МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра биомедицинской инженерии

УТВЕРЖДАЮ
Проректор по учебной работе
______________О.Г. Локтионова
«_25_» _08_________2023 г.

МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МЕДИКО-БИОЛОГИЧЕСКИХ ДАННЫХ

Методические указания по выполнению практических занятий
для студентов специальности
30.05.03 «Медицинская кибернетика»

Курск 2023

УДК 004.93:61

Составитель: О.В. Шаталова.

Рецензент
Кандидат технических наук, доцент *Ю.А. Халин*

Методы статистической обработки медико-биологических данных: методические указания по выполнению практических занятий для студентов специальности 30.05.03 «Медицинская кибернетика» / Юго-Зап. гос. ун-т; сост.: О.В. Шаталова. - Курск, 2023. - 264 с.

Предназначено для студентов специальности 30.05.03 «Медицинская кибернетика» по дисциплине «Методы статистической обработки медико-биологических данных». Может быть использована аспирантами, обучающимися по направлению подготовки 1.5.8. Математическая биология, биоинформатика.

Текст печатается в авторской редакции

Подписано в печать . Формат 60×84 1/16.
Усл. печ. л. 15,35. Уч.-изд. л. 13,89. Тираж 100 экз. Заказ 984. Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

3 семестр

Практическое занятие №1
Моделирование случайных чисел с заданным законом распределения

Целью занятия является 1) практическое ознакомление с алгоритмами моделирования случайных чисел с заданным законом распределения; 2) изучение основных способов статистической оценки характеристик случайных чисел.

Краткие теоретические сведения

Дискретные случайные величины

Слова "случайная величина" в обыденном смысле употребляют тогда, когда хотят подчеркнуть, что неизвестно, каким будет конкретное значение этой величины. Причем иногда за этими словами скрывается просто незнание, какова эта величина.

Математик употребляет эти же слова "случайная величина", вкладывая в них определенное содержание.

«Действительно, - говорит он, - мы не знаем, какое значение примет эта величина в данном конкретном случае, но мы знаем, какие значения она может принимать, и знаем, каковы вероятности тех или иных значений. На основании этих данных мы не можем точно предсказать результат одного испытания, связанного с этой случайной величиной, но можем весьма надежно предсказать совокупность результатов большого числа испытаний. Чем больше испытаний, тем точнее будут наши предсказания».

Итак, чтобы задать случайную величину, надо указать, какие значения она может принимать, и каковы вероятности этих значений.

Случайная величина X называется дискретной, если она может принимать дискретное множество значений $x_1, x_2, ..., x_n$.

Формально случайная дискретная величина X определяется таблицей

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}. \qquad (1.1)$$

где $x_1, x_2, \dots, x_n$ - возможные значения величины X;

$p_1, p_2, \dots, p_n$ - соответствующие вероятности.

Точнее говоря, вероятность $P\{X = x_i\}$ того, что случайная величина X примет значение x_i, равна:

$$P\{X = x_i\} = p_i. \qquad (1.2)$$

Таблица (1) называется распределением случайной дискретной величины.

Числа $x_1, x_2, \dots, x_n$ могут быть вообще говоря, любыми. Однако вероятности $p_1, p_2, \dots, p_n$ должны удовлетворять двум условиям:

$$p_i > 0 \qquad (1.3)$$

и

$$p_1 + p_2 + \dots + p_n = 1. \qquad (1.4)$$

Последнее условие означает, что X обязана в каждом случае принять одно из значений $x_1, x_2, \dots, x_n$.

Кроме распределения случайной величины, которая является исчерпывающей характеристикой, вводятся числовые характеристики, основными среди которых являются математическое ожидание и дисперсия.

Получение случайных величин на ЭВМ

Сама постановка вопроса "получение случайных чисел на ЭВМ" иногда вызывает недоумение: ведь все, что делает компьютер, должно быть заранее запрограммировано; откуда же может появиться случайность?

Специалисты считают, что в этом вопросе есть определенные трудности, но они относятся скорее к философии, так что мы на них останавливаться не будем. Отметим лишь, что случайные величины, о которых шла речь в предыдущем разделе это идеальные математические понятия.

Вопрос о том, можно ли с их помощью описать какое-либо явление природы, решается опытным путем. Такое описание всегда является приближенным. Более того, случайная величина, которая вполне удовлетворительно описывает какую-то физическую величину в одном классе явлений, может оказаться плохой характеристикой этой же величины при исследовании других явлений. Точно так же дорога, которую на карте страны можно считать прямой (идеальной математической прямой "без ширины"), становится полосой с изгибами на крупномасштабном плане населенного пункта.

Обычно различают три способа получения случайных величин:

- из заранее составленных таблиц случайных чисел;
- физические генераторы случайных чисел;
- с помощью формул (генераторов или датчиков) псевдослучайных чисел.

Поскольку "качество" используемых в имитационном моделировании случайных чисел проверяется с помощью специальных тестов, можно не интересоваться тем, как эти числа получены: лишь бы они удовлетворяли принятой системе тестов.

Числа, получаемые по какой-либо формуле и имитирующие значения случайной величины X, называются псевдослучайными числами. Под словом "имитирующие" подразумевается, что эти числа удовлетворяют ряду тестов так, как если бы они были значениями этой случайной величины.

Основой или «сырьем» для моделирования случайных величин с заданным законом распределения являются так называемые базовые случайные числа. Совокупность $\{R_i\}, i = 1, 2, ...$ независимых равномерно распределенных на отрезке [0, 1] случайных величин называется последовательностью базовых случайных чисел.

Мы называем эти числа псевдослучайными потому, что фактически они остаются полностью детерминированными в том смысле, что если каждое обращение к соответствующей формуле (точнее, к алгоритму) начинается с одними и теми же исходными данными (константами и начальными значениями), то на выходе получаются одинаковые последовательности чисел R.

В настоящее время почти все стандартные библиотечные программы вычисления равномерных случайных чисел основаны на конгруэнтных методах, разработанных Лемером.

Основная формула мультипликативного конгруэнтного метода Лемера имеет вид:

$$R_{i+1} = aR_i (\mathrm{mod}\, m), \tag{1.5}$$

где a и m – неотрицательные целые числа.

Согласно этому выражении, нужно взять случайное число R_i, умножить его на постоянный коэффициент a и взять модуль полученного числа m (т.е. разделить на aR_i и остаток считать как R_{i+1}). Поэтому для вычисления (или генерирования) последовательности R_i нам необходимы начальные значения R_0, множитель a и модуль m. Выбираются a, R_0 и m так, чтобы обеспечить максимальную длину (или, как говорят, период) неповторяющейся последовательности R_i и минимальную корреляцию между генерируемыми числами.

На рисунке 1.1 показан фрагмент среды MathCad, на котором проиллюстрирована математическая реализация этого метода.

Переменной A присваивается значение $a = 5^{13} = 1220703125$, переменной m – значение $m = 2^{31} + 1 = 2147483649$. Функция $\mathrm{mod}(x_1, x_2)$ вычисляет остаток от целочисленного деления первого аргумента во второй. Получаем последовательность {X} псевдослучайных чисел, равномерно распределенных от 0 до m. Делим каждый член этой последовательности на m, получаем базовую последовательность { R_i } – числа равномерно распределенные от 0 до 1.

Методы генерации псевдослучайных чисел с заданным законом распределения

Базовые случайные числа позволяют генерировать новые случайные последовательности, подчиняющиеся любому закону распределения.

$N := 50 \qquad n := 1..N$

$A := 5^{13} \qquad m := 2^{31} + 1$

$A = 1.220703125 \cdot 10^{9} \qquad m = 2.147483649 \cdot 10^{9} \qquad X_0 := 111111111$

$X_n := \mathrm{mod}(A \cdot X_{n-1}, m) \qquad R_0 := \dfrac{X_0}{m} \qquad R_n := \dfrac{X_n}{m}$

$X =$

	0
0	$1.11111111 \cdot 10^{8}$
1	$4.23519126 \cdot 10^{8}$
2	$1.365173986 \cdot 10^{9}$
3	$1.459022034 \cdot 10^{9}$
4	$1.442895947 \cdot 10^{9}$
5	$1.06708695 \cdot 10^{9}$
6	$6.7659776 \cdot 10^{8}$
7	$1.160823598 \cdot 10^{9}$
8	$1.586936277 \cdot 10^{9}$
9	$1.222270366 \cdot 10^{9}$

$R =$

	0
0	0.052
1	0.197
2	0.636
3	0.679
4	0.672
5	0.497
6	0.315
7	0.541
8	0.739
9	0.569
10	0.433
11	0.298
12	0.95
13	0.128

Рисунок 1.1 – Моделирование базовой последовательности мультипликативным конгруэнтным методом. Фрагмент среды MathCad

Существует два основных пути преобразования базовых случайных чисел $\{R_i\}$, в случайные числа $\{y_i\}$, распределенные по заданному закону распределения.

Один из них, который называется методом инверсии, состоит в реализации определенных арифметических операций над базовым числом R_i, чтобы получить y_i.

Второй метод основывается на моделировании условий соответствующей предельной теоремы теории вероятностей. Кроме указанных двух основных подходов можно также выделить эвристические способы генерирования случайных чисел.

Метод инверсии

Моделирование случайной величины, равномерной на (a, b)

Предположим, что нам необходимо составить программу для моделирования входного потока заявок распределенного по равномерному закону в интервале (a, b).

Уравнение метода инверсии (1.6) для рассматриваемого случая выглядит так:

$$\int_a^y \frac{dy}{b-a} = R, \tag{1.6}$$

где R – равномерно распределенное случайное число на (0; 1), т.е. базовое число. Это интегральное уравнение решается легко и ответ ясен:

$$\frac{y-a}{b-a} = R. \tag{1.7}$$

Отсюда мы имеем явное выражение для y:

$$y = a + R(b-a), \tag{1.8}$$

где R – как обычно, базовое случайное число.

Моделирование экспоненциальной случайной величины

Как известно, случайная величина x, распределенная по экспоненциальному закону описывается следующей плотностью распределения:

$$p(x) = \lambda e^{-\lambda x} \tag{1.9}$$

На рисунке 1.2 построены графики экспоненциальных плотностей распределения при различных параметрах λ.

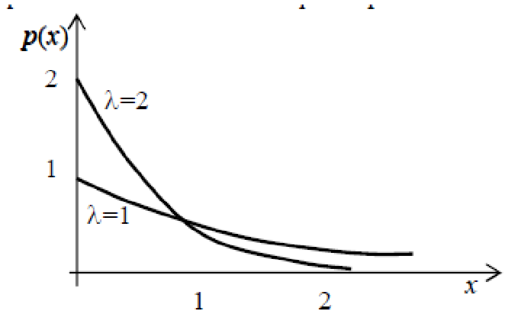


Рисунок 1.2 – Экспоненциальная плотность вероятностей $p(x) = \lambda e^{-\lambda x}$ с разными значениями параметра λ

Экспоненциальному распределению, как правило, подчиняется случайный интервал времени τ между поступлениями заявок в систему массового обслуживания. Поэтому весьма важно уметь моделировать потоки заявок разной интенсивности λ.

Напомним, что математическое $M[\tau]$ ожидание экспоненциально распределенной случайной величины τ равно:

$$M[\tau] = 1/\lambda,$$

а дисперсия: $D[\tau] = 1/\lambda^2$.

Чтобы найти алгоритм имитации экспоненциально распределенных чисел τ, применим метод инверсии:

$$\int_0^\tau \lambda e^{-\lambda x} = R \tag{1.10}$$

$$1 - e^{-\lambda\tau} = R, \tag{1.11}$$

откуда

$$\tau = -\frac{1}{\lambda}\ln(1 - R), \tag{1.12}$$

но, поскольку случайная величина (1 - R) распределена точно так же, как R, и находится в том же интервале (0,1), то (1.12) можно заменить на более удобную формулу:

$$\tau = -\frac{1}{\lambda}\ln R, \tag{1.13}$$

что дает искомый ответ.

Моделирование нормальной случайной величины на основе центральной предельной теоремы

Нормальное (или гауссово) распределение (рисунок 1.3) - это, несомненно, один из наиболее важных и часто используемых в имитационном моделировании видов непрерывных распределений.

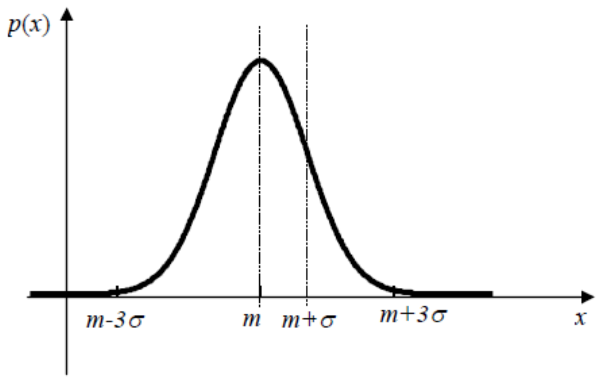


Рисунок 1.3 – Нормальная (гауссовская) плотность вероятностей

Плотность вероятности нормально распределенной случайной величины записывается так:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{1.14}$$

где m и σ - параметры нормального распределения $m = M_x$ - математическое ожидание; σ - среднеквадратическое отклонение.

Интегральная функция распределения нормальной случайной величины равна

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\sigma^2}} dx. \qquad (1.15)$$

Поэтому алгоритмы моделирования нормальных случайных чисел базируются на предельных теоремах теории вероятностей. Центральная предельная теорема говорит о том, что сумма n одинаково распределенных независимых случайных величин x со средним M_x и дисперсией D_x стремится к нормально распределенной величине с параметрами nM_x и nD_x при бесконечном увеличении n. Следствием теоремы является, в частности, и то, что для получения нормальной выборки, можно воспользоваться базовыми случайными числами R. Идея алгоритма состоит в следующем. Определим новую случайную величину s в виде суммы базовых чисел R_i, (i=1, 2, 3, …, n):

$$s = R_1 + R_2 + ... + R_n. \qquad (1.16)$$

Тогда, согласно утверждению центральной предельной теоремы, случайная величина s является асимптотически нормальной величиной с математическим ожиданием M_s и дисперсией D_s равными соответственно:

$$M_s = n / 2, \qquad (1.17)$$

и

$$D_s = n / 12. \qquad (1.18)$$

Для практического использования формула (1.16) неудобна (поясните почему), поэтому введем вспомогательную случайную величину z равную

$$z = \frac{(s - n/2)}{\sqrt{n/12}} \qquad (1.19)$$

Из (1.19) следует, что z – случайная величина, распределенная по нормальному закону с нулевым средним и единичной дисперсией. Тогда для любого нормального распределения со средним μ и дисперсией σ^2 случайное отклонение y, соответствующее указанным выше n случайным числам, получается из формулы

$$\frac{(y - \mu)}{\sigma} = z = \frac{\left(s - \frac{n}{2}\right)}{\sqrt{n/12}} \qquad (1.20)$$

Следовательно,

$$y = \mu + \frac{\sigma(s + n/2)}{\sqrt{n/12}} = \mu + \frac{\sigma}{\sqrt{n/12}}\left(\sum_{i=1}^{n} R_i + n/2\right). \qquad (1.21)$$

Согласно той же предельной теореме, нормальность достигается быстро даже при сравнительно небольших значениях n. В практических задач n обычно принимается равным 12. При этом последняя формула упрощается и принимает вид:

$$y = \mu + \sigma\left(\sum_{i=1}^{12} R_i + 6\right). \qquad (1.22)$$

Формула (1.22) и дает алгоритм моделирования нормальных случайных чисел с требуемыми параметрами μ и σ.

Описанный метод считается малоэффективным, так как требует генерации нескольких случайных базовых чисел R для получения одного нормального выборочного значения y.

Оценка статистических характеристик случайных величин

При решении многих прикладных задач необходимые вероятностные характеристики соответствующих случайных величин неизвестны исследователю и должны определяться по экспериментальным данным. Такое статистическое описание результатов наблюдений, построение и проверка различных математических моделей, использующих понятие вероятности, составляют основное содержание математической статистики. Фундаментальными понятиями статистической теории являются понятия генеральной совокупности и выборки.

Генеральная совокупность - совокупность всех мыслимых (возможных) результатов наблюдений над случайной величиной, которые в принципе могут быть проведены при данных условиях.

Содержательный смысл этого понятия состоит в том, что предполагается существование некоторых вполне определенных свойств, неслучайных закономерностей, присущих данной совокупности. Эти свойства и должны быть определены исследователем. Фактически эти свойства являются объективным отображением вероятностных свойств изучаемого объекта, которые могут быть охарактеризованы с помощью соответствующих законов распределения вероятностей или связанных с ними числовых параметров. Как правило, считается, что указанные свойства не изменяются во времени.

Выборка - это конечный набор $x_1, x_2, ..., x_N$ значений случайной величины, полученный в результате наблюдений. Число элементов N выборки называется ее объемом или размером.

Заметим, что выборка может иметь и совпадающие значения x_i случайной величины X. Интуитивно понятно, что чем больше объем выборки, тем более точно она должна отражать статистические свойства случайной величины. Определение. Выборка называется репрезентативной (представительной), если она достаточно полно характеризует свойства генеральной совокупности.

Для обеспечения репрезентативности выборки чаще всего используют метод случайного выбора элементов. Предполагается, что при таком выборе каждая возможная выборка фиксированного объема имеет одну и ту же вероятность выбора, а последовательные наблюдения взаимно независимы.

Оцениванием в статистике называется указание приближенного значения интересующего нас параметра (или функции от некоторых параметров) на основе наблюдаемых (экспериментальных) данных, представленных в виде выборки ограниченного объема.

Оценка - это правило вычисления приближенного значения параметра (или функции от некоторых параметров) по наблюдаемым данным.

При многократном извлечении выборок одного и того же объема и последующем нахождении множества оценок одного и того же параметра получаются различные числовые значении этих оценок, изменяющиеся от одной выборки к другой случайным образом.

Иными словами, любая оценка произвольного параметра есть случайная величина. В этом состоит принципиальное отличие оценки от самого параметра.

Элементарные статистические процедуры

В случае гауссовского распределения для истинного математического ожидания m_x существует его оценка $\tilde{m}_x$, вычисляемая по выборке объема n случайной величины $X = (x_1, x_2, ..., x_n)$:

$$\tilde{m}_x = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{1.23}$$

Для истинной дисперсии D_x (характеристика рассеивания случайной величины око ее математического ожидания) ее оценка $\tilde{D}_x$ при известном математическом ожидании m_x вычисляется так:

$$\tilde{D}_x = \tilde{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_x)^2, \tag{1.24}$$

где σ_x является среднеквадратическим отклонением.

В случае неизвестного математического ожидания дисперсию $\tilde{D}_x$ нужно вычислять по формуле:

$$\tilde{D}_x = \tilde{\sigma}_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) \qquad (1.25)$$

Приведенные оценки являются несмещенными и асимптотически эффективными.

Не будем забывать о том, что оценки сами являются случайными величинами, а значит, обладают некоторым разбросом, который оценивается дисперсией. Дисперсии D {} вышеуказанных оценок соответственно таковы:

дисперсия оценки среднего:

$$D\{\tilde{m}_x\} = \sigma_x^2 / n; \qquad (1.26)$$

дисперсия оценки $\tilde{D}_x$ в случае известного математического ожидания:

$$D\{\tilde{D}_x\} = 2\sigma_x^4 / n; \qquad (1.27)$$

в случае, если не известно математическое ожидание:

$$D\{\tilde{D}_x\} = 2\sigma_x^4 /(n-1). \qquad (1.28)$$

После вычисления точечных оценок обычно переходят к построению вариационного ряда, диаграммы накопленных частот и гистограммы выборки.

Пусть имеется набор (выборка) экспериментальных данных $x_1, x_2, ..., x_n$. Вариационный ряд (или ряд распределения) $z_1, z_2, ..., z_n$ получают из исходных данных путем расположения x_m $(m = 1, 2, ..., n)$ в порядке возрастания от x_{min} до x_{max} так, чтобы $x_{min} = z_1 \le z_2 \le ... \le z_n = x_{max}$.

Диаграмма накопленных частот $P_n(x)$ является эмпирическим аналогом интегрального закона распределения P(x) и ее строят в соответствии с формулой

$$P_n(x) = \sum_{j=1}^{\mu_n(x)} \frac{1}{n} \qquad (1.29)$$

где $\mu_n(x)$ - число элементов в выборке, для которых значение $x_j < x$.

Практически это делается так. На оси абсцисс указывают значения наблюдений x_{min} (или z_1). Значение по оси ординат равно нулю левее точки x_{min}; в точке x_{min} и далее во всех других точках x_m диаграмма имеет скачок, равный 1/n. Если существует λ совпадающих значений x_m, то в этом месте на диаграмме происходит скачок, равный λ / n. Ясно, что для величин $x > x_{max}$ значение диаграммы накопленных частот равно 1. Отметим, что если $n \to \infty$, то $P_n(x) \to P(x)$.

Пример. Пусть имеется выборка объема 5:

$$x_1 = 5; x_2 = 2; x_3 = 4; x_4 = 5; x_5 = 7.$$

Вариационный ряд для данной выборки будет таким:

$$z_1 = 2; z_2 = 4; z_3 = 5; z_4 = 5; z_5 = 7.$$

Соответствующая диаграмма накопительных частот представлена на рисунке 1.4.

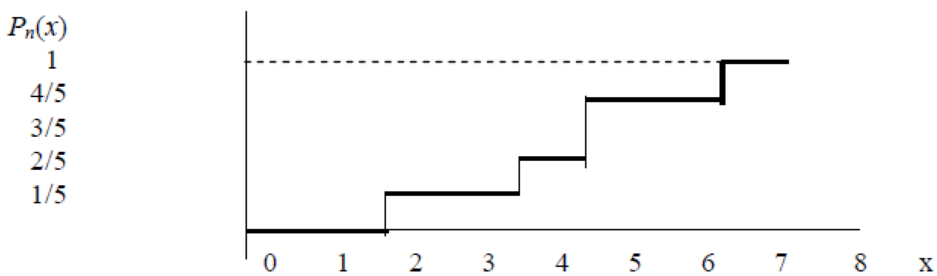


Рисунок 1.4 – Диаграмма накопленных частот

Гистограмма $f_n(x)$ является эмпирическим аналогом функции плотности распределения f(x). Последовательность построения гистограммы такова.

По оценочной формуле находят предварительное количество квантов (интервалов) K, на которое нужно разбить на ось Ox:

$$K=1+3.2\lg n;$$

найденное значение K округляется до ближайшего целого числа.

Формула для K является эмпирической, что означает примерное значение. Ее величина связана с той целью, чтобы в один квант попало хотя бы одно выборочное значение x_i. Интересно, что зависимость количества интервалов K от объема выборки n равна (таблица 1.1)

Таблица 1.1 - Зависимость количества интервалов K от объема выборки n

n	100	200	300	500	1000
K	7	8	9	10	11

Далее определяют длину каждого кванта (интервала):

$$\Delta x = (x_{max} - x_{min}) / K,$$

которую для удобства построений можно несколько округлить в ту или иную сторону.

Середину области изменения выборки (центр распределения)

$$(x_{max} + x_{min}) / 2$$

принимают за центр некоторого интервала, после чего находят границы и окончательное количество указанных интервалов так, чтобы в совокупности они перекрывали всю область от x_{min} и x_{max}.

Далее подсчитывают количество наблюдений n_m, попавшее в каждый квант: n_m равно числу членов вариационного ряда, для которых справедливо неравенство

$$x_m \le z_1 < x_m + \Delta x.$$

Здесь x_m и $x_m + \Delta x$ - границы m-го интервала. Отметим, что при использовании этой формулы значения z_1, попавшие на границу между (m-1) и m-м интервалами, относят к m-му интервалу.

Далее подсчитывают относительное количество (относительную частоту) наблюдений n_m / n, попавших в данных квант.

Наконец, строят гистограмму, представляющую собой ступенчатую кривую, значение которой на m-й интервале $(x_m, x_m + \Delta)$ (m=1, 2, …, K) постоянно и равно n_m / n, или с учетом условия $\int\limits_{-\infty}^{\infty} p_n(z)\,dz = 1$, равно $(n_m / n) \cdot \Delta x$.

Практическая часть

1. Смоделируйте базовую последовательность объемом N=1000 мультипликативным конгруэнтным методом.
2. Напишите одну комплексную программу моделирования выборки случайных чисел, оценки математического ожидания и дисперсии для всех ниже перечисленных распределений:
   a) равномерное на интервале (a, b);
   b) экспоненциальное с параметром λ;
   c) нормальное с параметрами μ и σ, используя метод суммирования или какой-либо один из эвристических методов.
3. Самостоятельно задав параметры распределений, промоделируйте выборки всех вышеуказанных распределений. Объем каждой выборки принять N=1000.
4. Вычислите оценки математического ожидания и дисперсии каждой из полученных в п. 2 последовательностей случайных чисел для следующих объемов выборки N_1 =10, N_2 =20, N_3 =50, N_4 =100 и N_5 =1000. Сравните полученные оценки с заданными в пп. 2 параметрами. Постройте графики зависимостей оценок от объема выборки. Оцените относительные погрешности для какой-либо одной выборки.
5. Для всех выборок разных распределений, рассчитайте и постройте:
   - диаграммы накопленных частот;
   - гистограммы распределений.

6. Сравните гистограммы с графиками теоретических распределений. Для сравнения постройте также вышеуказанные гистограммы на одном графике с функциями распределений.

Варианты практического занятия. Параметры распределений

№ вар	a	b	λ	μ	σ
1	3	11	1	0	1
2	10	16	5	0	2
3	1	6	1	0	3
4	4	7	2	0	4
5	9	17	4	0	5
6	1	11	1	0	6
7	8	13	4	1	1
8	3	9	1	1	2
9	2	7	1	1	3
10	9	14	5	1	4
11	6	11	3	1	5
12	3	7	2	1	6
13	7	12	3	2	1
14	5	12	3	2	2
15	7	13	4	2	3
16	3	12	1	2	4
17	2	12	2	2	5
18	4	13	2	10	6
19	4	8	5	10	1
20	10	14	4	10	2
21	9	16	1	10	3
22	0	7	3	10	4
23	6	6	5	10	5
24	10	11	3	10	6
25	6	11	4	5	1
26	9	18	2	5	2
27	4	7	1	5	3
28	2	6	2	5	4
29	4	8	5	5	5
30	9	13	4	5	6

Пример построения гистограммы и диаграммы накопленных частот в среде MathCad для базовой последовательности R. Оценки математического ожидания и дисперсии приведены ниже на рисунках.

Вычисляем количество квантов (интервалов), на которые разбиваем ось ох

$K := floor(1 + 3.2 \cdot \log(N))$ $\quad 1 + 3.2 \cdot \log(N) = 10.6$ $\quad K = 10$ $\quad k := 1..K$

границы области построения гистограммы $\quad max(R) = 0.999$ $\quad min(R) = 8.945 \cdot 10^{-4}$

ширина одного интервала $\quad \Delta x := \dfrac{max(R) - min(R)}{K}$

массив границ интервалов $\quad x_k := k \cdot \Delta x + min(R)$ $\quad x_0 := min(R)$

$G := hist(x, R)$

G =	x =
110	$8.945 \cdot 10^{-4}$
87	0.101
86	0.2
99	0.3
114	0.4
85	0.5
109	0.6
115	0.699
99	0.799
96	0.899
	0.999

Функция построения гистограммы. Вычисляет сколько значений массива R попадает в интервалы, определенные массивом x.
Возвращает массив, размерность которого на единицу меньше размерности массива x, поскольку количество интервалов на единицу меньше количества точек, ограничивающих эти интервалы

Рисунок 1.5 – Количество квантов, на которые разбивается ось OX

Теоретическое значение закона распределения (плотности вероятности)

$p(x) := if(x < 0, 0, if(x > 1, 0, 1))$ $\quad xx := -0.2, -0.19..1.2$

Теоретическое значение интегральной функции распределения

$P(x) := if(x < 0, 0, if(x > 1, 1, x))$

График, на котором построена теоретическая функция распределения, совмещенная с гистограмой. Гистограмма нормирована, относительная частота попадания в интервал делится на ширину интервала, тогда высота столбика может быть соотнесена со значением теоретической функци распределения в пределах каждого интервала.

