
1	3.21	10.30	5.10	26.01	16.37
2	2.21	4.88	4.02	16.16	8.88
3	2.11	4.45	3.22	10.37	6.79
4	2.49	6.20	2.57	6.60	6.40
5	3.19	10.18	3.62	13.10	11.55
6	3.29	10.82	4.93	24.30	16.22

7	4.41	19.45	6.53	42.64	28.80
8	2.67	7.13	4.59	21.07	12.26
9	0.75	0.56	2.84	8.07	2.13
10	2.20	4.84	3.46	11.97	7.61
$\sum$	**26.53**	**78.81**	**40.88**	**180.29**	**117.01**

Из таблицы получим: $\bar{x} = \dfrac{\sum x}{n} = \dfrac{26{,}53}{10} \approx 2{,}65$, $\bar{y} = \dfrac{\sum y}{n} = \dfrac{40{,}88}{10} \approx 4{,}09$,

$$\overline{xy} = \frac{\sum xy}{n} = \frac{117}{10} = 11{,}7 .$$

$$\sigma_x^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{78{,}81}{10} - (2{,}65)^2 \approx 0{,}86 \Rightarrow \sigma_x = \sqrt{0{,}86} \approx 0{,}93,$$

$$\sigma_y^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2 = \frac{180{,}3}{10} - (4{,}09)^2 \approx 1{,}30 \Rightarrow \sigma_y = \sqrt{1{,}3} \approx 1{,}14 .$$

Итак, коэффициент корреляции $\rho_r = \dfrac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x \sigma_y} = \dfrac{11{,}7 - 2{,}65\cdot 4{,}09}{0{,}93\cdot 1{,}14} = 0{,}81$.

Из полученного значения коэффициента корреляции очевидна сильная линейная связь.

Построить уравнение регрессии значения угла, измеренного на штативе (Y), в зависимости от значения угла, полученного на сигнале (X), – это уравнение зависимости $y_x - \bar{y} = \rho_r \dfrac{\sigma_y}{\sigma_x}(x - \bar{x})$, тогда $y_x - 4{,}09 = 0{,}81\dfrac{1{,}14}{0{,}93}(x - 2{,}65)$.

Итак, уравнение прямой линии регрессии Y на X имеет вид

$$y = 0{,}99x + 1{,}47 .$$

Данные в задачах на корреляцию могут быть сгруппированы и представлены в виде табл. 5.

Данные имеет смысл группировать в том случае, если во всем наборе упорядоченных пар чисел есть пары, которые встречаются несколько раз. Например, в следующем примере пара (X=45;Y=15) встретилась во всем наборе данных 10 раз и т. д.

Пример 4. Данные наблюдений системы случайных величин представлены в таблице.

n_{ij}		Y				
		11	13	15	17	19
X	25	4	3	1		
	35	3	5	2	2	
	45	1	4	10	4	
	55		3	4	5	2
	65			1	3	3

1. Установить силу линейной зависимости между значениями случайных величин.

2. Найти уравнения прямых линий регрессий и построить соответствующие прямые.

Решение.

1. Составим вспомогательную таблицу, найдем суммы частот по строкам и столбцам

n_{ij}		Y					
		11	13	15	17	19	$\sum$
X	25	4	3	1			8
	35	3	5	2	2		12
	45	1	4	10	4		19
	55		3	4	5	2	14
	65			1	3	3	7
$\sum$		8	15	18	14	5	60

Из таблицы вычислим числовые характеристики:

$$\overline{x} = \frac{\sum x_i n_i}{n} = \frac{25\cdot 8 + 35\cdot 12 + 45\cdot 19 + 55\cdot 14 + 65\cdot 7}{60} = \frac{2700}{60} = 45,$$

$$\overline{x^2} = \frac{\sum x_i^2 n_i}{n} = \frac{25^2\cdot 8 + 35^2\cdot 12 + 45^2\cdot 19 + 55^2\cdot 14 + 65^2\cdot 7}{60} = \frac{130100}{60} \approx 2168{,}33,$$

$$\overline{y} = \frac{\sum y_j n_j}{n} = \frac{11\cdot 8 + 13\cdot 15 + 15\cdot 18 + 17\cdot 14 + 19\cdot 5}{60} = \frac{886}{60} \approx 14{,}77,$$

$$\overline{y^2} = \frac{\sum y_j^2 n_j}{n} = \frac{11^2\cdot 8 + 13^2\cdot 15 + 15^2\cdot 18 + 17^2\cdot 14 + 19^2\cdot 5}{60} = \frac{13404}{60} = 223{,}4,$$

$$\overline{xy} = \frac{\sum x_i y_j n_{ij}}{n} = \frac{11\cdot 25\cdot 4 + 11\cdot 35\cdot 3 + 11\cdot 45\cdot 1 + 13\cdot 25\cdot 3 + \ldots}{60} = \frac{40970}{60} \approx 682{,}83,$$

$$\sigma_x^2 = \left(\overline{x^2}\right) - \left(\overline{x}\right)^2 = 2168{,}33 - (45)^2 = 143{,}33 \Rightarrow \sigma_x = \sqrt{143{,}33} \approx 11{,}97,$$

$$\sigma_y^2 = \left(\overline{y^2}\right) - \left(\overline{y}\right)^2 = 223{,}4 - (14{,}77)^2 = 5{,}25 \quad \sigma_y = \sqrt{5{,}25} \approx 2{,}29.$$

Итак, коэффициент корреляции $\rho_r = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \sigma_y} = \dfrac{682{,}83 - 45 \cdot 14{,}77}{11{,}97 \cdot 2{,}29} \approx 0{,}66$.

Из полученного значения коэффициента корреляции очевидна средняя линейная связь.

Найдем уравнения регрессий:

- X на Y: $y_x - \overline{y} = \rho_r \dfrac{\sigma_y}{\sigma_x}(x - \overline{x})$, тогда $y_x - 14{,}77 = 0{,}66 \dfrac{2{,}29}{11{,}97}(x - 45)$, из чего получим уравнение $\underline{y = 0{,}13x + 9{,}09}$;
- Y на X: $x_y - \overline{x} = \rho_r \dfrac{\sigma_x}{\sigma_y}(y - \overline{y})$, тогда $x_y - 45 = 0{,}66 \dfrac{11{,}97}{2{,}29}(y - 14{,}77)$, из чего получим уравнение $\underline{x = 3{,}45y - 5{,}95}$.

Далее укажем на рис. 23 графики уравнений регрессий с начальными точечными данными. Заметим, что на рисунке не виден “вес” каждой из точек.

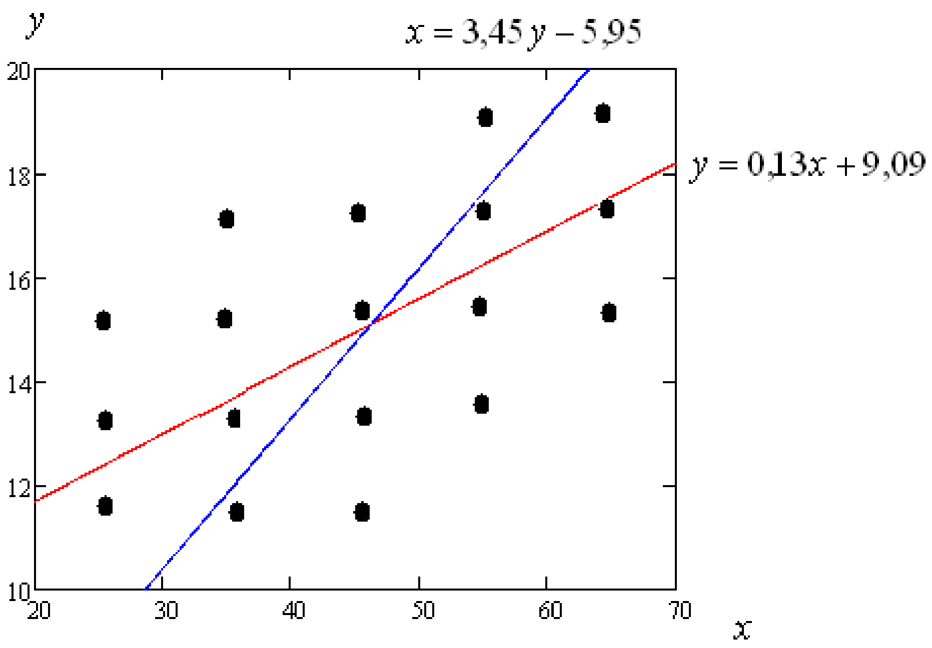


Рис. 23. Прямые линии регрессий задачи 4

Замечание. Вычисления можно существенно упростить, если воспользоваться методом условных вариант (см. стр. 85-86).

5. Элементы математической статистики

5.1. Понятие генеральной совокупности, выборки

Математическая статистика – это наука о массовых случайных явлениях.

Математическая статистика опирается на теоремы и определения теории вероятностей, однако отличие состоит в том, что выводы носят приблизительный характер, что следует из основного метода науки – выборочности данных. Тем не менее, разработаны критерии, позволяющие оценить степень надежности полученных результатов.

Пусть рассматривается совокупность однородных объектов, обладающих признаком *X*, изучение которого следует провести. Изучаемая характеристика объекта может быть сразу представлена в числовом виде или, если это фразеологическое выражение, то одинаковым фразам математически присваиваются одинаковые числа, что переводит данные на язык математики. Может быть так, что нет возможности изначально знать (перечислить или пересчитать) все представленное для исследования множество значений признака *X*, т. е. составить полную группу значений изучаемой характеристики, но возможно одинаковым ответам присвоить одинаковые номера, и таким образом, перевести все значения на язык цифр.

Генеральной совокупностью называется все множество значений характеристики изучаемого объекта. Генеральные совокупности обозначаются прописными буквами латинского алфавита X, Y и т. д., а ее элементы обозначаются малыми буквами, например, $x_1, x_2, ..., x_n \in X$, или $y_1, y_2, ... \in Y$.

Число элементов N в генеральной совокупности называют ***мощностью.*** Из всех элементов генеральной совокупности можно найти ее минимальный $x_{\min}$ и максимальный $x_{\max}$ элементы, тогда разность между ними $R = x_{\max} - x_{\min}$ есть ***размах генеральной совокупности***.

В исследованиях участвуют реально существующие совокупности, к которым относят данные измерений параметров выпускаемых изделий в партии (их технические или физические характеристики), данные социологических опросов, данные измерения аппаратуры и т. д.

Также исследуются теоретически полученные совокупности, примерами которых являются совокупности возможных теоретических результатов значений какой-либо физической величины. Бывают случаи, что многократное исследование объекта провести невозможно, т. к. это связано с большими материальными затратами – исследование балки на излом, испытание на максимальной мощности дорогостоящего оборудования и т. д. Для испытаний, которые приводят к уничтожению объекта обычно применяется *выборочный метод*, который заключается в том, что обследованию подвергаются только некоторая случайно отобранная часть элементов генеральной совокупности.

Сама генеральная совокупность тоже является случайной величиной, а значит для ее изучения желательно определить закон ее распределения, а далее и ее числовые характеристики. Таким образом, сразу ясно, что чем больше элементов в совокупности, тем больших ресурсов (технических, временных) требуется на решение этой задачи. Поэтому для упрощения вычислений, что уже говорилось выше, применяют метод составления подмножества из всей генеральной совокупности и исследуется это подмножество, это и называется *выборочный метод исследования*. Далее из полученных выборочных оценок составляют вывод о всей генеральной совокупности, ее характеристиках. Разработаны критерии проверки достоверности результатов по выборочным данным для всей генеральной совокупности.

Выборкой называется подмножество элементов из генеральной совокупности. ***Объемом выборки*** n называют число элементов, входящих в выборку. Выборка всегда содержит конечное число элементов, это ее объем, он задается изначально.

Выборка, представляющая собой подмножество из генеральной совокупности, тоже является случайной величиной. По ней можно составить достаточно правильное представление о генеральной совокупности, если она удовлетворяет условию ***репрезентативности*** (***представительной***) [7].

Укажем пример нерепрезентативной выборки. Проводится перепись населения, и переписчик проводит ее в рабочие часы будних дней. Очевидно, в перепись попадет только та часть населения, которая не работает, а работающая часть населения там вообще не будет представлена. Репрезентативная выборка должна адекватно отражать пропорции всевозможных составляющих генеральной совокупности.

Проводя исследования, приходится многократно составлять выборки. Даже если они одного и того же объема, все они будут иметь разные характеристики. При этом важно, чтобы полученные результаты обладали достаточной достоверностью, что обеспечивается репрезентативностью случайных выборок.

Основной задачей математической статистики является изучение генеральной совокупности, ее закона распределения, ее параметров, к которым относятся ее числовые характеристики. Инструментом для изучения генеральной совокупности являются случайные выборки по изучаемой характеристике X.

С примерами задач, которые решает математическая статистика, мы встречаемся в повседневной жизни, вот малая часть из них:

- Какова успеваемость студентов по результатам сессии? Оцениваем связь успеваемости, посещаемости и т. д.

- Каков средний рост, возраст, продолжительность жизни населения определённой группы населения? Какова взаимосвязь между этими показателями?
- Какое количество транспорта надо направить на различные маршруты города?
- Какова гарантия службы того или иного прибора?

5.2. Дискретное и непрерывное описание вариационного ряда

Сразу нужно отметить, что есть дискретные и есть интервальные представления выборки.

Пусть имеется выборка объёма n, полученная отбором из генеральной совокупности некоторых ее показателей исследуемого признака, они называются ***вариантами***. Варианты $x_1, x_2, \ldots, x_n$, располагаемые в порядке возрастания, называют ***вариационным рядом.***

Частотой варианты x_k называется число n_k, которое показывает, сколько раз эта варианта повторилась в выборке, а отношение $w_k = \frac{n_k}{n}$ называется ***относительной частотой*** варианты x_k.

Обычно дискретное задание статистического распределения (выборки) задается с помощью таблицы, где указывается значение элемента в выборке и его количественное появление в этой выборке:

Таблица 6

X	x_1	x_2	...	x_i	...	x_k
n	n_1	n_2	...	n_i	...	n_k

где сумма частот n_i равна объему выборки

$$\sum_{i=1}^{k} n_i = n. \tag{79}$$

Дискретное задание статистического распределения (выборки) может быть задано с помощью таблицы, где указывается соотношение появления элемента в выборке с его относительной частотой:

Таблица 7

X	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	...	$\boldsymbol{x_i}$	...	$\boldsymbol{x_k}$
$w_i = \frac{n_i}{n}$	w_1	w_2	...	w_i	...	w_k

где относительные частоты w_i, очевидно, связаны равенством

$$\sum_{i=1}^{k} w_i = 1. \tag{80}$$

Соответствие, указанное в табл. 6, называется *статистическим распределением частот*, а в табл. 7 – *статистическим распределением относительных частот* выборки, последнее является аналогом закона распределения дискретной случайной величины.

Графически распределениям, представленным в табл. 6 или табл. 7, соответствует *полигон частот* или *полигон относительных частот* – это ломаные линии, отрезки которых последовательно соединяют две соседние точки $A_1(x_1;n_1)$, $A_2(x_2;n_2)$, $A_3(x_3;n_3),\ldots$, $A_k(x_k;n_k)$ на плоскости xOn или $B_1(x_1;w_1)$, $B_2(x_2;w_2)$, $B_3(x_3;w_3),\ldots$, $B_k(x_k;w_k)$ на плоскости xOw, соответственно.