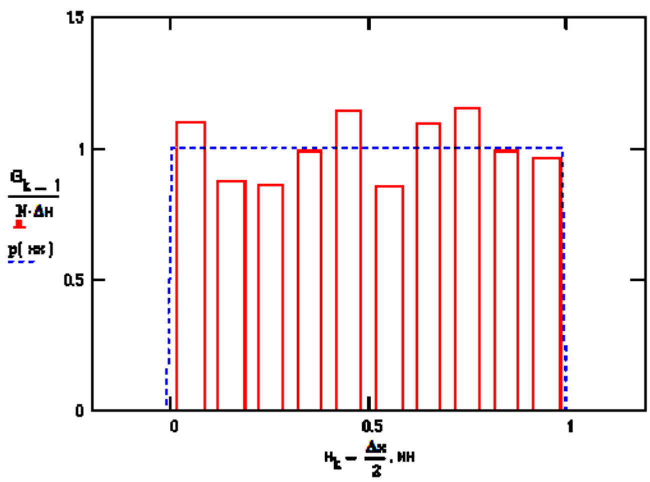


Рисунок 1.6 – Теоретическая функция распределения

$$Gn_k := \sum_{i=1}^{k} G_{i-1}$$

Вычисление массива значений для диаграммы накопленных частот

График, на котором диаграмма накопленных частот совмещена с интегральной функцией распределения

G =

110
87
86
99
114
85
109
115
99
96

Gn =

0
110
197
283
382
496
581
690
805
904
$1 \cdot 10^3$

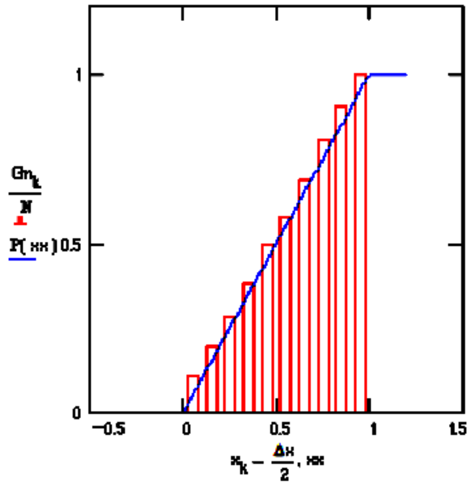


Рисунок 1.7 – Вычисление массива значений для диаграммы накопленных частот

оценка матожидания в зависимости от объема выборки

$$M(N,R) := \frac{1}{N} \cdot \sum_{i=1}^{N} R_i \qquad M(10,R) = 0.528 \qquad M(N,R) = 0.507 \qquad MR := M(N,R)$$

несмещенная оценка дисперсии в зависимости от объема выборки

$$D(N,R) := \frac{1}{N-1} \cdot \sum_{i=1}^{N} \left(R_i - MR\right)^2 \qquad m := 2..N$$

графики данных зависимостей

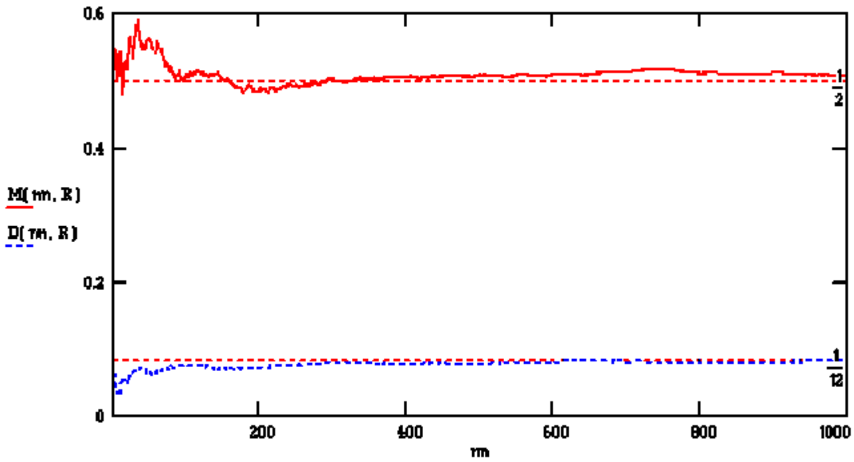


Рисунок 1.8 – Оценка математического ожидания в зависимости о объема выборки

Контрольные вопросы

1. Что такое распределение случайной дискретной величины?
2. Что такое дискретная случайная величина? Дайте развернутый ответ.
3. Каков алгоритм получения случайных величин на ЭВМ?
4. Раскройте понятие «методы генерации псевдослучайных чисел с заданным законом распределения».
5. Что такое метод инверсии?
6. Что значит «моделирование случайной величины равномерной на (a, b)»?
7. Что значит «моделирование экспоненциальной случайной величины»?
8. Что значит «моделирование нормальной случайной величины на основе центральной предельной теоремы»?
9. Что такое оценка статистических характеристик случайных величин?
10. Что такое элементарные статистические процедуры?

Практическое занятие №2
Элементарные задачи математической статистики

Цель занятия: Рассмотреть возможности MathCAD для решения элементарных задач математической статистики. Научиться использовать возможности MathCAD для ввода и вывода файловых данных. Познакомиться с расчетом основных выборочных характеристик в среде MathCAD. Научиться представлять графически выборку случайных величин в виде гистограмм и полигонов.

Краткие теоретические сведения

В большинстве статистических расчетов приходится иметь дело либо со случайными данными, полученными в ходе какого-либо эксперимента (которые выводятся из файла или печатаются непосредственно в документе), либо с результатами генерации случайных чисел.

Случайной выборкой называется случайный вектор, элементы которого независимы и одинаково распределены. Обычно под *выборкой* подразумевают результаты независимых измерений, которые проводятся в одинаковых условиях.

Ввод и вывод файлов данных

Важный компонент ввода-вывода — это ввод-вывод во внешние файлы. Ввод внешних данных в документы Mathcad применяется чаще вывода, поскольку Mathcad имеет гораздо лучшие возможности представления результатов расчетов, чем многие пользовательские программы. Для общения с внешними файлами данных в Mathcad имеется несколько разных способов.

Самый простой из них — использовать имеющееся семейство встроенных функций.

- READPRN (“file”) - чтение данных в матрицу из текстового файла;
- WRITEPRN (“file”) - запись данных в текстовый файл;

- APPENDPRN (“file”) - дозапись данных в существующий текстовый файл;
- file — путь к файлу.

Встроенная функция APPENDPRN может применяться и для создания нового файла. Иными словами, если файла с заданным именем не существовало, то он, после применения, будет создан и наполнен теми данными, которые Вами определены в документе.

Для удобства можно использовать функцию CWD - указания полигона, где необходимо создать файл или где находиться считываемый файл.

Можно задавать как полный путь к файлу, например, C:\Мои документы, так и относительный, имея в виду, что он будет отсчитываться от папки, в которой находится файл с документом Mathcad. В качестве имени файла можно использовать русские буквы.

Пример 1: Запись данных в файл "da.ta.txt"

x:=rpost(N, a, σ) создание выборки случайных величин распределенных по нормальном закону;

CWD:= "D:\tmp\" устанавливается текущий рабочий каталог.

WRITEPRN(data.txt") :=x запись в файл «data»созданной ранее выборки x.

Моделирование выборок из стандартных распределений

Mathcad обладает богатой библиотекой встроенных функций, предназначенных для генерации выборок из генеральных совокупностей с наиболее распространенными стандартными распределениями.

Вставку рассмотренных ранее статистических функций в программы удобно осуществлять с помощью диалогового окна *Insert Function* (Вставка функции).

Для этого необходимо выполнить следующие действия:

1. Установить курсор на место вставки функции в документе.

2. Вызвать диалоговое окно *Insert Functio*n нажатием кнопки f(x) на стандартной панели инструментов или командой меню *Insert/Function* (Вставка/Функция), или нажатием клавиш <Ctrl>+<E>.

3. Выбрать в списке *Function Category* (Категория функции) выберите одну из категорий статистических функций. Категория *Probability Density* (Плотность вероятности) содержит встроенные функции для плотности вероятности, Категория *Probability Distribution* (Функция распределения) — для вставки функций или квантилей распределения, Категория *Random Numbers* (Случайные числа) — для вставки функции генерации случайных чисел.

4. Выбрать в списке *Function Name* (Имя функции) функцию, соответствующую требующемуся закону распределения. При выборе элемента списка в текстовом поле в нижней части окна будет появляться информация о назначении выбранной функции и ее параметрах.

5. Вставить выбранную функцию в документ нажатием кнопки "Ок".

Функции Mathcad для расчета численных характеристик

В Mathcad имеется ряд встроенных функций для расчетов числовых статистических характеристик рядов случайных данных.

- mean(x) — выборочное среднее значение, оценка математического ожидания выборки;
- median (x) — выборочная медиана (median) — значение аргумента, которое делит гистограмму плотности вероятностей на две равные части;
- var(x) — выборочная дисперсия выборки (variance);
- stdev(x) — среднеквадратичное (или "стандартное") отклонение выборки (standard deviation);
- max(x), min(x) — максимальное и минимальное значения выборки;
- mode(x) — наиболее часто встречающееся значение выборки.

Построение гистограмм

Гистограммой называется график, аппроксимирующий по случайным данным плотность их распределения. При построении гистограммы область значений случайной величины (a, b) разбивается на некоторое количество *bin* сегментов, а затем подсчитывается процент попадания данных в каждый сегмент. Для построения гистограмм в Mathcad имеется несколько встроенных функций.

Гистограмма с произвольными сегментами разбиения

- hist,(intvls,x) - вектор частоты попадания данных в интервалы гистограммы;
- intvls - вектор, элементы которого задают сегменты построения гистограммы в порядке возрастания $a < \mathrm{int\,vis}_i < b$;
- x - вектор случайных данных.

Если вектор *intvls* имеет *bin* элементов, то и результат *hist* имеет столько же элементов.

1. Для построения гистограмм созданную случайную величину предварительно необходимо упорядочить. Для этого в Mathcad имеется встроенная функция.

sort(x) - сортировка выборки в порядке возрастания;

Для того, чтобы построить гистограмму, нужно сначала сгруппировать выборочные данные, записанные в массиве x, и сохранить граничные очки интервалов группировки в векторе *intvls, размерность которого равна числу интервалов.*

2. Сформировать вектор intvls границ интервалов.
3. Определить процент попадания данных в каждый сегмент.
4. Построить гистограмму.

<u>Пример 2.</u> Построение гистограммы

N:=1000

x:=binom (N, Δ, 0,5)

bin:=30; кол-во равных сегментов, на которые разбивается весь диапазон;

определение границы интервала построения гистограммы

lower:=floor (min(x)): наибольшее целое число $\le \min(x)$

upper:=ceil (max(x)); наименьшее целое число $\ge \max(x)$

$h := \frac{upper - lower}{bin}$; размер сегмента

j:=0…bin; счетчик сегментов

$int_j := lower + h \cdot j$

$f := \frac{1}{N \cdot h} \cdot hist(int, x)$; массив начальных точек каждого сегмента

int:=int+0.5h; от левой границы каждого сегмента к его центру; нормирование гистограммы для удобства отображения на одном графике вместе с плотностью распределения

В векторе *int* можно задать произвольные границы сегментов разбиения так, чтобы они имели разную ширину.

Недостаток упрощенной формы функции *hist* состоит в том, что необходимо дополнительно определять вектор сегментов построения гистограммы. От этого недостатка свободна функция *histogram.*

Гистограмма с разбиением на равные сегменты

histogram (bin, x) — матрица гистограммы размера bin*2, состоящая из столбца сегментов разбиения и столбца частоты попадания в них данных;

- bin — количество сегментов построения гистограммы;
- x — вектор случайных данных.

<u>Пример 3.</u> Построение гистограммы (упрощенный вариант)

N:=100; созданы выборки сл. величины

x:= exp (N, 1)

bin:=30; кол-во равных сегментов, на кот. разбивается весь диапазон

f := histogram(bin, x)

Аргументы у обоих процедур hist() и histogram() одинаковы: первый определяет интервалы для создания гистограммы, а второй - это выборка, на основе которой строится гистограмма. Первый аргумент может быть либо вектором конечных точек интервалов для группировки данных выборки, либо целым числом, задающим число интервалов. В последнем случаи весь диапазон значений в выборке разбивается на равные интервалы.

Создание графика гистограммы

Для того чтобы создать график в виде гистограммы необходимо:

1. Построить двумерный график, по оси x откладываются границы интегралов, по оси y частота(процент) попадания значения сл. величины в заданные интервалы.
2. Перейти в диалоговом окне *Formatting Currently Selected Graph* (Форматирование) выбранного графика (например, двойным щелчком мыши) в раздел *Traces* (Графики). Установить в поле *Type* (Тип) элемент списка *bar* (столбцы) или *solidbar* (гистограмма). Тип solidbar специально предназначен для гистограмм.
3. Нажать кнопку ОК.

Процедура *histogram()* инициализирует объект (матрицу), содержащий срединные точки интервалов гистограммы (первый столбец) и столбец частот, попадания в заданные интервалы.

Для построения таких графиков по оси x откладывается столбец срединные точки интервалов (столбец матрицы с нулевым индексом), а по оси y - столбец с частотами распределения данных по интервалам гистограммы (столбец с первым индексом). Индекс  столбца вводится с помощью соответствующей пиктограммы п панели Matrix или комбинации клавиш <Ctrl>+<6>.

Полигон частот

Иная форма графического представления группированных данных - полигон частот. *Полигон частот* - это ломанная линия, соединяющая точки с координатами ($\bar{x}_i, h_i$), т.е. с абсциссами, равными серединам интервалов группировки, и ординатами, равными соответствующим частотам. Если соединить центры элементарных сегментов гистограммы ломанной линией, то получится график полигона.

<u>Задание 1.</u> Создайте выборку из 100 случайных величин с нормальным распределением, среднее значение m=0,1*k (k - номер варианта) и стандартное отклонение σ =0,5. Запишите данную выборку в файл с произвольным именем. Рассчитайте с помощью встроенных функций MathCad числовые статистические характеристики созданной выборки. Постройте гистограмму двумя способами и полигон частот

Порядок выполнения анятия

1. С помощью встроенной функции из категории Random Numbers (Случайные числа) получить заданную выборку.
2. Записать полученную величину в файл с произвольным названием. В отчет вставьте фрагмент этого документа.
3. Упорядочить значения в выборке случайной величины по возрастанию.
4. С помощью стандартных функций Mathcad, получить числовые характеристики: min и max значения выборки, выборочное среднее, выборочную дисперсию, среднеквадратическое отклонение, выборочную медиану.
5. Используя функцию дозаписи, добавить в созданный ранее файл числовые характеристики выборки. Фрагмент вновь созданного файла привести в отчете.
6. Считать полученный файл.
7. Выполните расчет гистограммы с помощью функцию hist (int, x) Отобразить на графиках гистограмму и плотность распределения на одном и полигон частот на другом.
8. Выполните расчет гистограммы, используя функцию histogram (int, x).
9. Выведите на экран результаты процедур hist() и histogram(). Сравните их.
10. Понаблюдайте, как изменится внешний вид гистограммы, если изменить количество интервалов разбиения выборки. Сделать выводы.

В отчете представить все необходимые фрагменты, сделанные в Mathcad, и требуемые выводы.

Контрольные вопросы

1. Что такое случайная выборка?
2. Что такое выборка?
3. Как происходит ввод и вывод данных в MathCad?
4. Как в MathCad произвести моделирование выборок из стандартных распределений?
5. Что такое функция MathCad для расчета численных характеристик?
6. Как в MathCad построить гистограммы?
7. Что такое гистограмма с произвольным сегментом разбиения?
8. Что такое гистограмма с разбиением на равные сегменты?
9. Приведите алгоритм создания графика гистограммы.
10. Что такое полигон частот?

Практическое занятие №3
Числовые характеристики дискретных случайных величин

Цель занятия: изучить способы вычисления числовых характеристик случайных величин с использованием пакета MathCad.

Задание: решить представленную задачу.

Краткие теоретические сведения

Математическое ожидание

Математическим ожиданием дискретной случайной величины называется сумма произведений всех ее возможных значений на вероятности этих значений.

Если случайная величина принимает значения с разной вероятностью, математическое ожидание вычисляется по формуле

$$M(X) = \sum_{i=0}^{n} x_i \cdot p_i$$

Пример 1. Найти математическое ожидание дискретной случайной величины, закон распределения которой задан таблицей:

X	1	2	3	4	5
P	0,15	0,25	0,3	0,2	0,1

Зададим векторы

$$x := (1\,2\,3\,4\,5)^T \quad p := (0.15\,0.25\,0.3\,0.2\,0.1)^T$$

Найдем математическое ожидание

$$M := \sum_{i=0}^{last(x)} x_i \cdot p_i$$

$$M = 2.85$$

Если случайная величина принимает ряд значений с равной вероятностью, то математическое ожидание определяется как среднее арифметическое значение некоторого количественного признака выборки.

В MathCad среднее значение выборки можно подсчитать с помощью функции *mean(x)*.

Пример 2. При измерении величины силы тока были получены следующие значения: 0,45; 0,49; 0,44; 0,42; 0,48; 0,41; 0,44; 0,56; 0,47; 0,45; 0,52; 0,43. Вычислить выборочное среднее

$$X := (0.45\ 0.49\ 0.44\ 0.42\ 0.48\ 0.41\ 0.56\ 0.47\ 0.45\ 0.52\ 0.43)$$

$$\mathrm{mean}(X) = 0.463$$

При обработке экспериментальных данных среднее значение выборки считается равным значению параметра. Это утверждение верно только в том случае, если выборка является генеральной, т.е. содержит все возможные значения измеряемой величины. В реальной ситуации с генеральными совокупностями работать невозможно, а всегда приходится делать из них некоторые небольшие выборки. В зависимости от условий отбора и объема выборки она может передавать особенности генеральной совокупности с различной точностью. При этом такие характеристики, как среднее значение и дисперсия, приобретают случайный характер. Исследование особенностей поведения такого рода величин – очень сложная и важная статистическая задача.

Дисперсия и среднеквадратичное отклонение

В статистике дисперсией называется среднее арифметическое квадратов отклонений случайной величины от ее среднего значения:

$$D = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N}.$$

В общем случае дисперсия является характеристикой степени рассеяния значений выборки по сравнению с ее средней величиной.

В MathCad простая выборочная дисперсия вычисляется с помощью функции var(x). Кроме того, существует и функция Var(x), которая определяет исправленную дисперсию, котрая на практике используется для несмещенной оценки генеральной дисперсии при малом объеме выборки:

$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}.$$

На практике используют не саму дисперсию, а квадратный корень из нее, который называется среднеквадратичным отклонением. В MathCad существуют две функции для вычисления этого параметра: stdev(x) – выборочное стандартное отклонение и Stdev(x) – исправленное среднеквадратичное отклонение.

<u>Пример 3.</u> Подбрасывается игральный кубик. Случайная величина X – количество выпавших очков. Найти дисперсию и среднеквадратичное отклонение случайной величины X.

$$X := (1\,2\,3\,4\,5\,6)^T$$

$$\mathrm{var}(X) = 2.917 \quad \mathrm{stdev}(X) = 1.708$$

Аналогичные результаты получаются и при использовании формул:

$$D := \frac{1}{\mathrm{length}(X)} \sum_{i=0}^{\mathrm{last}(X)} (X_i - \mathrm{mean}(X))^2$$

$$\sigma := \sqrt{\frac{1}{\mathrm{length}(X)} \cdot \sum_{i=0}^{\mathrm{last}(X)} (X_i - \mathrm{mean}(X))^2},$$

$$D = 2.917 \quad \sigma = 1.708.$$

Мода и медиана

Модой в статистике называют варианту, которая встречается в выборке наиболее часто. В MathCad подсчитать моду выборки можно с помощью встроенной функции mode(x). В случае, если все варианты встречаются в выборке с одинаковой частотой, система выдаст сообщение: No value occurs more then any others (ни одна величина не встречается чаще, чем все остальные).

Медианой называется варианта, которая делит вариационный ряд (рассортированную выборку) на две части, равные по количеству вариант. То есть если количество элементов выборки нечетное и равно2k+1, то медианой будет являться (k+1)-й элемент. В случае четного количества вариант медиана определяется как среднее арифметическое между k-м и (k+1) элементами выборки. В MathCad медиана вычисляется с помощью встроенной функции median(x).

<u>Пример 4.</u> Вычисление моды и медианы

X:=(1 1 0 8 3 7)

mode(X)=1 median(X)=4

Статистические функции работают не только с векторами - столбцами, но и с векторами - строками.

Размах варьирования

Важная характеристика рассеяния вариационного ряда - размах варьирования может быть просто вычислена в MathCad с помощью двух специальных матричных функций: max(x) - находит максимальное значение в выборке, min(x) - функция находит минимальную величину в выборке. Используя описанные функции, размах варьирования можно задать как

R=max(x)-min(x).

<u>Пример 5.</u> Вычисление размаха варьирования. Для задания вектора выборки воспользуемся генератором случайных чисел, распределенных по показательному закону:

$$X:=\mathrm{rexp}(1000,4)$$
$$\mathrm{Max}(X)=2.892 \quad \min(X)=1.58\times10^{-5}$$
$$R=\max(X)-\min(X)=2.892.$$

Наибольший общий делитель и наименьшее общее кратное

Для решения некоторых задач в статистике бывает необходимым определить, на какое максимальное целое число делятся без остатка все величины в выборке. В MathCad очень просто вычислить такое число. Для этого необходимо воспользоваться встроенной функцией gcd(x) (от англ. Greatest common divisor - наибольший общий делитель).

Схожей с описанной является задача поиска наименьшего числа, которое делится без остатка на все значения элементов выборки. В MathCad ее можно решить с помощью встроенной функции lcm(x) (сокращение от Least common multiple – наименьшее общее кратное).

<u>Пример 6.</u> Найти наибольший общий делитель и наименьшее общее кратное.

$$X := (2\ 4\ 8\ 16\ 32\ 64)$$
$$\mathrm{gcd}(X) = 2 \quad \mathrm{lcm}(X) = 64$$

Задача. Для заданных случайных величин найти числовые характеристики (математическое ожидание, дисперсию, среднее квадратическое отклонение, моду, медиану), размах варьирования, а также наибольший делитель и наименьшее общее кратное элементов массива X.

№ варианта						
1	X	-1	0	2	4	7
	p	0,1	0,3	0,3	0,2	0,1
2	X	3	5	6	8	10
	p	0,2	0,2	0,3	0,2	0,1
3	X	-5	-4	-3	-1	1
	p	0,3	0,3	0,2	0,1	0,1
4	X	-1	2	3	4	6
	p	0,3	0,2	0,2	0,2	0,1
5	X	4	5	7	8	10
	p	0,1	0,4	0,3	0,1	0,1
6	X	-2	-1	0	2	5
	p	0,2	0,2	0,3	0,2	0,1
7	X	-4	-1	0	1	2
	p	0,1	0,1	0,2	0,5	0,1
8	X	-3	-2	-1	2	3
	p	0,2	0,4	0,1	0,1	0,2
9	X	6	7	9	10	12
	p	0,2	0,3	0,1	0,2	0,2
10	X	0	2	4	5	6
	p	0,4	0,2	0,2	0,1	0,1
11	X	-3	-2	0	1	2
	p	0,3	0,2	0,2	0,2	0,1
12	X	1	2	4	7	11
	p	0,1	0,1	0,3	0,3	0,2
13	X	-1	2	5	7	10
	p	0,2	0,2	0,3	0,1	0,2
14	X	4	7	9	11	13
	p	0,3	0,1	0,2	0,2	0,2

Отчет о выполненном занятии должен содержать:

1. Тему и цель работы.
2. Индивидуальное задание согласно варианту.
3. Решение предложенных задач.

Контрольные вопросы

1. Дайте определение основных числовых характеристик случайной величины (математическое ожидание, дисперсия, среднее квадратическое отклонение, мода, медиана). Каков их вероятностный смысл?
2. Что называется размахом выборки?

Практическое занятие №4
Вычисление числовых характеристик выборки

Цель занятия: изучить возможности Mathcad по вычислению числовых характеристик выборки

Задание: решить представленную задачу.

Краткие теоретические сведения

Моменты

В теории вероятностей и математической статистике, помимо математического ожидания и дисперсии, используются и другие числовые характеристики случайных величин. В первую очередь это начальные и центральные моменты.

Начальным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени случайной величины X, т.е. $\alpha_k = M(x^k)$.

Центральным моментом k-го порядка случайной величины ξ называется величина μ_k, определяемая формулой

$$\mu_k = M[(x - M)^k].$$

Заметим, что математическое ожидание случайной величины – начальный момент первого порядка $\alpha_1 = M$, а дисперсия – центральный момент второго порядка:

$$\mu_2 = M[(x - M)^2] = D(x).$$

Существуют формулы, позволяющие выразить центральные моменты случайной величины через ее начальные моменты. Одна из таких формул приведена выше:

$$D = M(x - M)^2 = \mu_2 - \alpha_1^2.$$

В дальнейшем будет использована формула

$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3.$$

Асимметрия

В теории вероятностей и математической статистике в качестве меры асимметрии распределения служит коэффициент асимметрии, который определяется формулой:

$$\beta = \frac{\mu_3}{\sigma^3}$$

где μ_3 - центральный момент третьего порядка;

$\sigma = \sqrt{D} = \sqrt{\mu_2}$ - среднеквадратичное отклонение

Коэффициент асимметрии – безразмерная величина, а по его знаку можно судить о характере асимметрии. Для симметричной СВ $\mu_3 = 0$.

Указание. Для того чтобы определить точность коэффициента асимметрии, выделенное выражение для него щелкните в строке Floating Point в меню Symbolics и укажите в окне диалога число десятичных знаков в выводе.

Эксцесс

Нормальное распределение наиболее часто используется в теории вероятностей и математической статистке, и поэтому график плотности вероятностей нормального распределения стал своего рода эталоном, с которым сравнивают другие распределения. Одним из параметров, определяющих отличие сравниваемого распределения от нормального, является эксцесс.

Эксцесс γ случайной величины ξ определяется равенством

$$\gamma = \frac{\mu_4}{(D)^2} - 3.$$

По известным нам свойствам математического ожидания и определению центрального момента получим формулу для μ_4

$$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4.$$

У нормального распределения, естественно, $\gamma = 0$. Если $\gamma > 0$, то это означает, что график плотности вероятностей p(x) сильнее «заострен», чем у нормального распределения, если же $\gamma < 0$, то «заостренность» графика p(x) меньше, чем у нормального распределения.

Порядок выполнения занятия

1. Изучить теоретический материал.
2. Вычислить для выборки заданной случайным образом выборочные моменты 3 и 4-го порядков выборочный эксцесс E, коэффициент асимметрии.
3. Оформить отчет.

Пример выполнения задания приведен на рисунке 4.1.

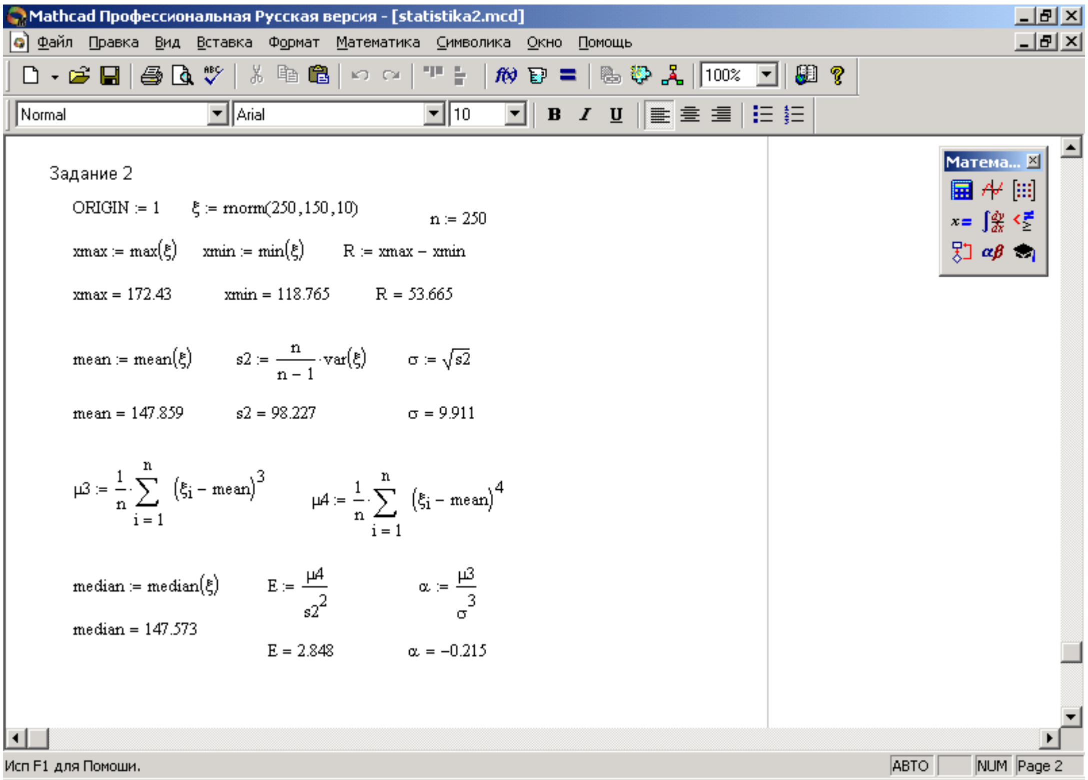


Рисунок 4.1 – Пример выполнения задания в MathCad

Контрольные вопросы

1. Раскройте понятие «моменты».
2. Что такое центральный момент?
3. Что такое начальный момент?
4. Что такое асимметрия?
5. Что такое коэффициент асимметрии?
6. Что такое эксцесс?
7. Что такое математическое ожидание?

Практическое занятие №5
Применения MathCad для решения задач теории вероятности

Цель занятия:

1. Закрепить знания об основных законах распределения случайной величины.
2. Научить применять MathCad для построения функций распределения и плотности распределения случайной величины.
3. Закрепить навыки расчета основных числовых характеристик случайных величин.
4. Научиться применять MathCad для расчета основных числовых характеристик случайных величин.

Функции и инструменты MathCad

Прежде чем приступать к решению задач теории вероятностей в MathCad, познакомимся с инструментами, которые предоставляет пакет для их решения.

Напомним, что дискретная случайная величина x, принимающая значения $x_1 < x_2 < ... < x_i < ...$ с вероятностями $p_1, p_2, ..., p_i, ...$, может быть задана распределением – таблицей вида (X_i - значение случайной величины, p – вероятность появления именно этого значения).

X	x_1	x_2	…	x_i	…	x_n
p	p_1	p_2	…	p_i	…	p_n

Для дальнейшей работы ряд распределения необходимо сделать вариационным.

Вариационный ряд – это дискретный ряд распределения, у которого значения X_i располагаются в порядке возрастания.

Такие таблицы в среде MathCad записывается в виде матрицы размерности $2 \times n$.

Указания. Распределение случайной величины в MathCad записывается в виде матрицы A

$$\left(\begin{array}{c|cccc} A_{1.1} & A_{1.2} & \ldots & \ldots & A_{1.n} \\ A_{2.1} & \ldots & \ldots & \ldots & A_{2.n} \end{array}\right),$$

где $A_{1.i}$ - значения случайной величины (x_i);

$A_{2.i}$ - соответствующие вероятности (p_i);

i=1, 2 , 3 , 4, 5.

Значения X в матрицу следует записывать в порядке возрастания.

По умолчанию первый элемент матрицы имеет индекс 0 (i=0). Используйте системную переменную ORIGIN для задания нижней границы индексации в матрице: ORIGIN:=1.

Функцией распределения случайной величины X называется функция F(x), значение которой в точке x равно вероятности того, что случайная величина X будет меньше этого значения x, то есть

$$F(x) = P\{X < x\}.$$

Функция F(x) для дискретной случайной величины вычисляется по формуле

$$F(x) = \sum_{x_i < x} p_i.$$

где суммирование ведется по всем значениям i, для которых $x_i < x$.

Функция распределения случайной величины, имеющей приведенное выше распределение, имеет вид:

$$F(x) = \begin{cases} 0, & x < x_x \\ p_1, & x_x < x < x_2 \\ p_1 + p_2, & x_2 < x < x_3 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ p_1 + p_2 + \ldots + p_{n-1}, & x_{n11} < x < x_n \\ 1, & x_n < x \end{cases}$$

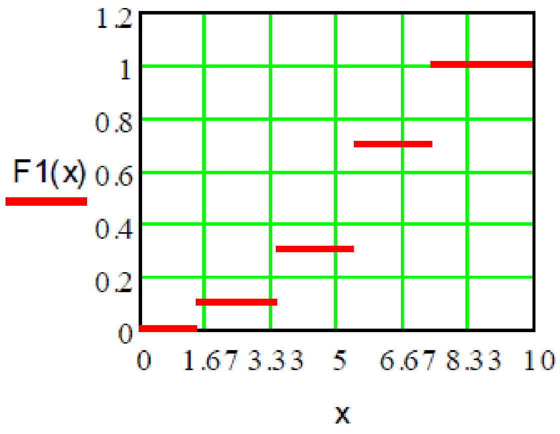
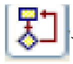


Рисунок 5.1 – Функция распределения

Функцию распределения, заданную разными выражениями на разных интервалах изменения аргументов, можно определить следующим образом:

1. Введите имя функции переменной x и знак присваивания (знак присваивания «:=»).
2. Используйте панель программных элементов (Programming) (кнопка ). С помощью кнопки Add Line добавьте необходимое число строк для задания функции распределения (рисунок 5.1).
3. Введите в помеченной позиции нуль, щелкните по кнопке if и введите неравенство, определяющие первый интервал изменения аргумента (символ ∞ можно ввести щелчком по соответствующей кнопке в панели  (Calculus)).
4. Затем перейдите во вторую строку определения функции, введите $A_{2,1}$ - имя переменной, содержащей значение p_1, или число 0.2 – значение.
5. Введите неравенство, определяющие второй интервал изменения аргумента (знак можно ввести щелчком по соответствующей кнопке в панели отношений (Boolean)); выделите, нажимая клавишу <SPACE>, вторую строку определения функции, щелкните по кнопке Add Line и введите, действуя, как описано выше определение функции на следующем интервале.

В результате у Вас должно получиться следующее:

$$F(x) := \begin{vmatrix} 0 \ \text{ if } \ -\infty < x < A_{1,1} \\ A_{2,1} \ \text{ if } \ A_{1,1} \le x < A_{1,2} \\ A_{2,1} + A_{2,2} \ \text{ if } \ \ldots \\ \ldots \ \ldots \ \ldots \\ \ldots \ \ldots \ \ldots \\ 1 \ \text{ if } \ A_{1,5} \le x < \infty \end{vmatrix}$$

Рисунок 5.2 – Функция распределения случайной величины

На приведенном рисунке 5.2 функция распределения определена с использованием имен переменных.

Замечание. Следует помнить, что MathCad не совсем корректно строит графики ступенчатых функций, соединяя отрезками прямых значения функции в точке скачка. Более точный график функции распределения представляет собой отрезки, параллельные оси абсцисс, с «выколотым» правым концом.

Случайные величины. Функции распределения

Каждая случайная величина полностью определяется своей функцией распределения.

Функция распределения любой случайной величины обладает следующими свойствами:

- F(x) определена на всей числовой прямой R;
- F(x) не убывает, т.е. если $x_1 \le x_2$, то $F(x_1) \le F(x_2)$;
- $F(-\infty) = 0$;
- $F(+\infty) = 1$;
- P(a<x<b)=F(b)-F(a).

Важно помнить, что функция распределения является «паспортом» случайной величины: она содержит всю информацию об этой случайной величине, и поэтому изучение случайной величины заключается в исследовании ее функции распределения, которую часто называют просто распределением.

Для проведения вычислений со случайными величинами (непрерывными и дискретными) в MathCad есть богатая библиотека встроенных функций наиболее распространенных стандартных распределений. Каждое распределение представлено в библиотеке тремя функциями – плотностью вероятностей, функцией распределения и функцией, обратной к функции распределения. Имена всех встроенных функций, определяющих плотности вероятностей, начинаются с буквы d, определяющих функции распределения – с буквы p.

Например, для работы с нормальным распределением предназначены функции dnorm(x, h, s), p(norm(x, h, s) и qnorm(x, h, s).

Наиболее распространенные распределения дискретных случайных величин

Познакомимся с дискретными случайными величинами, которые чаще всего используются при решении практических задач. Эти случайные величины имеют биномиальные, геометрические и пуассоновские распределения.

Биномиальное распределение (схема Бернулли)

Пусть проводится серия из n независимых испытаний, каждое из которых заканчивается либо «успехом», либо «неуспехом». Пусть в каждом испытании (опыте) вероятность успеха p, а вероятность неудачи – q=1-p. С таким испытанием можно связать случайную величину x, равную числу успехов в серии из n испытаний. Эта величина принимает целые значения от 0 до n.

Ее распределение называется биномиальным и определяется формулой Бернулли

$$p_k = P(\xi = k) = C_n^k p^k q^{n-k}$$

где 0<p1, q=1-p, k=0,1, …, n, $C_n^k = \dfrac{n!}{k!(n-k)!}$

В MathCad для вычисления плотности вероятности и функции распределения случайной величины, имеющей биномиальное распределение, предназначены функции dbinom(k, n,p) и pbinom(k, n, p), значения которых – соответственно p(k) и F(k).

Геометрическое распределение

Со схемой испытаний Бернулли можно связать еще одну случайную величину: h – число испытаний до первого успеха. Эта величина принимает бесконечное множество значений от 0 до $+\infty$, и ее распределение определяется формулой

$$p_k = P(\eta = k) = p^k q$$

где $0<p<1$, $q=1-p$, $k=0, 1, \ldots, n$.

В MathCad для вычисления плотности вероятности и функции распределения случайной величины, имеющей геометрическое распределение, предназначены функции dgeom(k, p) и pgeom(k, p), значения которых – соответственно p(k) и F(k).

Пуассоновское распределение

Пуассоновское распределение имеет случайная величина m, принимающая значения k=0, 1, 2, … с вероятностями

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

где $\lambda > 0$ - параметр пуассоновского распределения.

В MathCad для вычисления вероятности и функции распределения случайной величины, имеющей пуассоновское распределение, предназначены функции dpois(k, λ) и ppois(k, λ), значения которых – соответственно p(k) и F(k).

Наиболее распространенные частные распределения непрерывных случайных величин

Равномерное распределение

Непрерывная случайная величина ξ, принимающая значение на отрезке [a, b], распределена равномерно на [a, b], если плотность распределения p(x) и функция распределения случайной величины ξ имеют соответственно вид

$$p(x) = \begin{cases} 0, & x \notin [a, b] \\ \dfrac{1}{b - a}, & x \in [a, b] \end{cases} \qquad F(x) = \begin{cases} 0, & x \le a \\ \dfrac{x - a}{b - a}, & a < x \le b \\ 1, & x > b \end{cases}$$

В MathCad значения в точке x плотности распределения и функции распределения случайной величины имеющее равномерное распределение на отрезке [a, b], вычисляются встроенными функциями соответственно dunif(x, a, b) и punif(x, a, b).

Экспоненциальное (показательное) распределение

Непрерывная случайная величина ξ имеет показательное распределение с параметром $\lambda > 0$, если плотность распределения имеет вид

$$p(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \ge 0 \end{cases} \qquad F(x) = \begin{cases} 0, & x \le 0 \\ 1 - \lambda e^{-\lambda x}, & x > 0 \end{cases}$$

В MathCad значения в точке x плотности распределения и функции распределения случайной величины, имеющей экспоненциальное распределение с параметром λ, вычисляются встроенными функциями соответственно dexp(x, l) и pexp(x, l).

Нормальное распределение

Это распределение играет исключительно важную роль в теории вероятностей и математической статистике. Случайная величина ξ нормально распределена с параметрами m и σ, $(\sigma > 0)$, если ее плотность распределения имеет вид

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

В MathCad значения в точке x плотности распределения и функции распределения нормальной случайной величины с параметрами a, σ вычисляются встроенными функциями соответственно dnorm(x, a,s) и pnorm(x, a,s).

Распределение Стьюдента

Пусть случайная величина ξ имеет стандартное нормальное распределение, а случайная величина $\chi_n^2 - \chi^2$ распределение с n степенями свободы. Если ξ и χ_n^2 независимы, то про случайную величину $\tau_n = \dfrac{\xi}{\sqrt{\chi_n^2 / n}}$ говорят, что она имеет распределение Стьюдента с числом степеней свободы n. Доказано, что плотность вероятности этой величины вычисляется по формуле

$$p_\tau n(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}, x \in R$$

При больших n распределение Стьюдента практически не отличается от N (0, 1).

В MathCad значения в точке x плотности распределения и функции Стьюдента с n степенями свободы вычисляются встроенными функциями соответственно dt(x, n) и pt(x, n).

Числовые характеристики случайных величин

Каждая случайная величина полностью определяется своей функцией распределения. В то же время при решении практических задач достаточно знать несколько числовых параметров, которые позволяют представить основные особенности случайной величины в сжатой форме.

К таким величинам относятся, в первую очередь, математическое ожидание и дисперсия.

Математическое ожидание случайной величины

Математическое ожидание – число, вокруг которого сосредоточены значения случайной величины.

Если ξ - дискретная случайная величина с распределением,

ξ	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

то ее математическим ожиданием – оно обозначается M – называется величина

$$M[x] = \sum_{i=1}^{n} p_i x_i .$$

Математическое ожидание непрерывной случайной величины с плотностью вероятностей p(x) вычисляется по формуле

$$M[x] = \int_{-\infty}^{\infty} xp(x)dx.$$

Дисперсия случайной величины

Дисперсия случайной величины характеризует меру разброса значений случайной величины около ее математического ожидания. Если случайная величина ξ имеет математическое ожидание M, то дисперсией случайной величины ξ называется величина $D = M[(X - M)^2]$. Легко показать, что $D = M[x^2] - (M[x])^2$. Эта универсальная формула одинаково хорошо применима как для дискретных случайных величин, так и для непрерывных. Величина M^2 вычисляется по формулам:

$$M[x^2] = \sum_{i=1}^{n} p_i x_i^2 \quad M[x^2] = \int_{-\infty}^{\infty} x^2 p(x) dx,$$

для дискретных и непрерывных случайных величин соответственно.

Еще одним параметром для определения меры разброса значений случайной величины является среднеквадратическое отклонение σ, связанное с дисперсией соотношением $\sigma = \sqrt{D}$.

Моменты

В теории вероятностей и математической статистике, помимо математического ожидания и дисперсии, используются и другие числовые характеристики случайных величин. В первую очередь это начальные и центральные моменты.

Начальным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени случайной величины X, т.е. $\alpha_k = M(x^k)$.

Центральным моментом k-го порядка случайной величины ξ называется величина μ_k, определяемая формулой

$$\mu_k = M[(x - M)^k].$$

Заметим, что математическое ожидание случайной величины – начальный момент первого порядка $\alpha_1 = M$, а дисперсия – центральный момент второго порядка:

$$\mu_2 = M[(x - M)^2] = D(x).$$

Существуют формулы, позволяющие выразить центральные моменты случайной величины через ее начальные моменты. Одна из таких формул приведена выше:

$$D = M(x - M)^2 = \mu_2 - \alpha_1^2.$$

В дальнейшем будет использована формула

$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3.$$

Асимметрия

В теории вероятностей и математической статистике в качестве меры асимметрии распределения служит коэффициент асимметрии, который определяется формулой:

$$\beta = \frac{\mu_3}{\sigma^3}$$

где μ_3 - центральный момент третьего порядка;

$\sigma = \sqrt{D} = \sqrt{\mu_2}$ - среднеквадратичное отклонение

Коэффициент асимметрии – безразмерная величина, а по его знаку можно судить о характере асимметрии. Для симметричной СВ $\mu_3 = 0$.

Указание. Для того чтобы определить точность коэффициента асимметрии, выделенное выражение для него щелкните в строке Floating Point в меню Symbolics и укажите в окне диалога число десятичных знаков в выводе.

Эксцесс

Нормальное распределение наиболее часто используется в теории вероятностей и математической статистке, и поэтому график плотности вероятностей нормального распределения стал своего рода эталоном, с которым сравнивают другие распределения. Одним из параметров, определяющих отличие сравниваемого распределения от нормального, является эксцесс.

Эксцесс γ случайной величины ξ определяется равенством

$$\gamma = \frac{\mu_4}{(D)^2} - 3.$$

По известным нам свойствам математического ожидания и определению центрального момента получим формулу для μ_4

$$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4.$$

У нормального распределения, естественно, $\gamma = 0$. Если $\gamma > 0$, то это означает, что график плотности вероятностей p(x) сильнее «заострен», чем у нормального распределения, если же $\gamma < 0$, то «заостренность» графика p(x) меньше, чем у нормального распределения.

Порядок выполнения занятия

Задание 1.

Постройте с помощью MathCad график функции распределения для случайной величины:

X	1	0	7	4	-2
P	0.1	0.5	0.1	0.1	0.2

Порядок выполнения задания

1. Задайте случайную величину А в виде матрицы.
2. Определите функцию распределения случайной величины F(x).
3. Определите функцию распределения G(x) той же случайной величины с использованием конкретных значений переменных.
4. Постройте графики функций распределений F(x) и G(x). Отредактируйте и сравните графики.

В отчете представить два графика F(x) и G(x), сделать выводы о их сходстве и объясните, почему так произошло.

Задание 2.

Порядок выполнения задания

1. Задайте случайные величины, имеющие биномиальное распределение. Параметры распределения задайте самостоятельно. Постройте графики распределения и функции распределения случайной величины. (Подберите удобные параметры отображения графиков). Удобно строить оба графика в одной системе координат.
2. Проверьте для них $\sum_{k=0}^{n} p_k = 1$.
3. Вычислите вероятность попадания значений случайной величины в выбранный интервал.
4. Найдите значение k, для которого величина $P(\xi = k)$ максимальна (медиану). Исследуйте (понаблюдайте) зависимость этой вероятности от параметров распределения.

<u>*Указание.*</u> Для того, чтобы определить по графику распределения наиболее вероятное значение случайной величины, щелкните в меню Format (Формат) в пункте Graph (График) по строке Trace (Следование), установите перекрестье маркера на точке максимума распределения и выведите в рабочий документ вероятность значения, указанного в окне X-Value (Величина X).

5. Измените значения параметров распределения и повторите вычисления. Сравните полученные результаты. Выводы привести в отчете.

Повторить п. 1-5 для других распределений (геометрическое и пуассоновское).

В отчете представьте по одному варианту для каждого распределения: параметры распределения, графики вероятности и функции распределения, значение наиболее вероятного значения СВ и значение вероятности попадания СВ в указанный диапазон. В отчет вставлять фрагменты из MathCAD.

Задание 3.

Порядок выполнения задания

1. Введите параметры равномерного распределения. (Можно использовать любые).
2. Определите плотность вероятности и функцию распределения случайной величины.
3. Постройте графики.
4. Поэкспериментируйте с параметрами различных распределений.
5. Сделайте выводы о зависимости выходных данных от входных параметров.

Повторить п. 1-5 для других распределений (экспоненциального, нормального распределения и распределения Стьюдента).

В отчете представьте по одному варианту для каждого распределения: параметры распределения, графики вероятности и функции распределения и выводы о зависимости выходных данных от входных параметров. В отчет вставлять фрагменты из MathCad.

Задание 4.

Порядок выполнения задания

1. Вычислите математическое ожидание и дисперсии для случайных величин, имеющих дискретные распределения (биномиальные, геометрические и пуассоновские распределения). В качестве исходных параметров возьмите:

n=10 – число испытаний;

p=0.1*h – вероятность наступления события;

$\lambda = 4 + k$ - параметр пуассоновского распределения;

N=1000 – число испытаний стремится к ∞;

k – номер варианта студента по журналу;

2. Сверите полученные значения со справочными данными.

В отчете привести все расчетные формулы и результаты сравнения для каждого распределения в виде фрагментов из MathCad.

Задание 5.

Порядок выполнения задания

1. Вычислите математические ожидания и дисперсии для случайных величин, имеющих непрерывные распределения (равномерное, экспоненциальное, нормальное, распределение Стьюдента). В качестве исходных параметров возьмите (при желании можно использовать любые другие) следующие параметры:

a=1*h, b=3*h, n=5, λ =0.1*h, σ =0.05*h – необходимые параметры распределений;

h – порядковый номер студента по журналу.

2. Сверите полученные значения со справочными данными.

В отчете привести все расчетные формулы и результаты для каждого распределения в виде фрагментов из MathCad.

Задание 6 (Дополнительное).

Вычислите коэффициент асимметрии случайной величины X с равномерным распределением (или любого другого).

Порядок выполнения задания

1. Определите значения параметров распределения случайной величины.
2. Вычислите коэффициент асимметрии.
3. Сверьте полученное значение со справочными данными для данного распределения.
4. Постройте график плотности вероятности.
5. Сделайте выводы, опираясь на значение коэффициента асимметрии и форму распределения.

Задание 7. (Дополнительное)

Вычислите эксцесс случайной величины ξ с равномерным распределением (или любым другим).

Порядок выполнения задания

1. Определите значения параметров распределения случайной величины.
2. Вычислите коэффициент асимметрии.
3. Сверьте полученное значение со справочными данными для данного распределения.
4. Постройте график плотности вероятности.
5. Сделайте выводы, опираясь на значение коэффициента асимметрии и форму распределения.

Справочный материал приведен ниже.

Законы распределения случайных величин и их характеристики

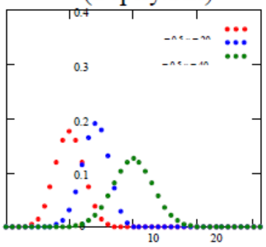
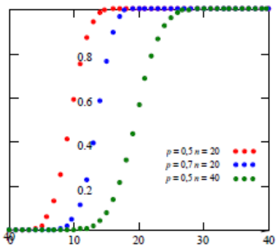
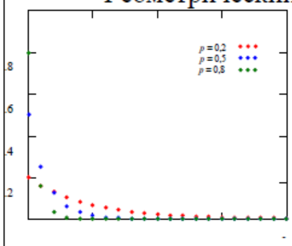
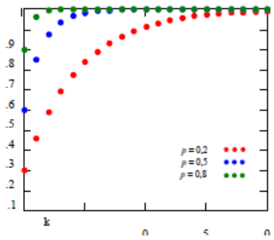
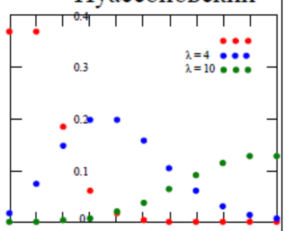
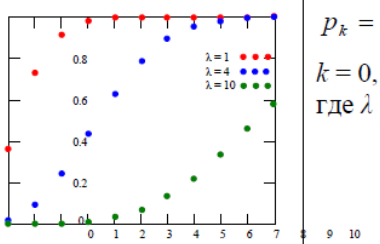
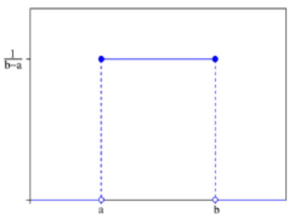
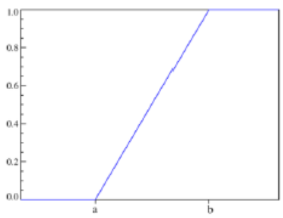
Вид закона распределения		Аналитическое выражение	Основн. число-вые хар-ки
Плотность распределения	Функция распределения		
Дискретные случайные величины			
Биноминальный (Бернулли) 		$p_k = C_n^k p^k q^{n-k}$, где $0 < p < 1,\ q = 1 - p$, $k = 0,\ 1,\ \ldots,\ n$, $C_n^k = \dfrac{n!}{k!(n-k)!}$	$M(X) = np$, $D(X) = npq$, $A = \dfrac{q-p}{\sqrt{npq}}$, $E = \dfrac{1-6pq}{npq}$
Геометрический 		$p_k = p^k q$ где $0 < p < 1,\ q = 1 - p$, $k = 0,\ 1,\ \ldots,\ n$	$M(X) = \dfrac{1-p}{p}$, $D(X) = \dfrac{q}{p^2}$, $A = \dfrac{2-p}{\sqrt{1-p}}$, $E = 6 + \dfrac{p^2}{1-p}$.
Пуассоновский 		$p_k = \dfrac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \ldots$, где $\lambda > 0$.	$M(X) = \lambda$, $D(X) = \lambda$, $A = \lambda^{-1/2}$, $E = \lambda^{-1}$
Непрерывные случайные величины			
 Равномерный		$p(x) = \begin{cases} 0, & x \notin [a,b] \\ \dfrac{1}{b-a}, & x \in [a,b] \end{cases}$ $F(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases}$	$M(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{(b-a)^2}{12}$, $A = 0$, $E = -\dfrac{6}{5}$.

Рисунок 5.2 – Законы распределения случайных величин

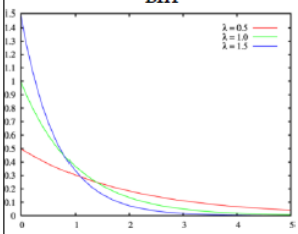
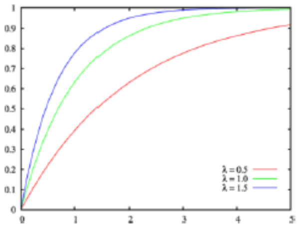
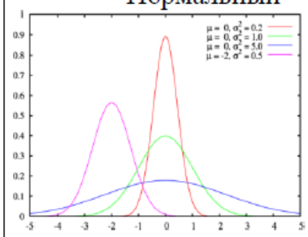
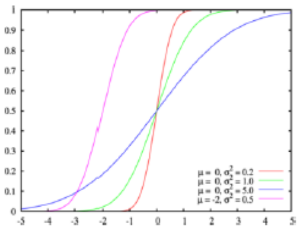
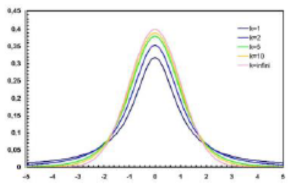
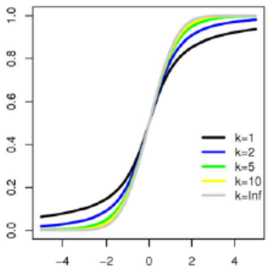
Экспоненциальный 		$p(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x}, & x\ge 0\end{cases}$ $F(x)=\begin{cases}0, & x\le 0\\ 1-\lambda e^{-\lambda x}, & x>0\end{cases}$ $\lambda > 0.$	$M(X)=\lambda^{-1}$, $D(X)=\lambda^{-2}$, $A=2,\ E=6$.
Нормальный 		$p(x)=$ $=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$ $F(x)=\Phi(x)$, где $\Phi(x)$ – функция Лапласа: $\Phi(x)=$ $=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}\exp\left(-\frac{(z)^2}{2\sigma^2}\right)dz$ где $z=x-m$	$M(X)=m$, $D(X)=\sigma^2$, $A=0,\ E=0$.
Стьюдента 		$p(x)=$ $=\frac{1}{\sqrt{n\pi}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$ $x\in R$ где $\Gamma\left(\frac{n}{2}\right)$ - гамма-функция Эйлера	$M(X)=0$, если n>1 $D(X)=\frac{n}{n-2}$, если n>2, $A=0$, если n>3, $E=\frac{6}{n-4}$, если n>4.

Рисунок 5.3 – Законы распределения случайных величин

Практическое занятие №6
Классические и геометрические вероятности

Краткие теоретические сведения

Основными в теории вероятностей являются понятия события и его вероятности.

Любое событие можно рассматривать как результат испытания (эксперимента, опыта).

Испытанием называется некоторое действие, явление, реализуемое при определенном комплексе условий.

<u>Примеры.</u> Бросание игральной кости (кубика), монеты, организация лотереи и т.д.

Каждое испытание характеризуется множеством возможных исходов (результатов).

<u>Пример.</u> Бросание монеты (испытание) характеризуется исходами: герб, решка (номинал монеты).

Событием называется любой исход испытания (эксперимента, опыта).

<u>Примеры.</u> Выпадение герба при бросании монеты, появление трех очков при бросании игральной кости, выигрыш в телебинго, покупка некачественного товара.

Все события могут быть разделены на три группы: 1. Достоверные. 2. Случайные. 3. Невозможные.

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

<u>Пример.</u> Извлечение белого шара из урны, содержащей только белые шары, является достоверным событием.

Событие называется *случайным*, если в результате опыта оно либо происходит, либо не происходит.

<u>Примеры.</u> Появление 5 очков при бросании игральной кости; выигрыш в лотерее; попадание в цель при стрельбе.

Событие называется *невозможным*, если в результате опыта, оно заведомо не происходит.

<u>Пример.</u> Появление 7 очков при бросании игральной кости.

Обычно события обозначают большими буквами латинского алфавита (с индексом или без них):

$$A, B, C, ...; \quad A_1, A_2, A_3, ..., A_n; \quad B_1, B_2, B_3, ..., B_n.$$

Достоверное событие обычно обозначают буквой Ω, а невозможное – символом $\varnothing$.

События A и B называются *несовместными*, если появление одного из них исключает появление другого в результате одного опыта.

Пример. Из урны, содержащей белые и черные шары, извлекли один шар. События: A – «извлечен белый шар» и B – «извлечен черный шар» являются несовместными событиями.

Любой опыт может быть охарактеризован множеством его возможных исходов (результатов). Например, если бросается игральная кость, то это множество следующее:

$$\{A_1, A_2, A_3, A_4, A_5, A_6\},$$

где событие A_i означает появление грани с i очками. Эти события называются *элементарными событиями*, а их множество образует *пространство элементарных событий*.

Любое событие может быть представлено как некоторое множество элементарных событий.

Пример. Событие A – «выпало четное число очков при бросании игральной кости» образовано элементарными событиями A_2, A_4 и A_6. Обозначение: $A = \{A_2, A_4, A_6\}$.

Говорят, что события A_2, A_4 и A_6 *благоприятствуют* событию A, так как событие A имеет место тогда, когда имеет место одно из элементарных событий A_2, A_4 и A_6. Эти события называются *благоприятствующими исходами для A*.