Пример 1. Построить полигон частот по данному статистическому распределению частот выборки (см. табл. 8):

Таблица 8

x_k	4	10	16	20	24	30	$\sum$
n_k	15	18	6	4	5	12	60

Решение. В одной таблице укажем вариационный ряд с частотами и с относительными частотами. Поскольку объем выборки равен 60, относительные частоты каждой из вариант $w_i = \dfrac{n_i}{60}$, тогда получим

x_k	4	10	16	20	24	30	$\sum$
n_k	15	18	6	4	5	12	60
w_k	0,25	0,3	0,1	0,067	0,083	0,2	1

Строим на координатной плоскости xOn точки (x_k, n_k) и соединяем их последовательно отрезками – получим ломаную линию (см. рис. 24):

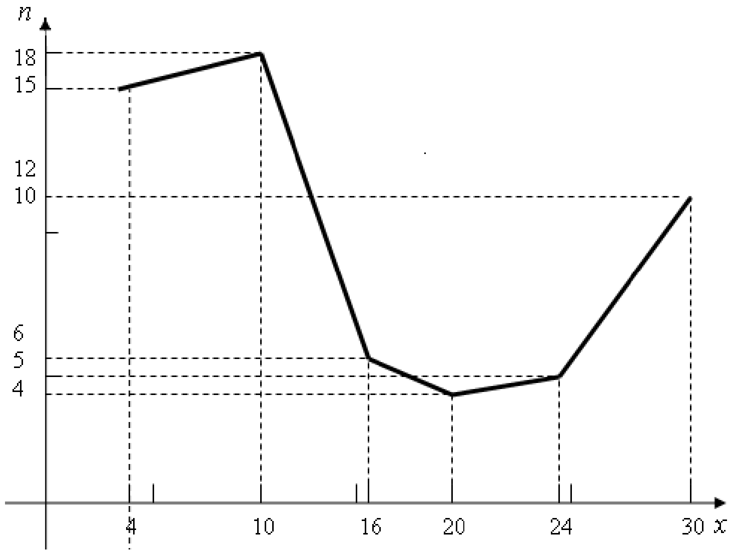


Рис. 24. Полигон частот по данным табл. 8

При большом числе k (например, при измерениях) для вариационного ряда с частотами, а в случае, когда генеральная совокупность распределена по непрерывному закону (причем чаще всего *нормальному*) применяют интервальное распределение выборки: разбивают $x_{max} - x_{min}$ на k равных интервалов, а количество вариант $x_{k_1}, x_{k_2}, ..., x_{k_s} \in (a_k, a_{k+1})$ дает частоту $\tilde{n}_k$ попадания в k-ый интервал. Значения, попадающие на границу интервала, относят к одному из соседних интервалов, а если их несколько, то делят пополам.

Рекомендуется количество интервалов выбирать по формуле Стерджерса, округляя полученное число до целого значения:

$$k \approx 1 + 3{,}2 \cdot \lg n \qquad (k \approx 1 + 1{,}4 \cdot \ln n)\,, \tag{81}$$

где n – объем выборки.

Длина интервала равна

$$h = \frac{x_{max} - x_{min}}{k}, \tag{82}$$

где x_{max} и x_{min}, соответственно, максимальное и минимальное значения случайной величины в данной выборке. Подсчитывая число значений, попавших в интервалы, получают значения частот, по которым заполняют таблицу.

Интервальное задание вариационного ряда Таблица 9

$(x_1;x_2)$	$(x_2;x_3)$	$(x_3;x_4)$	$\ldots$	$(x_i;x_{i+1})$	$\ldots$	$(x_k;x_{k+1})$
$\tilde{n}_1$	$\tilde{n}_2$	$\tilde{n}_3$	$\ldots$	$\tilde{n}_i$	$\ldots$	$\tilde{n}_k$

Частоты $\tilde{n}_i$ показывают, сколько раз встретилось значение вариант в интервале $(x_i; x_{i+1})$ с учетом кратности повторения каждой из них, формула (79) справедлива и в этом случае: $\sum_{i=1}^{k} \tilde{n}_i = n$.

Если изначально вариационный ряд был задан интервально, а поставлена задача определения среднего значения, то легко можно перейти к дискретному заданию, для этого в качестве вариант выбирают середины интервалов $(x_i; x_{i+1})$:

Таблица 10

.

$y_1 = \frac{x_1 + x_2}{2}$	$y_2 = \frac{x_2 + x_3}{2}$	$\ldots$	$y_i = \frac{x_i + x_{i+1}}{2}$	$\ldots$	$y_k = \frac{x_k + x_{k+1}}{2}$
$\tilde{n}_1$	$\tilde{n}_2$	$\ldots$	$\tilde{n}_i$	$\ldots$	$\tilde{n}_k$

5.3. Гистограмма, эмпирическая функция распределения

Графически интервальное распределение задается ***гистограммой*** – это ступенчатая фигура, основанием каждой ступеньки которой служат отрезки длины h: $[x_1; x_2]$, $[x_2; x_3]$,…, $[x_k; x_{k+1}]$, а соответствующие высоты ступенек равны $h_i = w_i / h$. Гистограмма относительных частот является аналогом функции плотности, так как её площадь равна единице.

Пример 2. Построить гистограмму частот распределения по интервально заданному распределению:

$h=3$	$(3;6)$	$(6;9)$	$(9;12)$	$(12;15)$	$(15;18)$	$(18;21)$	$(21;24)$
n_k	6	9	12	21	18	6	3

Определить частоту попадания вариант в интервал $W(4 < X < 16)$.

Решение. Объем выборки равен 75, а ширина ступенек $h = 3$, тогда высоты ступенек $h_i = \frac{w_i}{h} = \frac{\left(\frac{n_i}{75}\right)}{3}$:

$h=3$	$[3;6]$	$(6;9]$	$(9;12]$	$(12;15]$	$(15;18]$	$(18;21]$	$(21;24]$
n_k	6	9	12	21	18	6	3
$\frac{w_k}{h}$	$\frac{2}{75}$	$\frac{3}{75}$	$\frac{4}{75}$	$\frac{7}{75}$	$\frac{6}{75}$	$\frac{2}{75}$	$\frac{1}{75}$

Тогда гистограмма имеет вид:

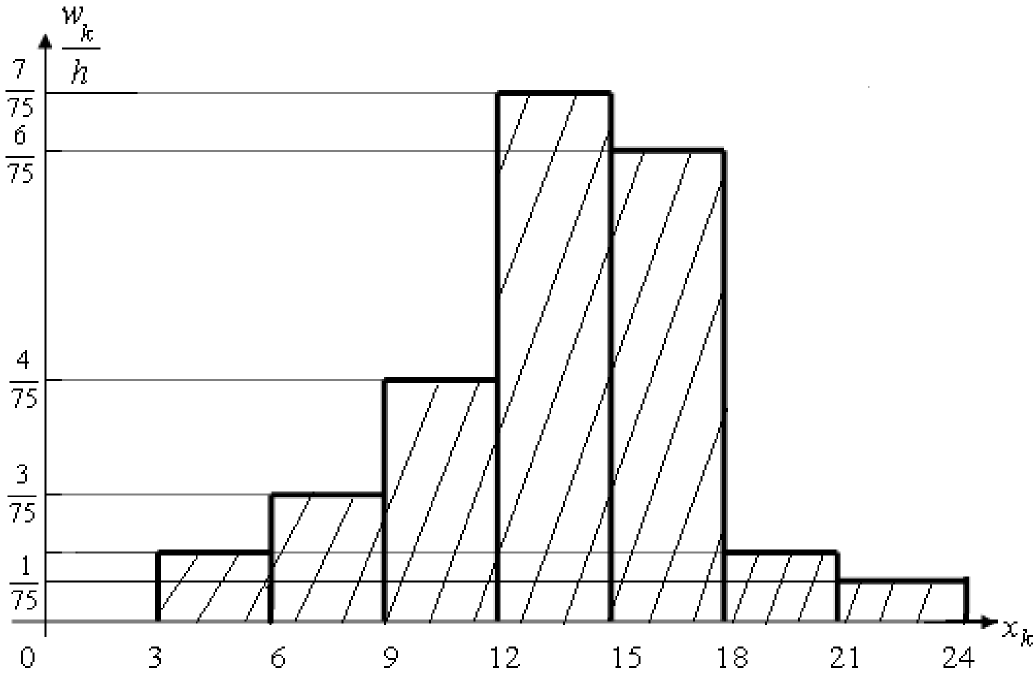


Рис. 25. Гистограмма, построенная по данным примера 2

С помощью гистограммы, площадь которой равна единицы, можно решить вторую часть задания. Площадь гистограммы, расположенной над отрезком $[4;16]$, численно равна частоте попадания вариант в заданный интервал. Укажем ее на рис. 26.

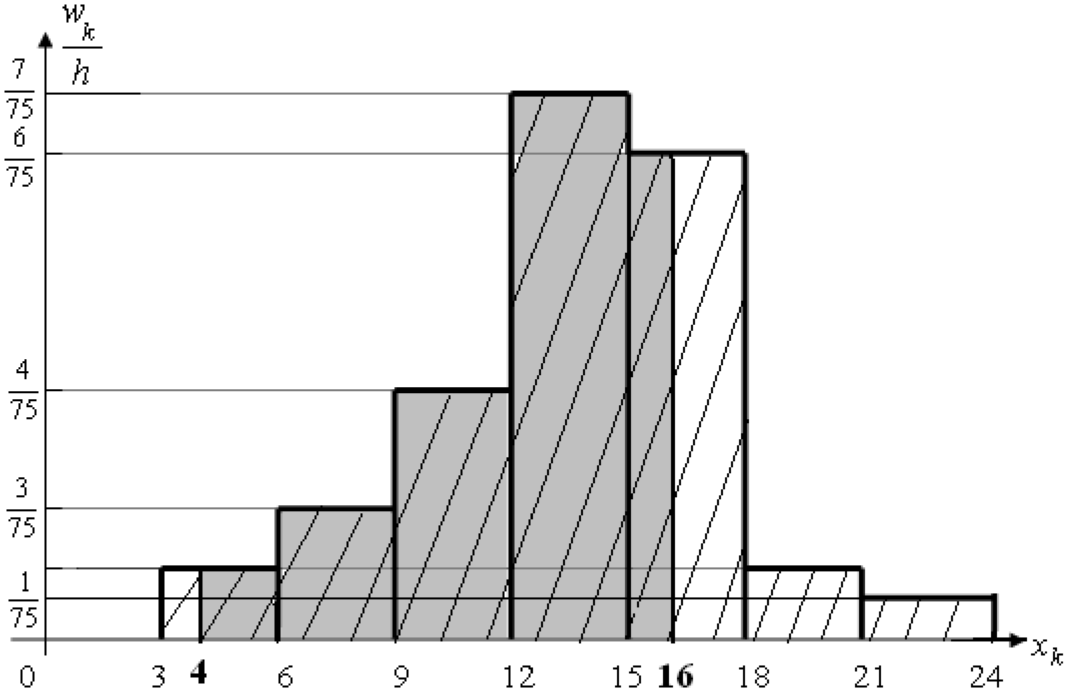


Рис. 26. Гистограмма с полезной (серой) площадью

Вычислим $W(4 < X < 16) = 2 \cdot \frac{2}{75} + 3 \cdot \frac{3}{755} + 3 \cdot \frac{4}{75} + 3 \cdot \frac{7}{75} + 1 \cdot \frac{6}{75} \approx 0{,}69$.

Статистическим аналогом функции распределения $F(x)$ дискретной случайной величины является ***эмпирическая функция*** распределения $F^*(x)$, задающая для каждого $x \in R$ относительную частоту события $X < x$, т. е. по определению,

$$F^*(x) = \frac{n_x}{n}, \qquad (\text{или } F^*(x) = W(x_k < x)), \tag{83}$$

где n_x число вариант, меньших x, а n – объём выборки.

По гистограмме и полигону частот судят о виде плотности распределения исследуемой непрерывной случайной величины или о распределении вероятностей дискретной случайной величины.

Эмпирическая функция распределения дает представление о функции распределения и используется в основном в статистической проверке гипотез. Кроме того, эмпирическая функция распределения используется для определения эмпирических (выборочных) квантилей. Квантилем x_p называют значение аргумента эмпирической функции при известном значении накопительной вероятности $F(x_p) = p$.

Пример 3. Для вариационного ряда относительных частот

x_k	4	10	16	20	24	30	$\sum$
w_k	0.25	0.3	0.1	0.05	0.1	0.2	1

построить эмпирическую функцию распределения.

Определить частоту попадания вариант в интервал $W(8 < X < 20)$.

Решение. Эмпирическая функция вычисляет накопительную частоту с увеличением переменной, тогда получим

$$F^*(x) = \begin{cases} 0 & x \le 4, \\ 0{,}25 & 4 < x \le 10, \\ 0{,}55 & 10 < x \le 16, \\ 0{,}65 & 16 < x \le 20, \\ 0{,}7 & 20 < x \le 24, \\ 0{,}8 & 24 < x \le 30 \\ 1 & 30 < x. \end{cases}$$

График функции укажем на рис. 27:

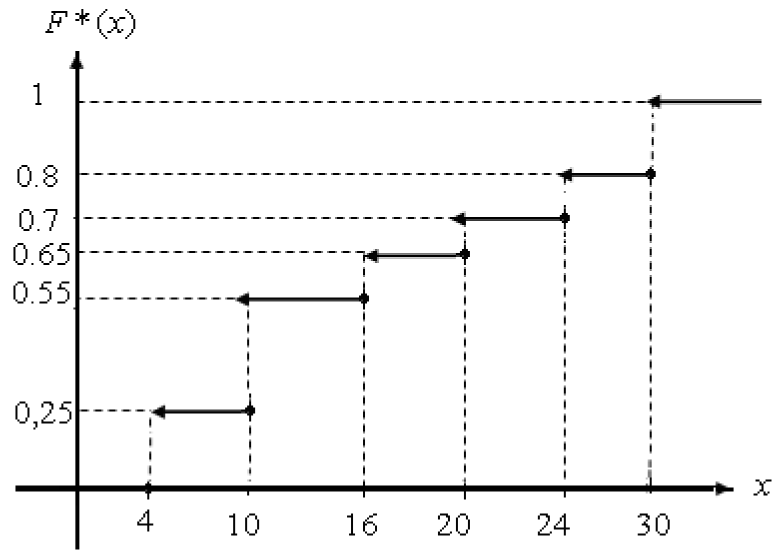


Рис. 27. График эмпирической функции примера 3

Функция $F^*(x)$ обладает такими же свойствами, что и функция распределения $F(x)$ дискретной случайной величины, по ней можно вычислить частоту попадания в интервал $W(8 < X < 20) = F^*(20) - F^*(8) = 0{,}65 - 0{,}25 = 0{,}4$.

5.4. Выборочные числовые характеристики вариационного ряда

Основными характеристиками, как и в теории вероятностей, являются:

- среднее значение генеральной совокупности $M(x)$, ее оценивают по выборочной средней вариационного ряда $\overline{x_в}$;
- дисперсия генеральной совокупности $D(x)$, ее оценивают по выборочной дисперсии вариационного ряда $D_в$;
- среднее квадратическое отклонение генеральной совокупности $\sigma(x)$, ее оценивают по среднему квадратическому отклонению вариационного ряда $\sigma_в$.

Формулы для вычисления выборочных характеристик аналогичны (36), (39), (41), отличие только в том, что в дискретных случайных величинах данные сгруппированы – у каждого значения случайной величины есть вероятность ее появления. А в статистике данные чаще представлены в виде совокупности значений изучаемой характеристики, например, как в таблице, представленной далее в примере 4.

Выборочное среднее

Выборочное среднее $\overline{x_в}$ есть среднее значение изучаемой характеристики выборки, что является аналогом математического ожидания дискретной случайной величины.

В данном случае для вычисления среднего естественно взять среднее арифметическое

$$\overline{x_в} = \frac{1}{n} \cdot \sum_{k=1}^{n} x_k \,. \tag{84}$$

Пример 4. Определить выборочные числовые характеристики по таблице.

-1,8	-0,35	+1,5	-4	+2
+1,5	+2	-1,8	+2	-1,8
-0,35	-4	+2	+0,4	-1

В заданном примере 4, состоящим из 15 значений, получим

$$\overline{x_в} = \frac{1}{n} \cdot \sum_{k=1}^{n} x_k = \frac{1}{15}\left(-1{,}8 + 1{,}5 - 0{,}35 + \ldots + 2 - 1{,}8 - 1\right) \approx -0{,}247 \,.$$

Данные можно сгруппировать. Для этого заметим, что в таблице имеется всего семь разных значений $x = \{-4; -1{,}8; -1; -0{,}35; +0{,}4; +1{,}5; +2\}$, а значит, их можно представить в виде таблицы

x_i	-4	-1,8	-1	-0,35	+0,4	+1,5	+2
n_i	2	3	1	2	1	2	4

Тогда для сгруппированных данных применима формула

$$\overline{x_B} = \frac{1}{n} \cdot \sum_{m=1}^{n} x_m n_m \,, \text{ (или } \overline{x_в} = \sum_{m=1}^{n} x_m w_m \text{)}, \tag{85}$$

где относительные частоты w_i являются для выборки аналогом вероятностей в формуле (36).

Итак, получим $\overline{x_B} = \frac{1}{n} \cdot \sum_{m=1}^{n} x_m n_m = \frac{1}{15}\left(-4 \cdot 2 - 1{,}8 \cdot 3 + \ldots + 2 \cdot 4\right) \approx -0{,}247$.

Полученное значение выборочной средней, как видим, не зависит от способа вычисления, но заданное статистическое распределение с помощью табл. 6 в большинстве случаев более выгодно.