Пусть опыт характеризуется n элементарными событиями $A_1, A_2, ..., A_n$. Они называются *возможными исходами* опыта(эксперимента, испытания).

Предположим, что некоторому событию A, которое может произойти в результате опыта, благоприятствуют m исходов. Пусть, из некоторых геометрических и физических соображений можно допустить, что исходы A_i равновозможны. Тогда

$$P(A) = \frac{m}{n}. \qquad (6.1)$$

Формула (6.1) определяет *классическую вероятность* события A.

Очевидно, для достоверного события Ω имеем $P(\Omega) = 1$, а для невозможного $\varnothing$, P($\varnothing$)=0. Для любого события

$$0 \le P(A) \le 1. \qquad (6.2)$$

При вычислении вероятности P(A) по формуле (6.1) очень полезными оказываются основные понятия *комбинаторики*.

А. Перестановки. Пусть задано некоторое множество M, которое содержит n элементов. Очевидно, его элементы можно упорядочить различными способами.

Пример. Множество чисел (1, 2, 3) можно упорядочитьшестью способами:

$$(1, 2, 3), \quad (2, 1, 3), \quad (3, 1, 2),$$
$$(3, 2, 1), \quad (1, 3, 2), \quad (2, 3, 1).$$

Каждое из полученных упорядоченных множеств называется *перестановкой* заданного множества. Если обозначить через P_n *число перестановок* некоторого множества из n элементов, тогда:

$$P_n = n! = 1 \cdot 2 \cdot ... \cdot (n-1) \cdot n. \qquad (6.3)$$

Пример. Сколько различных чисел можно образовать из цифр 1, 2, 3, 4 так, чтобы каждое число содержало все цифры, причем по одному разу каждую из них.

Решение. Очевидно, что из цифр 1,2, 3, 4 можно образовать $P_4 = 4! = 24$ таких чисел.

Б. Размещения. Пусть задано множество M, содержащее n элементов. Упорядоченные подмножества множества M которые содержат m элементов, где $m \le n$ называются *размещениями*.

Отметим, что одно размещение отличается от другого или порядком расположения, или хотя бы одним элементом. Число размещений из n элементов по m обозначается A_n^m. Формула для вычисления числа размещений:

$$A_n^m = n \cdot (n-1) \cdot \ldots \cdot (n-m+1) = \frac{n!}{(n-m)!}. \qquad (6.4)$$

Заметим, что $P_n = A_n^m$.

Пример. На консультацию пришли 7 студентов. В аудитории 15 стульев. Сколькими способами можно рассадить студентов?

Решение. Число способов выразится размещениями:

$$A_{15}^7 = 15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 = 32\,432\,400.$$

*В. Сочетания.*Пусть задано множество M, содержащее n элементов. Образуем неупорядоченные подмножества, которые содержат m элементов $(m \le n)$ и отличаются друг от друга хотя бы одним элементом. Говорят, что этиподмножества образуют *сочетания* из n элементов по m.Число таких сочетаний обозначается C_n^m и вычисляется по формуле:

$$C_n^m = \frac{A_n^m}{P_m} = \frac{n!}{m! \cdot (n-m)!}. \qquad (6.5)$$

Пример. На столе разложено 20 экзаменационных билетов. Пришедший на экзамен студент, случайно берет 2 билета. Сколько существует для него возможностей?

Решение. Очевидно, что число таких возможностей равно:

$$C_{20}^2 = \frac{20!}{2! \cdot 18!} = \frac{19 \cdot 20}{1 \cdot 2} = 190.$$

Имеет место формула $C_n^m = C_n^{n-m}$.

Г. Пусть задано множество, которое содержит N элементов. Среди них имеются K элементов, которые обладают каким-то свойством. Из заданного множества случайно делается выборка n элементов. Среди отобранных желательно иметь k элементов с указанным свойством. Число таких выборок, очевидно, будет следующим:

$$l = C_K^k \cdot C_{N-K}^{n-k}. \tag{6.6}$$

Д. Пусть задано множество, которое содержит N элементов. Из него случайно берут один элемент. Число способов такого выбора, очевидно, совпадает с числом элементов, т. е. оно равно N. Выбранный элемент возвращается обратно. Опять берут один элемент. Всего возможностей N. Для обоих выборов число возможностей равно $N^2 = N \cdot N$. Если эту процедуру повторить s раз, то будем иметь N^s возможностей.

<u>Пример.</u> Игральная кость бросается 4 раза. Каково число возможных исходов?

<u>Решение.</u> При одном бросании имеем шесть возможных исходов. Для двух бросаний будем иметь $6^2 = 6 \cdot 6$ возможных исходов. При четырех бросаниях будем иметь 6^4 возможностей.

Пусть испытание, в результате которого имеет место событие A, повторяется n раз. Если A имело место k раз, тогда отношение

$$f(A) = \frac{k}{n}. \tag{6.7}$$

называется *частотой* события A.

К. Пирсон (1857 - 1936) осуществил 24000 бросаний монеты. Герб (событие A) появился 12012 раз. Следовательно,

$$f(A) = \frac{12012}{24000} = 0{,}5005 \approx 0{,}5.$$

Частота события или число, вокруг которого она колеблется, называется его *статистической вероятностью*.

Пусть заданы области G и g. Область g находится внутри области G (рисунок 6.1).

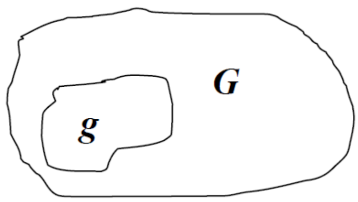


Рисунок 6.1 - Области G и g

Некоторая точка M наудачу проектируется на область G. Какова вероятность, что эта точка попадает в область g?

Предполагаем, что точка M может оказаться в любой точке области G. Обозначим через A событие «точка M попадает в области g». Очевидно, вероятность этого события будет прямо пропорциональна величине области g. Поэтому, естественно определить

$$P(A) = \frac{\text{mes } g}{\text{mes } G}. \tag{6.8}$$

В формуле (6.8) символ «mes» происходит от слова «мера» и означает длину, площадь или объем в зависимости от того, где расположены G и g (на прямой, на плоскости или в пространстве).

Вероятность, определения формулой (6.8), называется геометрической вероятностью события A.

Решение типовых задач

Задача 6.1. Набирая номер телефона своего товарища, студент забыл одну цифру и набрал ее случайно. Найти вероятность того, что был набран нужный номер.

Решение. Пусть A - событие «набран нужный номер». По классической формуле P(A)=m/n, где n - число всех возможных исходов, m - число исходов благоприятствующих A. Забытая цифра может быть любой из десяти: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, т. е. m=1, n=10.

Ответ.P(A)=1/10.

Задача 6.2. Набирая номер телефона, некто забыл последние 2 цифры, помня при этом лишь, что они различные. Найти вероятность того, что набран желаемый номер.

Решение. Пусть В - событие «набран желаемый номер». Очевидно, m=1, а n - число упорядоченных пар, которые можно образовать из 10 цифр: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, т. е.

$$n = A_{10}^{2} = 10 \cdot 9 = 90 \Rightarrow P(B) = \frac{1}{90}.$$

Ответ. Вероятность набора желаемого номера равна 1/90 .

Задача 6.3. В партии имеется 10 деталей, 7 из них высшего качества. Найти вероятность того, что из 6 деталей, взятых наудачу, 4 будут высшего качества.

Решение.Введем событие А - «из 6 отобранных случайно деталей, 4 будут высшего качества».
Всего в партии 10 деталей, берут из них наудачу 6. Таким образом, $n = C_{10}^{6}$. В партии 7 деталей высшего качества и 3 не являются таковыми. Среди 6 отобранных, 4 должны быть высшего качества и 2 не будут этого качества. По формуле (1.6) находим $m = C_7^4 \cdot C_3^2$.Следовательно,

$$P(A) = \frac{C_7^4 \cdot C_3^2}{C_{10}^{6}} = \frac{1}{2}.$$

Ответ. P(A)=1/2.

Задача 6.4. В коробке находятся 75 спичек, среди которых 5 без головок. Берут наугад 15 спичек. Найти вероятность того, что среди них будем иметь 3 спички без головок.

Решение.Введем событие А - «среди 15 отобранных спичек 3 будут без головок». $n = C_{75}^{15}, m = C_5^3 \cdot C_{70}^{12}$.

Следовательно, $P(A) = \frac{m}{n} = \frac{C_5^3 \cdot C_{70}^{12}}{C_{75}^{15}} \approx 0{,}047.$

Ответ. Вероятность того, что среди 15 отобранных спичек 3 будут без головок, равна 0,047.

Задача 6.5. В круг радиуса R вписан правильный треугольник. Найти вероятность того, что точка, брошенная наудачу в круг, попадет в треугольник.

Решение. Пусть A событие «точка попадет в треугольник». Очевидно, P(A) будет равна отношению между площадями треугольника и круга.

$$P(A)=\frac{\frac{3}{4}\sqrt{3}R^2}{\pi R^2}=\frac{3\sqrt{3}}{4\pi}\approx 0{,}414.$$

Ответ. Вероятность того, что точка попадает в треугольник, равна 0,414.

Задача 6.6. из интервала [0; 2] случайно берут два числа x и y. Найти вероятность того, что эти числа удовлетворяют неравенству

$$x^2/4\le y\le x.$$

Решение. Пусть x и y декартовы координаты некоторой точки плоскости. Условие задачи приводит нас к следующим областям:

$$G=\{(x;y):0\le x\le 2;0\le y\le 2\}\text{ и } g=\left\{(x;y):\frac{x^2}{4}\le y\le x\right\}.$$

Вводим событие A – «числа x и y удовлетворяют неравенству $\frac{x^2}{4}\le y\le x$». Формула (1.8) для нашего случая будет выглядеть так: $P(A)=\frac{S_g}{S_G}$. На координатной плоскости xOy изображаем области G и g, вычисляем площади и берем их отношение. Получаем

$$P(A)=\frac{\int_0^2\left(x-\frac{x^2}{4}\right)dx}{4}=\frac{1}{3}.$$

Ответ. Вероятность того, что числа удовлетворяют заданному неравенству, равна 1/3.

Задачи для упражнений

1. Деревянный куб, все грани которого окрашены, распилили на 1000 одинаковых кубиков, которые положили в мешок и тщательно перемешали. Затем извлекли один кубик. Найти вероятности событий:

A – «кубик имеет три окрашенные грани».

B – «кубик имеет две окрашенные грани».

C – «кубик имеет одну окрашенную грань».

D – «кубик имеет ни одной окрашенной грани».

2. В ящике находятся 25 деталей первого сорта и 5 деталей второго сорта. Из него наудачу извлекают 10 деталей. Найти вероятность того, что среди извлеченных деталей окажутся две детали второго сорта.

3. В коробке шесть одинаковых, занумерованных кубиков. Наудачу, по одному извлекают все кубики. Какова вероятность того, что номера извлеченных кубиков появятся в возрастающем порядке?

4. Цифры 0; 1; 2; 3; 4; 5; 6; 7; 8; 9 написаны на карточках, которые помещаются в урну и тщательно перемешиваются. Из урны наудачу извлекают две карточки. Найти вероятность того, что из них можно образовать число, которое делится на 18.

5. В группе 25 студентов, среди которых 15 ребят. Было куплено и распределено 5 билетов в театр. Найти вероятность того, что в театр пойдут 3 ребят.

6. Площадь разграфлена параллельными прямыми, находящимися друг от друга на расстоянии 2a. На нее случайно бросается игла длиной 2l ($l<a$). Найти вероятность того, что игла пересечет какую-нибудь прямую.

7. Заданы два неотрицательных числа, которые не превосходят единицу. Найти вероятность того, что сумма этих чисел не превзойдет единицу, а их произведение будет не больше 2/9.

8. Два студента условились встретиться в определенном месте в течение часа. Первый пришедший ждет второго 15 минут, после чего уходит. Найти вероятность того, что встреча состоится, если каждый может прийти на нее случайно в любой момент указанного часа.

9. Деревянный прут, длиной l случайно обломали в двух местах. Найти вероятность того, что из трех образовавшихся кусков можно составить треугольник.

10. На книжную полку ставят 10 книг. Найти вероятность того, что 3 определенные книги окажутся рядом.

11. Десять друзей случайным образом садятся в десять вагонов поезда. Найти вероятность того, что ни один из них не окажется в вагоне со своим другом.

12. В урне находятся 3 белых и 7 черных шаров. Наудачу извлекли 2 шара. Найти вероятность того, что они будут белыми.

13. Номер телефона состоит из 6 цифр. Найти вероятность того, что у случайно набранного номера все цифры различны.

14. На складе находятся 100 пар сапог: 10 пар имеют черный цвет, остальные - коричневый. Наудачу взяли 8 пар. Какова вероятность, что все взятые пары будут коричневыми?

15. Некто, набирая номер телефона, забыл последние 2 цифры, помня лишь, что они различны и нечетны. Найти вероятность того, что будет набран нужный номер.

16. В некоторый момент времени в лифте 9 - этажного дома находятся 7 пассажиров. Найти вероятность того, что 2 пассажира выйдут на одном этаже, а остальные 5 - на разных этажах.

17. Семь путешественников случайным образом садятся в поезд, имеющий 12 вагонов. Найти вероятность того, что в каждый из вагонов сядет не более одного путешественника.

18. На спортивной студенческой базе находятся 30 ребят и 25 девчат. Для изучения мнения относительно графика тренировок наудачу отбирают 10 студентов. Найти вероятность того, что в эту группу попадет 8 ребят и 2 девчат.

19. Трамвай имеет 3 вагона. В него случайно садятся 8 человек. Найти вероятность того, что в первый вагон сядет 3 человека.

20. В урне находятся 20 белых и 10 черных шаров. Извлекают наудачу 11 шаров. Найти вероятность того, что среди них окажутся 7 белых.

21. Имеется шесть прутиков длиной в 2, 4, 6, 8, 10 и 12 см. Из них наудачу берутся три. Найти вероятность того, что из них можно построить треугольник.

22. Бросают 5 игральных костей. Найти вероятность того, что одно очко выпадет, по крайней мере, на одной кости.

23. В чемпионате страны по футболу участвуют 16 команд, которые разбиты на 2 одинаковые подгруппы. Найти вероятность того, что 2 лучшие команды попадут в разные подгруппы.

24. В ящике находятся 200 яблок, среди которых 25 испорчены. Найти вероятность того, что среди 10 случайно отобранных яблок будут испорченные.

25. В лифт 9 - этажного здания на первом этаже село 5 человек. Каждый из них может выйти на любом этаже, начиная со второго. Найти вероятности событий:

а) все пассажиры выйдут на одном этаже;

б) все пассажиры выйдут на восьмом этаже;

в) все пассажиры выйдут на разных этажах.

26. В библиотеке имеется 10 различных книг. 5 книг стоят по 4 лея каждая, 3 книги стоят по 1 лею, а 2 книги - по 3 лея. Найти вероятность того, что 2 книги, взятые наугад, стоят 5 лей.

Практическое занятие №7
Сложение и умножение вероятностей

Краткие теоретические сведения

Пусть заданы два события A и B .

Суммой событий A и B называется событие $C = A \cup B$, которое происходит тогда, когда происходит хотя бы одно из событий A или B.

Пример. Пусть A - «попадание в мишень при первом выстреле», B - «попадание в мишень при втором выстреле». Тогда $C = A \cup B$ - «попадание в мишень по крайней мере один раз при двух выстрелах».

Произведением событий A и B называется событие $D = A \cap B$, которое происходит тогда, когда происходят оба события A и B .

Пример. Пусть A - «извлечение белого шара первый раз из урны», B - «извлечение белого шара второй раз из урны». Тогда $D = A \cap B$ - «последовательное извлечение двух белых шаров из урны».

Замечание. Иногда сумму событий A и B обозначают через A+B , а произведение - через $A \cdot B$.

Если A и B любые события, то

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \qquad (7.1)$$

Для двух событий A и B ($A \cap B = \varnothing$) имеем

$$P(A \cup B) = P(A) + P(B). \qquad (7.2)$$

Пусть A, B, C – три произвольных события. Тогда

$$\begin{gathered} P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - \\ - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{gathered} \qquad (7.3)$$

Если события A, B, C попарно несовместны, то

$$P(A \cup B \cup C) = P(A) + P(B) + P(C). \qquad (7.4)$$

Формулы (7.3) и (7.4) могут быть распространены на n событий. Например, если события $A_1, A_2, ..., A_n$ попарно несовместны, то

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n). \qquad (7.5)$$

Говорят, что события $A_1, A_2, ..., A_n$ образуют *полную группу событий*, если в результате испытания, одно из них обязательно происходит, т.е. $A_1 \cup A_2 \cup ... \cup A_n = \Omega$.

Широкое применение на практике имеет полная группа попарно несовместных событий.

Пример. При бросании игральной кости события A_i - «выпало i очков» - образуют полную группу попарно несовместных событий. Для таких событий справедливо равенство

$$P(A_1) + P(A_2) + ... + P(A_n) = 1. \qquad (7.6)$$

Частным случаем полной группы является группа, состоящая из двух несовместных событий

$$A_1 = A \quad \text{и} \quad A_2 = \overline{A}.$$

Например, появление числа, выражающего номинал монеты (A) и появление герба ($\overline{A}$) при бросании монеты являются событиями, образующими полную группу.

Событие $\overline{A}$ называется *противоположным событием* для события A.

Очевидно, что противоположным для $\overline{A}$ является событие A. Поэтому говорят обычно о двух противоположных друг другу событиях. Легко выводится формула

$$P(\overline{A}) = 1 - P(A). \qquad (7.7)$$

Переход к противоположному событию зачастую существенно облегчает подсчет вероятностей.

Условной вероятностью P(A/B) называется вероятность события A, вычисленная в предположении, что событие B уже наступило.

Имеет место формула умножения вероятностей:

$$P(A \cap B) = P(A) \cdot P(B/A), \qquad (7.8)$$

или

$$P(A \cap B) = P(B) \cdot P(A/B). \qquad (7.9)$$

Формула (7.8) (или (7.9)) может быть распространена на n произвольных событий $A_1, A_2, ..., A_n$.

$$P\left(\bigcap_{i=1}^{n} A_i)\right) = P(A_1) \cdot P(A_2 / A_1) \cdot ... \cdot P\left(A_n \Big/ \bigcap_{i=1}^{n} A_i\right). \qquad (7.10)$$

Говорят, что событие A не зависит от события B, если

$$P(A/B) = P(A). \qquad (7.11)$$

Независимость имеет взаимный характер: если A не зависит от B, тогда и событие B не зависит от события A. Говорят, поэтому, о независимых событиях A, B. Для таких событий имеем:

$$P(A \cap B) = P(A) \cdot P(B). \qquad (7.12)$$

События $A_1, A_2, ..., A_n$ называются*независимыми в совокупности*, если каждое из них и любая комбинация остальных независимы.

Если события $A_1, A_2, ..., A_n$ независимы в совокупности, то

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2) \cdot ... \cdot P(A_n). \qquad (7.13)$$

Решение типовых задач

Задача 2.1. В коробке находятся катушки с нитками. 40% из них белые, 20% - красные, 25% - синие и 15% - зеленые. Найти вероятность того, что взятая наугад катушка будет с красными или зелеными нитками.

Решение. Пусть А событие - «взяли катушку с красными или зелеными нитками». A_1 - «взяли катушку с красными нитками». A_2 - «взяли катушку с зелеными нитками». Очевидно, A_1 и A_2 несовместны и $A = A_1 \cup A_2$. Следовательно,

$$P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2),$$
$$P(A_1) = \frac{20}{100} = \frac{1}{5}, \quad P(A_2) = \frac{15}{100} = \frac{3}{20}.$$

Поэтому $P(A) = \frac{1}{5} + \frac{3}{20} = \frac{4+3}{20} = \frac{7}{20}.$

Ответ.Вероятность того, что взятая наудачу катушкабудет с красными или зелеными нитками, равна 7/20.

Задача 2.2. Для того чтобы сдать экзамен студент должен ответить на 3 вопроса. Программа содержит 30 вопросов, из которых студент подготовил лишь 25. Найти вероятность того, что студент сдаст экзамен.

Решение.Пусть А событие - «студент сдаст экзамен». A_1 - «студент ответит на первый вопрос». A_2 - «студент ответит на второй вопрос». A_3 - «студент ответит на третий вопрос».

$$P(A) = P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 / A_1) \cdot P(A_3 / A_1 \cap A_2) =$$
$$= \frac{25}{30} \cdot \frac{24}{29} \cdot \frac{23}{28} = \frac{115}{203} \approx 0{,}57.$$

Ответ. Вероятность того, что студент сдаст экзамен, равна 0,57.

Задача 2.3. В урне находятся 8 белых и 6 черных шаров. Последовательно извлекают два шара без возврата. Найти вероятность того, что будут извлечены шары: белый, черный.

<u>Решение.</u> Обозначим через В событие «извлеклипоследовательно шары: белый, черный». B_1 - «первый раз извлекли белый шар». B_2 - «второй раз извлекли черный шар». Тогда, $B = B_1 \cap B_2$. Поэтому

$$P(B) = P(B_1 \cap B_2) = P(B_1) \cdot P(B_2 / B_1) = \frac{8}{14} \cdot \frac{6}{13} = \frac{4}{7} \cdot \frac{6}{13} = \frac{24}{91} \approx 0{,}26.$$

<u>Ответ.</u> Вероятность того, что будут извлечены последовательно белый и черный шар, равно 0,26.

<u>Задача 2.4.</u> Три стрелка производят по одному выстрелу по мишени. Вероятности попадания в мишень соответственно равны 0,7; 0,8 и 0,9. Найти вероятности событий:

А - в мишень попадут все стрелки;

В - в мишень попадет один стрелок;

С - в мишень попадет хотя бы один стрелок.

<u>Решение.</u> Вводим события: A_1 - «первый стрелок попадает в мишень». A_2 - «второй стрелок попадает в мишень». A_3 - «третий стрелок попадает в мишень». Имеем:

$$P(A_1) = 0{,}7 \Rightarrow P(\overline{A}_1) = 1 - P(A_1) = 1 - 0{,}7 = 0{,}3;$$
$$P(A_2) = 0{,}8 \Rightarrow P(\overline{A}_2) = 1 - P(A_2) = 1 - 0{,}8 = 0{,}2;$$
$$P(A_3) = 0{,}9 \Rightarrow P(\overline{A}_3) = 1 - P(A_3) = 1 - 0{,}9 = 0{,}1;$$
$$A = A_1 \cap A_2 \cap A_3,$$
$$P(A) = P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3),$$

так как события A_1, A_2 и A_3 независимы в совокупности (следует из условий задачи). Тогда

$$P(A) = 0{,}7 \cdot 0{,}8 \cdot 0{,}9 = 0{,}504.$$

Находим P(B). Имеем:

$$B = (A_1 \cap \overline{A}_2 \cap \overline{A}_3) \cup (\overline{A}_1 \cap A_2 \cap \overline{A}_3) \cup (\overline{A}_1 \cap \overline{A}_2 \cap A_3). \quad (7.14)$$

Слагаемые суммы (7.14) являются несовместными событиями. В свою очередь, события в скобках независимы, поэтому

$$\begin{aligned}P(B) &= P(A_1 \cap \overline{A}_2 \cap \overline{A}_3) + P(\overline{A}_1 \cap A_2 \cap \overline{A}_3) + P(\overline{A}_1 \cap \overline{A}_2 \cap A_3) = \\ &= P(A_1) \cdot P(\overline{A}_2) \cdot P(\overline{A}_3) + P(\overline{A}_1) \cdot P(A_2) \cdot P(\overline{A}_3) + P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot P(A_3) = \\ &= 0{,}7 \cdot 0{,}2 \cdot 0{,}1 + 0{,}3 \cdot 0{,}8 \cdot 0{,}1 + 0{,}3 \cdot 0{,}2 \cdot 0{,}9 = 0{,}092.\end{aligned}$$

Находим $P(\overline{C})$. $\overline{C}$ - «в мишень не попадает ни один стрелок».

$$\overline{C} = \overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3,$$

$$P(\overline{C}) = P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) = P(\overline{A}_1) \cdot P(\overline{A}_2) \cdot P(\overline{A}_3) = 0{,}1 \cdot 0{,}2 \cdot 0{,}3 = 0{,}006.$$

Следовательно,

$$P(C) = 1 - 0{,}006 = 0{,}994.$$

Ответ. $P(A) = 0{,}504$; $P(B) = 0{,}092$; $P(C) = 0{,}994$.

Задача 2.5. Охотник попадает в мишень с вероятностью 0,4. Сколько раз он должен стрелять, чтобы с вероятностью не меньше чем 0,9 можно было бы утверждать о поражении мишени хотя бы раз.

Решение. Пусть A событие – «охотник попал в мишень хотя бы раз». $\overline{A}$ - «охотник не попал в мишень ни разу». Обозначим через n число выстрелов. Тогда, $P(\overline{A}) = (0{,}6)^n$.

$$P(A) = 1 - P(\overline{A}) = 1 - (0{,}6)^n.$$

Из условий задачи следует неравенство

$$1 - (0{,}6)^n \geq 0{,}9 \Leftrightarrow 0{,}1 \geq (0{,}6)^n.$$

Прологарифмируем по основанию 10 обе части полученного неравенства. Получим $n \lg 0{,}6 \leq \lg 0{,}1$. Делим на $\lg 0{,}6<0$ обе части и приходим к результату $n \geq \frac{\lg 0{,}1}{\lg 0{,}6} = 4{,}5$. Так как n должно быть целым, то $n \geq 5$.

Ответ. Охотник должен стрелять хотя бы 5 раз.

Задача 2.6. На полке в магазине находятся 80 пачек сигарет. Среди них имеются 4 пачки с порченными сигаретами. Найти вероятность того, что кто-то, покупая 4 пачки, получит хотя бы две пачки с порченными сигаретами.

Решение. Пусть А – «было куплено хотя бы две пачки с порченными сигаретами». A_1 - «было куплено две пачки с порченными сигаретами». A_2 - «было куплено три пачки с порченными сигаретами». A_3 - «было куплено четыре пачки с порченными сигаретами». События A_1, A_2, A_3 несовместимы и $A = A_1 \cup A_2 \cup A_3$.

По классической формуле находим вероятности

$$P(A_1) = \frac{C_4^2 \cdot C_{76}^2}{C_{80}^4}; \quad P(A_2) = \frac{C_4^3 \cdot C_{76}^1}{C_{80}^4}; \quad P(A_3) = \frac{C_4^4}{C_{80}^4}.$$

Следовательно,

$$P(A) = P(A_1) + P(A_2) + P(A_3) = \frac{1}{C_{80}^4} \cdot (C_4^2 \cdot C_{76}^2 + C_4^3 \cdot C_{76}^1 + 1) \approx 0{,}01.$$

Ответ. $P(A) \approx 0{,}01$.

Задача 2.7. Из двух орудий стреляют по цели. Для первого орудия вероятность попадания в цель равна 0,7, а для второго 0,8. Найти вероятность того, что цель будет поражена хотя бы раз, если из каждого орудия стреляют по одному разу.

Решение. Пусть A_1 - «первое орудие попадает в цель». A_2 - «второе орудие попадает в цель». А- «мишень поражена хотябы один раз». Очевидно, $A = A_1 \cup A_2$. События A_1 и A_2 совместны и независимы, поэтому

$$P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1) \cdot P(A_2) =$$
$$= 0{,}7 + 0{,}8 - 0{,}7 \cdot 0{,}8 = 1{,}5 - 0{,}56 = 0{,}94$$

Ответ. Вероятность того, что цель будет поражена хотя бы один раз, равна 0,94.

Задача 2.8. Завод выпускает электролампы. Из них производственный брак составляет 3%, а 5% - монтажный брак. Найти вероятность того, что отобранная наугадэлектролампа будет бракованной.

Решение. Пусть A событие - «электролампа имеет брак». A_1 - «электролампа имеет брак производства». A_2 - «электролампа имеет монтажный брак». Тогда $A = A_1 \cup A_2$. События A_1 и A_2 совместны и независимы, поэтому

$$P(A) = P(A_1) + P(A_2) - P(A_1) \cdot P(A_2) = 0{,}03 + 0{,}05 - 0{,}0015 \approx 0{,}079.$$

Ответ. Вероятность того, что отобранная электролампа будет бракованной, равна 0,079.

Задачи для упражнений

1. На стадионе установлены три экрана. Вероятности того, что в данный момент горят экраны, равны соответственно 0,9; 0,8; и 0,7. Найти вероятности событий:
   а) в данный момент горят два экрана;
   б) в данный момент горит не более одного экрана;
   в) в данный момент горят 3 экрана.
2. На сборку поступают детали с трех станков. Первый станок дает 20% деталей, второй - 30%, а третий - 50%. Найти вероятности того, что из трех взятых наугад деталей:
   а) все будут выпущены разными станками;
   б) все будут выпущены третьим станком;
   в) две детали выпущены вторым станком.
3. Вероятность того, что студент сдаст первый экзамен, равна 0,9; для второго экзамена эта вероятность равна 0,7; для третьего - 0,6. Найти вероятности событий:

А - «студент сдаст два экзамена»;

В - «студент сдаст хотя бы два экзамена»;

С - «студент сдаст не более двух экзаменов».

4. Стрелок стреляет по цели до первого попадания. Вероятность попадания при одном выстреле равна 0,8. Какова вероятность, что будет сделано 4 выстрела?

5. Шесть охотников заметили лису и выстрелили одновременно. В предположении, что каждый охотник попадает и убивает лису с вероятностью 1/3, найти вероятность того, что лиса будет убита.

6. Слово «вектор» образовано из букв, написанных на карточках, которые перемешиваются и опускаются в урну. Затем достают по одной 4 карточки и располагают их в24ряд слева направо. Найти вероятность того, что получилось слово «крот».

7. Для прохождения производственной практики студентам было предложено 15 мест г. Курск, 8 – в Железногорске и 7 – в Курчатове. Найти вероятность того, что 2 студента пройдут практику в одном городе.

8. На складе находятся 40 пар обуви черного цвета, 25 пар коричневого, 23 пары белого и 12 пар обуви красного цвета. Внешне коробки обуви ничем не отличаются. Найти вероятность того, что взятая наугад коробка содержит обувь белого или красного цвета.

9. Бросается игральная кость. Найти вероятности следующих событий:

а) при однократном бросании игральной кости получим четное или кратное трём число очков;

б) при трёхкратном бросании игральной кости только один раз получим четное или кратное трём число очков;

в) при четырёхкратном бросании игральной кости хотя бы один раз получим четное или кратное трём число очков.

10. Случайно образуется двузначное число. Найти вероятность того, что это число будет делиться:

а) на 2 или на 5;

б) на 2 и на 5.

11. Написанные на 30 карточках числа от 11 до 40, помещаются в мешочек и перемешиваются. Найти вероятность того,

что при случайном отборе одной карточки получим число, которое делится на 2 или на 3.

12. Найти вероятность того, что при бросании трех игральных костей, хотя бы на одной выпадет два очка.

13. Преподаватель подготовил для экзамена 20 билетов, из которых 5 по алгебре, 5 по геометрии и 10 по математическому анализу. Студент берет последовательно три билета, без возвращения. Найти вероятности событий:

А - «студент взял 3 билета по алгебре»;

В - «студент взял один билет по геометрии»;

С - «студент взял хотя бы один билет по математическому анализу».

14. Два стрелка стреляют по цели по одному разу. Вероятность поражения цели для первого стрелка равна 0,8, для второго – 0,6. Найти вероятность того, что цель будет поражена один раз.

15. Рабочий обслуживает три станка. Вероятность того, что в течение смены потребует его внимания первый станок, равна 0,7; второй - 0,75; третий - 0,8. Найти вероятность того, что в течение смены внимания рабочего потребуют какие-либо два станка.

16. Радист трижды вызывает корреспондента. Вероятность того, что будет принят первый вызов, равна 0,2; второй - 0,3; третий - 0,4. События, состоящие в том, что вызов будет услышан, независимы. Найти вероятность того, что корреспондент услышит вызов радиста.

17. Для сообщения об аварии установлены два независимо работающих сигнализатора - автомата. Вероятность того, что при аварии сработает первый сигнализатор, равна 0,95; второй - 0,9. Найти вероятность того, что при аварии поступит сигнал: а) хотя бы от одного сигнализатора; б) только от одного сигнализатора.

18. Вероятность выигрыша по одному билету лотереи равна 1/7. Какова вероятность того, что обладатель пяти билетов лотереи выиграет: а) по всем пяти билетам ? б) ни по одному ? в) хотя бы по одному билету ?

19. Экзаменационный билет содержит три вопроса. Вероятность того, что студент ответит на первый, второй вопросы,

равна по 0,9; на третий - 0,8. Найти вероятность того, что студент сдаст экзамен, если для этого необходимо ответить:

а) на все вопросы;

б) по крайней мере, на два вопроса билета.

20. В экзаменационные билеты включено по два теоретических вопроса и по одной задаче. Всего составлено 28 билетов. Вычислить вероятность того, что при случайном выборе билета, студент ответит на него, если он подготовил 50 теоретических вопросов и 22 задачи.

21. Имеется 60 предметов, среди которых 4 с дефектами. Эти предметы разделены на 4 одинаковые группы. Найти вероятность того, что в каждую группу попадет по одному предмету с дефектом.

22. Заводом послана автомашина за различными материалами на четыре базы. Вероятность наличия нужного материала на первой базе равна 0,9, на второй - 0,95, на третьей - 0,8, на четвертой - 0,6. Найти вероятность того, что только на одной базе не окажется нужного материала.

23. Три команды A_1, A_2, A_3 спортивного общества A состязаются соответственно с тремя командами B_1, B_2, B_3 общества B. Вероятности того, что команды общества A выиграют матчи у команд общества B, таковы:

при встрече A_1 с B_1 - 0,8;

при встрече A_2 с B_2 - 0,4;

при встрече A_3 с B_3 - 0,4.

Для победы необходимо выиграть не менее двух матчей из трех (ничьи во внимание не принимаются). Найти вероятность победы общества A.

Практическое занятие №8
Формула полной вероятности. Формула Байеса

Краткие теоретические сведения

Пусть событие A может произойти при условии появления одного из событий $B_1, B_2, ..., B_n$, которые образуют полную группу, т.е.

$$B_1 \cup B_2 \cup ... \cup B_n \qquad (8.1)$$

Предположим также, что события попарно несовместны

$$B_i \cap B_j = \varnothing,\ i \neq j,\ i, j=1, 2, \ldots, n. \qquad (8.2)$$

Тогда

$$\begin{aligned} P(A) = P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2) + ... + \\ + P(B_n) \cdot P(A/B_n). \end{aligned} \qquad (8.3)$$

Формула (8.3) может быть переписана более компактно

$$P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A/B_i). \qquad (8.4)$$

Формула (8.3) (или (8.4)) называется формулой *полной вероятности*. События $B_1, B_2, ..., B_n$ называются обычно *гипотезами*. Вероятности гипотез

$$P(B_1), P(B_2), ..., P(B_n) \qquad (8.5)$$

заданы. Производится испытание, чьим результатом является событие A. Появление A может изменить заданные вероятности. Поэтому можно поставить задачу определения условных вероятностей

$$P(B_1/A), P(B_2/A), ..., P(B_n/A). \qquad (8.6)$$

Имеем

$$P(B_k/A) = \frac{P(B_k)\cdot P(A/B_k)}{\sum_{i=1}^{n} P(B_i)\cdot P(A/B_i)}, \quad k = 1, 2, ..., n. \qquad (8.7)$$

Формула (8.7) называется *формулой Байеса*.

Вероятности (8.5), которые известны до испытания, называются *вероятностями априори*, а вероятности (8.6), которые вычисляются по (8.7), называются *вероятностями апостериори*.

Решение типовых задач

Задача 8.1. Управление лотереи организует специальный тираж с четырьмя категориями выигрышей. Вероятности того, что купленный билет будет принадлежать соответствующей категории, приведены в таблице.

Категория	1	2	3	4
Вероятность принадлежности	0,46	0,24	0,16	0,14

Даны также вероятности выигрыша для каждой категории:

Категория	1	2	3	4
Вероятность выигрыша	0,55	0,25	0,06	0,14

Найти вероятность того, что купленный билет будет выигрышным.

Решение. Пусть А событие «купленный билет будет выигрышным». Вводим события - гипотезы:

B_1 - «билет относится к первой категории»,

B_2 - «билет относится к второй категории»,

B_3 - «билет относится к третьей категории»,

B_4 - «билет относится к четвертой категории».

Тогда,

$$P(B_1) = 0{,}46; \quad P(B_2) = 0{,}24; \quad P(B_3) = 0{,}16; \quad P(B_4) = 0{,}14.$$

Даны также условные вероятности

$$P(A/B_1) = 0{,}55; \quad P(A/B_2) = 0{,}25; \quad P(A/B_3) = 0{,}06; \quad P(A/B_4) = 0{,}14.$$

Используя формулы (8.3), получаем

$$P(A) = 0{,}46 \cdot 0{,}55 + 0{,}24 \cdot 0{,}25 + 0{,}16 \cdot 0.06 + 0{,}14 \cdot 0{,}14 \approx 0{,}34.$$

Ответ. Вероятность того, что купленный билет будет выигрышным, равна 0,34.

Задача 8.2. В спортклубе имеется 15 мячей, из которых 9 новые. Для первой игры случайно взяли 3 мяча, которые после нее вернули. Для второй игры снова взяли 3 мяча. Найти вероятность того, что мячи, взятые для второй игры, все новые.

Решение. Обозначим через А событие «для второй игры было взято 3 новых мяча». Неизвестно, какие мячи были взяты для первой игры, поэтому вводим события-гипотезы: B_i - «взято i новых мячей для первой игры», i=0, 1, 2, 3. Применим формулу полной вероятности:

$$P(A) = P(B_0) \cdot P(A/B_0) + P(B_1) \cdot P(A/B_1) + \\ + P(B_2) \cdot P(A/B_2) + P(B_3) \cdot P(A/B_3). \qquad (8.8)$$

Вычислим все вероятности, присутствующие в этой формуле, используя классическое определение (формула (8.1))

$$P(B_0) = \frac{C_6^3}{C_{15}^3}; \quad P(B_1) = \frac{C_9^1 \cdot C_6^2}{C_{15}^3}; \quad P(B_2) = \frac{C_9^2 \cdot C_6^1}{C_{15}^3}; \quad P(B_3) = \frac{C_9^3}{C_{15}^3};$$

$$P(A/B_0) = \frac{C_9^3}{C_{15}^3}; \quad P(A/B_1) = \frac{C_8^3}{C_{15}^3}; \quad P(A/B_2) = \frac{C_7^3}{C_{15}^3}; \quad P(A/B_3) = \frac{C_6^3}{C_{15}^3}.$$

Подстановка этих данных в формулу (8.8) дает:

$$P(A)=\frac{C_6^3}{C_{15}^3}\cdot\frac{C_9^3}{C_{15}^3}+\frac{C_9^1\cdot C_6^2}{C_{15}^3}\cdot\frac{C_8^3}{C_{15}^3}+\frac{C_9^2\cdot C_6^1}{C_{15}^3}\cdot\frac{C_7^3}{C_{15}^3}+\frac{C_9^3}{C_{15}^3}\cdot\frac{C_6^3}{C_{15}^3}=0{,}089.$$

Ответ. Вероятность того, что для второй игры взяли три новых мяча, равна 0,089.

Задача 8.3. В группе спортсменов 20 лыжников, 6 велосипедистов и 4 бегуна. Для того чтобы попасть на соревнование каждый спортсмен должен выполнить квалификационную норму. Вероятность выполнения нормы для велосипедиста равна 0,8; для лыжника - 0,9; для бегуна - 0,75. Найти вероятность того, что спортсмен, выбранный наугад, выполнит квалификационную норму.

Решение. Пусть А событие «спортсмен выполнит норму»; B_1 - «спортсмен является лыжником»; B_2 - «спортсмен является велосипедистом»; B_3 - «спортсмен является бегуном».

$$P(B_1)=\frac{20}{30}=\frac{2}{3};\quad P(B_2)=\frac{6}{30}=\frac{1}{5};\quad P(B_3)=\frac{4}{30}=\frac{2}{15}.$$

$$P(A/B_1)=0{,}9;\quad P(A/B_2)=0{,}8;\quad P(A/B_3)=0{,}75.$$

$$P(A)=P(B_1)\cdot P(A/B_1)+P(B_2)\cdot P(A/B_2)+P(B_3)\cdot P(A/B_3).$$

Следовательно,

$$P(A)=\frac{2}{3}\cdot\frac{9}{10}+\frac{1}{5}\cdot\frac{4}{5}+\frac{2}{15}\cdot\frac{3}{4}=\frac{43}{50}=0{,}86.$$

Ответ. Вероятность того, что взятый наугад спортсмен, выполнит квалификационную норму, равна 0,86.

Задача 8.4. С подводной лодки пускают по военному кораблю четыре торпеды. Вероятность того, что торпеда попадет в корабль, равна 0,3. Для потопления корабля достаточно двух торпед, а если одна торпеда попадает в корабль, то он погружается с вероятностью 0,6. Найти вероятность потопления корабля.

Решение. Вводим события:

A - «корабль будет потоплен»;

B_i - «в корабль попало i торпед», i=0, 1, 2, 3, 4;

$\overline{A}$ - «корабль не будет потоплен».

$$P(\overline{A}) = P(B_0) \cdot P(\overline{A} / B_0) + P(B_1) \cdot P(\overline{A} / B_1) + P(B_2) \cdot P(\overline{A} / B_2) +$$
$$+ P(B_3) \cdot P(\overline{A} / B_3) + P(B_4) \cdot P(\overline{A} / B_4) = P(B_0) \cdot P(\overline{A} / B_0) + P(B_1) \cdot P(\overline{A} / B_1),$$

так как $P(\overline{A} / B_2) = P(\overline{A} / B_3) = P(\overline{A} / B_4) = 0$,

$P(B_0)$ - вероятность того, что ни одна торпеда не попадет в корабль. Так как события независимы, имеем:

$$P(B_0) = (1 - 0{,}3)^4 = (0{,}7)^4 \approx 0{,}24.$$

$P(B_1)$ - вероятность того, что одна из четырех торпед попадет в корабль. Ее можно вычислить по формуле Бернулли.

$$P(B_1) = C_4^1 \cdot 0{,}3 \cdot (0{,}7)^3 \approx 0{,}41.$$

Условные вероятности

$$P(\overline{A} / B_0) = 1; \quad P(\overline{A} / B_1) = 1 - 0{,}6 = 0{,}4.$$

Следовательно,

$$P(\overline{A}) = 0{,}24 \cdot 1 + 0{,}41 \cdot 0{,}4 \approx 0{,}4.$$
$$P(A) = 1 - P(\overline{A}) \approx 0{,}6.$$

<u>Ответ.</u> Вероятность потопления корабля равна 0,6.

<u>Задача 8.5.</u> Магазин получает обувь с трех фабрик. 35% обуви поставляет магазину фабрика N_1, 40% - фабрика N_2, а остальная обувь производится на фабрике N_3. Фабрика N_1 дает 5% брака, вторая - 12% брака и третья дает в среднем 8% брака. Некто купил пару обуви, которая оказалась качественной. Найти вероятность того, что эта пара обуви выпущена фабрикой N_2.

Решение. Вводим события - гипотезы:

B_i - « пара обуви выпущена фабрикой N_i», i=1,2,3. Тогда,

$$P(B_1) = 0{,}35; \quad P(B_2) = 0{,}4; \quad P(B_3) = 0{,}25.$$

Вводим также событие А – «купленная пара обуви качественная». Нужно найти $P(B_2 / A)$. Используя формулу Байеса, получим

$$P(B_2 / A) = \frac{P(B_2) \cdot P(A / B_2)}{\sum_{i=1}^{3} P(B_i) \cdot P(A / B_i)} =$$

$$= \frac{0{,}4 \cdot 0{,}88}{0{,}35 \cdot 0{,}95 + 0{,}4 \cdot 0{,}88 + 0{,}25 \cdot 0{,}92} \approx 0{,}385.$$

Ответ. Вероятность того, что купленная пара обуви выпущена фабрикой N_2, равна 0,385.

Задача 8.6. Для участия в соревнованиях из трех групп берут студентов: из первой группы - 4 студента, из второй - 6 и из третьей группы - 5. Вероятности попадания студента из групп N_1, N_2 и N_3 в сборную Университета равны 0,9; 0,7; и 0,8 соответственно. Выбранный наудачу студент попал в сборную. Найти вероятность того, что он из первой группы.

Решение.Пусть А - событие «студент, выбранный наудачу, попал в сборную Университета». Вводим события - гипотезы: B_1 - «студент из первой группы», B_2 - «студент из второй группы», B_3 - «студент из третьей группы». Нужно вычислить $P(B_1 / A)$. Используя формулу Байеса, будем иметь

$$P(B_1 / A) = \frac{P(B_1) \cdot P(A / B_1)}{\sum_{i=1}^{3} P(B_i) \cdot P(A / B_i)};$$

где $P(B_1) = 4/15; \quad P(B_2) = 2/5; \quad P(B_3) = 1/3;$

$$P(A / B_1) = 0{,}9; \quad P(A/B_2) = 0{,}7; \quad P(A/B_3) = 0{,}8.$$

Поэтому,

$$P(B_1/A)=\frac{\frac{4}{15}\cdot\frac{9}{10}}{\frac{4}{15}\cdot\frac{9}{10}+\frac{2}{5}\cdot\frac{7}{10}+\frac{1}{3}\cdot\frac{8}{10}}=\frac{4\cdot 9}{4\cdot 9+6\cdot 7+5\cdot 8}\approx 0{,}31.$$

Ответ. Вероятность того, что студент, включенный в сборную, из первой группы, равна 0,31.

Задача 8.7. Имеются 3 урны со следующим составом: в первой урне 5 белых и 5 черных шаров, во второй урне - 6 белых и 8 черных шаров, а в третьей урне - 3 белых и 10 черных шаров. Из урны, выбранной наугад, извлечен белый шар. Найти вероятность того, что извлечение производилось из второй урны.

Решение. Вводим события - гипотезы: B_1 - «извлечениепроизводилось из первой урны», B_2 - «извлечениепроизводилось из второй урны», B_3 - «извлечение производилось из третьей урны». Имело место событие А - «был извлечен белый шар». Вычислим:

$$P(B_2/A)=\frac{P(B_2)\cdot P(A/B_2)}{\sum_{i=1}^{3}P(B_i)\cdot P(A/B_i)}. \qquad (8.9)$$

Имеем $P(B_1)=P(B_2)=P(B_3)=\frac{1}{3}$,

$$P(A/B_1)=\frac{5}{10}=\frac{1}{2};\quad P(A/B_2)=\frac{6}{14}=\frac{3}{7};\quad P(A/B_3)=\frac{3}{13}.$$

Подставим эти вероятности в формулу (8.9). Получим

$$P(B_2/A)=\frac{\frac{1}{3}\cdot\frac{3}{7}}{\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{3}{7}+\frac{1}{3}\cdot\frac{3}{13}}=\frac{\frac{1}{3}\cdot\frac{3}{7}}{\frac{1}{3}\cdot\left(\frac{1}{2}+\frac{3}{7}+\frac{3}{13}\right)}\approx 0{,}37.$$

Ответ. Вероятность того, что извлечение производилось из второй урны, равно 0,37.

Задачи для упражнений

1. На сборку поступает 40% деталей с первого станка, 30% - со второго, 20% - с третьего и с четвертого станка поступает 10% от общего количества деталей. Известно, что первый станок дает 2% брака, второй - 1%, третий - 0,5% и четвертый - 0,2%. Найти вероятность того, что случайно поступающая на сборку деталь - качественная.

2. По самолету производится три выстрела, с вероятностями попадания, равными 0,4; 0,5 и 0,7 соответственно. Для сбития самолета достаточно трех попаданий. В результате одного попадания самолет будет сбит с вероятностью 0,2; а для двух попаданий эта вероятность равна 0,6.

а) Найти вероятность того, что в результате трех выстрелов самолет будет сбит.

б) Самолет был сбит. Найти вероятность того, что это произошло в результате двух попаданий.

3. В урне находятся 10 шаров, из которых 7 белых. Из нее выпал один шар. Найти вероятность извлечения после этого из урны белого шара.

4. Две урны имеют одинаковый состав: а белых шаров и b черных. Из первой урны во вторую переложили один шар, перемешали, затем из второй в первую урну опять переложили шар. После этих процедур из первой урны извлекают один шар. Найти вероятность того, что извлеченный шар будет черным.

5. В составе студенческого отряда две бригады студентов второго курса и одна бригада с первого курса. В каждой бригаде второкурсников 4 мальчика и 6 девочек, а в бригаде первокурсников 5 мальчиков и 3 девочки. Из бригады, выбранной наудачу, посылается в город один студент.

а) Найти вероятность того, что в город поедет мальчик.

б) В город поехал мальчик. Найти вероятность того, что это студент первого курса.

6. В специализированную больницу поступают больные, которые страдают болезнями B_1, B_2 и B_3. 50% больныхстрадают болезнью B_1, 30% имеют болезнь B_2 и 20% - болезнь B_3. Вероятности излечения этих болезней равны 0,7; 0,8; 0,9 соответственно. Больной, поступивший в больницу, был выписан здоровым. Найти вероятность того, что он страдал болезнью B_2.

7. В группе из 20 студентов, пришедших на экзамен, восемь подготовлены отлично, шесть - хорошо, четыре - посредственно, и два - плохо. В экзаменационных билетах имеется 40 вопросов. Студент, подготовленный отлично, может ответить на все вопросы, хорошо - на 35, посредственно - на 25, плохо - на 10 вопросов. Вызванный наугад студент ответил на три произвольно заданных вопроса. Найти вероятность того, что этот студент подготовлен: а) хорошо; б) плохо.

8. Три завода выпускают электролампы. Первый завод выпускает 45% всей продукции, второй - 40% и третий - 15%. 70% электроламп первого завода стандартны, 80% - со второго и 81% - с третьего завода. Найти вероятность того, что купленная в магазине электролампа будет стандартной.

9. Два стрелка производят по некоторой цели по одному выстрелу. Вероятность попадания в цель для первого стрелка равна 0,2; а для второго - 0,6. В цель попали один раз. Найти вероятность того, что промахнулся первый стрелок.

10. Имеются две урны с шарами. В первой урне находятся 5 белых и 6 черных шаров. Во второй - 4 белых и 8 черных шаров. Из первой урны перенесли во вторую наудачу 3 шара, после чего из второй урны извлекли 4 шара. Найти вероятность того, что все извлеченные шары белые.

11. Имеем 6 урн следующего состава: в двух урнах находится по 2 белых и 4 черных шара, три урны содержат по 2 белых и 8 черных шаров, а в одной урне имеется 6 белых и 2 черных шара. Из урны, взятой наудачу, извлекают один шар. Найти:

а) вероятность того, что извлеченный шар белый;

б) вероятность того, что извлекли белый шар из одного из составов.

12. В группе из 20 стрелков имеются четыре отличных, десять хороших и шесть посредственных стрелков. Вероятность попадания

в цель при одном выстреле для отличного стрелка равна 0,9, для хорошего - 0,7, для посредственного - 0,5. На линию огня вызывают наугад два стрелка. Они производят по одному выстрелу. Найти вероятность того, что стрелки попадут в цель.

13. На торговой базе находятся электрические лампочки, изготовленные на двух заводах. Среди них 60% изготовлено первым заводом и 40% - вторым. Известно, что из каждых 100 лампочек первого завода, 90 соответствуют стандарту, а из 100 штук, изготовленных на втором заводе, стандартны 80. Определить вероятность того, что взятая наудачу с базы лампочка, будет стандартная.

14. Двадцать пять экзаменационных билетов содержат по два вопроса, которые не повторяются. Студент может ответить только на 40 вопросов. Определить вероятность того, что экзамен будет сдан, если для этого достаточно ответить на два вопроса из одного билета или на один вопрос из первого билета и на указанный дополнительно вопрос из другого билета.

15. Радиолампа, поставленная в телевизор, может принадлежать к одной из трех партий с вероятностями $p_1 = 0{,}25$; $p_2 = 0{,}5$ и $p_3 = 0{,}25$. Вероятности того, что лампа проработает определенное количество часов, для этих партий соответственно равны: 0,1; 0,2 и 0,4. Найти вероятность того, что лампа проработает заданное число часов.

16. Среди поступающих на сборку деталей с первого станка 0,1% бракованных, со второго - 0,2%, с третьего - 0,25%, с четвертого - 0,5%. Производительности их относятся как4:3:2:1 соответственно. Взятая наудачу деталь оказалась стандартной. Найти вероятность того, что она изготовлена: а) на первом; б) на втором; в) на третьем; г) на четвертом станке. Как проверить правильность вычислений этих вероятностей ?

17. Имеется 10 одинаковых по виду урн. Из них в 9 находятся по 2 черных и два белых шара, а в одной - 5 белых и один черный шар. Из наугад взятой урны извлечен шар. Найти вероятность того, что этот шар взят из урны, содержащей 5 белых шаров, если он оказался белым.

18. Группа из 20 стрелков упражняется в стрельбе. Пять стрелков попадают в мишень с вероятностью 0,8; семь стрелков - с

вероятностью 0,7; шесть - с вероятностью 0,6 и двое - с вероятностью 0,5. Наудачу выбранный стрелок произвел выстрел, но в мишень не попал. Какой группе, вероятнее всего, принадлежит этот стрелок?

19. Пассажир может обратиться для получения билета в одну из трех касс. Вероятности обращения в каждую кассу зависят от их местоположения и равны соответственно: 0,3; 0,45 и 0,25. Вероятность того, что к моменту прихода пассажира не будут билетов, равна для первой кассы 0,4, для второй - 0,2 и для третьей - 0,1. Пассажир направился за билетом в одну из касс и приобрел билет. Найти вероятность того, что это была первая касса.

20. Для участия в спортивных соревнованиях были привлечены студенты из двух групп: из первой - 10 студентов, из второй - 8. Вероятность участия в соревнованиях для студента первой группы равна 0.9, а для студента второй группы - 0.7. Найти вероятности событий:

A - «студент будет участвовать в соревнованиях»;

B - «студент не будет участвовать в соревнованиях».

Практическое занятие №9
Формула Бернулли. Формулы Муавра-Лапласа

Краткие теоретические сведения

Пусть испытания (опыт), в результате которого появляется событие A или $\overline{A}$, повторяются n раз. Заданы вероятности

$$P(A) = p; \quad P(\overline{A}) = 1 - p = q.$$

Предположим, что испытания независимы. Обозначим через $P_n(m)$ вероятность того, что в n испытаниях событие A произойдет m раз, где m=1, 2, …, n.

Известно, что эта вероятность вычисляется по формуле Бернулли:

$$P_n(m) = C_n^m p^m q^{n-m}. \qquad (9.1)$$

Число m_0 называется *наивероятнейшим числом,* если

$$P_n(m_0) = \max_{0 \le m \le n} P_n(m) \qquad (9.2)$$

Для числа m_0 имеет место неравенство

$$n\,p - q \le m_0 \le n\,p + p \qquad (9.3)$$

Имеем

$$n\,p + p - (np - q) = np + p + np + q = p + q = 1.$$

Если np-q (а также np+p) целое число, то неравенству (9.3) удовлетворяют два числа m_0 и $m_0 - 1$. Для дробных np-q и np+p имеем одно наивероятнейшее число m_0.

Применение формулы (9.1) становится довольно трудным когда n и m являются большими числами. В таких случаях, обычно, делают приближенные вычисления, в основе которых локальная формула Муавра-Лапласа.

$$P_n(m) \approx \frac{1}{\sqrt{npq}}\varphi(x_{m,n}), \qquad (9.4)$$

где

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad (9.5)$$

$$x_{m,n} = \frac{m - np}{\sqrt{npq}}. \qquad (9.6)$$

Формула (9.4) применяется следующим образом. Для заданных значений m, n, p, q вычисляется $x_{m,n}$. Затем из таблиц находят $\varphi(x_{m,n})$. Значения функции (9.5) можно найти в любом учебнике или задачнике по теории вероятностей и математической статистике. Найденное в таблице значение функции, умножается на число $\frac{1}{\sqrt{npq}}$. В результате получим приближенное значение вероятности $P_n(m)$.

Замечание. Функция (9.5) табулирована для $x \in [0;4]$. $\varphi(x) \approx 0$ для x>4. Для нахождения $\varphi(x)$ при x<0 используют ее четность, т.е. $\varphi(-x) = \varphi(x)$.

Иногда нужно найти вероятность того, что в серии из n независимых испытаний, число появлений события A будет заключено между двумя заданными значениями m_1 и m_2, т.е. $m_1 \le m \le m_2$.

Введем обозначение

$$P_n(m_1;m_2) = P_n(m_1 \le m \le m_2). \qquad (9.7)$$

Очевидно,

$$P_n(m_1;m_2) = P_n(m_1) + P_n(m_1+1) + ... + P_n(m_2). \qquad (9.8)$$

Таким образом, можно вычислить $P_n(m_1;m_2)$, вычислив каждое слагаемое суммы (9.8) по формуле Бернулли или по локальной формуле Муавра-Лапласа.