Выборочная дисперсия и среднее квадратичное отклонение

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонений значений признака X выборки от выборочной средней $\overline{x_в}$.

Если значения $x_1, x_2, ..., x_n$ признака X выборки объёма n не сгруппированы, то $D_в = \frac{1}{n}\sum_{m=1}^{n}(x_m - \bar{x}_в)^2$. Если же значения признака сгруппированы и число различных вариант равно k, то применима формула

$$D_в = \frac{1}{n}\sum_{m=1}^{k}(x_m - \bar{x}_в)^2 n_m = \sum_{m=1}^{k}(x_m - \bar{x}_в)^2 \frac{n_m}{n} = \sum_{m=1}^{k}(x_m - \bar{x}_в)^2 w_m .$$

Отметим, что свойства выборочной дисперсии аналогичны свойствам дисперсии дискретной случайной величины. В частности, более удобной для вычисления выборочной дисперсии является формула:

$$D_в = \overline{x_в^{\,2}} - (\bar{x}_в)^2 , \qquad (86)$$

где $\bar{x}_в$ вычисляется по (85), а $\overline{x_в^2} = \frac{1}{n}\sum_{m=1}^{k} x_m^2 n_m$ (или $\overline{x_в^2} = \sum_{m=1}^{k} x_m^2 w_m$).

Для вычисления $D_в$ в примере 4, добавим в таблицу промежуточные вычисления

x_i	-4	-1,8	-1	-0,35	+0,4	+1,5	+2	$\sum$
n_i	2	3	1	2	1	2	4	15
$w_i = \frac{n_i}{n}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{4}{15}$	1
$x_i w_i$	-0,533	-0,36	-0,067	-0,047	0,027	0,2	0,533	-0,247
$x_i^2 w_i$	2,133	0,648	0,067	0,016	0,011	0,3	1,067	4,24

При вычислении числовых характеристик $\bar{x}_в$, $D_в$ округление w_i может привести к погрешностям, поэтому правильнее работать с частотам $\bar{n}_i$, а затем делить на n.

Получим выборочную дисперсию $D_в = \overline{x_в^{\,2}} - (\bar{x}_в)^2 = 4{,}24 - (-0{,}247)^2 \approx 4{,}18$.

Выборочным средним квадратичным отклонением $\sigma_в$ называется средний разброс значений вариант вокруг выборочного среднего, его значение равно корню квадратному из выборочной дисперсии, т.е.

$$\sigma_в = \sqrt{D_в} . \qquad (87)$$

В примере 4 значение выборочного среднего квадратичного отклонения равно $\sigma_в = \sqrt{D_в} = \sqrt{4{,}18} \approx 2{,}04$.

5.5. Связь числовых оценок генеральной совокупности с выборочными характеристиками

Предположим, что генеральная совокупность с признаком X задана функцией плотности $\varphi(x,\theta)$, в которой θ – это оцениваемый параметр распределения. Оцениваемым параметром может являться λ в распределении Пуассон или математическое ожидание a в нормальном распределении и т. д.

Далее предположим, что исследовать все множество значений X невозможно, т. е. или их очень много или это очень дорого, тогда выбираем подмножество $x_1, x_2, ..., x_n$, то что называется выборкой.

Обозначим через θ_n (полученную из выборки $x_1, x_2, ..., x_n$) оценку исследуемого параметра θ. Отметим, что оценка θ_n сама будет являться случайной величиной, т. к. ее значение для различных выборок того же объема n будет отличаться. Сама выборка – тоже случайная величина.

Оценка θ_n – будет давать хорошее приближение к исследуемой величине, если она удовлетворяет требованиям:

1. Оценка θ_n параметра θ является ***несмещённой***, т. е. если $M(\theta_n) = \theta$ – это значит, что ее математическое ожидание равно оцениваемому параметру. Условие гарантирует отсутствие систематических ошибок при оценивании.

2. Оценка является ***состоятельной***, т. е. если для любого наперёд заданного, сколь угодно малого, положительного числа $\varepsilon > 0$ вероятность события

$$P\{|\theta_n - \theta| < \varepsilon\} \underset{n\to\infty}{\longrightarrow} 1 .$$

Данное условие обозначает, что при больших n с очень большой вероятностью можно утверждать, что выборочная характеристика θ_n отличается от значения характеристики θ меньше, чем на величину ε.

Заметим, что несмещённая оценка θ_n будет состоятельной, если её дисперсия стремится к нулю: $D(\theta_n) \underset{n\to\infty}{\longrightarrow} 0$.

3. Оценка θ_n называется ***эффективной***, если при заданном n она имеет наименьшую дисперсию, т.е. $D(\theta_n) = D_{\min}$.

Рассмотрим основные выборочные характеристики, проверив их соответствие требованиям несмещенности, состоятельности и эффективности.

Найдём $M(\overline{x_в})$, где $\overline{x_в} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Пользуясь тем, что для всех вариант выборки $k = 1,2,...,n$ верно $M(x_k) = a$ и учитывая свойства математического ожидания, получим

$$M(\overline{x_в}) = M\left(\frac{x_1 + x_2 + ... + x_n}{n}\right) = \frac{M(x_1 + x_2 + ... + x_n)}{n} =$$

$$= \frac{M(x_1) + M(x_2) + ... + M(n_n)}{n} = \frac{a + a + ... + a}{n} = \frac{1}{n} \cdot na = a.$$

Таким образом, $M(\overline{x_в}) = a$, а значит, математическое ожидание средней выборочной совпадает с генеральной средней, т. е. $\overline{x_в}$ *– выборочная средняя есть несмещенная оценка математического ожидания* $M(X)$.

Найдём выборочную дисперсию $D_в(X)$, пользуясь тем, что для всех $k = 1,2,\ldots,n$ верно $D(x_k) = D(x)$ и учитывая свойства дисперсии, получим

$$D_в(X) = D\left(\frac{x_1 + \ldots + x_n}{n}\right) = \frac{D(x_1 + \ldots + x_n)}{n^2} = \frac{D(x_1) + \ldots + D(x_n)}{n^2} = \frac{1}{n^2} nD(X) = \frac{D(X)}{n},$$

отсюда следует, что дисперсия является *состоятельной*, так как $\lim\limits_{n\to\infty} D_в(X) = \lim\limits_{n\to\infty} \frac{D(X)}{n} = D(X) \lim\limits_{n\to\infty} \frac{1}{n} = 0.$

К сожалению, $D_в(X)$ оказалась смещенной влево оценкой: $M\big(D_в(X)\big) = \frac{n-1}{n} D(X)$, поэтому вводится ещё одна оценка – *исправленная дисперсия*

$$s^2 = \frac{n}{n-1} D_в,$$

она тоже является оценкой генеральной дисперсии, но является уже *несмещённой* оценкой генеральной дисперсии, так как

$$M(s^2) = M\left(\frac{n}{n-1} D_в\right) = \frac{n}{n-1} M(D_в) = \frac{n}{n-1} \cdot \frac{n-1}{n} D(X) = D(X).$$

Для оценки генерального среднего квадратического отклонения используют *исправленное среднее квадратическое отклонение*, которое равно квадратному корню из исправленной дисперсии:

$$s = \sqrt{\frac{n}{n-1} D_в} = \sqrt{\frac{n}{n-1}} \sigma_в. \qquad (88)$$

6. Оценка параметров генеральной совокупности

6.1. Основные определения

Рассмотренные выше оценки, которые вычисляются по данным некоторой выборки для оценки неизвестных числовых параметров генеральной совокупности, сами являются *точечными оценками*. Любая точечная оценка элементов выборки сама является случайной величиной (даже при одном и том же объеме выборки n числовые характеристики от разных выборок будут различны), и мы не имеем никаких данных о степени ее близости к истинному значению оцениваемого числового параметра.

Предположим, что мы получили точечную оценку $\bar{\theta}_n$ параметра θ. Естественно возникает вопрос о точности этой оценки, т. е. можно ли утверждать, что полученная оценка достаточно близка к оцениваемой характеристике генеральной совокупности. Далее, переводя этот вопрос на язык математических формул, будем проверять утверждение $\left|\theta-\overline{\theta_n}\right|<\delta$, где $\delta>0$ некоторая малая положительная величина, которая называется *точностью*. Например, если говорят вычислить с точностью до одной сотой, то это значит $\delta=0{,}01$.

Качественное отличие вероятностного мира состоит в том, что утверждать наступление проверяемого неравенства невозможно, а можно говорить только о *доверительной вероятности* γ наступления этого события. Итак, изучаемое событие состоит в том, что полученная оценка $\overline{\theta}_n$ удовлетворяет точности δ с вероятностью γ: $p\left(\left|\theta-\overline{\theta_n}\right|<\delta\right)=\gamma$, или, что тоже самое,

$$p\left(\overline{\theta_n}-\delta<\theta<\overline{\theta_n}+\delta\right)=\gamma, \tag{89}$$

то сам интервал $\left(\overline{\theta_n}-\delta;\overline{\theta_n}+\delta\right)$ называют *доверительным интервалом*, отвечающим *доверительной вероятности* $p=\gamma$, доверительную вероятность называют еще *надежностью*.

В литературе доверительная вероятность обозначается $\gamma=1-\alpha$, где коэффициент α называется *уровнем значимости*.

Границы доверительного интервала для выборки являются случайными – это значит, что понимать равенство (89) нужно так: *интервал* $\left(\overline{\theta_n}-\delta;\overline{\theta_n}+\delta\right)$ *с вероятностью* γ *содержит неизвестное значение параметра* θ.

Не трудно заметить, что оценка θ будет точнее, если значение величины δ будет меньше – интервал будет уже.

Знание закона распределения генеральной совокупности тоже упрощает задачу. Ранее указывалась, что основным и наиболее часто применяемым законом распределения является *нормальный*, что абсолютно логично, т. к. именно это распределение уже указывалось как основное в анализе массовых закономерностей, поэтому построение доверительных интервалов далее будет ориентироваться именно на этот закон распределения. Конечно же, если сразу известно или в процессе обработки информации по конкретной выборке видно, что данные не подчиняются предполагаемому закону (такое не часто, но бывает), то указанные далее оценки не работают.

6.2. Доверительный интервал для математического ожидания при известном среднем квадратичном отклонении σ

Пусть для генеральной совокупности изучаемый признак X подчиняется нормальному закону распределения, где a и σ – его параметры, причем среднеквадратическое отклонение σ **известно**. Оценим границы неизвестного математического ожидания a генеральной совокупности с помощью доверительного интервала с заданной надёжностью γ.

Установленным фактом является то, что для генеральной совокупности X, распределённой по нормальному закону, выборочная средняя $\overline{x_в}$, вычисленная по независимым выборкам, тоже имеет нормальное распределение, и параметрами этого распределения являются $M(\overline{x_в}) = a$ и $\sigma(\overline{x_в}) = \dfrac{\sigma}{\sqrt{n}}$, что указано на стр. 79.

Равенством $P(|\overline{X} - a| < \delta) = \gamma$, где $\gamma -$ заданная точность, устанавливают доверительный интервал математического ожидания генеральной совокупности.

В нашем случае для случайной величины X распределенной нормально, справедлива формула (58) $p(|X - a| < \delta) = 2\Phi\left(\dfrac{\delta}{\sigma}\right)$, в которой, заменив X на $\overline{x_в}$ и σ на $\sigma(\overline{x_в}) = \dfrac{\sigma}{\sqrt{n}}$, получим $p(|\overline{x_в} - a| < \delta) = 2\Phi\left(\dfrac{\delta\sqrt{\text{n}}}{\sigma}\right)$. Обозначив $t = \dfrac{\delta\sqrt{\text{n}}}{\sigma}$, имеем $\delta = \dfrac{t\sigma}{\sqrt{n}}$, из чего получим

$$p(|\overline{x_в} - a| < \frac{t\sigma}{\sqrt{n}}) = 2\Phi(t).$$

Требование попасть в доверительный интервал с доверительной вероятность $p = \gamma$ обозначает $2\Phi(t) = \gamma$, отсюда получим оценку для среднего значения генеральной совокупности:

$$p(\overline{x_в} - \frac{t \cdot \sigma}{\sqrt{n}} < a < \overline{x_в} + \frac{t \cdot \sigma}{\sqrt{n}}) = \gamma.$$

Полученное обозначает, что доверительный интервал

$$(\overline{x_в} - \frac{t \cdot \sigma}{\sqrt{n}}; \overline{x_в} + \frac{t \cdot \sigma}{\sqrt{n}}) \tag{90}$$

с надёжностью γ покрывает неизвестный параметр a, где $\overline{x_в}$ находится по данным выборки, $n -$ объем выборки, $\sigma -$ известная величина, а число t находим по таблице функции Лапласа из равенства $\Phi(t) = \gamma/2$.

Замечание. Из оценки $p(|\overline{x_в} - a| < \dfrac{t\sigma}{\sqrt{n}}) = 2\Phi(t)$, где $\Phi(t)$ – интегральная функция Лапласа, которая возрастает с ростом аргумента, очевидным является факт – чем больше t, тем больше $\Phi(t) = \dfrac{\gamma}{2}$ надежность, но тем больше интервал $|\overline{x_в} - a| < \dfrac{t\sigma}{\sqrt{n}}$, что уменьшает точность вычислений. Обратная связь тоже очевидна: с ростом точности $t \to 0$ уменьшается надежность $\Phi(t) \underset{t\to 0}{\to} 0$. Итак, надежность и точность вычислений находятся в обратной зависимости.

Пример 1. Произведено 9 независимых наблюдений случайной величины X, распределенной по нормальному закону $N(a,\sigma)$, где результаты наблюдений: $x_1=-25,\ x_2=34,\ x_3=-20,\ x_4=10,\ x_5=21,\ x_6=6,\ x_7=10,\ x_8=-2,\ x_9=2.$

Найти оценку для математического ожидания и построить для него доверительный интервал с надежностью $\gamma=0{,}95$ при условии, что у генеральной совокупности известно $\sigma=20$.

Решение. Находим $\overline{x_в}=\dfrac{-25+34-20+10+21+6+10-2+2}{9}=4$. Учитывая, что $\gamma=0{,}95$, получаем $\Phi(t)=\dfrac{\gamma}{2}=0{,}475$. Найти значение аргумента можно по таблице интегральной функции Лапласа $t=1{,}96$ (см. [1], прил. 2), тогда $\delta=\dfrac{t\sigma}{\sqrt{n}}=\dfrac{1{,}96\cdot 20}{\sqrt{9}}\approx 13{,}1.$

Получаем доверительный интервал: (4–13,1; 4+13,1), т.е. (–9,1; 17,1). Читается полученный вывод так: генеральная совокупность с вероятностью 0,95 имеет математическое ожидание в пределах (–9,1; 17,1).

Заметим, что такая оценка не является качественной, что связано с малым объем выборки $n=9$. Чтобы повысить точность в 10 раз, что возможно за счет $\sqrt{n}$ в формуле (90), надо увеличить объем выборки в 100 раз.

Пример 2. У генеральной совокупности, распределенной по нормальному закону, $x_{\min}=20$, $x_{\max}=32$. Требуется определить минимальный объем выборки, чтобы оценить математическое ожидание с точностью $\delta=0{,}1$ и надежностью $\gamma=0{,}95$.

Решение. Разрешив формулу $\delta=\dfrac{t\sigma}{\sqrt{n}}$ (см. стр. 81) относительно объема выборки получим $n=\dfrac{t^2\sigma^2}{\delta^2}$. Так как $n\in \mathrm{N}$, то естественно выбрать наименьшее натуральное число, удовлетворяющее неравенству $n\ge\dfrac{t^2\sigma^2}{\delta^2}$. Для $\gamma=0{,}95$ в предыдущем примере было найдено $t=1{,}96$. По правилу 3-х сигм найдем $\sigma=\dfrac{32-20}{6}=2$. Подставив все данные в неравенство, найдем необходимый объем выборки $n\ge\dfrac{t^2\sigma^2}{\delta^2}=\dfrac{1{,}96^2 2^2}{0{,}1^2}=1536{,}64$.

Ответ: для получения адекватной оценки требуется выборка объемом не менее $n_{\min}=1537$.

6.3. Доверительный интервал для математического ожидания при неизвестном среднем квадратическом отклонении σ

Пусть для генеральной совокупности изучаемый признак X подчиняется нормальному закону распределения, где a и σ – его параметры, причем среднеквадратическое отклонение σ **неизвестно**. Оценим границы неизвестного математического ожидания a генеральной совокупности с помощью доверительного интервала. Найдем для выборки $\overline{x_в}$, $\sigma_в$, и по формуле $s=\sqrt{\frac{n}{n-1}}\sigma_в$ исправленное среднее квадратическое отклонение.