Существует, к счастью, более рациональный путь, а именно, применение интегральной формулы Муавра-Лапласа.

$$P_n(m_1;m_2) \approx \Phi(x_2) - \Phi(x_1), \qquad (9.9)$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-\frac{t^2}{2}} dt, \qquad (9.10)$$

$$x_1 = \frac{m_1 - np}{\sqrt{npq}}, \;\; x_2 = \frac{m_2 - np}{\sqrt{npq}}. \qquad (9.11)$$

Формула (9.9) применяется следующим образом.Для заданных n, m_1, m_2, p и q вычисляем значения x_1 и x_2 аргумента функции Ф(х) по формулам (9.11). Функция (9.10) называется *функцией Лапласа*. Она табулирована для значений $x \in [0;4]$. После вычисления x_1 и x_2 из таблицы находим $\Phi(x_1)$ и $\Phi(x_2)$. Разность $\Phi(x_2)$-$\Phi(x_1)$ дает приближенное значение вероятности $P_n(m_1;m_2)$. Если нужно вычислить значение Ф(х) для х>4, то можно положить Ф(х)=0,5. Для отрицательных значений аргумента используем свойство нечетности, т.е.

$$\Phi(-x)=-\Phi(x).$$

Пусть в результате n независимых испытаний (экспериментов) вероятность появления события А изменяется от одного испытания к другому и принимает значения $p_1, p_2, ..., p_n$. Тогда вероятность того, что событие A произойдет m раз в этой серии будет равно коэффициенту при x^m многочлена

$$\varphi(x) = (q_1 + p_1 x) \cdot \ldots \cdot (q_n + p_n x), \qquad (9.12)$$

где $q_k = 1 - p_k$, k=1, 2, …, n.

Пусть в результате опыта, который повторяется n раз, может произойти одно из событий $A_1, A_2, ..., A_k$. Обозначим $p_1 = P(A_1),\ p_2 = P(A_2), ..., p_k = P(A_k)$. Очевидно,

$$p_1 + p_2 + \ldots + p_k = 1. \qquad (9.13)$$

Если обозначить через $P_n(m_1; m_2; ...; m_k)$ - вероятность того, что в этих n испытаниях A_1 произойдет m_1 раз; A_2 произойдет m_2 раз; … A_k произойдет m_k раз, то

$$P_n(m_1; m_2; ...; m_k) = \frac{n!}{m_1! ... m_k!} p_1^{m_1} p_2^{m_2} ... p_k^{m_k}. \qquad (9.14)$$

В некоторых испытаниях вероятность появления A, т.е. P(A)=p мала, (близка к 0). Тогда, для больших n вероятность $P_n(m)$ может быть вычислена по формуле *Пуассона*:

$$P_n(m) \approx \frac{\lambda^m}{m!} \cdot e^{-\lambda}, \qquad (9.15)$$

где $\lambda = np$.

Если m – число появлений события A в n независимых испытаниях, то отношение m/n называется *относительной частотой* события A.

Известно, что

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \approx 2 \cdot \Phi\left(\varepsilon \cdot \sqrt{\frac{n}{pq}}\right), \qquad (9.16)$$

где $\varepsilon > 0$ заданное число, Φ(x) – функция Лапласа, p=P(A), $q = P(\overline{A}) = 1 - p$.

Решение типовых задач

Задача 9.1. Вероятность сдать один из 5 экзаменов сессии равна 0,7. Найти вероятности того, что:

а) будут сданы три экзамена;

б) будут сданы два экзамена;

в) будут сданы хотя бы два экзамена.

Решение. а) Испытание состоит в сдаче экзамена, т.е. n=0,7; p=0,7; q=1-p=0,3. Нужно вычислить $P_5(3)$. По формуле (4.1) будем иметь

$$P_5(3) = C_5^3(0{,}7)^3 \cdot (0{,}3)^2 = \frac{5 \cdot 4}{1 \cdot 2} \cdot (0{,}7)^3 \cdot (0{,}3)^2 = 10 \cdot (0{,}7)^3 \cdot (0{,}3)^2 \approx 0{,}31.$$

б) Необходимо вычислить $P_5(2)$. Применим опять (9.1):

$$P_5(2) = C_5^2(0{,}7)^2 \cdot (0{,}3)^3 = \frac{5 \cdot 4}{1 \cdot 2} \cdot 0{,}49 \cdot 0{,}027 \approx 0{,}13.$$

в) $P_5(m \geq 2) = P_5(2) + P_5(3) + P_5(4) + P_5(5).$

Задача, однако, может быть решена более рационально, переходом к противоположному событию.

Пусть A событие – «$m \geq 2$», тогда $\overline{A}$ - «m<2». Следовательно, $P(\overline{A}) = P_5(0) + P_5(1).$

$$P(\overline{A}) = C_5^0(0{,}7)^0 \cdot (0{,}3)^5 + C_5^1 0{,}7 \cdot (0{,}3)^4 = (0{,}3)^5 + 5 \cdot 0{,}7 \cdot 0{,}0081 \approx 0{,}03.$$

$$P(A) = 1 - P(\overline{A}) = 1 - 0{,}03 = 0{,}97.$$

Ответ. $P_5(3) \approx 0{,}31;\ P_5(2) \approx 0{,}13;\ P(m \geq 2) \approx 0{,}97.$

Задача 9.2. По данным метеослужбы вероятность того, что 1-го июня в Курске пойдет дождь равна 4/17. Найти наивероятнейшее число дождевых дней 1-го июня в Курске на протяжении 50 лет.

Решение. Имеем: n=50; $p=\dfrac{4}{17}$; $q=\dfrac{13}{17}$;

$$np+p=50\cdot\frac{4}{17}+\frac{4}{17}=\frac{200}{17}+\frac{4}{17}=\frac{204}{17}=12;$$

$$np-q=\frac{200}{17}-\frac{13}{17}=\frac{187}{17}=11.$$

Поэтому, $11\le m_0\le 12$.

Ответ. $m_0=12; m_0-1=11.$

Задача 9.3. Стреляют 4 раза по цели. Вероятности попаданий соответственно равны 0,4; 0,3; 0,2; и 0,1. Найти вероятность того, что в результате стрельбы:

а) не попадут ни разу, попадут один, два, три раза;

б) попадут хотя бы раз;

в) попадут хотя бы два раза.

Решение. Число опытов n=4.

$$p_1=0{,}4;\quad p_2=0{,}3;\quad p_3=0{,}2 \text{ и } p_4=0{,}1.$$

Следовательно,

$$q_1=1-p_1=1-0{,}4=0{,}6;\quad q_2=1-p_2=1-0{,}3=0{,}7;$$
$$q_3=1-p_3=1-0{,}2=0{,}8;\quad q_4=1-p_4=1-0{,}1=0{,}9.$$

а) Образуем функцию

$$\varphi(x)=(q_1+p_1x)(q_2+p_2x)(q_3+p_3x)(q_4+p_4x)=$$
$$=(0{,}6+0{,}4x)(0{,}7+0{,}3x)(0{,}8+0{,}2x)(0{,}9+0{,}1x)=$$
$$=0{,}302+0{,}460x+0{,}205x^2+0{,}031x^3+0{,}002x^4.$$

Таким образом, получим следующий

Ответ.

а) $P_4(0)=0{,}302; P_4(1)=0{,}46; P_4(2)=0{,}205; P_4(3)=0{,}031;$ $P_4(4)=0{,}002;$

б) $P_4(m\ge 1)=1-P_4(0)=0-0{,}302=0{,}698;$

в) $P_4(m \geq 2) = 1 - [P_4(0) + P_4(1)] = 1 - 0{,}302 - 0{,}46 = 0{,}238.$

Задача 9.4. Цель разделена на три сектора. Вероятность попадания в эти сектора в результате одного выстрела соответственно равны 0,5; 0,3; и 0,2. Найти вероятность того, что в результате 6 выстрелов попадем 3 раза в первый сектор, 2 раза – во второй и один раз в третий сектор.

Решение. Имеем: n=6; $p_1 = 0{,}5; p_2 = 0{,}3; p_3 = 0{,}2.$

$$m_1 = 3; m_2 = 2; m_3 = 1.$$

По формуле (9.4) получим

$$P_6(3;2;1) = \frac{6!}{3! \cdot 2! \cdot 1!} \cdot (0{,}5)^3 \cdot (0{,}3)^2 \cdot 0{,}2 \approx 0{,}135.$$

Ответ. $P_6(3;2;1) \approx 0{,}135.$

Задача 9.5. Найти вероятность того, что событие А произойдет 80 раз в результате повторения опыта 400 раз, если А происходит в одном опыте с вероятностью 0,2.

Решение. По данным задачи:

$$n=400;\ m=80;\ p=P(A)=0{,}2;\ q = p(\overline{A}) = 1 - p = 0{,}8.$$

Так как n и m числа большие, то применим локальную формулу Муавра-Лапласа (9.4) для вычисления вероятности $P_{400}(80)$. Имеем:

$$P_{400}(80) \approx \frac{1}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} \cdot \varphi(x_{80,\,400}),$$

где

$$x_{80,\,400} = \frac{80 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = \frac{80 - 80}{8} = 0.$$

Следовательно,

$$P_{400}(80) \approx \frac{1}{8} \cdot \varphi(0) = \frac{0,3989}{8} \approx 0,05$$

Ответ. $P_{400}(80) \approx 0,05$.

Задача 9.6. В магазин поступает обувь некоторой фирмы. Вероятность того, что пара обуви качественная, равна 0,8. Найти вероятность того, что из 400 полученных пар некачественными будут не менее 70 и не более 100 пар.

Решение. Необходимо вычислить $P_{400}(70;100)$. Для этого применим интегральную формулу Муавра-Лапласа.

$$x_1 = \frac{70 - 400 \cdot 0,2}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = -\frac{10}{8} = -1,25; \; x_2 = \frac{100 - 400 \cdot 0,2}{\sqrt{400 \cdot 0,2 \cdot 0,82}} = \frac{20}{8} = 2,5.$$

Следовательно,

$$P_{400}(70;100) \approx \Phi(2,5) - \Phi(-1,25) = \Phi(2,5) + \Phi(1,25) =$$
$$= 0,4938 + 0,3944 = 0,8882.$$

Ответ. $P_{400}(70;100) \approx 0,89$.

Задача 9.7. В каждом из 500 независимых испытаний событие A происходит с вероятностью p=P(A)=0,4. Найти вероятность того, что число m появлений события A будет находится между 180 и 240.

Решение. Имеем: n=500; p=0,4; q=0,6. Нужно найти $P_{500}(180 \le m \le 240) = P_{500}(180; 240)$. Используем интегральную формулу Муавра-Лапласа (9.9).

$$\sqrt{npq} = \sqrt{500 \cdot 0,4 \cdot 0,6} = \sqrt{120} \approx 11.$$

$$x_1 = \frac{180 - 500 \cdot 0,4}{\sqrt{500 \cdot 0,4 \cdot 0,6}} \approx -1,82;$$

$$x_2 = \frac{240 - 500 \cdot 0,4}{11} \approx 3,64.$$

Следовательно,

$$P_{500}(180;240) \approx \Phi(3{,}64) - \Phi(-1{,}82) = \Phi(3{,}64) + \Phi(1{,}82) =$$
$$= 0{,}4999 + 0{,}4656 = 0{,}9655.$$

Ответ. $P_{500}(180;240) \approx 0{,}96.$

Задача 9.8. На телефонной станции неправильные соединения случаются с вероятностью p=1/200. Найти вероятность того, что из 200 соединений неправильными будут:

а) одно соединение; б) более двух соединений.

Решение. а) Так как n велико, а p мало, то применим формулу Пуассона (9.15).

$$n = 200;\ p = \frac{1}{200};\ m = 1;\ \lambda = np = 200 \cdot \frac{1}{200} = 1.$$

Из таблицы находим $P_{500}(1) = e^{-1} \approx 0{,}3679.$

б) Дано $n = 200;\ p = \dfrac{1}{200};\ m > 2;\ \lambda = 1.$ Имеем

$$P_{200}(m > 2) = 1 - P_{200}(m \le 2) = 1 - [P_{200}(0) + P_{200}(1) + P_{200}(2)] \approx$$
$$\approx 1 - [0{,}3679 + 0{,}3679 + 0{,}1839] \approx 1 - 0{,}9197 \approx 0{,}0803.$$

Следовательно, $P_{200}(m > 2) \approx 0{,}0803.$

Ответ. $P_{200}(1) \approx 0{,}37;\ P_{200}(m > 2) \approx 0{,}08.$

Задача 9.9. Вероятность появления события A в каждом из 10000 независимых испытаний равна 0,75. Найти вероятность того, что относительная частота события отклонится от его вероятности по абсолютной величине не более чем на 0,001.

Решение. Воспользуемся формулой (9.16). Подставляя сюда данные задачи, будем иметь:

$$P\left(\left|\frac{m}{10000} - 0{,}75\right| < 0{,}001\right) \approx 2\Phi\left(0{,}001 \cdot \sqrt{\frac{10000}{0{,}75 \cdot 0{,}25}}\right) =$$
$$= 2\Phi\left(0{,}001 \cdot \sqrt{\frac{10000}{(3/4)\cdot(1/4)}}\right) =$$
$$= 2\Phi(0{,}001 \cdot 400 \cdot 1/\sqrt{3} = 2\Phi(0{,}4/\sqrt{3}) = 2\Phi(0{,}23) \approx 0{,}182.$$

Ответ. Вероятность отклонения равна 0,182.

Задачи для упражнений

1. Процент всхожести семян пшеницы составляет 80 из 100. Найти вероятность того, что из 6 семян взойдут:

а) три семени;

б) не менее трех семян;

в) четыре семени.

2. В семье четверо детей. Считая, что рождение мальчика или девочки равновозможное, найти вероятности событий:

а) в семье три мальчика;

б) в семье хотя бы три мальчика;

в) в семье два мальчика.

3. Торговая база обслуживает 6 магазинов. Вероятность получения заказа от магазина равна 0,6. Найти вероятности того, что:

а) база получит заказы от 5 магазинов;

б) база получит не менее пяти заказов;

в) база получит не менее четырех заказов.

4. Цель имеет форму квадрата со стороной а, в который вписан круг. По этой цели производится 4 выстрела. Найти вероятность того, что попадем 3 раза в круг.

5. Вероятность поражения цели равна 0,3. Каким должно быть наименьшее число выстрелов, чтобы наивероятнейшее число попаданий было бы равно 25?

6. По статистическим данным 20% населения имеет черный цвет волос, 30% - темный, 40% - светлый и 10% - рыжий. Случайно формируется группа из 6 человек. Найти вероятности того, что:

а) в группе будет хотя бы 3 светловолосых;

б) в группе будет хотя бы один рыжий;

в) в группе будет 2 светловолосых, 2 рыжих, один с темными волосами и один с черными.

7. Вероятность выигрыша по лотерейному билету равна 1/3. Найти вероятности того, что из 6 билетов выигрышными будут:

а) два билета; б) хотя бы два билета.

8. Производится пять независимых выстрелов по цели. Вероятность попадания в цель при одном выстреле равна 0,2. Для

уничтожения цели достаточно трех попаданий. Найти вероятность того, что цель будет уничтожена.

9. В поезд с 6 вагонами садятся 12 пассажиров, которые могут сесть в любой вагон. Найти вероятности того, что:

а) в каждый вагон сядет по 2 пассажира;

б) в один вагон не сядет ни один пассажир, в другой сядет один пассажир, в два вагона сядет по 2 пассажира, а в остальные сядут 3 и 4 пассажира.

10. Магазин получил 1000 бутылок с минеральной водой. Вероятность того, что при транспортировке одна бутылка будет разбита равна 0,003. Найти вероятности того, что магазин получит:

а) две разбитые бутылки;

б) более двух разбитых бутылок.

11. Вероятность того, что пара сапог является высшего качества, равна 0,5. В магазин поступило 400 пар сапог. Найти вероятность того, что среди них будет не менее 194 и не более 208 пар высшего качества.

12. Вероятность появления события А в одном из 900 независимых испытаний равна 0,5. Найти вероятность того, что отклонение относительной частоты события от его вероятности, по модулю не превысит 0,02.

13. Вероятность появления события А в одном из 1000 независимых опытов равна 0,75. Найти $P\left(\left|\frac{m}{n}-p\right|<0{,}01\right)$.

14. Вероятность появления события А в результате одного опыта равна 0,5. Найти число необходимых опытов для того, чтобы с вероятностью 0,7698, можно было бы утверждать: относительная частота отклонится от вероятности события по модулю не более чем на 0,02.

15. Вероятность появления события А в каждом из 600независимых опытов равна 0,85. Вычислить $P\left(\left|\frac{m}{n}-p\right|<0{,}0055\right)$.

16. Вероятность того, что покупателю нужна обувь 40 размера, равна 0,2. Найти вероятность того, что из пяти покупателей обувь этого размера необходима: а) одному, б) по крайней мере одному покупателю.

17. Вероятность изготовления стандартной детали равна 0,9. Найти вероятность того, что среди 10 деталей окажется не более одной нестандартной.

18. Вероятность того, что пассажир опоздает к отплытию катера, равна 0,02. Найти наивероятнейшее число опоздавших, если общее число пассажиров равно 855.

19. Оптовая база обслуживает 12 магазинов. От каждого из них заявка на товары на следующий день может поступать с вероятностью 0,3. Найти наивероятнейшее число заявок и вероятность получения базой такого числа заявок.

20. Вероятность попадания в цель при одном выстреле равна 0,4. Найти вероятность того, что из 320 выстрелов будет 100 попаданий.

21. Вероятность поражения цели при одном выстреле равна 0,4. Найти вероятность того, что цель будет поражена не менее 200 и не более 250 раз при 600 выстрелах.

22. На прядильной фабрике работница обслуживает 720 веретен. При вращении веретена пряжа рвется в случайные моменты времени с вероятностью 0,008. Найти вероятность того, что за некоторое время произойдет не более 10 обрывов.

23. При штамповке металлических клемм получается в среднем 90% годных. Найти вероятность того, что среди 900 клемм будет от 790 до 820 годных.

24. Всхожесть семян данного растения составляет 90%. Найти вероятность того, что из 800 посеянных семян взойдет не менее 700.

25. Найти вероятность того, что из 500 посеянных семян не взойдет 130, если всхожесть семян оценивается вероятностью $p=0,75$.

26. Вероятность появления события A в одном испытании равна 0,2. Найти вероятность того, что в 400 испытаниях событие A наступит 104 раза.

27. Мебельная фабрика выпускает столы. Вероятность того, что стол является высшего качества, равна 0,75. Найти наивероятнейшее число столов высшего качества среди 150 выпущенных фабрикой.

28. Сколько раз с вероятностью 0,048 можно ожидать появление события А в 100 независимых испытаниях, если

вероятность его появления в отдельном испытании равна 0,5?

29. Было посажено 400 деревьев. Вероятность того, что дерево приживется равна 0,8. Найти вероятность того, что число прижившихся деревьев будет больше 250.

30. Вероятность наступления события A в каждом испытании равна 0,8. Испытания независимы. Сколько нужно произвести испытаний, чтобы с вероятностью 0,95 можно было бы ожидать отклонение относительной частоты события от его вероятности не более чем на 0,04?

Практическое занятие №10
Дискретные случайные величины

Краткие теоретические сведения

Случайной величиной называется величина, которая в результате эксперимента (опыта, испытания) принимает одно значение из множества возможных.

Примеры. Число выпавших очков при бросании игральной кости; число студентов группы, которые сдадут экзамен на 8; дальность полета снаряда при стрельбе; число дождливых дней в году; ошибка, допущенная при измерении.

Общей характеристикой любой случайной величины является ее *закон распределения*.

Законом распределения случайной величины называется соотношение, устанавливающее связь между ее возможными значениями и соответствующими этим значениям вероятностями.

Хорошо известны два типа случайных величин: *дискретные* и *непрерывные*.

Случайная величина называется *дискретной*, если множество ее возможных значений конечно или счетно.

Примеры. Число выпавших гербов при десятикратном бросании монеты является дискретной случайной величиной с возможными значениями: {0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10}.

В магазин завезли партию из 100 костюмов. Среди них имеются костюмы с дефектами. Очевидно, что число костюмов с дефектами является дискретной случайной величиной с множеством возможных значений {0; 1; 2; ...; 100}.

Будем обозначать случайные величины греческими буквами: $\xi, \eta, \delta, \theta, \ldots$ (иногда употребляя индексы).

Закон распределения дискретной величины может иметь вид *ряда распределения* или *функции распределения*.

В дальнейшем будем рассматривать случайные величины с конечным множеством значений.

Пусть $x_1, x_2, \ldots, x_n$ - возможные значения дискретной случайной величины ξ. Обозначим

$$p_i = P(\xi = x_i), \quad i = 1, 2, ..., n. \tag{10.1}$$

Матрица

$$\xi : \begin{pmatrix} x_1, & x_2, & ..., & x_n \\ p_1, & p_2, & ..., & p_n \end{pmatrix} \tag{10.2}$$

называется *рядом распределения* дискретной случайной величины ξ , $(x_1 < x_2 < ... < x_n)$.

Имеет место соотношение

$$\sum_{i=1}^{n} p_i = 1. \tag{10.3}$$

<u>Пример.</u> Для случайной величины ξ - числа выпавших очков при бросании игральной кости, закон распределения следующий:

$$\xi : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}.$$

Функция распределения дискретной случайной величины имеет вид:

$$F_\xi(x) = P(\xi < x) = \sum_{i:\, x_i < x} P(\xi = x_i) = \sum_{i:\, x_i < x} p_i. \tag{10.4}$$

Математическое ожидание дискретной величины вычисляется по формуле:

$$M\xi = \sum_{i=1}^{n} x_i p_i. \tag{10.5}$$

Дисперсия дискретной величины ξ:

$$D\xi = M(\xi - M\xi)^2 = \sum_{i=1}^{n} (x_i - M\xi)^2 p_i. \qquad (10.6)$$

Среднее квадратичное отклонение

$$\sigma\xi = \sqrt{D\xi}. \qquad (10.7)$$

Используя понятие математического ожидания, можно определить моменты случайной величины.

Начальный момент порядка k, $k \in N$:

$$a_k = M\xi^k = \sum_{i=1}^{n} x_i^k p_i. \qquad (10.8)$$

Две случайные величины называются *независимыми*, если закон распределения одной из них не зависит от того, какие значения приняли другая величина.

Математическое ожидание $M\xi$ обладает свойствами:

$$Mc = c, \quad c\text{ - константа.}$$
$$M(c\xi) = cM\xi.$$
$$M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2.$$

Если ξ_1 и ξ_2 независимы, то $M(\xi_1 \cdot \xi_2) = M\xi_1 \cdot M\xi_2$.

Дисперсия $D\xi$ имеет свойства:

$$Dc = 0.$$
$$D(c\xi) = c^2 D\xi, \quad c\text{ - вещественная константа.}$$
$$D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2, \quad \text{если } \xi_1 \text{ и } \xi_2 \text{ независимы.}$$

Центральный момент порядка k, $k \in N$.

$$\mu_k = M(\xi - M\xi)^k = \sum_{i=1}^{n} (x_i - M\xi)^k p_i. \qquad (10.9)$$

Очевидно,

$$M\xi = \alpha_1, \quad D\xi = \mu_2. \qquad (10.10)$$

Для вычисления дисперсии часто применяется формула:

$$D\xi = M\xi^2 - (M\xi)^2. \qquad (10.11)$$

Решение типовых задач

Задача 10.1. Случайная величина ξ имеет закон распределения:

$$\xi : \begin{pmatrix} 3 & 5 & 7 & 11 \\ 0{,}14 & 0{,}20 & 0{,}49 & 0{,}17 \end{pmatrix}.$$

Найти функцию распределения $F_\xi(x)$, математическое ожидание $M\xi$, дисперсию $D\xi$ и среднее квадратичное отклонение $\sigma\xi$.

Решение. Используя формулу (5.4), находим выражение для функции распределения:

$$F_\xi(x) = \begin{cases} 0; & x \le 3 \\ 0{,}14; & 3 < x \le 5 \\ 0{,}34; & 5 < x \le 7 \\ 0{,}83; & 7 < x \le 11 \\ 1; & x > 11 \end{cases}. \qquad (10.12)$$

График функции распределения $F_\xi(x)$ (рисунок 10.1).

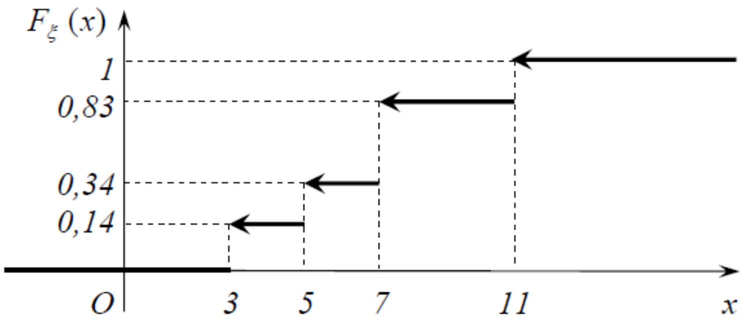


Рисунок 10.1 - График функции распределения $F_\xi(x)$

Вычислим математическое ожидание по формуле (10.5):

$$M\xi = 3\cdot 0{,}14 + 5\cdot 0{,}20 + 7\cdot 0{,}49 + 11\cdot 0{,}17 = 6{,}72.$$

Для дисперсии будем иметь:

$$D\xi = \sum_{i=1}^{n}(x_i - M\xi)^2 p_i = (3-6{,}72)^2\cdot 0{,}14 + (5-6{,}72)^2\cdot 0{,}2 +$$
$$+(7-6{,}72)^2\cdot 0{,}49 + (11-6{,}72)^2\cdot 0{,}17 =$$
$$=(-3{,}72)^2\cdot 0{,}14 + (-1{,}72)^2\cdot 0{,}2 + (0{,}28)^2\cdot 0{,}49 + (4{,}28)^2\cdot 0{,}17 =$$
$$=1{,}94 + 0{,}59 + 0{,}04 + 3{,}11 = 5{,}68.$$
$$\sigma\xi = \sqrt{D\xi} = \sqrt{5{,}68} \approx 2{,}38.$$

Ответ. Функция распределения имеет вид (10.12). $M\xi = 6{,}72$, $D\xi = 5{,}68$, $\sigma\xi = 2{,}38$.

Задача 10.2. Бросают 2 игральные кости. Пусть ξ - сумма выпавших очков на обеих костях. Написать закон распределения ξ и найти математическое ожидание $M\xi$.

Решение. Множество возможных значений ξ совпадает с множеством: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. Для нахождения соответствующих вероятностей посмотрим, как получаются эти значения суммы. Для этого изобразим таблицу, где под каждым значением суммы напишем ее составляющие: первое слагаемое это число очков на первой кости, а второе слагаемое - очки, которые выпали на второй кости.

2	3	4	5	6	7	8	9	10	11	12
1+1	1+2 2+1	1+3 2+2 3+1	1+4 2+3 3+2 4+1	1+5 2+4 3+3 4+2 5+1	1+6 2+5 3+4 4+3 5+2 6+1	2+6 3+5 4+4 5+3 6+2	3+6 4+5 5+4 6+3	4+6 5+5 6+4	5+6 6+5	6+6

Каждой сумме из этой таблицы соответствует вероятность 1/36. Тем самым, приходим к распределению:

$$\xi:\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 1/36 & 2/36 & 3/36 & 4/36 & 5/36 & 6/36 & 5/36 & 4/36 & 3/36 & 2/36 & 1/36 \end{pmatrix}.$$

После необходимых преобразований будем иметь:

$$\xi:\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 1/36 & 1/18 & 1/12 & 1/9 & 5/36 & 1/6 & 5/36 & 1/9 & 1/12 & 1/18 & 1/36 \end{pmatrix}. \quad (10.13)$$

Следовательно,

$$M\xi=\sum_{i=1}^{n}x_i p_i=2\cdot\frac{1}{36}+3\cdot\frac{1}{18}+4\cdot\frac{1}{12}+5\cdot\frac{1}{9}+6\cdot\frac{5}{36}+$$
$$+7\cdot\frac{1}{6}+8\cdot\frac{5}{36}+9\cdot\frac{1}{9}+10\cdot\frac{1}{12}+$$
$$+11\cdot\frac{1}{18}+12\cdot\frac{1}{36}=\frac{126}{18}=7.$$

Ответ. Закон распределения задан матрице (10.13). $M\xi=7$.

Задача 10.3. Производят три выстрела по цели. Вероятность попадания в цель при одном выстреле равна 0,3. Найти закон распределения случайной величины ξ - числа попаданий. Вычислить математическое ожидание, дисперсию. Найти также функцию распределения величины ξ.