Введем новую случайную величину $T=\frac{\overline{x_в}-a}{s/\sqrt{n}}$, которая имеет распределение Стьюдента (см. [1], стр. 146) с $k=n-1$ степенями свободы, их число есть разность между объемом выборки и числом оцениваемых параметров по этой выборке (у нас один - a). Распределение Стьюдента задается параметром n и не зависит от параметров a и σ.

Определяется интервал следующим образом:

$$\left(\overline{x_в}-t_\gamma\cdot\frac{s}{\sqrt{n}};\,\overline{x_в}+t_\gamma\cdot\frac{s}{\sqrt{n}}\right). \tag{91}$$

Параметр $t_\gamma=t(\gamma,n)$ находят по таблице распределения Стьюдента для заданных n и γ (см., например, [1], приложение 3). Предлагается сравнить формулу (91) с формулой (90) при известном σ.

6.4. Доверительный интервал для среднего квадратического отклонения σ нормального распределения

Пусть для генеральной совокупности изучаемый признак X распределён по нормальному закону. Определим границы неизвестного генерального среднего квадратичного отклонение σ по известному исправленному выборочному среднему квадратическому отклонению s, т. е. потребуем выполнения равенства $p(|\sigma-s|<\delta)=\gamma$, которое запишем в виде

$$p(s-\delta<\sigma<s+\sigma)=\gamma,$$

равенство позволяет найти доверительный интервал, покрывающий параметр σ с заданной надёжностью γ.

Из условия $s-\delta<\sigma<s+\delta$, очевидно, получим равносильное неравенство $s(1-\frac{\delta}{s})<\sigma<s(1+\frac{\delta}{s})$, в котором, обозначив $\frac{\delta}{s}=q$, получим

$$s(1-q)<\sigma<s(1+q), \tag{92}$$

где q находится по специальной таблице $q=q(\gamma,n)$, представленной в приложениях.

Пример 3. Количественный признак X генеральной совокупности распределён нормально. По выборке объёма $n = 25$ найдено исправленное среднее квадратичное отклонение $s = 0{,}8$. Найти доверительный интервал, покрывающий генеральное среднее квадратичное отклонение σ с надёжностью 0,95.

Решение. По таблице (см. [1], приложение 4) по данным $\gamma = 0{,}95$ и $n = 25$ найдём $q = q(\gamma; n) = q(0{,}95; 25) = 0{,}32$.

Искомый доверительный интервал для σ: 0,8(1– 0,32)$< \sigma <$ 0,8 (1+0,32), т. е. 0,544$< \sigma <$ 1,056.

Пример 4. Задано интервальное распределение выборки:

$(y_i; y_{i+1})$	12-15	15-18	18-21	21-24	24-27	27-30
n_i	2	6	12	19	7	4

Требуется:

1) построить гистограмму относительных частот;

2) перейти к вариантам, записать эмпирическую функцию распределения и построить ее график;

3) найти точечные оценки $\bar{x}_в$ и $\sigma_в$;

4) считая генеральную совокупность нормально распределенной, найти доверительные интервалы для a и σ с надежностью $\gamma = 0{,}95$.

Решение.

1) Объем выборки $n = \sum_{i=1}^{6} n_i = 2 + 6 + 12 + 19 + 7 + 4 = 50$. Найдем относительные частоты и выпишем таблицу с относительными частотами:

$(y_i; y_{i+1})$	12-15	15-18	18-21	21-24	24-27	27-30
w_i	0,04	0,12	0,24	0,38	0,14	0,08

Гистограмма относительных частот представляет собой ступенчатую функцию, где на оси абсцисс откладываются интервалы задания выборки, их длина $h = 3$, а высоты ступенек на каждом из интервалов равны $h_i = w_i / h$. Для лучшей наглядности на рис. 28 выберем разный масштаб по осям координат.

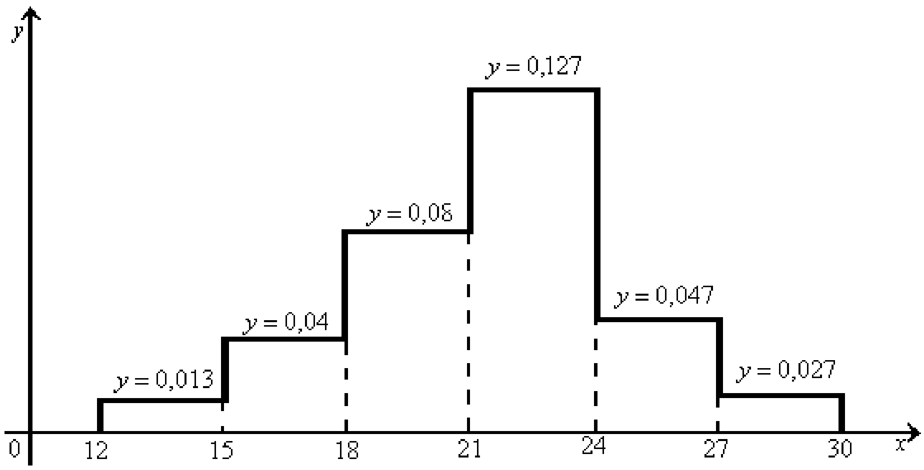


Рис. 28. Гистограмма

2) Перейдем к точечному заданию выборки, где значения вариант вычисляются как середины интервалов.

x_i	13,5	16,5	19,5	22,5	25,5	28,5
n_i	2	6	12	19	7	4
w_i	0,04	0,12	0,24	0,38	0,14	0,08

Составим эмпирическую функцию $F^*(x)$ – функцию накопительной частоты:

$F^*(x) = 0$, если $X \le 13{,}5$;

$F^*(x) = 0{,}04$, если $13{,}5 < X \le 16{,}5$;

$F^*(x) = 0{,}04 + 0{,}12 = 0{,}16$, если $16{,}5 < X \le 19{,}5$;

$F^*(x) = 0{,}16 + 0{,}24 = 0{,}4$, если $19{,}5 < X \le 22{,}5$;

$F^*(x) = 0{,}4 + 0{,}38 = 0{,}78$, если $22{,}5 < X \le 25{,}5$;

$F^*(x) = 0{,}78 + 0{,}14 = 0{,}92$, если $25{,}5 < X \le 28{,}5$;

$F^*(x) = 0{,}92 + 0{,}08 = 1$, если $28{,}5 < X$.

График функции $F^*(x)$ представлен на рис. 29.

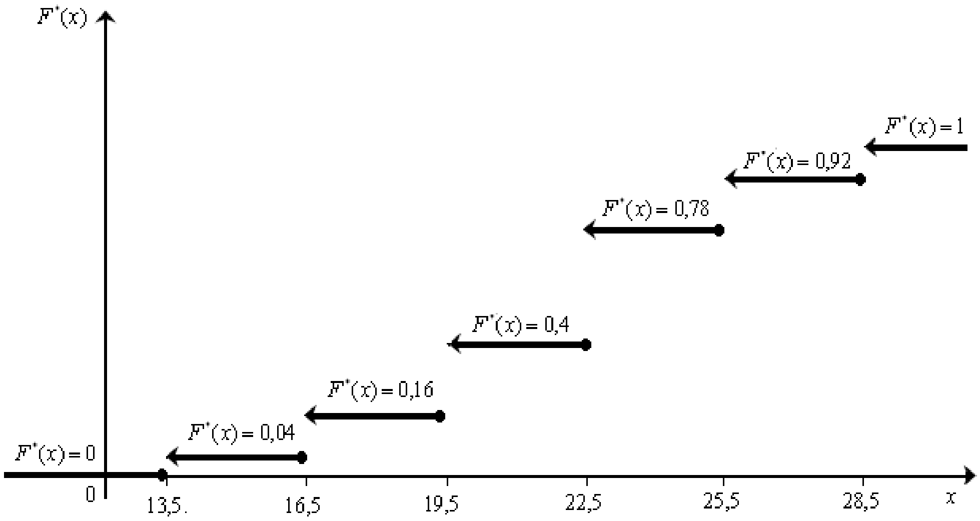


Рис. 29

3) Найдем точечные оценки $x_в$ и $\sigma_в$ по формулам (85)-(87):

$$x_в = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{3,5\cdot 2 + 16,5\cdot 6 + 19,5\cdot 12 + 22,5\cdot 19 + 25,5\cdot 7 + 28,5\cdot 4}{50} = 21,6;$$

$$\sigma_в(x) = \sqrt{\left(\overline{x_в^2}\right) - \left(x_в\right)^2} =$$

$$= \sqrt{\frac{(3,5)^2\cdot 2 + (16,5)^2\cdot 6 + (19,5)^2\cdot 12 + (22,5)^2\cdot 19 + (25,5)^2\cdot 7 + (28,5)^2\cdot 4}{50} - (21,6)^2}$$

$$= 3,6.$$

Обратим внимание на то, что вычисления достаточно трудоемки, несмотря на небольшой объем выборки. Укажем другой способ вычисления числовых характеристик выборки – **метод условных вариант**.

Сделаем замену переменных $u_i = \dfrac{x_i - x_{ср}}{h}$. В нашем случае $h = 3$, а из двух значений в середине таблицы выгодно выбрать среднее значение варианты $x_{ср} = 22,5$ с большей частотой, тогда $u_i = \dfrac{x_i - 22,5}{3}$ и таблица частот имеет вид

u_i	–3	–2	–1	0	1	2
n_i	2	6	12	19	7	4

Найти точечные оценки такого вариационного ряда значительно легче:

$$u_в = \frac{\sum_{i=1}^{k} u_i n_i}{n} = \frac{-3\cdot 2 - 2\cdot 6 - 1\cdot 12 + 0\cdot 19 + 1\cdot 7 + 2\cdot 4}{50} = -0,3;$$

$$\sigma_в(u)=\sqrt{\left(\overline{u_в^2}\right)-\left(\overline{u}_в\right)^2}=$$

$$=\sqrt{\frac{(-3)^2\cdot 2+(-2)^2\cdot 6+(-1)^2\cdot 12+0^2\cdot 19+1^2\cdot 7+2^2\cdot 4}{50}-(-0{,}3)^2}\approx 1{,}2.$$

Вернемся к первоначальной варианте, из замены $u_в=\frac{x_в-x_{ср}}{h}$ следует $x_в=x_{ср}+\overline{u}_в\cdot h=22{,}5+(-0{,}3)\cdot 3=21{,}6$.

Среднее квадратичное отклонение исходной варианты определяется по формуле $\sigma_в(x)=h\cdot\sigma_в(u)\approx 3\cdot 1{,}2=3{,}6$.

4) В предыдущем пункте были найдены точечные оценки $x_в=21{,}6$, $\sigma_в(x)=3{,}6$. Поскольку σ неизвестно, то по формулам (91), (92) (см. стр. 84) определим доверительный интервал для математического ожидания и среднего квадратичного отклонения.

Подставив значение объема выборки $n=50$ в формулу (91), получим исправленное среднее квадратичное отклонение $s=\sqrt{\frac{50}{49}}\cdot 3{,}6\approx 3{,}636$. По таблице значений функции $t_\gamma(\gamma,n)$ [1. прил. 3] найдем $t_\gamma(0{,}95;50)=2{,}009$.

Подставив найденные значения $x_в=21{,}6$, $n=50$ и $s\approx 3{,}636$ в формулу (91), получим доверительный интервал для математического ожидания a:

$$21{,}6-\frac{2{,}009\cdot 3{,}636}{\sqrt{50}}<a<21{,}6+\frac{2{,}009\cdot 3{,}636}{\sqrt{50}}.$$

После вычислений получим $20{,}57<a<22{,}63$.

По формуле (92) определим доверительный интервал для среднего квадратичного отклонения σ. Из [1, прил. 4] найдем $q(\gamma;n)=q(0{,}95;50)=0{,}21$.

Подставив найденные значения $q(\gamma;n)=0{,}21$, $s\approx 3{,}636$ в формулу (92), получим доверительный интервал для среднего квадратичного отклонения σ:

$$3{,}636\cdot(1-0{,}21)<\sigma<3{,}636\cdot(1+0{,}21).$$

Окончательно получим $2{,}87<\sigma<4{,}4$.

Ответ: $x_в=21{,}6$; $\sigma_в(x)=3{,}6$; $20{,}57<a<22{,}63$; $2{,}87<\sigma<4{,}4$.

7. Статистические гипотезы и их проверка

7.1. Понятие статистических гипотез

Гипотезой в широком понимании считается предположение какого-либо события, которое может случиться. В повседневной жизни появление события оценивается принципом практической частоты его появления, т. е. если вероятность появления события A в испытании считается очень малой, то при одном испытании мы уверенны в том, что событие A не произойдёт, и считаем его невозможным. Этот подход к пониманию вероятности не доказывается математически, он проверен практическим опытом человека, мы постоянно бессознательно им руководствуемся. Например, отправляясь в отпуск, мы не рассчитываем на возможность аварии в дороге, хотя вероятность такого события (очень маленькая) всё же имеется, оцениваем возможность выигрыша в лотерею как очень маленькую при покупке единственного билета.

Если испытаний, в каждом из которых вероятность события A очень мала, проводить много раз, то вероятность того, что событие A все-таки произойдёт хотя бы один раз во всем наборе испытаний будет значительно выше. Многократное повторение испытаний уже не позволяет считать событие A невозможным.

Граница между маловероятным событием и случайным определяется индивидуально, исходя из оценки последствий, которые влечет за собой выпадение этого события. Если событие ведет к разрушению какого-либо здания, моста и т. д., то игнорировать такое событие нельзя, а вероятность его появления должна быть очень маленькой.

Еще одной важной задачей математической статистики является задача определения закона распределения случайной величины, наилучшим образом описывающего имеющиеся данные, а также определения оценок неизвестных параметров закона распределения и его числовых характеристик – математического ожидания $M(x)$, дисперсии $D(x)$ и др.

Полученные данные генеральной совокупности обрабатывают:

- если объем данных большой, то из них составляют выборку в соответствии с условиями представительности;
- группируют;
- помещают в специальные таблицы (дискретные или интервальные);
- вычисляют выборочные числовые характеристики.

Часто случайные величины связывают с измерениями. Например, геодезическими данными являются многократные измерения расстояний, углов, размеров объектов – из этих данных и составляется генеральная совокупность. Однако эти измерения содержат ошибки, полученные из-за влияния внешних условий, неточности изготовления и юстировки приборов, неточности выполнения операций наблюдателем и т. д. Далее возникает вопрос, насколько достоверными можно считать эти данные. По имеющимся измерениям определяют выборочные характеристики $\overline{x_в}$ и $\overline{\sigma_в}$ и составляют по ним теоретические данные, которые сравнивают с полученными эмпирическими измерениями. Чтобы сделать вывод более надежным, используют метод исследования, называемый *статистической проверкой гипотез*, где оценивают существенно или случайно расхождения между эмпирическими и теоретическими данными.

Гипотезой называется научное предположение, требующее проверки на опыте.

В математической статистике ***статистической гипотезой*** называется любое предположение о виде или параметрах закона распределения генеральной совокупности.

Гипотезы разделяют на *нулевую выдвинутую* (*основную*) H_0 и *альтернативную* H_1 (конкурирующую). Нулевая – это проверяемая гипотеза, а *альтернативная* – это логическое отрицание нулевой гипотезы. Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки статистических гипотез. Например, H_0 – случайная величина имеет нормальное распределение, то H_1 – распределение случайной величины не является нормальным законом; H_0 – случайная величина имеет среднее значение $a<10$, то H_1 – среднее значение $a \geq 10$.

Различают *простые* гипотезы, содержащие только одно предположение, и *сложные*, которые состоят из конечного или бесконечного числа простых. Например, $\lambda = 7$ в показательном распределении – простая гипотеза, а математическое ожидание $a = 7$ в нормальном распределении при неизвестном σ – сложная гипотеза.

Выделяют два вида ошибок:

- ошибка первого рода заключается в том, что будет отвергнута правильная гипотеза;
- ошибка второго рода заключается в принятии неправильной гипотезы.

Что из этих двух видов ошибок хуже, однозначно сказать нельзя, все зависит от конкретной ситуации.

Если нулевая гипотеза неверна и генеральная совокупность не обладает проверяемым по H_0 свойством, то расхождения между гипотетическим и полученным из опыта данными будут значимыми. Для решения вопроса о значимости расхождения, т. е. для ответа на вопрос о принятии или непринятии выдвинутой гипотезы необходимо построить *меру расхождения,* которую называют *критерием* и обозначить $t_{кр}$ – критическое значение критерия.

Одной из важнейших задач математической статистики является выбор критерия, который зависит от вида распределения, от вида поставленной в проверке задачи: сравнить, оценить, проверить. Значение критерия зависит от требуемой вероятности γ, а вероятность, которой решено пренебрегать в данной области исследования (при ошибке первого рода), называется уровнем значимости α ($\alpha = 1-\gamma$). В технических исследованиях принимают α = 0,10; 0,05; 0,01, таблицы значения критерия настроены именно на эти показатели уровня значимости.

7.2. Проверки статистических гипотез. Критерий Пирсона χ^2

Рассмотрим задачу о неизвестном законе распределения, т. е. проверим, действительно ли вариационный ряд (и сама генеральная совокупность) подчиняется этому закону распределения. Если да, то полученные числовые характеристики являются действительными характеристиками исследуемого объекта.

Очевидно, что случайные ошибки измерений являются результатом суммирования большого числа независимых элементарных ошибок. На основании центральной предельной теоремы (теорема Ляпунова) можно считать, что случайные ошибки измерений подчиняются нормальному закону распределения. После вычисления $x_в$ и $\sigma_в$ естественно проверить правило “3-х сигм”, если при этом вид гистограммы напоминает кривую Гаусса, то логичным будет выдвинуть нулевую гипотезу о том, что имеющиеся данные подчиняются нормальному закону распределения

Одним из методов сравнения эмпирического и теоретического распределений является критерий согласия Пирсона χ^2. В этом критерии вычисляют значение случайной величины “хи квадрат” по данной выборке:

$$\chi^2_{набл} = \sum_{i=1}^{n} \frac{\left(n_i - n_i'\right)^2}{n_i'}, \tag{93}$$

где n_i – эмпирические частоты, а n_i' – теоретические частоты выборочных интервалов, вычисленные для нормального закона распределения с параметрами $a = x_в$ и $\sigma = s$.

Аргументами критерия являются заданный уровень значимости α (например, наиболее часто применимо α = 0,05) и число степеней свободы $k = n - r - 1$, где r – число параметров теоретического распределения. Для нормального закона распределения $k = n - r - 1 = n - 2 - 1 = n - 3$, где n – число частичных интервалов выборки, а $r = 2$, т. к. у нормального распределения есть два свободных параметра a и σ.

Если значение накопленных расхождений теоретических и эмпирических данных не превосходит некоторого критического значения $\chi^2_{кр} = \chi^2_{кр}(\alpha, k)$ (которое определяется по таблице значений критерия “хи квадрат”, в [1] это приложение 5)

$$\chi^2_{набл} < \chi^2_{кр}, \tag{94}$$

то имеем нормальный закон распределения. Из чего следуют выводы о незначительности ошибок в измерениях и о верности полученного результата.

Например, очень часто с целью уменьшения влияния систематических ошибок устанавливают закон их появления, после чего ошибки уменьшают введением поправок в данные измерений.

Примером случайной величины являются случайные ошибки – они являются частью показаний единичных измерений. Закономерности их появления обнаруживаются только в массовом проявлении проводимых испытаний.

Приведем пример исследования ошибок на нормальное распределение.

Пример 1. В качестве примера выполним статистический анализ невязок треугольников триангуляции. При измерении углов треугольника, сумма углов которого равна 180^0, получаем ряд невязок $\{\Delta_i\}$, где $\Delta_i = 180^0 - \sum_{k=1}^{3} \beta_k$.

Полученные данные невязки в рядах триангуляции представлены в таблице

№	Δ_i	№	Δ_i	№	Δ_i	№	Δ_i
1	–0,76	9	+1,29	17	+0,71	25	+0,22
2	+1,52	10	+0,38	18	+1,04	26	+0,06
3	–0,24	11	–1,03	19	–0,38	27	+0,43
4	+1,31	12	+0,00	20	+1,16	28	–1,28
5	–1,27	13	–1,23	21	–0,19	29	–0,41
6	–1,88	14	–1,38	22	+2,28	30	–2,50
7	+0,01	15	–0,25	23	+0,07	31	+1,92
8	–0,69	16	–0,73	24	–0,95	32	–0,62

Проверить соответствие ряда невязок $\{\Delta_i\}$ нормальному закону распределения.

Решение.

Перейдем к интервальному описанию, для этого заметим $x_{\min} = \min_i \Delta_i = -2.5$, $x_{\max} = \max_i \Delta_i = 2.28$, тогда $\Delta_i \in [-2.5; 2.28] \subset [-2.5; 2.5]$.

Будем считать длину интервала равную пяти, тогда по формуле Стерджеса (81) число интервалов $k \approx 1 + 1{,}4 \cdot \ln 32 \approx 5$, а длина интервала равна $h = 1$: $(-2.5; -1.5)$, $(-1.5; -0.5)$, $(-0.5; 0.5)$, $(0.5; 1.5)$, $(1.5; 2.5)$.

Определим для дальнейшего исследования по формулам (84) - (88) числовые характеристики – среднее, исправленное среднее квадратичное отклонение:

$$x_в = \frac{1}{n} \sum_{m=1}^{n} x_m = 0.3875;$$

$$s = \sqrt{\frac{n}{n-1} D_в} = \sqrt{\frac{(-0{,}76)^2 + (1{,}52)^2 + \ldots + (-0{,}62)^2}{32} - (-0{,}106)^2} \cdot \sqrt{\frac{32}{31}} = 1{,}11.$$

Составим таблицу для сравнения эмпирических и теоретических данных:

- n_i - равно числу вариант, попавших на интервал $(x_i; x_{i+1})$;

- $z_i = \dfrac{x_i - \overline{x}_в}{s}$ - теоретические границы интервалов, соответствующие характеристикам выборки $x_в$ и s (например, $z_2 = \dfrac{-2{,}5 - (-0{,}106)}{1{,}11} = -2{,}16$ и т. д.);
- вычислить $\Phi(z_i)$ – значение интегральной функции Лапласа на границах теоретического интервала; левее и правее отрезка $[-2{,}5; 2{,}5]$, значения этой функции равно $(-0{,}5)$ слева и $(+0{,}5)$ справа, на самом интервале значения вычисляются по таблице $\Phi(z_2) = \Phi(-2{,}16) = -0{,}485$ и т. д.;
- вычисляется теоретическая частота попадания в интервал $(x_i; x_{i+1}]$: $p_i = \Phi(z_{i+1}) - \Phi(z_i)$, $n_i' = n \cdot p_i$ (например, $p_1 = \Phi(-2{,}16) - \Phi(-0{,}5) = -0{,}485 - (-0{,}5) = 0{,}015$, тогда $n_1' = 32 \cdot 0{,}015$ и т. д.).

Теоретические n_i' и эмпирические n_i частоты собраны в таблице. Последняя колонка таблицы – это вычисление статистики $\chi^2 = \sum_{i=1}^{n} \dfrac{(n_i - n_i')^2}{n_i'}$.

i	x_i	x_{i+1}	n_i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$p_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n_i' = n \cdot p_i$	$\dfrac{(n_i - n_i')^2}{n_i'}$
1	$-\infty$	-2,5	1		-2,16	-0,5	-0,485	0,014	0,448	0,680
2	-2,5	-1,5	1	-2,16	-1,26	-0,485	-0,396	0,088	2,816	1,171
3	-1,5	-0,5	10	-1,26	-0,35	-0,396	-0,137	0,258	8,256	0,368
4	-0,5	0,5	12	-0,35	0,55	-0,137	0,209	0,346	11,072	0,078
5	0,5	1,5	5	0,55	1,45	0,209	0,426	0,216	6,912	0,529
6	1,5	2,5	3	1,45	2,35	0,426	0,491	0,064	2,048	0,443
7	2,5	$+\infty$	0	2,35		0,491	0,5	0,014	0,448	0,448
$\sum$			**32**					**1**	**32**	**3,717**

Значение критерия получено, далее сравним его с критическим значением, которое найдем по таблице значений критерия Пирсона. В соответствии со степенями свободы и требуемой точностью получим $\chi^2_{кр} = \chi^2(4; 0{,}05) = 9{,}5$.

Итак, $3{,}717(\chi^2_{набл}) < 9{,}5(\chi^2_{кр}(4; 0{,}05))$.

Таким образом, можно сделать вывод о верности *нулевой гипотезы* H_0: принимается гипотеза о нормальном распределении невязок.

ИНДИВИДУАЛЬНЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ

1. Классическое определение вероятности

1. Сборная конструкция состоит из шести элементов, два из которых одинаковы. Элементы конструкции завозятся на стройку в произвольном порядке. Какова вероятность того, что элементы будут завезены в том порядке, в котором они должны быть смонтированы?
2. В аудитории 8 двухместных парт. Какова вероятность того, что при случайном рассаживании два данных студента из группы в 16 человек окажутся за одной партой?
3. Студент идет на экзамен, выучив 25 вопросов из 36. Какова вероятность ответить на три вопроса, задаваемых преподавателем поочерёдно?
4. На подмостках лежат три белых и два красных кирпича. Каменщик случайным образом выкладывает их в один ряд. Какова вероятность того, что цвет кирпичей будет чередоваться?
5. Куратор назначает трех наугад выбранных студентов из своей группы делегатами на профсоюзную конференцию. Какова вероятность того, что делегация будет состоять из одного студента и двух студенток, если в группе 15 студентов и 5 студенток?
6. Бросают три игральных кубика. Какова вероятность того, что сумма выпавших на них очков будет равна трем?
7. Игроку сдают 6 карт из 36. Какова вероятность того, что среди них два туза?
8. Управляющий проверяет наугад 4 СМУ. Какова вероятность того, что номера этих СМУ будут идти в порядке возрастания?
9. На складе имеется 30 мешков цемента марки 300, 50 – марки 400 и 20 – марки 500. Наугад берется один мешок цемента и привозится на стройку. Какова вероятность того, что его придется обменивать на складе на другой мешок, если цемент марки 300 не подходит для планируемой работы?
10. Какова вероятность того, что из группы в 20 студентов при случайном выборе старостой станет студент Иванов, а профоргом – Петров, если в группе есть два студента по фамилии Иванов и один по фамилии Петров?
11. Бросают два игральных кубика. Какова вероятность того, что сумма выпавших на них очков окажется равной шести?
12. Из двух бригад по 10 человек наугад выбирают по одному человеку. Два брата входят в состав различных бригад. Найти вероятность того, что оба брата окажутся выбранными?
13. Бросают два игральных кубика. Какова вероятность того, что произведение выпавших на них очков окажется равным тридцати шести?
14. На строительных лесах лежат 12 красных и 8 белых кирпичей. Наугад берут 3 кирпича. Какова вероятность того, что два из них красные, а один белый?
15. Бросаются две игральные кости. Какова вероятность того, что на них выпадет одинаковое число очков?

16. Из шести карточек с буквами сложено слово “КАРЕТА”. Ребенок перемешивает карточки и снова раскладывает их в ряд. Какова вероятность того, что получится слово “РАКЕТА”?
17. В бригаде каменщиков 6 мужчин и 4 женщины. Случайным образом отбирают бригаду из 7 человек. Какова вероятность того, что среди отобранных окажется 5 мужчин?
18. Бросаются две игральные кости. Какова вероятность того, что выпадет дубль (одинаковое число очков)?
19. Ребенок на компьютере случайно нажимает подряд три разные клавиши. Какова вероятность, что будет напечатано слово “МИР”, если имеется всего 106 клавиш?
20. В аудитории 12 двухместных парт. Какова вероятность того, что при случайном рассаживании два данных студента из группы в 24 человека окажутся за одной партой?
21. Игрок в “Спортлото” зачеркивает 5 чисел из 36. Какова вероятность угадать все 5 номеров?
22. Человек забыл две последние цифры телефона, но помнит, что среди них есть одна цифра “5”. Какова вероятность набрать правильный номер с одной попытки?
23. В партии из 40 смесителей 5 бракованных. Наудачу берут 6 смесителей. Какова вероятность того, что из выбранных есть один бракованный?
24. Ребенок играет с 33 карточками разрезной азбуки, на которых написаны разные буквы русского алфавита. Какова вероятность того, что из наугад сложенных в ряд четырех карточек азбуки получится слово “НЕБО”?
25. Какова вероятность, что при бросании двух игральных костей произведение выпавших на них очков равно 12?

2. Вероятность суммы и произведения событий

26. Вероятность при одиночном выстреле поразить цель равна 0,4. Сколько нужно провести выстрелов, чтобы с вероятностью не менее 0,9 цель была поражена?
27. Вероятность отказа первого узла прибора – 1/8, а второго – 1/7. Найти вероятность безотказной работы прибора, состоящего из этих двух узлов.
28. При заезде на стройку вероятность прокола шины самосвала равна 0,005. Какова вероятность, что прокола шины не будет при трехкратном заезде на стройку?
29. Стрелок набирает не менее 9 очков в 75% случаев, а вероятность попадания в десятку равна 0,4. Какова вероятность попадания в девятку?
30. Вероятность заболеть во время эпидемии гриппа равна 0,1. Какова вероятность того, что в бригаде из восьми человек никто не заболел?
31. Какова вероятность для студента быть отличником, если вероятность сдать на отлично первый экзамен – 80%, второй – 90% и третий – 95%?

32. Вероятность разрушения у двух конкретных домов при 6 балльном землетрясении равны соответственно 1/10 и 1/15. Найти вероятность того, что оба дома при землетрясении устоят.
33. Баскетболист выполняет два штрафных броска, при этом вероятность попасть в первый раз равна 0,8, а во второй – 0,9. Найти вероятность получить только одно очко из двух.
34. Прибор состоит из трех жизненно важных узлов, отказ каждого из них выводит из строя весь прибор. Какова вероятность безотказной работы всего прибора, если вероятности отказа узлов равны 0,3; 0,2 и 0,1?
35. На стройке работают два крана. Один из них занят 70% всего рабочего времени, а другой – 80%. Какова вероятность того, что в данный момент работает только один кран?
36. Два стрелка стреляют по одной мишени, делая по одному выстрелу. Вероятность попадания в мишень при одном выстреле первым стрелком равна 0,8 и вторым – 0,7. Какова вероятность того, что в цель попадет только один стрелок?
37. 90% продукции завода составляют стандартные изделия, из них 70% – высшего сорта. Какова вероятность, что взятое наугад изделие высшего сорта?
38. Вероятность отказа насоса на атомной станции равна 0,001, что приводит к аварии. Сколько надо поставить запасных насосов, чтобы вероятность аварии по вине насосов была 0,000000001?
39. Три орудия стреляют по цели с вероятностями попаданий 0,5, 0,8 и 0,9 соответственно. Цель поражена, если есть хотя бы одно попадание. Найти вероятность поражения цели.
40. Вероятность того, что в течение одной смены возникнет неполадка башенного крана, равна 0,05. Какова вероятность безаварийной работы крана в течение трех смен?
41. По статистическим данным из 100000 десятилетних детей до 40 лет доживают 82277 человек, а до 70 лет – 37965 человек. Найти вероятность для сорокалетних дожить до 70 лет.
42. Стрелок попадает в десятку в 15% случаев и в девятку в 25%. Найти вероятность набора не более 8 очков.
43. Два стрелка стреляют по одной мишени, делая по одному выстрелу. Вероятность попадания в мишень при одном выстреле первым стрелком равна 0,7 и вторым – 0,8. Какова вероятность того, что оба попадут в цель?
44. При изготовлении детали заготовка должна пройти 4 операции. Предполагая появление брака на отдельных операциях событиями независимыми, найти вероятность изготовления стандартной детали, если вероятность брака на первой операции 0,02; на второй – 0,01; на третьей – 0,03 и на четвертой – 0,04.
45. Брошены две игральные кости. Чему равна вероятность того, что хотя бы на одной из них выпадет 6 очков?

46. Вероятность завести двигатель автомобиля зимой с одной попытки равна 0,6. Какова вероятность того, что двигатель заведется со второй попытки?
47. Какова вероятность того, что две сданные карты окажутся одной масти, если в колоде 36 карт?
48. У двоих работающих вместе строителей простои составляют соответственно 20% и 30% от всего рабочего времени. Найти вероятность того, что в данный момент работает хотя бы один из них.
49. Сколько раз нужно бросить игральную кость, чтобы с вероятностью не меньшей 0,3 хотя бы один раз выпала шестерка?
50. Экзаменационный билет содержит три вопроса. Вероятности по отдельности ответить на каждый из этих вопросов равна 0,7, 0,8 и 0,9 соответственно. Найти вероятность того, что студент сдаст экзамен, если для этого необходимо ответить на все три вопроса?