<u>Решение.</u> Очевидно, что возможными значениями ξ будут 0, 1, 2, 3. Соответствующие этим значениям вероятности вычислим по формуле Бернулли (4.1): n=3; p=0,3; q=0,7. Поэтому,

$$P(\xi = 0) = C_3^0(0;3)^0 \cdot (0,7)^3 = 0,343.$$

$$P(\xi = 1) = C_3^1(0;3)^1 \cdot (0,7)^2 = 3 \cdot 0,3 \cdot 0,49 = 0,441.$$

$$P(\xi = 2) = C_3^2(0;3)^2 \cdot (0,7)^1 = 3 \cdot 0,009 \cdot 0,7 = 0,189.$$

$$P(\xi = 3) = C_3^3(0;3)^3 \cdot (0,7)^0 = 0,027.$$

Таким образом, случайная величина ξ имеет распределение:

$$\xi : \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0,343 & 0,441 & 0,189 & 0,027 \end{pmatrix} \qquad (10.14)$$

$$M\xi = 0 \cdot 0,343 + 1 \cdot 0,441 + 2 \cdot 0,189 + 3 \cdot 0,027 = 0,441 + 0,378 + 0,081 = 0,9.$$

Для вычисления дисперсии случайной величины ξ воспользуемся формулой (10.11). Величина ξ^2 будет иметь распределение:

$$\xi : \begin{pmatrix} 0^2 & 1^2 & 2^2 & 3^2 \\ 0,343 & 0,441 & 0,189 & 0,027 \end{pmatrix}$$

$$M\xi^2 = 0 \cdot 0,343 + 1 \cdot 0,441 + 4 \cdot 0,189 + 9 \cdot 0,027 = 0,441 + 0,756 + 0,243 = 1,44.$$

$$D\xi = 1,44 - (0,9)^2 = 1,44 - 0,81 = 0,63.$$

Функция распределения величины ξ будет ступенчатой функцией:

$$F_{\xi}(x)=\begin{cases}0; & x\le 0\\ 0{,}343; & 0<x\le 1\\ 0{,}784; & 1<x\le 2.\\ 0{,}973; & 2<x\le 3\\ 1; & x>3\end{cases} \qquad (10.15)$$

Ответ. Закон распределения ξ задан матрицей (10.14). $M\xi=0{,}9$. $D\xi=0{,}63$. Функция распределения имеет вид (10.15).

Задача 10.4. В урне находятся 3 белых и 5 черных шаров. Из нее извлекают 3 шара. Найти закон распределения и математическое ожидание числа извлеченных черных шаров.

Решение. Пусть ξ - число извлеченных черных шаров. Возможными значениями ξ будут: 0, 1, 2, 3. Вычислим соответствующие вероятности. Имеем:

$$P(\xi=0)=\frac{C_3^3}{C_8^3}=\frac{1}{56};\quad P(\xi=1)=\frac{C_5^1\cdot C_3^2}{C_8^3}=\frac{15}{56};$$

$$P(\xi=2)=\frac{C_5^2\cdot C_3^1}{C_8^3}=\frac{30}{56};\quad P(\xi=3)=\frac{C_5^3}{C_8^3}=\frac{10}{56}.$$

Закон распределения величины ξ:

$$\xi:\begin{pmatrix}0 & 1 & 2 & 3\\ 1/56 & 15/56 & 30/56 & 10/56\end{pmatrix}. \qquad (10.16)$$

$$M\xi=0\cdot\frac{1}{56}+1\cdot\frac{15}{56}+2\cdot\frac{30}{56}+3\cdot\frac{10}{56}=\frac{105}{56}=1{,}875.$$

Ответ. Закон распределения имеет вид (10.16). $M\xi=1{,}875$.

Задача 10.5. Случайная величина ξ принимает значения: $x_1=-1; x_2=0; x_3=1; M\xi=0{,}1; M\xi^2=0{,}9$. Найти вероятности p_1, p_2 и p_3.

Решение. Закон распределения величины ξ:

$$\xi : \begin{pmatrix} -1 & 0 & 1 \\ p_1 & p_2 & p_3 \end{pmatrix}.$$

Так как в законе распределения сумма вероятностей равна 1, получаем соотношение $p_1 + p_2 + p_3 = 1$. Второе и третье соотношения получим, использовав формулы для $M\xi$ и $M\xi^2$:

$$(-1) \cdot p_1 + 0 \cdot p_2 + 1 \cdot p_3 = 0{,}1;$$
$$(-1)^2 \cdot p_1 + 0^2 \cdot p_2 + 1^2 \cdot p_3 = 0{,}9.$$

После этого приходим к системе уравнений для p_1, p_2 и p_3.

$$\begin{cases} p_1 + p_2 + p_3 = 1 \\ -p_1 + p_3 = 0{,}1 \\ p_1 + p_3 = 0{,}9 \end{cases}$$

Решив эту систему, получим значения вероятностей.

Ответ. $p_1 = 0{,}4;\ p_2 = 0{,}1;\ p_3 = 0{,}5.$

Задачи для упражнений

1. Монета бросается 7 раз. Вычислить математическое ожидание и дисперсию числа выпавших гербов.
2. Два стрелка производят независимо друг от друга по два выстрела по цели. Вероятность попадания в цель для первого стрелка равна 0,5; а для второго - 0,6. Найти закон распределения, математическое ожидание и дисперсию случайной величины ξ - числа попаданий в цель.
3. В блоке из 50 пачек сигарет находятся 6 пачек с дефектом. Некто покупает 5 пачек. Найти закон распределения, математическое ожидание и дисперсию случайной величины ξ - числа купленных пачек с дефектом.
4. Даны дискретные случайные величины:

$$\xi_1:\begin{pmatrix}-4 & 6 & 10\\ 0,2 & 0,3 & 0,5\end{pmatrix};\quad \xi_2:\begin{pmatrix}0,21 & 0,54 & 0,61\\ 0,1 & 0,5 & 0,4\end{pmatrix}.$$

Найти математические ожидания $M\xi_1$ и $M\xi_2$.

5. Производятся 4 выстрела по цели. Вероятности попадания в цель равны соответственно p_1 =0,6; p_2 =0,4; p_3 =0,5; p_4 =0,2. Найти закон распределения, вычислить математическое ожидание, дисперсию и среднее квадратичное отклонение числа попаданий в цель.

6. Проверяется функционирование некоторого устройства. Вероятность того, что оно не функционирует, равна 0,2. Найти закон распределения числа испорченных устройств, если проверяется 10 устройств. Найти математическое ожидание, дисперсию и среднее квадратичное отклонение этого числа.

7. Бросаются 2 игральные кости. Пусть ξ - произведение числа выпавших очков на обеих костях. Найти закон распределения, математическое ожидание, дисперсию и среднее квадратное отклонение случайной величины ξ.

8. Куплено 10 лотерейных билетов. Вероятность выигрыша по одному билету равна 0,4. Найти закон распределения, математическое ожидание и дисперсию числа выигрышных билетов.

9. Случайная величина ξ принимает значения x_1 и x_2. $P(\xi = x_1) = 0,3; P(\xi = x_2) = 0,7$. Найти x_1 и x_2, еслиизвестно, что $M\xi$ =2,7; $D\xi$ =0,21.

10. Делаются попытки открыть замок, имея в распоряжение 8 ключей, которые берутся наудачу. Найти закон распределения, математическое ожидание и дисперсию числа попыток в случае когда:

а) Ключ, которым сделана неудачная попытка, исключается из дальнейшего выбора.

б) Ключ, которым сделана неудачная попытка, не исключается из дальнейшего выбора.

11. Из урны, содержащей 5 белых и 3 черных шара, извлекают последовательно по одному шару до появления белого. Найти закон распределения, математическое ожидание и дисперсию числа вынутых черных шаров, если вынутый шар обратно в урну не возвращается.

12. Из партии деталей берется последовательно для контроля по одной детали. Если взятая деталь имеет дефект, то контроль на этом завершается. Исходя из этих условий, были проверены не более

5 деталей. Вероятность того, что деталь имеет дефект, равна 0,1. Найти закон распределения, математическое ожидание и дисперсию числа проверенных деталей.

13. В партии содержатся 10 деталей, среди которых 3 имеют дефект. Случайно берут 3 детали. Найти закон распределения, математическое ожидание и дисперсию числа деталей, оказавшихся с дефектом среди выбранных.

14. Случайная величина ξ принимает два значения x_1 и x_2 ($x_1 < x_2$). $P(\xi = x_1) = 0{,}6; M\xi = 0{,}4; D\xi = 0{,}24$. Найти закон распределения случайной величины ξ.

15. Автомобиль должен проехать по улице, на которой установлено три светофора, дающие независимо друг от друга зеленый сигнал в течение 1,5 минут, желтый в течение 0,3 минут и красный в течение 1,2 минут. Написать закон распределения ξ - числа остановок автомобиля на этой улице.

16. Из партии в 25 деталей, среди которых имеется 6 нестандартных, выбраны случайным образом 3 детали. Построить закон распределения случайного числа ξ нестандартных деталей, содержащихся в выборке.

17. Два баскетболиста по очереди забрасывают мяч в корзину с вероятностью попадания при каждом броске для первого 0,8, для второго - 0,7. Всего производится 5 бросков. Составить законы распределения числа попаданий для каждого игрока, если начинает бросать первый баскетболист.

18. Монету подбрасывают 6 раз. Составить ряд распределения и построить функцию распределения отношения числа появления герба к числу появления номинала монеты.

19. Стрелок производит шесть выстрелов по мишени. Вероятность попадания при одном выстреле равна 0,2. Найти функцию распределения ξ - числа попаданий в мишень. С ее помощью найти $P(1 \le \xi \le 5)$.

20. В некотором цехе брак составляет 5% всех изделий. Составить закон распределения ξ - числа бракованных изделий из шести взятых наудачу. Найти математическое ожидание и дисперсию случайной величины ξ.

4 семестр

Практическое занятие №1
Интервальное оценивание

1.1 Понятие доверительного интервала

Пусть имеется выборка $X = (X_1, \ldots, X_n)$ из распределения F_θ с неизвестным параметром $\theta \in \Theta \subseteq R$. Задача интервального оценивания заключается в том, чтобы найти интервал, который накрывает оцениваемый параметр с заданной наперед вероятностью.

Интервал (θ^-, θ^+) называется доверительным интервалом для параметра θ уровня доверия $1-\varepsilon$, если для любого $\theta \in \Theta$

$$P_\theta\left(\theta^- < \theta < \theta^+\right) \geq 1-\varepsilon \qquad (1.1)$$

Если в соотношении (1.1) вероятность в точности равна $1-\varepsilon$ (или стремится к $1-\varepsilon$), то интервал называют точным (или асимптотически точным) доверительным интервалом уровня доверия $1-\varepsilon$.

1.2 Построение точных доверительных интервалов для параметров нормального распределения

Пусть $X = (X_1, \ldots, X_n)$ - выборка объема n из нормального распределения N_{a,σ^2}. Рассмотрим возможные задачи интервального оценивания:

1. Пусть известно σ^2, а $a \in R$ - неизвестный параметр. Требуется построить точный доверительный интервал для параметра a.

$$P_{a,\sigma^2}\left(\overline{X} - \frac{\tau_{1-\varepsilon/2} \cdot \sigma}{\sqrt{n}} < a < \overline{X} + \frac{\tau_{1-\varepsilon/2} \cdot \sigma}{\sqrt{n}}\right) = 1-\varepsilon, \qquad (1.2)$$

где $\tau_{1-\varepsilon/2}$ - квантиль уровня $1-\varepsilon/2$ стандартного нормального распределения (см. Приложение Б, таблица 1).

2. Пусть известен параметр a, требуется построить доверительный интервал для σ^2.

$$P_{a,\sigma^2}\left(\frac{n\cdot s_1^2}{g_2}<\sigma^2<\frac{n\cdot s_1^2}{g_1}\right)=1-\varepsilon, \qquad (1.3)$$

где s_1^2 - выборочная дисперсия, g_1 и g_2 - квантили распределения "хи- квадрат" с n степенями свободы ($\chi^2_{\alpha,n}$) уровня $a=\varepsilon/2$ и $a=1-\varepsilon/2$ соответственно (см. Приложение Б, таблица 3).

3. Доверительный интервал для σ^2 при неизвестном a:

$$P_{a,\sigma^2}\left(\frac{(n-1)\cdot s_0^2}{g_2}<\sigma^2<\frac{(n-1)\cdot s_0^2}{g_1}\right)=1-\varepsilon, \qquad (1.4)$$

где s_0^2 - несмещенная выборочная дисперсия.

4. Доверительный интервал для a при неизвестном σ^2:

$$P_{a,\sigma^2}\left(\overline{X}-\frac{t_{1-\varepsilon/2,n-1}\cdot s_0}{\sqrt{n}}<a<\overline{X}+\frac{t_{1-\varepsilon/2,n-1}\cdot s_0}{\sqrt{n}}\right)=1-\varepsilon, \qquad (1.5)$$

где $t_{1-\varepsilon/2,n-1}$ - квантиль распределения Стьюдента с $n-1$ степенью свободы уровня $1-\varepsilon/2$ (см. Приложение Б, таблица 2).

Приведем пример построения доверительного интервала.

Пусть имеется выборка объема N из нормального распределения, для которого известен параметр $\sigma^2=4$, а параметр a неизвестен. Требуется построить точный доверительный интервал для параметра a. На рисунке 1.1 приведен текст программы, реализующей метод построения доверительного интервала в среде Mathcad.

Ввод исходных данных
x - вектор, представляющий выборку из нормального распределения
dx - известное среднеквадратическое отклонение
q=1-e - уровень доверия

x :=

C:\..\v1.txt

вектор, представляющий выборку считывается из файла

$x^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.432	-0.501	2.332	0.494	3.503	1.233	-1.313	1.969	2.207	1.188

вывод на экран элементов выборки (их можно просмотреть, щелкнув по вектору x и воспользовавшись линейкой прокрутки)

$dx := 2$ $\qquad q := 0.9$ $\qquad$ ввод значений среднеквадратического отклонения и уровня доверия

$N := \mathrm{length}(x)$ $\qquad N = 700$ $\qquad$ нахождение обьема выборки

$Mx := \dfrac{\sum_{i=0}^{N-1} x_i}{N}$ $\qquad Mx = 1.97$ $\qquad$ вычисление среднего выборочного

$p := 1 - \dfrac{1-q}{2}$

$v := \mathrm{qnorm}(p,0,1)$ $\qquad v = 1.645$ $\qquad$ нахождение квантиля стандартного нормального распределения уровня 1-e/2

$upper := Mx + \dfrac{v \cdot dx}{\sqrt{N}}$ $\qquad upper = 2.094$

нахождение границ доверительного интервала

$lower := Mx - \dfrac{v \cdot dx}{\sqrt{N}}$ $\qquad lower = 1.846$

$length := upper - lower$ $\qquad length = 0.249$ $\qquad$ вычисление длины доверительного интервала

Рисунок 1.1. Построение доверительного интервала для параметра а нормального распределения при известном σ

1.3 Задание к практическому занятию

Даны две выборки одной случайной величины с нормальным распределением N_{a,σ^2} объема n_1 и n_2 соответственно.

Для вариантов с нечетным номером:

1. Для обеих выборок построить точный доверительный интервал уровня доверия q_0 для параметра a, считая:

а) σ неизвестным,

б) σ известным и равным σ_0.

2. В одной системе координат построить графики зависимости длины доверительного интервала от уровня доверия q для всех четырех случаев (объем выборки равен n_1, σ неизвестно; объем выборки равен n_1, σ известно; объем выборки равен n_2, σ неизвестно; объем выборки равен n_2, σ известно). При этом q придать минимум 50 разных значений через равные промежутки.

Проанализировать взаимное расположение полученных графиков и объяснить его.

Для вариантов с четным номером:

1. Для обеих выборок построить точный доверительный интервал

уровня доверия q_0 для параметра σ^2, считая:

а) a неизвестным,

б) a известным и равным a_0.

2. В одной системе координат построить графики зависимости длины доверительного интервала от уровня доверия q для всех четырех случаев (объем выборки равен n_1, a неизвестно; объем выборки равен n_1, a известно; объем выборки равен n_2, a неизвестно; объем выборки равен n_2, a известно). При этом q придать минимум 50 разных значений через равные промежутки.

Проанализировать взаимное расположение полученных графиков и объяснить его.

Указания: Выборки необходимо считать с двух текстовых файлов - "di-V.txt "di-V-1 .txt"(V - номер вашего варианта).

Для написания программы можно пользоваться следующими встроенными функциями: функцией length(x) для определения объема выборки, функциями для нахождения значений квантилей распределений qnorm(p,a,a) (возвращает значение квантиля нормального распределения N, a, σ^2 уровня p), qchisq(p,d) (возвращает значение квантиля распределения "хи-квадрат" с d степенями свободы уровня p), qt(p,d) (возвращает значение квантиля распределения Стьюдента с d степенями свободы уровня p). Остальные статистические функции должны быть запрограммированы.

Варианты заданий

1. $\sigma_0 = 2; q_0 = 0,9$
2. $\sigma_0 = 0; q_0 = 0,8$
3. $\sigma_0 = 3; q_0 = 0,7$
4. $\sigma_0 = 2; q_0 = 0,5$
5. $\sigma_0 = 1; q_0 = 0,6$
6. $\sigma_0 = 3; q_0 = 0,9$
7. $\sigma_0 = 0,5; q_0 = 0,8$
8. $\sigma_0 = -1; q_0 = 0,8$
9. $\sigma_0 = 1,5; q_0 = 0,7$
10. $\sigma_0 = 0,5; q_0 = 0,8$
11. $\sigma_0 = 1; q_0 = 0,5$
12. $\sigma_0 = -5; q_0 = 0,6$
13. $\sigma_0 = 1,2; q_0 = 0,7$
14. $\sigma_0 = 4; q_0 = 0,8$
15. $\sigma_0 = 2,5; q_0 = 0,75$
16. $\sigma_0 = 10; q_0 = 0,6$
17. $\sigma_0 = 3,2; q_0 = 0,9$
18. $\sigma_0 = 0; q_0 = 0,75$
19. $\sigma_0 = 3; q_0 = 0,75$
20. $\sigma_0 = 3; q_0 = 0,5$

Практическое занятие №2. Проверка гипотезы о виде распределения с помощью критерия согласия Смирнова

2.1 Статистическая гипотеза и статистический критерий. Критерий согласия

Пусть в результате некоторого эксперимента получена выборка $X = (X_1, X_2, \ldots, X_n)$ из некоторого распределения F.

Статистической гипотезой H называется любое утверждение о виде неизвестного распределения, или о параметрах известного распределения наблюдаемой в эксперименте случайной величины.

Гипотеза H называется простой, если она однозначно определяет распределение выборки: $H = \{F = F_1\}$, иначе H называют сложной, например, $H = \{F \in \{F\}\}$, где $\{F\}$ - некоторое семейство распределений: $\{F\} = \{F(x, \theta), \theta \in \Theta\}$.

Если выдвигаются всего две гипотезы, то одну из них принято называть основной (H_0), а другую альтернативной (конкурирующей) (H_1).

Правило, согласно которому проверяемая гипотеза H_0 принимается или отвергается, называется статистическим критерием. Дадим формальное определение критерия.

Статистическим критерием для проверки гипотез $H_1, \ldots, H_k$ называется любое измеримое отображение $\rho : R^n \to \{H_1, \ldots, H_k\}$.

Если $\rho(X) = H_i$, то мы принимаем гипотезу H_i (или считаем $\theta = \theta_i$ в параметрическом случае).

Качество критерия характеризуется набором вероятностей ошибочных решений.

Будем говорить, что произошла ошибка i-го рода, если гипотеза H_i отвергнута, когда она верна. Вероятностью ошибки i-го рода критерия р называется

$$q_i(\delta) = P_{H_i}(p(X) \neq H_i).$$

Если удастся выбрать критерий ρ так, что все числа $q_i(\rho)$ малы, то мы будем объявлять, что верна гипотеза H_k, если $\rho(X) = H_k$. При этом мы будем ошибаться примерно в доле случаев q_i, если верна гипотеза H_i.

Уровнем значимости статистического критерия называют вероятность ошибочно отвергнуть основную проверяемую гипотезу, когда она верна (вероятность ошибки первого рода).

Рассмотрим случай, когда о распределении наблюдений X имеется две гипотезы: $H_0 = \{F = F_1\}$ при альтернативе $H_1 = \{F \neq F_1\}$. В этом случае любой критерий $\rho(X): R^n \to \{H_0, H_1\}$ принимает не более двух значений, то есть область R^n делится на две части

$$R^n = S \cup \left(R^n \setminus S\right)$$

так, что

$$\rho(X) = \begin{cases} H_0, & X \in R^n \setminus S, \\ H_1, & X \in S. \end{cases}$$

Область S, в которой принимается альтернативная гипотеза, называется критической областью критерия ρ.

Обозначим q = q(ρ) - уровень значимости критерия ρ, тогда

$$q = q_1(\rho) = P_{H_0}\left(\rho(X) \neq H_0\right) = P_{H_0}\left(\rho(X) = H_1\right) = P_{H_0}\left(X \in S\right).$$

По своему смыслу критическая область должна строиться так, чтобы событие X ∈ S было маловероятным. В конкретных задачах ρ выбирают обычно равной 0,1; 0,05; 0, 01 и так далее.

Критериями согласия называют критерии для проверки простой гипотезы H_0 при сложной альтернативе H_1 = {H_0 неверна}.

Критерии согласия принимают или отвергают основную гипотезу исходя из величины некоторой статистики T(X), характеризующей отклонение эмпирических данных от соответствующих (гипотезе H_0) гипотетических значений, распределение которой в случае справедливости H_0 можно было бы определить.

Предположим, такая статистика и ее распределение при гипотезе H_0 найдены. Пусть T - множество всевозможных значений статистики T. Определим для фиксированного заранее достаточно малого числа α > 0 подмножество $\tilde{T} \subset T$, так чтобы вероятность осуществления события T (X) ∈ $\tilde{T}$ в случае справедливости гипотезы H_0 удовлетворяла условию

$$P\left(T(X) \in \tilde{T} \mid H_0\right) \leq \alpha$$

Тогда правило проверки гипотезы H_0 можно сформулировать следующим образом: Если для данной выборки ***X*** значение статистики $T(X) \in \tilde{T}$**,** то в предположении справедливости гипотезы ***H_0*** произошло маловероятное событие и гипотеза должна быть отвергнута как противоречащая статистическим данным. В противном случае нет основания отказываться от принятия гипотезы ***H_0. На языке критериев это можно записать так:***

$$\rho(X) = \begin{cases} H_0, & T(X) \in T \setminus \tilde{T} \\ H_1, & T(X) \in \tilde{T} \end{cases}$$

2.2 Критерий Смирнова для проверки гипотезы о виде распределения

Имеется выборка $X = (X_1, X_2, \ldots, X_n)$ из некоторого распределения F. Проверяется простая гипотеза $H_0 = \{F = F_1\}$ против сложной альтернативы $H_1 = \{F \neq F_1\}$. В том случае, когда распределение F_1 имеет непрерывную функцию распределения F_1, для проверки гипотезы можно воспользоваться критерием Смирнова.

Пусть $F_n^*(x)$ - эмпирическая функция распределения. Рассмотрим статистику

$$T_n = \sqrt{n}\sup_x\left(F_n^*(x)\right) = \sqrt{n}\max_{X_i}\left(F_n^*(X_i) - F_1(X_i)\right).$$

Теорема 2.1 ***В случае справедливости гипотезы*** H_0 при $n \to +\infty$ ***статистика*** T_n ***имеет распределение Смирнова:***

$$\lim_{n\to\infty} P(T_n < x) = S(x),$$

здесь $S(x) \equiv 1 - e^{-2x^2}$ - функция Смирнова.

Зададим некоторый уровень значимости q и найдем пороговое значение статистики C_q из условия $S(C_q) = 1 - q$. Построим критерий Смирнова для проверки гипотезы о виде распределения:

$$\rho(X) = \begin{cases} H_0, & T_n(X) \le C_q, \\ H_1, & T_n(X) > C_q. \end{cases}$$

Приведем пример применения этого критерия. Пусть в результате эксперимента получена выборка объема N=200 из распределения некоторой случайной величины. Проверим гипотезу о том, что эта случайная величина имеет стандартное нормальное распределение. Текст программы, реализующий проверку данной гипотезы с помощью критерия Смирнова, приведен на рисунке 2.1.

2.3 Задание к практическому занятию

В файле smirn-V.txt (V - номер вашего варианта) задана выборка из некоторого распределения. Задав некоторый уровень значимости q, с помощью критерия Смирнова проверить следующие гипотезы:

Варианты заданий

1. а) $H_0 = \{F = N_{1,2}\}$ против $H_1 = \{F \ne N_{1;2}\}$;

   б) $H_0 = \{F = E_3\}$ против $H_1 = \{F \ne E_3\}$;

   в) $H_0 = \{F = F_{3,5}\}$ против $H_1 = \{F \ne F_{3,5}\}$.

$n := 200$ — Задается объем выборки

$x :=$ C:\..\smirn.txt — Выборка считывается из файла

$$F(y) := \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{y} e^{\frac{-t^2}{2}}\,dt$$

Задается теоретическая функция распределения.

$$I(t,y) := t < y \qquad Fn(y) := \frac{1}{n} \cdot \sum_{i=0}^{n-1} I(x_i, y)$$

Определяется эмпирическая функция распределения

$$MAX(x,F) := \left| \begin{array}{l} c \leftarrow Fn(x_0) - F(x_0) \\ \text{for } i \in 1..n-1 \\ \quad c \leftarrow Fn(x_i) - F(x_i) \ \text{ if } Fn(x_i) - F(x_i) > c \\ c \end{array} \right.$$

$$T(x) := \sqrt{n} \cdot MAX(x,F) \qquad T(x) = 0.507$$

Строится статистика и вычисляется ее значение

$$C(q) := \sqrt{\frac{-\ln(q)}{2}} \qquad q := 0.05 \qquad C(q) = 1.224$$

Находится пороговое значение C(q), где q- уровень значимости. В данном случае T(x)<C(q), поэтому гипотеза о распределении случайной величины по нормальному закону $N_{0,1}$ принимается

Рисунок 2.1. Проверка гипотезы о распределении случайной величины по нормальному закону $N_{0,1}$ с помощью критерия Смирнова

2. а) $H_0 = \{F = U_{2,4}\}$ против $H_1 = \{F \neq U_{2,4}\}$;
   б) $H_0 = \{F = K_{0,1}\}$ против $H_1 = \{F \neq K_{0,1}\}$;
   в) $H_0 = \{F = \Gamma_{2,2}\}$ против $H_1 = \{F \neq \Gamma_{2,2}\}$.
3. а) $H_0 = \{F = \beta_{3,2}\}$ против $H_1 = \{F \neq \beta_{3,2}\}$;
   б) $H_0 = \{F = K_{1,1}\}$ против $H_1 = \{F \neq K_{1,1}\}$;
   в) $H_0 = \{F = E_4\}$ против $H_1 = \{F \neq E_4\}$.
4. а) $H_0 = \{F = \beta_{3,4}\}$ против $H_1 = \{F \neq \beta_{3,4}\}$;
   б) $H_0 = \{F = F_{5,5}\}$ против $H_1 = \{F \neq F_{5,5}\}$;
   в) $H_0 = \{F = U_{-1,1}\}$ против $H_1 = \{F \neq U_{-1,1}\}$.
5. а) $H_0 = \{F = K_{2,4}\}$ против $H_1 = \{F \neq K_{2,4}\}$;
   б) $H_0 = \{F = N_{2,4}\}$ против $H_1 = \{F \neq N_{2,4}\}$;

в) $H_0 = \{F = U_{0,1}\}$ против $H_1 = \{F \neq U_{0,1}\}$.