3. **Формула полной вероятности**

51. Стреляющий первым во время дуэли может попасть в противника с вероятностью 0,8. При ответном выстреле второй дуэлянт может попасть в первого с вероятность 0,9, если первый промахнулся, и с вероятностью 0,3, если первый попал в него. Найти вероятность второму дуэлянту попасть в первого при ответном выстреле?
52. Вероятности сдать или не сдать экзамен без дополнительных вопросов равны соответственно 0,3 и 0,2. В остальных случаях студенту задаются дополнительные вопросы, вероятность ответить на которые равна 0,6. Какова вероятность того, что студент сдаст экзамен?
53. Пассажир электрички вышел на платформу в пункте А и знает, в каком направлении находится пункт В. Какова вероятность того, что он прибудет в пункт В, если в каждой развилке путник выбирает дорогу в нужном направлении случайным образом? (Схема дорог указана на рис. 30).

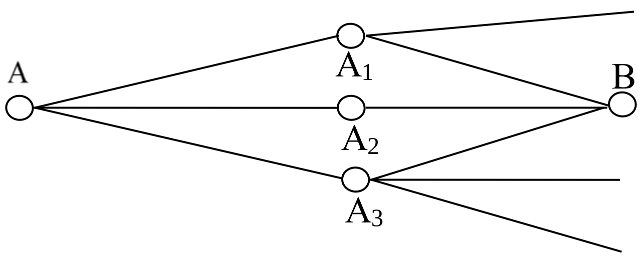


Рис. 30

54. Вероятность обнаружить брак в цехе равна 0,3. Вероятность того, что в дальнейшем брак будет обнаружен в ОТК завода, равна 0,6. Какова вероятность того, что брак будет обнаружен?
55. На стройку поступают плиты с трех железобетонных заводов: 200 плит с первого завода, 400 плит со второго и 900 с третьего. Процент брака изделий

этих железобетонных заводов равен соответственно 1,5%, 2% и 2,5%. Найти вероятность того, что плита, поднимаемая краном, стандартная.

56. При оценке остаточных знаний процент справившихся с заданиями составляет 80% у отличников, 60% у хорошистов и 20% у остальных. В потоке из 96 человек 12 отличников и 24 хорошиста по тестируемому предмету. Какова вероятность для наугад выбранного студента справиться с заданием?
57. Прибор нормально работает в 90% случаев. Найти вероятность выхода прибора из строя за время T, если вероятность его выхода из строя за указанное время T в нормальном режиме составляет 0,1, а в противном случае 0,7.
58. У передвижной бетономешалки остановилось вращение барабана. Есть возможность довести бетон, прежде чем он успеет схватиться, до любой из двух ближайших строек с вероятностями 0,9 и 0,8 соответственно. Кроме того, к первой стройке ведет одна дорога, а ко второй – две. Какова вероятность успешно довезти бетон, если дорога выбрана случайно?
59. На сборку попадают детали с трех станков. Известно, что брак с первого станка составляет 0,3%, со второго – 0,2% с третьего – 0,4%. Найти вероятность попадания на сборку бракованной детали, если с первого и третьего станков поступило по 2000 деталей, а со второго – 1000 деталей.
60. Нормальный режим работы прибора наблюдается в 80% случаев. Вероятность выхода прибора из строя за время T в нормальном режиме составляет 0,2, а в противном случае – 0,7. Найти вероятность безотказной работы прибора за время T.
61. В группе из 20 человек 3 отличника и 7 хорошистов. Отличник наверняка сдаст экзамен на “отлично” или “хорошо”; хорошист в 10% случаев сдаст на “отлично” и в 60% случаев сдаст “хорошо”; а остальные только в 20% случаев получат хорошую оценку. Какова вероятность наугад выбранному студенту этой группы сдать экзамен на “хорошо” или “отлично”?
62. У передвижной бетономешалки остановилось вращение барабана. Есть возможность довести бетон, прежде чем он успеет схватиться, до любой из трех ближайших строек с вероятностями 0,9; 0,7 и 0,8 соответственно. Кроме того, к первой стройке ведет одна дорога, а ко второй и третьей стройкам по две дороги. Какова вероятность успешно довезти бетон, если дорога выбрана случайно?
63. На столе лежит 25 билетов, из них 16 “счастливых” для данного студента. Изменится ли вероятность вытащить “счастливый” билет, если студент идет сдавать экзамен не первым, а вторым?
64. В первой коробке лежит 20 дюбелей, из которых 15 стандартных. Из первой коробки во вторую, содержащую 24 дюбеля (из них 19 стандартных), переложен один дюбель. Какова вероятность после этого достать из второй коробки стандартный дюбель?
65. На стройку поставляют партии кирпичей с трех заводов: с первого – 25000, со второго – 35000 и с третьего – 40000 штук. Процент брака у партий кир-

пича с этих заводов составляет соответственно 2%, 3% и 4%. Какова вероятность того, что каменщик возьмет из пакета стандартный кирпич, если крановщик поднимает ближайший к нему в данный момент пакет кирпичей?

66. На столе лежит 30 билетов, из них 25 "счастливых" для данного студента. Изменится ли вероятность вытащить "счастливый" билет, если студент идет сдавать экзамен не первым, а вторым?
67. При оценке остаточных знаний процент справившихся с заданиями составляет 70% у отличников, 50% у хорошистов и 10% у остальных. В потоке из 150 человек 16 отличников и 36 хорошиста по тестируемому предмету. Какова вероятность для наугад выбранного студента справиться с заданием?
68. На стройку поступают плиты с трех железобетонных заводов: 100 плит с первого завода, 200 плит со второго и 500 – с третьего. Процент брака изделий железобетонных заводов составляет соответственно 3%, 4% и 2%. Найти вероятность того, что плита, поднимаемая краном, стандартная.
69. Вероятность обнаружить брак в цехе равна 0,6. Вероятность того, что в ОТК завода в дальнейшем брак будет обнаружен равна 0,85. Какова вероятность того, что брак будет обнаружен?
70. Пассажир электрички вышел на платформу в пункте А и знает, в каком направлении находится пункт В. Какова вероятность того, что он прибудет в пункт В, если после отдыха в одном из пунктов A_1, A_2 или A_3 он может пойти в любом направлении, в том числе и в обратном? (Схема дорог указана на рис. 30)
71. Вероятности сдать или не сдать экзамен без дополнительных вопросов равны соответственно 0,4 и 0,05. В остальных случаях студенту задаются дополнительные вопросы, вероятность ответить на которые равна 0,7. Какова вероятность того, что студент сдаст экзамен?
72. Стреляющий первым во время дуэли может попасть в противника с вероятностью 0,9. При ответном выстреле второй дуэлянт может попасть в первого с вероятность 0,7, если первый промахнулся и с вероятностью 0,2, если первый попал в него. Найти вероятность второму дуэлянту попасть в первого при ответном выстреле.
73. На столе лежит 20 билетов, из них 12 "счастливых" для данного студента. Изменится ли вероятность вытащить "счастливый" билет, если студент идет сдавать экзамен не первым, а вторым?
74. В первой коробке лежит 36 дюбелей, из которых 20 стандартных. Из первой коробки во вторую, содержащую 19 дюбеля (из них 14 стандартных), переложен один дюбель. Какова вероятность после этого достать из второй коробки стандартный дюбель?
75. Пассажир электрички вышел на платформу в пункте А и знает, в каком направлении находится пункт В. Какова вероятность того, что он прибудет в пункт В, если в каждой развилке путник выбирает дорогу в нужном направлении случайным образом? (Схема дорог указана на рис. 31).

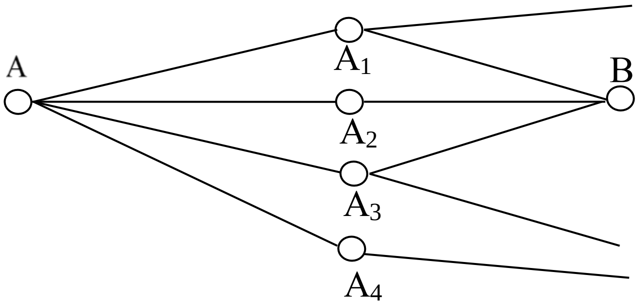


Рис. 31

4. Повторение испытаний (схема Бернулли)

Задачи №№76-100. Производится n независимых испытаний, в каждом из которых событие A происходит с вероятностью p. Найти вероятность того, что:

а) событие A произойдет ровно k раз;

б) событие A произойдет не менее k_1 и не более k_2 раз.

№	n	p	k	k_1	k_2
76.	390	0,6	240	230	235
77.	7	0,7	5	4	6
78.	9	0,4	3	2	4
79.	290	0,7	200	205	216
80.	6	0,7	4	3	6
81.	110	0,03	4	3	5
82.	180	0,7	125	130	140
83.	8	0,6	5	4	7
84.	195	0,6	115	120	130
85.	142	0,02	3	4	5
86.	6	0,4	2	3	5
87.	625	0,6	380	360	370
88.	250	0,01	3	1	4
89.	8	0,7	5	4	6
90.	120	0,04	5	4	6
91.	540	0,4	200	245	260
92.	6	0,9	4	5	6
93.	7	0,3	3	2	4
94.	350	0,6	170	175	190
95.	7	0,4	3	4	6
96.	200	0,8	165	155	170
97.	170	0,03	5	3	5
98.	240	0,7	160	165	175
99.	9	0,8	7	5	7
100.	150	0,02	2	2	4

5. **Дискретные случайные величины**

Три плотника сделали по одному экземпляру одного и того же изделия. Вероятность предоставить готовое изделие без брака для них соответственно равны p_1, p_2, p_3. Составить закон распределения случайной величины X - числа готовых изделий без брака, найти ее математическое ожидание, дисперсию и среднее квадратическое отклонение.

№	p_1	p_2	p_3
101.	0,9	0,6	0,4
102.	0,6	0,7	0,5
103.	0,7	0,4	0,6
104.	0,4	0,7	0,8
105.	0,9	0,7	0,5
106.	0,6	0,3	0,8
107.	0,9	0,7	0,1
108.	0,8	0,6	0,5
109.	0,5	0,4	0,8
110.	0,9	0,2	0,8
111.	0,6	0,3	0,7
112.	0,3	0,6	0,9
113.	0,8	0,1	0,6
114.	0,6	0,3	0,9
115.	0,9	0,4	0,7
116.	0,7	0,1	0,8
117.	0,2	0,9	0,7
118.	0,8	0,3	0,9
119.	0,5	0,9	0,2
120.	0,2	0,4	0,9
121.	0,2	0,8	0,7
122.	0,4	0,3	0,8
123.	0,7	0,9	0,3
124.	0,9	0,8	0,4
125.	0,8	0,5	0,7

6. **Непрерывные случайные величины, их способы задания и основные числовые характеристики**

Задача I.

Задана функция $f(x)$ на указанных промежутках, для нее нужно найти:

1. Константу A, при которой функция $f(x)$ может быть плотностью распределения некоторой случайной величины X;
2. Соответствующую функцию распределения $F(x)$;
3. Математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

126. $f(x)=\begin{cases} 0, & x\le -1; \\ A(x+1), & -1<x\le 1; \\ 0, & 1<x. \end{cases}$

127. $f(x)=\begin{cases} 0, & x\le 0; \\ Ax^2, & 0<x\le 2; \\ 0, & 2<x. \end{cases}$

128. $f(x)=\begin{cases} 0, & x\le -2; \\ A(2-x), & -2<x\le 0; \\ 0, & 0<x. \end{cases}$

129. $f(x)=\begin{cases} 0, & x\le 0; \\ Ax^3, & 0<x\le 3; \\ 0, & 3<x. \end{cases}$

130. $f(x)=\begin{cases} 0, & x\le 0; \\ A(3x+1), & 0<x\le 1/3; \\ 0, & 1/3<x. \end{cases}$

131. $f(x)=\begin{cases} 0, & x\le 0; \\ Ax^4, & 0<x\le 1; \\ 0, & 1<x. \end{cases}$

132. $f(x)=\begin{cases} 0, & x\le -3; \\ A(2-3x), & -3<x\le -1; \\ 0, & -1<x. \end{cases}$

133. $f(x)=\begin{cases} 0, & x\le 0; \\ Ax^5, & 0<x\le 2; \\ 0, & 2<x. \end{cases}$

134. $f(x)=\begin{cases} 0, & x\le 1; \\ A(2x+3), & 1<x\le 2; \\ 0, & 2<x. \end{cases}$

135. $f(x)=\begin{cases} 0, & x\le 0; \\ Ax^7, & 0<x\le 1; \\ 0, & 1<x. \end{cases}$

136. $f(x)=\begin{cases} 0, & x\le 3; \\ A(2x-3), & 3<x\le 4; \\ 0, & 4<x. \end{cases}$

137. $f(x)=\begin{cases} 0, & x\le 0; \\ Ax^6, & 0<x\le 3; \\ 0, & 3<x. \end{cases}$

138. $f(x)=\begin{cases} 0, & x\le -4; \\ A(3-x), & -4<x\le -2; \\ 0, & -2<x. \end{cases}$

139. $f(x)=\begin{cases} 0, & x\le -2; \\ Ax^2, & -2<x\le 0; \\ 0, & 0<x. \end{cases}$

140. $f(x)=\begin{cases} 0, & x\le 0; \\ A(x^2+x), & 0<x\le 2; \\ 0, & 2<x. \end{cases}$

141. $f(x)=\begin{cases} 0, & x\le -3; \\ Ax^3, & -3<x\le 0; \\ 0, & 0<x. \end{cases}$

142. $f(x)=\begin{cases} 0, & x\le -1; \\ A(x^2-x), & -1<x\le 0; \\ 0, & 0<x. \end{cases}$

143. $f(x)=\begin{cases} 0, & x\le -3; \\ Ax^4, & -3<x\le 0; \\ 0, & 0<x. \end{cases}$

144. $f(x)=\begin{cases} 0, & x\le 1; \\ A(x^2-3x+2), & 1<x\le 2; \\ 0, & 2<x. \end{cases}$

145. $f(x)=\begin{cases} 0, & x\le -2; \\ Ax^5, & -2<x\le 0; \\ 0, & 0<x. \end{cases}$

146. $f(x)=\begin{cases}0, & x\le -2;\\ A(x^2-4x+3), & -2<x\le 0;\\ 0, & 0<x.\end{cases}$ 147. $f(x)=\begin{cases}0, & x\le -3;\\ Ax^6, & -3<x\le 0;\\ 0, & 0<x.\end{cases}$

148. $f(x)=\begin{cases}0, & x\le 0;\\ A(x^3+x), & 0<x\le 1;\\ 0, & 1<x.\end{cases}$ 149. $f(x)=\begin{cases}0, & x\le -1;\\ Ax^7, & -1<x\le 0;\\ 0, & 0<x.\end{cases}$

150. $f(x)=\begin{cases}0, & x\le -1;\\ A(x^3-x), & -1<x\le 0;\\ 0, & 0<x.\end{cases}$

Задача II.

Случайная величина X задана функцией распределения $F(x)$. Найти функцию плотности распределения $f(x)$. Построить графики функций $f(x)$ и $F(x)$. Найти вероятность попадания случайной величины X в заданный интервал $(\alpha;\beta)$.

151. $F(x)=\begin{cases}0, & x\le 0;\\ 2\sin x, & 0<x\le \pi/6;\\ 1, & \pi/6<x.\end{cases}$ $\alpha=-1,\ \beta=\pi/9.$

152. $F(x)=\begin{cases}0, & x\le 3\pi/2;\\ 1+\sin x, & 3\pi/2<x\le 2\pi;\\ 1, & 2\pi<x.\end{cases}$ $\alpha=-1,\ \beta=5\pi/3.$

153. $F(x)=\begin{cases}0, & x\le \pi;\\ 1+\cos x, & \pi<x\le 3\pi/2;\\ 1, & 3\pi/2<x.\end{cases}$ $\alpha=5\pi/4,\ \beta=2\pi.$

154. $F(x)=\begin{cases}0, & x\le -\pi;\\ 2+2\cos x, & -\pi<x\le -2\pi/3;\\ 1, & -2\pi/3<x.\end{cases}$ $\alpha=-5\pi/6,\ \beta=-\pi/3.$

155. $F(x)=\begin{cases}0, & x\le \pi;\\ 2\cos^2(x/2), & \pi<x\le 3\pi/2;\\ 1, & 3\pi/2<x.\end{cases}$ $\alpha=\pi/3,\ \beta=4\pi/3.$

156. $F(x)=\begin{cases}0, & x\le 3\pi/2;\\ 2+2\sin x, & 3\pi/2<x\le 11\pi/6;\\ 1, & 11\pi/6<x.\end{cases}$ $\alpha=5\pi/3,\ \beta=2\pi.$

157. $F(x)=\begin{cases}0, & x\le 0;\\ \sin x, & 0<x\le \pi/2;\\ 1, & \pi/2<x.\end{cases}$ $\alpha=\pi/3,\ \beta=2\pi/3.$

158. $F(x)=\begin{cases}0, & x\le \pi/2;\\ -\cos x, & \pi/2<x\le \pi;\\ 1, & \pi<x.\end{cases}$ $\alpha=\pi/6,\ \beta=3\pi/4.$

159. $F(x)=\begin{cases}0, & x\le \pi/2;\\ -2\cos x, & \pi/2<x\le 2\pi/3;\\ 1, & 2\pi/3<x.\end{cases}$ $\alpha=7\pi/12,\ \beta=3\pi/4.$

160. $F(x)=\begin{cases}0, & x\le -\pi;\\ -\sin x, & -\pi<x\le -\pi/2;\\ 1, & -\pi/2<x.\end{cases}$ $\alpha=-5\pi/6,\ \beta=2\pi.$

161. $$F(x)=\begin{cases}0, & x\le -2\pi;\\ 2\sin x, & -2\pi<x\le -11\pi/6;\\ 1, & -11\pi/6<x.\end{cases}$$
$$\alpha=-23\pi/12,\ \beta=0.$$

162. $$F(x)=\begin{cases}0, & x\le 0;\\ 1-\cos x, & 0<x\le \pi/2;\\ 1, & \pi/2<x.\end{cases}$$
$$\alpha=-\pi/3,\ \beta=\pi/3.$$

163. $$F(x)=\begin{cases}0, & x\le 2\pi;\\ 1-\cos x, & 2\pi<x\le 5\pi/2;\\ 1, & 5\pi/2<x.\end{cases}$$
$$\alpha=9\pi/4,\ \beta=3\pi.$$

164. $$F(x)=\begin{cases}0, & x\le \pi/2;\\ 1-\sin x, & \pi/2<x\le \pi;\\ 1, & \pi<x.\end{cases}$$
$$\alpha=3\pi/4,\ \beta=3\pi/2.$$

165. $$F(x)=\begin{cases}0, & x\le -2\pi;\\ \sin x, & -2\pi<x\le -3\pi/2;\\ 1, & -3\pi/2<x.\end{cases}$$
$$\alpha=-3\pi,\ \beta=-7\pi/4.$$

166. $$F(x)=\begin{cases}0, & x\le -\pi/2;\\ \cos x, & -\pi/2<x\le 0;\\ 1, & 0<x.\end{cases}$$
$$\alpha=-\pi/6,\ \beta=\pi/3.$$

167. $$F(x)=\begin{cases}0, & x\le 5\pi/2;\\ -\cos x, & 5\pi/2<x\le 3\pi;\\ 1, & 3\pi<x.\end{cases}$$
$$\alpha=2\pi,\ \beta=11\pi/4.$$

168. $$F(x)=\begin{cases}0, & x\le \pi;\\ -2\sin x, & \pi<x\le 7\pi/6;\\ 1, & 7\pi/6<x.\end{cases}$$
$$\alpha=\pi/2,\ \beta=13\pi/12.$$

169. $$F(x)=\begin{cases}0, & x\le -\pi/2;\\ 1+\sin x, & -\pi/2<x\le 0;\\ 1, & 0<x.\end{cases}$$
$$\alpha=-\pi/6,\ \beta=\pi/2.$$

170. $$F(x)=\begin{cases}0, & x\le \pi;\\ 2+2\cos x, & \pi<x\le 4\pi/3;\\ 1, & 4\pi/3<x.\end{cases}$$
$$\alpha=7\pi/6,\ \beta=2\pi.$$

171. $$F(x)=\begin{cases}0, & x\le -3\pi/2;\\ -2\cos x, & -3\pi/2<x\le -4\pi/3;\\ 1, & -4\pi/3<x.\end{cases}$$
$$\alpha=-17\pi/12,\ \beta=-\pi.$$

172. $$F(x)=\begin{cases}0, & x\le \pi/2;\\ 2-2\sin x, & \pi/2<x\le 5\pi/6;\\ 1, & 5\pi/6<x.\end{cases}$$
$$\alpha=3\pi/4,\ \beta=\pi.$$

173. $$F(x)=\begin{cases}0, & x\le -\pi/2;\\ 2+2\sin x, & -\pi/2<x\le -\pi/6;\\ 1, & -\pi/6<x.\end{cases}$$
$$\alpha=-\pi/3,\ \beta=\pi/2.$$

174. $$F(x)=\begin{cases}0, & x\le \pi/2;\\ -2\cos x, & \pi/2<x\le 2\pi/3;\\ 1, & 2\pi/3<x.\end{cases}$$
$$\alpha=3\pi/4,\ \beta=4\pi/3.$$

175. $$F(x)=\begin{cases}0, & x\le -\pi/2;\\ 2\cos x, & -\pi/2<x\le -\pi/3;\\ 1, & -\pi/3<x.\end{cases}$$
$$\alpha=-5\pi/12,\ \beta=0.$$

7. **Равномерный, показательный и биномиальный законы распределения**

Задан закон распределения и его основные параметры:

- для равномерного распределения – интервал $(a;b)$;
- для показательного распределения – параметр λ;
- для биномиального распределения – вероятность $p = p(A)$ и число испытаний n.

Найти:

1. Для равномерного и показательного распределения функцию плотности распределения и построить ее график.
   Для биномиального распределения найти закон распределения.
2. Для всех видов распределения написать функции распределения, построить их графики.
3. Числовые характеристики случайных величин – математическое ожидание, дисперсию и среднее квадратическое отклонение.
4. Вероятность попадания случайной величины в заданный интервал $(\alpha;\beta)$.

№	Закон распределения	a	b	λ	p	n	α	β
176.	Равномерный	6	15				3	12
177.	Показательный			6			0	1/12
178.	Биномиальный				0,7	3	0,3	10
179.	Равномерный	7	18				2	11
180.	Показательный			1,5			0	1/3
181.	Биномиальный				0,8	3	2,5	5
182.	Равномерный	7	16				3	10
183.	Показательный			2,5			0	1/5
184.	Биномиальный				0,9	3	0,5	6
185.	Равномерный	9	17				4	12
186.	Показательный			1			0	0,5
187.	Биномиальный				0,2	3	1	2,5
188.	Равномерный	3	10				6	13
189.	Показательный			2			0	0,25
190.	Биномиальный				0,3	3	0,5	2,5
191.	Равномерный	4	13				8	16
192.	Показательный			3			0	1/6
193.	Биномиальный				0,4	3	1,5	4
194.	Равномерный	10	17				15	20
195.	Показательный			4			0	1/8
196.	Биномиальный				0,1	3	2	3,5

197.	Равномерный	1	10				8	13
198.	Показательный			5			0	0,1
199.	Биномиальный				0,6	3	0,3	38
200.	Равномерный	2	11				5	15

8. Нормальный закон распределения

Даны математическое ожидание a и среднее квадратичное отклонение σ нормально распределенной случайной величины X.

Требуется:

1. Записать функцию плотности данной нормально распределенной случайной величины и построить ее график.
2. Найти вероятность того, что случайная величина X примет значения из заданного интервала $(\alpha;\beta)$.
3. Найти вероятность того, что абсолютная величина отклонения от среднего значения окажется меньше δ.

№	a	σ	α	β	δ
201.	11	4	5	25	9
202.	13	3	6	21	5
203.	14	5	7	22	11
204.	15	4	11	21	6
205.	13	5	9	19	4
206.	11	3	7	19	6
207.	9	5	15	18	8
208.	7	3	3	16	6
209.	11	5	7	19	9
210.	13	4	7	20	8
211.	15	5	8	21	7
212.	7	2	3	12	5
213.	9	3	5	11	4
214.	10	3	6	15	8
215.	12	4	2	15	3
216.	14	4	10	20	10
217.	12	5	8	18	10
218.	10	4	6	16	10
219.	8	2	4	14	5
220.	10	5	8	20	8
221.	12	3	10	15	4
222.	14	6	5	24	11
223.	16	6	5	28	12
224.	8	4	3	15	8
225.	10	5	3	20	9

9. **Корреляционный анализ. Уравнение регрессии**

Задачи №№226-250. Для представленных ниже данных вычислить коэффициент корреляции и построить уравнения регрессии.

№ 226

n_{ij}		*Y*					
		2	7	12	17	22	27
	10	2	4				
	20		6	2			
X	30			3	50	2	
	40			1	10	б	
	50				4	7	3

№ 227

n_{ij}		*Y*					
		10	15	20	25	30	35
	30	2	6				
	40		4	4			
X	50			7	35	8	
	60			2	10	8	
	70				5	б	3

№ 228

n_{ij}		*Y*					
		10	15	20	25	30	35
	30	2	6				
	40		4	4			
X	50			7	35	8	
	60			2	10	8	
	70				5	б	3

№ 229

n_{ij}		*Y*					
		6	12	18	24	30	36
	5					3	3
	10				4	5	
X	15		2	40	8	2	
	20	8	2	10			
	25	9	12	4			

№ 230

n_{ij}		*Y*					
		8	13	18	23	28	33
	18	2					
	28	2	10			4	
X	38		10	30	5		
	48			15	7	8	2
	58						3

№ 231

n_{ij}		*Y*					
		4	9	14	19	24	29
	10	2	3				
	20		7	3			
X	30			2	5	2	
	40				10	6	
	50				1	б	4

№ 232

n_{ij}		*Y*				
		15.5	16.0	16.5	17.0	17.5
	12	5	2			
	13	3	15	6		
Y	14		8	20	7	
	15			9	15	2
	16				2	6

№ 233

n_{ij}		*Y*					
		15	20	25	30	35	40
	5	4	2				
	8		6	4			
X	11		5	6	45	2	
	14			2	3	6	
	17				4	7	4

№ 234

n_{ij}		Y				
		40	50	60	70	80
X	25	2				
	40	6	15	10		4
	55		8	15	5	
	70			7	15	8
	85	2				3

№ 235

n_{ij}		Y					
		5	10	15	20	25	30
X	8	2	4				
	12		3	7			
	16			5	30	10	
	20			7	10	8	
	24				5	б	3

№ 236

n_{ij}		Y					
		5,5	6,0	6,5	7,0	7,5	8,0
X	4	6	4				
	6		10	8			
	8			4	35	5	
	10			б	10	4	
	12				3	3	2

№ 237

n_{ij}		Y					
		65	75	85	95	105	115
X	80	5	3				
	85	3	30	8			
	90			70	10		
	95			10	40	10	
	100					6	5

№ 238

n_{ij}		Y					
		10	20	30	40	50	60
X	60	5	3				
	75		18	7			
	90	2		30	10		
	105				12	3	4
	120					2	4

№ 239

n_{ij}		Y				
		2,4	2,8	3,2	3,6	4,0
X	2,5	10	3			4
	3,0	7	30	15		
	3,5		10	50	15	
	4,0			10	30	6
	4,5					10

№ 240

n_{ij}		Y				
		50	55	60	65	70
X	30				5	5
	40	4	6			
	50		15	40		
	60			10	5	
	70				5	
	80				3	2

№ 241

n_{ij}		Y				
		25	30	35	40	45
X	20					2
	22			5	4	3
	24			35	10	
	26	5	10	5		
	28		7			
	30	6	8			

№ 242

n_{ij}		Y					
		10	20	30	40	50	60
X	30	5	10				
	40			15		55	
	50			20	80		
	60		5		10	35	
	70	5					10

№ 243

n_{ij}		Y					
		10	20	30	40	50	60
X	30					10	5
	40				15		
	50			80	20		
	60		35	10		5	
	70	10	3			2	

№ 244

n_{ij}		Y					
		6	12	18	24	30	36
X	5					3	3
	10				4	5	
	15		2	40			
	20		6	10	5	б	
	25	3	7	4			

№ 245

n_{ij}		Y					
		-2	0	4	6	12	24
X	1	4	5				
	2	1	4	9			
	4		1	20	21		
	6			4	17	6	
	7				5	10	8

№ 246

n_{ij}		Y					
		15	18	21	24	27	30
X	20	2	4				
	25		6	3	35	4	
	30			6	8	6	
	35			2	4	7	3
	40				6	4	

№ 247

n_{ij}		Y					
		-6	0	6	12	18	24
X	1	8	1				
	2	1	4	9			
	3		1	10	31		
	4			4	16	7	
	5				2	13	8

№ 248

n_{ij}		Y					
		7	9	12	18	22	23
X	10	5	3				
	20	2	17	8			
	30			20	20		
	40				10	5	4
	50					2	4

№ 249

n_{ij}		Y				
		2,5	2,9	3,4	3,9	4,0
X	2,5	10	3			
	3,0	7	30	15		
	3,5		20	40	15	
	4,0			20	20	6
	4,5	4				10

№ 250						
n_{ij}		Y				
		2,5	3,0	3,5	4,0	4,5
X	15	10	3			
	20	9	30	13		4
	25		20	40	15	
	30		10	10	20	8
	35					8

10. **Математическая статистика**

Задачи №№251-275. Задано интервальное распределение выборки.
Требуется:

а) построить гистограмму относительных частот;

б) перейти к вариантам, выписать эмпирическую функцию распределения и построить ее график;

в) методом условных вариант найти точечные оценки $\overline{x_в}$ и $\sigma_в$;

г) считая генеральную совокупность нормально распределенной, найти доверительные интервалы для a и σ с надежностью $\gamma = 0{,}95$.

№ 251	(–3, –1)	(–1, 1)	(1, 3)	(3, 5)	(5, 7)	(7, 9)
	2	8	19	15	5	1

№ 252	(–12, –10)	(–10, –8)	(–8, –6)	(–6, –4)	(–4, –2)	(–2, 0)
	2	9	14	15	8	2

№ 253	(–7, –5)	(–5, –3)	(–3, –1)	(–1, 1)	(1, 3)	(3, 5)
	3	4	18	20	4	1

№ 254	(0, 2)	(2, 4)	(4, 6)	(6, 8)	(8, 10)	(10, 12)
	1	4	16	18	8	3

№ 255	(5, 7)	(7, 9)	(9, 11)	(11, 13)	(13, 15)	(15, 17)
	3	5	18	17	6	1

№ 256	(1, 3)	(3, 5)	(5, 7)	(7, 9)	(9, 11)	(11, 13)
	3	5	16	17	6	3

№ 257	(0, 2)	(2, 4)	(4, 6)	(6, 8)	(8, 10)	(10, 12)
	2	4	18	17	6	3

№ 258	(–8, –6)	(–6, –4)	(–4, –2)	(–2, 0)	(0, 2)	(2, 4)
	1	4	21	19	3	2

№ 259	(5, 7)	(7, 9)	(9, 11)	(11, 13)	(13, 15)	(15, 17)
	1	5	18	19	4	3

№ 260	(–2, 0)	(0, 2)	(2, 4)	(4, 6)	(6, 8)	(8, 10)
	2	9	15	14	8	2

№ 261	(4, 6)	(6, 8)	(8, 10)	(10, 12)	(12, 14)	(14, 16)
	2	10	12	13	10	3

№ 262	(–5, –3)	(–3; –1)	(–1, 1)	(1, 3)	(3, 5)	(5, 7)
	1	4	18	20	5	2

№ 263	(7, 9)	(9, 11)	(11, 13)	(13, 15)	(15, 17)	(17, 19)
	2	6	17	19	5	1

№ 264	(1, 3)	(3, 5)	(5, 7)	(7, 9)	(9, 11)	(11, 13)
	3	5	18	16	6	2

№ 265	(–7, –5)	(–5, –3)	(–3, –1)	(–1, 1)	(1, 3)	(3, 5)
	1	5	18	19	4	3

№ 266	(–6, –4)	(–4, –2)	(–2, 0)	(0, 2)	(2, 4)	(4, 6)
	2	6	17	18	4	3

№ 267	(0, 2)	(2, 4)	(4, 6)	(6, 8)	(8, 10)	(10, 12)
	1	3	19	21	4	2

№ 268	(–4, –2)	(–2, 0)	(0, 2)	(2, 4)	(4, 6)	(6, 8)
	3	8	14	15	9	1

№ 269	(–2, 0)	(0, 2)	(2, 4)	(4, 6)	(6, 8)	(8, 10)
	1	4	20	19	4	2

№ 270	(0, 2)	(2, 4)	(4, 6)	(6, 8)	(8, 10)	(10, 12)
	1	5	18	19	4	3

№ 271	(2, 4)	(4, 6)	(6, 8)	(8, 10)	(10, 12)	(12, 14)
	1	6	17	18	7	1

№ 272	(0, 2)	(2, 4)	(4, 6)	(6, 8)	(8, 10)	(10, 12)
	3	5	16	17	6	3

№ 273	(3, 5)	(5, 7)	(7, 9)	(9, 11)	(11, 13)	(13, 15)
	3	8	15	16	7	1

№ 274	(–5, –3)	(–3, –1)	(–1, 1)	(1, 3)	(3, 5)	(5, 7)
	3	6	17	18	4	2

№ 275	(4, 6)	(6, 8)	(8, 10)	(10, 12)	(12, 14)	(14, 16)
	1	3	19	21	4	2

11. **Проверка статистических гипотез. Критерий согласия Пирсона**

Проверить соответствие данных выборки нормальному закону распределения. Ниже перечислены данные для проведения статистической обработки.

№ 276	26,0	31,3	33,6	34,6	37,1	38,1	39,5	41,4	43,2	47,2
	28,0	31,5	33,8	34,8	37,3	38,1	39,7	41,7	43,5	47,8
	28,6	31,8	33,8	34,9	37,4	38,3	39,9	42,0	43,7	48,6
	29,2	32,0	34,0	35,1	37,5	38,4	40,1	42,3	43,9	50,2
	29,7	32,3	34,0	35,3	37,6	38,6	40,2	42,6	44,2	51,0

№ 277	86,0	95,0	99,6	102,4	104,8	107,8	110,6	114,2	119,5	128,6
	87,0	95,5	100,2	102,5	105,1	108,1	110,9	114,5	120,5	130,0
	88,0	96,0	100,4	102,8	105,3	108,4	111,2	114,8	121,0	131,5
	90,5	96,5	100,6	103,0	105,5	108,5	111,5	115,1	121,5	133,5
	91,0	96,8	100,8	103,2	105,7	108,7	111,8	115,4	122,0	136,0

№ 278	28	32,5	34,5	36,4	37,1	38,5	39,6	40,8	42,0	44,0
	28,5	32,9	34,5	36,4	37,1	38,8	39,6	40,8	42,4	44,4
	29,0	33,2	35,0	36,4	37,1	38,8	40,0	40,8	42,5	44,5
	30,0	33,2	35,0	36,4	37,5	38,8	40,0	41,2	42,8	45,2
	31,0	33,5	35,3	36,8	37,5	38,8	40,0	41,2	42,8	48,0

№ 279	82,0	93,0	98,0	101,6	104,0	106,6	109,5	113,0	117,0	125,0
	83,0	93,5	98,4	101,8	104,2	106,9	109,7	113,3	117,5	126,0
	84,0	94,0	98,8	102,0	104,4	107,2	109,9	113,6	118,0	126,5
	85,0	94,5	99,2	102,2	104,6	107,5	110,3	113,9	119,0	127,4
	86,0	95,0	99,6	102,4	104,8	107,8	110,6	114,2	119,5	132,0

№ 280	27,0	29,4	30,8	31,8	32,2	33,1	33,8	34,2	35,5	37,5
	27,4	29,4	31,1	31,8	32,2	33,1	33,8	34,4	35,8	37,8
	27,7	29,6	31,1	32,0	32,5	33,1	33,8	34,4	35,8	38,4
	28,0	29,6	31,3	32,0	32,5	33,3	34,0	34,6	36,2	40,4
	28,0	29,8	31,3	32,0	32,5	33,3	34,0	34,6	36,2	42,0

№ 281	80,0	96,0	100,4	102,8	105,3	108,4	111,2	114,8	121,0	131,5
	86,5	96,5	100,6	103,0	105,5	108,5	111,5	115,1	121,5	133,5
	91,0	96,8	100,8	103,2	105,7	108,7	111,8	115,4	122,0	135,3
	91,5	97,2	101,0	103,4	106,0	108,9	112,1	115,7	123,0	137,6
	92,0	97,5	101,2	103,6	106,3	109,1	112,4	116,0	123,5	140,0

№ 282	90,0	99,0	105,0	110,3	112,6	116,2	119,5	125,0	131,5	140,5
	91,5	100,3	105,5	110,4	112,9	116,5	119,8	125,5	132,5	142,0
	92,0	100,5	106,0	110,6	113,5	116,8	120,4	126,0	133,5	144,0
	92,5	101,0	106,5	110,8	113,8	117,1	121,0	126,5	134,5	147,0
	93,0	101,5	107,0	111,0	114,1	117,4	121,5	127,0	135,0	150,0

№ 283	80,0	93,5	98,4	101,8	104,2	106,9	109,7	113,3	117,5	126,0
	82,5	94,0	98,8	102,0	104,4	107,2	109,9	113,6	118,0	126,5
	85,0	94,5	99,2	102,2	104,6	107,5	110,3	113,9	119,0	127,4
	86,0	95,0	99,6	102,4	104,8	107,8	110,6	114,2	119,5	128,6
	87,0	95,5	100,2	102,5	105,1	108,1	110,9	114,5	120,5	130,0

№ 284	93,0	97,7	100,9	102,0	103,3	104,8	107,4	111,0
	94,0	98,0	101,0	102,2	103,4	105,2	107,8	111,5
	95,3	98,3	101,2	102,3	103,5	105,4	108,2	112,2
	95,6	99,0	101,3	102,4	103,6	105,6	108,5	113,0
	95,9	99,5	101,4	102,5	103,7	105,8	108,8	113,0

№ 285	-20,0	-9,0	-3,0	0,4	2,4	4,6	6,8	9,1	12,6	18,0
	-18,0	-8,5	-2,5	0,6	2,6	4,9	7,0	9,5	12,9	18,5
	-17,0	-7,5	-2,2	0,8	2,8	5,2	7,2	10,3	13,3	19,0
	-16,0	-7,0	-2,1	1,0	3,0	5,4	7,4	10,6	13,8	19,5
	-14,0	-6,0	-2,0	1,2	3,2	5,6	7,6	10,9	14,5	20,0

№ 286	0,50	0,70	0,87	1,03	1,15	1,28	1,46	1,64	1,82	2,01
	0,54	0,71	0,88	1,05	1,16	1,30	1,48	1,65	1,83	2,05
	0,58	0,73	0,90	1,07	1,17	1,31	1,50	1,67	1,85	2,09
	0,61	0,74	0,92	1,08	1,18	1,32	1,52	1,68	1,87	2,17
	0,63	0,76	0,93	1,09	1,21	1,33	1,54	1,69	1,89	2,50

№ 287	90,0	96,8	100,5	101,7	102,9	104,1	106,6	109,6
	91,0	97,1	100,6	101,8	103,0	104,3	106,8	109,8
	92,0	97,4	100,8	101,9	103,2	104,5	107,1	110,5
	96,2	100,2	101,5	102,6	103,8	106,2	109,2	114,5
	96,5	100,3	101,6	102,8	103,9	106,4	109,4	115,0

№ 288	1,2	6,0	8,0	9,6	10,9	11,9	13,5	14,7	16,4	18,7
	1,5	6,2	8,2	9,8	11,0	12,0	13,6	14,9	16,6	19,0
	2,0	6,4	8,4	10,1	11,1	12,2	13,7	15,2	16,8	19,1
	2,5	6,6	8,6	10,2	11,2	12,3	13,9	15,3	17,0	19,3
	3,5	6,8	8,7	10,3	11,3	12,5	14,0	15,4	17,1	21,2

№ 289	-13,0	-5,5	-1,8	1,4	3,4	5,8	7,8	11,2	15,5	22,0
	-12,0	-4,5	-1,6	1,6	3,6	6,0	8,0	11,5	16,0	23,0
	-11,0	-4,0	-1,4	1,8	3,8	6,2	8,2	11,8	16,5	24,0
	-10,5	-3,5	-0,8	2,0	4,0	6,4	8,4	12,0	17,0	26,4
	-9,5	-3,5	0,2	2,2	4,3	6,6	8,7	12,3	17,5	27,0

№ 290	0,20	0,65	0,79	0,94	1,10	1,22	1,35	1,56	1,71	1,91
	0,25	0,66	0,81	0,96	1,11	1,23	1,37	1,58	1,72	1,93
	0,32	0,67	0,83	0,98	1,12	1,24	1.41	1,60	1,73	1,95
	0,41	0,68	0,85	0,99	1,13	1,25	1,43	1,61	1,76	1,97
	0,45	0,69	0,86	1,01	1,14	1,26	1,45	1,63	1,79	2,20

№ 291	4,0	7,0	8,8	10,4	11,4	12,7	14,1	15,6	17,3	19,9
	5,2	7,2	8,9	10,5	11,5	12,8	14,3	15,8	17,5	21,0
	5,4	7,4	9,0	10,6	11,6	13,0	14,4	16,0	17,8	22,0
	5,6	7,6	9,2	10,7	11,7	13,2	14,5	16,1	18,1	23,0
	5,8	7,8	9,4	10,8	11,8	13,4	14,6	16,2	18,4	24,0

№ 292	32,0	40,2	42,5	44,2	45,6	47,7	48,6	50,3	52,8	54,6
	34,5	40,4	42,7	44,3	45,9	47,7	48,8	50,4	52,9	55,0
	35,8	40,6	42,9	44,4	46,2	47,8	48,9	50,6	53,0	55,5
	36,6	40,8	43,1	44,4	46,5	47,8	49,1	50,8	53,1	56,0
	37,0	41,1	43,4	44,6	46,8	47,9	49,3	51,1	53,1	57,0

№ 293	5,0	7,3	9,0	10,4	11,8	13,4	15,1	16,8	18,6	20,9
	5,7	7,4	9,2	10,6	11,9	13,5	15,2	17,0	18,8	22,2
	6,2	7,6	9,3	10,8	12,2	13,6	15,3	17,1	19,0	23,4
	6,4	7,8	9,4	10,9	12,4	13,8	15,5	17,2	19,2	24,7
	6,6	7,9	9,6	11,0	12,6	13,9	15,7	17,4	19,4	25,0

№ 294	10,00	13,20	14,50	15,35	15,73	16,15	16,57	16,87	17,90	19,20
	11,20	13,40	14,60	15,40	15,76	16,20	16,60	16,91	18,00	19,40
	11,40	13,50	14,70	15,50	15,78	16,25	16,63	16,95	18,10	19,60
	11,60	13,60	14,80	15,52	15,80	16,30	16,66	17,20	18,20	19,80
	11,80	13,70	15,05	15,54	15,83	16,35	16,70	17,30	18,30	20,00

№ 295	1,20	5,10	7,10	9,10	10,05	10,70	11,35	12,05	13,30	16,30
	1,50	5,30	7,30	9,20	10,10	10,80	11,40	12,10	13,60	17,70
	2,00	5,50	7,50	9,30	10,15	10,85	11,45	12,20	13,90	18,40
	2,50	5,70	7,70	9,40	10,20	10,90	11,50	12,30	14,20	19,80
	3,00	5,90	7,90	9,50	10,25	10,95	11,60	12,40	14,50	21,20

№ 296	7,5	8,2	8,7	9,1	9,3	9,4	9,7	9,8	10,1	10,5
	7,6	8.3	8,8	9,1	9,3	9,5	9,7	9,9	10,1	11,6
	7,7	8,4	8,8	9,1	9,3	9,5	9,7	9,9	10,2	11,7
	7,8	8,5	8,8	9,2	9,3	9,5	9,7	9,9	10,2	12,5
	7,9	8,6	8,9	9,2	9,4	9,5	9,8	9,9	10,3	12,5

№ 297	36,0	41,3	43,6	44,6	47,1	48,1	49,5	51,4	53,2	57,2
	38,0	41,5	43,8	44,8	47,3	48,1	49,7	51,7	53,5	57,8
	38,6	41,8	43,8	44,9	47,4	48,3	49,9	52,0	53,7	58,6
	39,2	42,0	44,0	45,1	47,5	48,4	50,1	52,3	53,9	60,2
	39,7	42,3	44,0	45,3	47,6	48,6	50,2	52,6	54,2	61,0

№ 298	4,0	6,8	8,2	9,8	11,1	12,8	14,2	15,8	17,6	18,6
	4,2	6,9	8,4	9,9	11,2	12,9	14,4	15,9	17,7	19,2
	4,5	7,0	8,6	10,1	11,4	13,0	14,6	16,2	17,9	19,5
	4,8	7,1	8,8	10,2	11,6	13,1	14,8	16,4	18,0	20,0
	5,1	7,2	8,9	10,3	11,7	13,2	15,0	16,6	18,4	20,0

№ 299	5,0	8,8	11,0	12,2	13,2	14,4	15,7	17,4	19,6	24,0
	5,5	9,0	11,1	12,3	13,3	14,6	15,8	17,6	19,9	24,5
	6,0	9,3	11,2	12,4	13,4	14,8	15,9	17,8	20,2	25,0
	6,5	9,5	11,3	12,5	13,5	15,0	16,0	18,0	20,5	26,0
	6,9	9,8	11,4	12,6	13,6	15,1	16,2	18,2	20,8	27,0

№ 300	45,0	56,5	62,3	64,8	66,8	68,3	69,3	70,8	72,5	77,0
	49,0	57,5	62,6	65,0	67,0	68,4	69,4	71,0	72,7	77,5
	51,0	58,5	62,9	65,1	67,2	68,5	69,5	71.2	72,9	78,5
	51,5	59,0	63,2	65,2	67,4	68,6	69,6	71,4	73,3	79,0
	52,5	59,5	63,5	65,4	67,6	68,7	69,7	71,6	73,7	80,0

ЗАКЛЮЧЕНИЕ

Теория вероятностей и математическая статистика обычно завершают курс математики в ВУЗах. В зависимости от уровня овладения предметом Вам, возможно, станет интересно пересчитывать вероятности событий, которые встречаются в повседневной жизни – вероятность выигрыша в спортлото, вероятность получения на экзамене счастливого билета и т. д. Есть несколько уровней овладения любым предметом. Первоначальный, он же низший уровень, который позволит решать примеры по образу и подобию, подставив в образец решения свои числовые данные. Средний, к которому стремится студент, является достаточным для успешной сдачи экзамена или зачета, он требует умения самостоятельно решать стандартные примеры. И, наконец, самый высший уровень – он позволит применять полученные знания на практике в самых различных сферах окружающего нас мира, самостоятельно подбирать необходимую в вычислениях формулу. Именно на этом уровне Вам захочется самостоятельно, просто для себя, пересчитать вероятности происходящих вокруг Вас событий. Теория вероятностей и математическая статистика дают широкий простор в силу простоты математического аппарата.

Более подробное изложение материала можно найти в книгах, учебных пособиях указанных ниже в библиографическом списке.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Гмурман, В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. – М.: Высшее образование, 2008. – 478 с.
2. Карасев, А.И. Теория вероятностей и математическая статистика / А.И. Карасев. – М.: Статистика, 1979. – 279 с.
3. Вентцель, Е.С. Теория вероятностей и ее инженерные приложения / Е.С. Вентцель. – М.: Высшая школа, 2000. – 479 с.
4. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. – М.: Высшее образование, 2008. – 403 с.
5. Кущев, А.Б. Теория вероятностей и математическая статистика. ВГАСА / А.Б. Кущев, В.Г. Король. / Воронеж: ВГАСА, 1995. – 44 с.
6. Демин, С.Е. Математическая статистика. М-во образования и науки РФ; ФГАОУ ВО «УрФУ им. первого Президента России Б.Н. Ельцина», Нижнетагил. технол. ин-т (фил.). / Демин С.Е., Демина Е.Л. / Нижний Тагил : НТИ (филиал) УРФУ, 2016. – 284 с.
7. Беликов, А.Б. Математическая обработка результатов геодезических измерений : учебное пособие / А.Б. Беликов, В.В. Симонян / М-во образования и науки Рос. Федерации, Нац. исследоват. Моск. гос. строит. ун-т. Москва : НИУ МГСУ, 2016. – 432 с

ОГЛАВЛЕНИЕ

Учебное пособие

Акчурина Людмила Васильевна
Кущев Анатолий Борисович
Сумера Светлана Сергеевна

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Учебное пособие

Редактор

Подписано в печать _______.2022. Формат 60x84
1/16. Бумага для множительных аппаратов
Уч.-изд.л.____ . Тираж 100 экз. Заказ №______.

ФГБОУ ВО «Воронежский государственный технический университет»
394026 Воронеж, Московский проспект, 14

Участок оперативной полиграфии издательства ВГТУ
394026 Воронеж, Московский проспект, 14