6. а) $H_0 = \{F = \beta_{5,4}\}$ против $H_1 = \{F \neq \beta_{5,4}\}$;
   б) $H_0 = \{F = F_{2,4}\}$ против $H_1 = \{F \neq F_{2,4}\}$;
   в) $H_0 = \{F = E_3\}$ против $H_2 = \{F \neq E_3\}$.
7. а) $H_0 = \{F = U_{3,6}\}$ против $H_1 = \{F \neq U_{3,6}\}$;
   б) $H_0 = \{F = K_{1,1}\}$ против $H_1 = \{F \neq K_{1,1}\}$;
   в) $H_0 = \{F = \Gamma_{2,3}\}$ против $H_1 = \{F \neq \Gamma_{2,3}\}$.
8. а) $H_0 = \{F = N_{-1,3}\}$ против $H_1 = \{F \neq N_{-1,3}\}$;
   б) $H_0 = \{F = E_{1.5}\}$ против $H_1 = \{F \neq E_{1.5}\}$;
   в) $H_0 = \{F = \beta_{3,5}\}$ против $H_1 = \{F \neq \beta_{3,5}\}$.
9. а) $H_0 = \{F = F_{2,3}\}$ против $H_1 = \{F \neq F_{2,3}\}$;
   б) $H_0 = \{F = K_{0,2}\}$ против $H_1 = \{F \neq K_{0,2}\}$;
   в) $H_0 = \{F = \Gamma_{2,2}\}$ против $H_1 = \{F \neq \Gamma_{2,2}\}$.
10. а) $H_0 = \{F = \beta_{3,4}\}$ против $H_1 = \{F \neq \beta_{3,4}\}$;
    б) $H_0 = \{F = E_5\}$ против $H_1 = \{F \neq E_5\}$;
    в) $H_0 = \{F = U_{0,5}\}$ против $H_1 = \{F \neq U_{0,5}\}$.
11. а) $H_0 = \{F = \Gamma_{4,4}\}$ против $H_1 = \{F \neq \Gamma_{4,4}\}$;
    б) $H_0 = \{F = F_{2,4}\}$ против $H_1 = \{F \neq F_{2,4}\}$;
    в) $H_0 = \{F = E_3\}$ против $H_1 = \{F \neq E_3\}$.
12. а) $H_0 = \{F = N_{5,1}\}$ против $H_1 = \{F \neq N_{5,1}\}$;
    б) $H_0 = \{F = E_{2.5}\}$ против $H_1 = \{F \neq E_{2.5}\}$;
    в) $H_0 = \{F = U_{3,5}\}$ против $H_1 = \{F \neq U_{3,5}\}$.
13. а) $H_0 = \{F = \beta_{3,2}\}$ против $H_1 = \{F \neq \beta 3{,}2\}$;
    б) $H_0 = \{F = K_{0,1}\}$ против $H_1 = \{F \neq K_{0,1}\}$;
    в) $H_0 = \{F = E_2\}$ против $H_1 = \{F \neq E_2\}$.
14. а) $H_0 = \{F = U_{0,2}\}$ против $H_1 = \{F \neq U_{0,2}\}$;

б) $H_0 = \{F = \Gamma_{3,3}\}$ против $H_1 = \{F \neq \Gamma_{3,3}\}$;

в) $H_0 = \{F = K_{2,1}\}$ против $H_1 = \{F \neq K_{2,1}\}$.

15. а) $H_0 = \{F = F_{5,3}\}$ против $H_1 = \{F \neq F_{5,3}\}$;

б) $H_0 = \{F = E_2\}$ против $H_1 = \{F \neq E_2\}$;

в) $H_0 = \{F = \Gamma_{2,4}\}$ против $H_1 = \{F \neq \Gamma_{2,4}\}$.

16. а) $H_0 = \{F = N_{0,4}\}$ против $H_1 = \{F \neq N_{0,4}\}$;

б) $H_0 = \{F = E_{3.5}\}$ против $H_1 = \{F \neq E_{3.5}\}$;

в) $H_0 = \{F = F_{4,5}\}$ против $H_1 = \{F \neq F_{4,5}\}$.

17. а) $H_0 = \{F = \Gamma_{4,2}\}$ против $H_1 = \{F \neq \Gamma_{4,2}\}$;

б) $H_0 = \{F = E_5\}$ против $H_1 = \{F \neq E_5\}$;

в) $H_0 = \{F = U_{-2,2}\}$ против $H_1 = \{F \neq U_{-2,2}\}$.

18. а) $H_0 = \{F = \beta_{3,6}\}$ против $H_1 = \{F \neq \beta_{3,6}\}$;

б) $H_0 = \{F = F_{2,2}\}$ против $H_1 = \{F \neq F_{2,2}\}$;

в) $H_0 = \{F = E_{4.5}\}$ против $H_1 = \{F \neq E_{4.5}\}$.

19. а) $H_0 = \{F = F_{3,4}\}$ против $H_1 = \{F \neq F_{3,4}\}$;

б) $H_0 = \{F = N_{5,5}\}$ против $H_1 = \{F \neq N_{5,5}\}$;

в) $H_0 = \{F = U_{-1,0}\}$ против $H_1 = \{F \neq U_{-1,0}\}$.

20. а) $H_0 = \{F = K_{2,3}\}$ против $H_1 = \{F \neq K_{2,3}\}$;

б) $H_0 = \{F = E_{1.5}\}$ против $H_1 = \{F \neq E_{1.5}\}$;

в) $H_0 = \{F = U_{0,6}\}$ против $H_1 = \{F \neq U_{0,6}\}$.

Практическое занятие №3
Проверка параметрической гипотезы о виде распределения с помощью критерия согласия χ^2 Пирсона

3.1 Критерий согласия Пирсона

Пусть имеется выборка $X = (X_1, X_2, \ldots, X_n)$ из некоторого распределения F. Будем проверять гипотезу о принадлежности, наблюдаемой в опыте случайной величины некоторому семейству распределений. То есть будем проверять сложную гипотезу $H_0 = \{F \in \{F_\theta\}\}$ против сложной альтернативы $H_1 = \{F \notin \{F_\theta\}\}$.

Чтобы воспользоваться критерием Пирсона, выборочные данные предварительно группируют. Разделим область выборочных данных на интервалы $\Delta_1, \Delta_2, \ldots, \Delta_k$. Обозначим за $v_j (j = 1,2,\ldots k)$ число элементов выборки, попавших в интервал $\Delta_j (v_1 + \ldots + v_k = n)$. Эмпирические вероятности попадания элементов выборки в Δ_j обозначим q_j:

$$q_j = \frac{v_j}{n}, j = 1,..., k.$$

За $p_j(\theta)$ обозначим теоретические вероятности попадания значения случайной величины в интервал группировки Δ_j в случае, если выполняется гипотеза H_0.

Составим статистику, характеризующую отклонение выборочных данных (т.е. вероятностей q_j) от соответствующих гипотетических значений (p_j):

$$\chi_n^2 = \chi_n^2(\theta) = n \sum_{j=1}^{n} \frac{(q_j - p_j(\theta))^2}{p_j(\theta)}.$$

Непосредственно использовать эту статистику для построения критерия нельзя - для начала надо исключить неопределенность, связанную с неизвестным параметром θ. Для этого поступают следующим образом - заменяют θ некоторой оценкой $\tilde{\theta}_n = \tilde{\theta}_n(X)$ найденной по выборке X.

$$\tilde{\chi}_n^2 = \tilde{\chi}_n^2(\tilde{\theta}_n) = n\sum_{j=1}^{n}\frac{\left(q_j - p_j(\tilde{\theta}_n)\right)^2}{p_j(\tilde{\theta}_n)}.$$

Такую статистику уже можно однозначно вычислить для каждой заданной реализации выборки X.

Теорема 3.1 Если верна гипотеза ***H_0*** и ***l*** - размерность векторного параметра θ, то при фиксированном ***k*** и при $n \to \infty$ ***то***

$$\tilde{\chi}_n^2 \to \xi,$$

где случайная величина ξ имеет распределение χ^2 с $k-l-1$ степенями свободы.

Приведем схему использования критерия согласия χ^2:

1. По заданной выборке X найти оценку векторного параметра θ.
2. Найти теоретические вероятности попадания значений случайной величины в интервалы группировки Δ_j.
3. Вычислить значение статистики $\tilde{\chi}_n^2$ (см. Приложение Б, таблица 3).
4. По заданному уровню значимости q найти пороговое значение статистики C_q из условия $\chi_{k-l-1}^2(C_q) = 1 - q$.
5. Если $\tilde{\chi}_n^2 < C_q$ то гипотезу H_0 принимаем, в противном случае отклоняем.

Замечание: Малочисленные частоты ($v_j < 5$), следует объединить, объединив соответствующие интервалы группировки. Теоретические частоты следует вычислять уже после объединения интервалов. При определении числа степеней свободы в этом случае вместо k следует взять число интервалов, получившихся после объединения.

$N := 500$ Задается объем выборки

$x :=$ C:\..\pears2.txt Вектор, задающий выборку считывается из файла

$x^T =$

	0	1	2	3	4	5	6	7	8	9
0	8.85	2.992	3.925	4.055	5.859	5.455	1.031	1.435	0.4	1.99

$i := 0..N-1$

$lower := floor(min(x))$ $lower = 0$ определяются границы области выборочных данных

$upper := floor(max(x)) + 1$ $upper = 11$

$k := upper - lower$ $k = 11$ Задается число интервалов группировки данных

$i := 0..k-1$ $v_{i,0} := lower + i$

$v_{i,1} := lower + (i+1)$

$I(i,y) := y < v_{i,1} \wedge y \geq v_{i,0}$

$$v_{i,2} := \sum_{j=0}^{N-1} I(i, x_j)$$

Формируется таблица частот.
Для i-го столбца
В 1-й строке записано начало i-го интервала группировки
во 2-й строке - конец этого интервала,
в 3-й строке - количество попаданий значений случайной величины в i-й интервал группировки

$v^T =$

	0	1	2	3	4	5	6	7	8	9	10
0	0	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10	11
2	41	141	105	78	55	32	29	10	4	4	1

$$z := \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 10 \\ 42 & 124 & 131 & 86 & 61 & 34 & 14 & 7 \end{pmatrix} \qquad z := z^T$$

Формируется новая таблица частот, в которой объединены интервалы группировки, количества попаданий значений случайной величины в которые незначительны

$L := rows(z)$ $L = 8$

$$Mx := \frac{1}{N} \cdot \sum_{i=0}^{N-1} x_i \qquad \lambda := Mx \qquad \lambda = 3.08$$

Методом моментов находится оценка параметра λ

$$i := 0..L-1 \qquad q_i := \frac{z_{i,2}}{N}$$

Находятся эмпирические частоты попадания случайной величины в i-й интервал группировки

$q^T = (0.084\ \ 0.248\ \ 0.262\ \ 0.172\ \ 0.122\ \ 0.068\ \ 0.028\ \ 0.014)$

$$\alpha := 1 \qquad p_i := \int_{z_{i,0}}^{z_{i,1}} \frac{\alpha^{\lambda}}{\Gamma(\lambda)} \cdot t^{\lambda-1} \cdot e^{-\alpha \cdot t}\, dt$$

Находятся теоретические частоты попадания значений случайной величины в i-й интервал группировки

$p^T = (0.072\ \ 0.233\ \ 0.253\ \ 0.19\ \ 0.119\ \ 0.067\ \ 0.035\ \ 0.029)$

$$h := N \cdot \sum_{i=0}^{L-1} \frac{(q_i - p_i)^2}{p_i} \qquad h = 7.233$$

$h\_kp := qchisq(0.95, L-2-1)$ $h\_kp = 11.07$

В данном случае h<h_kp, поэтому гипотеза о том, что случайная величина имеет гамма-распределение с параметрами α=1, λ=2.921 принимается

Рисунок 3.1. Проверка гипотезы о том, что случайная величина имеет гамма-распределение $\Gamma_{1,\lambda}$

Приведем пример использования критерия χ^2 для проверки параметрической гипотезы о виде распределения.

Пусть имеется выборка объема n = 500 из некоторого распределения F. Требуется проверить гипотезу $H_0 = \{F = \Gamma_{1,\lambda}\}$. На рисунке 3.1 приведен текст программы для проверки этой гипотезы с помощью критерия χ^2. Обратите внимание на то, что так как один из параметров гамма- распределения уже задан, то по выборке оценивается только параметр λ.

3.2 Задание к практическому занятию

Варианты заданий

1. а) В ходе испытания 400 ламп накаливания была получена выборка, элементы которой - длительности их горения (в часах). Данные приведены в файле pearson-1a.txt. С помощью критерия Пирсона проверить гипотезу о распределении времени горения ламп по показательному закону с уровнем значимости $q = 0.05$.

б) В течение 3 месяцев (90 дней) в супермаркете вели статистику о количестве проданных за день буханок белого хлеба. Полученная выборка приведена в файле pearson-1b.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества проданных в день буханок по закону Пуассона с уровнем значимости $q = 0.05$.

2. а) В тепличном хозяйстве проводился контроль урожайности томатов некоторого сорта. Было измерено, сколько килограммов томатов было собрано за сезон с каждого из 100 выбранных кустов. Полученные данные приведены в файле pearson-2a.txt. С помощью критерия Пирсона проверить гипотезу о равномерном распределении урожайности томатов с уровнем значимости $q = 0.02$.

б) В результате испытаний 300 дискет на количество циклов перезаписывания, которые дискета выдерживает без выхода из строя, была получена выборка случайной величины, значение которой - количество циклов, выдержанных дискетой до поломки. Выборка приведена в файле pearson-2b.txt с помощью критерия

Пирсона проверить гипотезу о распределении этой случайной величины по закону Пуассона с уровнем значимости $q = 0.02$.

3. а) В некоторой местности в течение 300 суток регистрировалась среднесуточная температура воздуха (в градусах Цельсия). В результате была получена выборка, приведенная в файле pearson-3a.txt. С помощью критерия Пирсона проверить гипотезу о равномерном распределении температуры воздуха с уровнем значимости $q = 0.03$.

б) Для проверки качества работы заводского оборудования было подсчитано количество нестандартных деталей, изготовленных каждым из 200 станков за неделю. Полученная выборка приведена в файле pearson- 3b.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества нестандартных изделий по закону Пуассона с уровнем значимости $q = 0.05$.

4. а) В ходе школьного медицинского осмотра был измерен рост 100 первоклассников. Полученные данные приведены в файле pearson-4a.txt. С помощью критерия согласия Пирсона проверить гипотезу о распределении роста учеников по нормальному закону с уровнем значимости $q = 0.05$.

б) Опыт, состоящий в одновременном подбрасывании 5 игральных костей, повторили 150 раз. Событие А в единичном испытании состоит в том, что на кости выпало не более двух очков. Данные о том, сколько раз при каждом из 150 подбрасываний повторилось событие А приведены в файле pearson-4b.txt. С помощью критерия Пирсона проверить гипотезу о распределении исследуемой случайной величины по биномиальному закону с уровнем значимости $q = 0.05$.

5. а) В результате испытания 300 батареек на длительность их работы (в часах) была получена выборка, приведенная в файле pearson-5a.txt. С помощью критерия Пирсона проверить гипотезу о распределении длительности работы батареек по показательному закону с уровнем значимости $q = 0.03$.

б) Опыт состоял в том, что человек одновременно вытягивал из 3-х карточных колод по 1 карте и складывал их обратно. Опыт было повторен 200 раз и каждый раз подсчитывалось сколько вынималось карт пиковой масти. Данные приведены в файле pearson-5b.txt. С помощью критерия Пирсона проверить гипотезу о распределении исследуемой случайной величины по биномиальному закону с уровнем значимости $q = 0.04$.

6. а) В результате взвешивания 500 бильярдных шаров была получена выборка, представленная в файле pearson-6a.txt (вес задан в граммах). С помощью критерия Пирсона проверить гипотезу о равномерном распределении веса шаров с уровнем значимости $q = 0.01$.

б) После поступления на склад 200 коробок стеклянных елочных игрушек кладовщик проверил, сколько игрушек в каждой коробке было повреждено в результате транспортировки. Результаты его проверки приведены в файле pearson-6b.txt. С помощью критерия Пирсона проверить гипотезу распределении количества поврежденных игрушек в 1 коробке по закону Пуассона с уровнем значимости $q = 0.03$.

7. а) По 100 магазинам города был проведен мониторинг цен на красную икру. В файле pearson-7a.txt приведены цены за 100 граммов икры в каждом из магазинов (в рублях). С помощью критерия Пирсона проверить гипотезу о нормальном распределении цен на икру с уровнем значимости $q = 0.01$.

б) Среди студентов третьего курса (250 человек) был проведен тест по теории вероятностей, состоявший из 20 вопросов. Случайная величина ξ - количество вопросов, на которые каждый студент ответил верно. В итоге была получена выборка из распределения случайной величины ξ, которая приведена в файле pearson-7b.txt. С помощью критерия согласия Пирсона проверить гипотезу о распределении случайной величины ξ по биномиальному закону с уровнем значимости $q = 0.02$ (параметр p биномиального закона оценить по выборке).

8. а) Работники рыбного хозяйства выясняли пригодность некоторого озера для разведения карпов, запустив в это озеро 200 меченых мальков и отследив продолжительность их жизни. Длительность жизни карпов (в месяцах) приведена в файле pearson-8a.txt. С помощью критерия Пирсона проверить гипотезу о распределении продолжительность жизни карпов по

показательному закону с уровнем значимости $q = 0.05$.

б) Опыт, состоящий в одновременном подбрасывании 4 монет, повторили 200 раз. Данные о том, сколько "гербов" выпало при каждом из повторений опыта, приведены в файле pearson-8b.txt. С помощью критерия Пирсона проверить гипотезу о распределении числа одновременно выпадающих "гербов" по биномиальному закону с уровнем значимости $q = 0.03$.

9. а) Производителем компьютерной техники проводился эксперимент по выявлению длительности безотказной работы жестких дисков. Было исследовано 300 жестких дисков, результаты эксперимента представлены в файле pearson-9a.txt. С помощью критерия Пирсона проверить гипотезу о распределении длительности безотказной работы дисков по показательному закону с уровнем значимости $q = 0.02$.

б) В обувном магазине решили выявить наиболее ходовой размер женской обуви. Для этого продавцы фиксировали размер каждой проданной пары обуви. Данные приведены в файле pearson-9b.txt. С помощью критерия Пирсона проверить гипотезу о распределении предпочтений покупателей относительно размера обуви по закону Пуассона с уровнем значимости $q = 0.04$.

10. а) Была собрана статистика о продолжительности жизни жителей Красноярска. Данные о продолжительности жизни 1000 человек приведены в файле pearson-10a.txt. С помощью критерия Пирсона проверить гипотезу о распределении продолжительности жизни красноярцев по нормальному закону с уровнем значимости $q = 0.2$.

б) Среди семян ржи имеется некоторое количество семян сорняков. В 300 пакетах семян по 500 штук посчитали количество семян сорняков. Данные приведены в файле pearson-10b.txt. С помощью критерия Пирсона проверить гипотезу о распределении предпочтений покупателей относительно размера обуви по биномиальному закону с уровнем значимости $q = 0.05$ (параметр p оценить по выборке).

11. а) В одном из родильных домов была собрана статистика о весе новорожденных мальчиков, родившихся в течение года. Данные о весе 500 детей приведены в файле pearson-11a.txt. С

помощью критерия Пирсона проверить гипотезу о равномерном распределении веса новорожденных мальчиков с уровнем значимости $q = 0.01$.

б) Оператор сотовой связи ведет учет количества смс, отправленных за день ее абонентами. Данные о количестве смс, отправленных 600 абонентами за 1 день приведены в файле pearson-11b.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества отправленных смс по закону Пуассона с уровнем значимости $q = 0.01$.

12. а) Имеются данные о количестве электроэнергии, потребленной за 1 месяц жителями 500 квартир. Данные приведены в файле pearson- 12a.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества потребленной энергии по равномерному закону с уровнем значимости $q = 0.05$

б) Любительница бразильских сериалов стала записывать, сколько серий продолжался каждый из просмотренных ею фильмов. Данные о количестве серий в 120 сериалах приведены в файле pearson-12b.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества серий в сериалах по биномиальному закону с уровнем значимости $q = 0.05$.

13. а) На молочной ферме в течении недели регистрировали удои коров. В файле pearson-13a.txt приведены средние удои за неделю 120коров. С помощью критерия Пирсона проверить гипотезу о нормальном распределении удоев коров на молочной ферме с уровнем значимости $q = 0.04$.

б) Студентами 1 курса (400) человек была написана контрольная по русскому языку. Результаты контрольной (по 100 бальной шкале) приведены в файле pearson-13b.txt. С помощью критерия Пирсона проверить гипотезу о распределении баллов студентов по биномиальному закону с уровнем значимости $q = 0.02$.

14. а) На некотором месторождении было взято 100 проб руды на содержание железа. Данные о процентном содержании железа в каждой пробе приведены в файле pearson-14a.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества потребленной энергии по равномерному закону с уровнем значимости $q = 0.05$.

б) Данные о количестве заказных писем, отправленных через некоторое почтовое отделение за 1 день в течение последних 3 месяцев (90 дней) приведены в файле pearson-14b.txt. С помощью критерия Пирсона проверить гипотезу о распределении числа отправляемых в день писем по закону Пуассона с уровнем значимости $q = 0.05$.

15. а) На мелькомбинате было переработано 800 тонн зерна, данные о количестве муки, получившейся из каждой тонны зерна, приведены в файле pearson-15a.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества полученного при переработке зерна муки по равномерному закону с уровнем значимости $q = 0.02$.

б) На автозаправочной станции 5 суток регистрировали количество автомобилей, заправившихся на бензоколонке в течение каждого часа. Данные приведены в файле pearson-15b.txt. С помощью критерия Пирсона проверить гипотезу о распределении количества заправляющихся на бензоколонке автомобилей в течение часа по биномиальному закону с уровнем значимости $q = 0.05$.

16. а) Таксист в течение 3 месяцев (90 дней) измерял расход бензина на своей машине (в литрах на 100 километров). Данные его измеренийприведенывфайлеpearson-16a.txt.СпомощьюкритерияПирсона проверить гипотезу о распределении количества потребляемого топлива по нормальному закону с уровнем значимости $q = 0.05$.

б) В библиотеке проверили 400 книг на наличие недостающих страниц. В файле pearson-16b.txt приведены данные о том, сколько страниц не доставало в каждой книге. С помощью критерия Пирсона проверить гипотезу о распределении числа недостающих в книгах страниц по закону Пуассона с уровнем значимости $q = 0.01$.

17. а) Интернет-провайдер ведет статистику интернет-трафика для каждого абонента. В файле pearson-17a.txt приведен объем входящего трафика (в мегабайтах) для каждого из 500 абонентов сети за 1 неделю. С помощью критерия Пирсона проверить гипотезу о распределении количества потребленной энергии по показательному закону с уровнем значимости $q = 0.05$.

б) В файле pearson-17b.txt приведены результаты социологического опроса о том, сколько книг за год прочитал каждый из опрошенных (всего 1000 человек). С помощью критерия Пирсона проверить гипотезу о распределении числа недостающих в книгах страниц по закону Пуассона с уровнем значимости $q = 0.02$.

18. а) Оператор сотовой связи ведет учет времени разговоров своих абонентов по исходящей связи. Данные о количестве минут, которые каждый из 500 абонентов проговорил по исходящей связи за 1 сутки, приведены в файле pearson-18a.txt. С помощью критерия Пирсона проверить гипотезу о длительности разговоров абонентов по исходящей связи по показательному закону с уровнем значимости $q = 0.03$.

б) В файле pearson-18b.txt. приведены данные о том, сколько ограблений совершалось ежедневно за последние полгода (180 дней) в городе N. С помощью критерия Пирсона проверить гипотезу о распределении числа ограблений, совершаемых за одни сутки по биномиальному закону с уровнем значимости $q = 0.04$.

19. а) В файле pearson-19a.txt приведены данные о курсе некоторой валюты (в рублях) за последние полгода (180 дней). С помощью критерия Пирсона проверить гипотезу о распределении стоимости этой валюты по нормальному закону с уровнем значимости $q = 0.05$.

б) На консультации перед экзаменом преподаватель решил выяснить степень готовности студентов. Он спросил каждого о том, сколько вопросов студент уже подготовил (всего на экзамен вынесено 40 вопросов). Данные о готовности 120 студентов приведены в файле pearson- 19a.txt. С помощью критерия Пирсона проверить гипотезу о распре-делении числа подготовленных студентами вопросов по биномиальному закону с уровнем значимости $q = 0.05$.

20. а) В файле pearson-20a.txt приведена стоимость 1 килограмма говядины в 150 магазинах города. С помощью критерия Пирсона проверить гипотезу о распределении стоимости говядины по нормальному закону с уровнем значимости $q = 0.04$

б) Среди студентов университета (300 человек) был проведен опрос о количестве фильмов, просмотренных ими в кинотеатрах. Данные опроса приведены в файле pearson-20b.txt. С помощью критерия Пирсона проверить гипотезу о распределении числа просмотренных в кинотеатрах фильмов по закону Пуассона с уровнем значимости $q = 0.03$.

Практическое занятие №4
Проверка гипотезы однородности

Одной из важных задач прикладной статистики является задача проверки однородности статистического материала.

Пусть имеются две независимые выборки

$$X = (X_1, X_2, ..., X_n) \text{ и } Y = (Y_1, Y_2, ..., Y_m),$$

описывающие один и тот же процесс, явление и т.д., но полученные в разное время или в разных условиях. Требуется установить, являются ли они выборками из одного и того же распределения.

Пусть X - выборка из распределения F, а Y - выборка из распределения G. Требуется проверить гипотезу однородности $H_1 = \{F = G\}$: против альтернативы $H_2 = \{H_1$ неверна$\}$.

4.1 Критерий однородности Колмогорова-Смирнова

Этот критерий применяется для непрерывных случайных величин и основан на статистике

$$T_{nm} = \sqrt{\frac{nm}{n+m}} \sup_{t \in R} \left| F_n^*(t) - G_m^*(t) \right|,$$

где $F_n^*(t)$ и $G_m^*(t)$ - эмпирические функции распределения, построенные по выборкам X и Y.

Теорема 4.1 Если гипотеза H_1 верна, то $T_{nm} \to \xi$ при $n,m \to 0$, где случайная величина ξ имеет распределение Колмогорова с функцией распределения ***K (t).***

По заданному уровню значимости q найдем C_q из условия $K(C_q) = 1$-q. Построим критерий согласия Колмогорова-Смирнова :

$$\rho(X, Y) = \begin{cases} H_1, & T_{n,m} < C_q, \\ H_2, & T_{n,m} \geq C_q. \end{cases}$$

Таким образом, для поверки гипотезы однородности по критерию Колмогорова-Смирнова необходимо следовать следующему алгоритму:

1. По выборкам X и Y построить соответствующие эмпирические функции распределения $F_n^*(t)$и $G_m^*(t)$
2. Найти значение статистики

$$T_{nm} = \sqrt{\frac{nm}{n+m}} \max_{t=X_{1,\ldots,}X_n, Y_1,\ldots,Y_m} \left|F_n^*(t) - G_m^*(t)\right|$$

3. По заданному уровню значимости q по таблице распределения Колмогорова (см. Приложение Б, таблица 5) найдем C_q из условия $K(C_q) = 1 - q$.
4. Если $T_{nm} < C_q$ то гипотезу однородности принимаем, в противном случае отклоняем.

4.2 Критерий однородности χ^2

Этот критерий используется для проверки однородности любых данных (как дискретных случайных величин, так и непрерывных) и позволяет сравнивать любое количество выборок.

Рассмотрим дискретный случай. Пусть осуществлено k серий наблюдений, состоящих из $n_1,\ldots,n_k$ наблюдений соответственно, т.е. имеются выборки

$$X^1 = \left(X_1^1, X_2^1, \ldots, X_{n_1}^1\right)$$
$$X^2 = \left(X_1^2, X_2^2, \ldots, X_{n_2}^2\right)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$X^k = \left(X_1^k, X_2^k, \ldots, X_{n_k}^k\right).$$

При каждом опыте наблюдается одно из s различных значений, задаваемых вектором $Y = (Y_1, Y_2, \ldots, Y_s)$.

Пусть v_{ij} - число реализаций исхода Y_i в j-й серии, так что $\sum_{i=1}^{s} v_{ij} = n_j, \; j = 1,2,\ldots k$. Требуется проверить гипотезу о том, что все k наблюдений проводились над одной случайной величиной. Другими словами, если p_{ij} - вероятность появления i-го исхода в испытаниях j-й серии, то гипотеза однородности означает утверждение:

$$(p_{ij},\ldots,p_{sj}) = (p_1,\ldots,p_s), \; j = 1,2,\ldots,k,$$

где $p = (p_1,\ldots,p_s)$ - некоторый неизвестный вектор вероятностей $(p_1 + \ldots + p_s = 1)$.

Следуя принципу χ^2, в качестве меры отклонения опытных данных от их гипотетических значений следовало бы выбрать статистику

$$\chi_n^2(p) = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(v_{ij} - n_j p_i)^2}{n_j p_i}.$$

Но так как p_i неизвестны, то их нужно предварительно оценить. Оцениваем эти вероятности методом максимального правдоподобия. Получаем следующие оценки:

$$\hat{p}_i = \frac{\mu_i}{n}, \quad \text{где} \quad \mu_i = \sum_{j=1}^{k} v_{ij}, i = 1,\ldots,s, n = n_1 + n_2 + \ldots + n_k \tag{4.1}$$

Таким образом получена следующая статистика критерия:

$$\chi_n^2(p) = n\left(\sum_{i=1}^{s}\sum_{j=1}^{k} \frac{v_{ij}^2}{n_j \mu_i} - 1\right). \tag{4.2}$$

Теорема 6.2 ***При*** $n \to \infty$ $\chi_n^2(p) \to \xi$ ***где случайная величина*** ξ ***имеет распределение*** χ^2 с (s-1)(k-1) ***степенями свободы.***

Запишем алгоритм проверки гипотезы однородности с помощью критерия χ^2:

1. По выборкам $X^1,\ldots,X^k$ строим вектор наблюдаемых значений Y.

2. Для каждого исхода $Y_1,\ldots,Y_s$ вычисляем число его реализаций в j-й серии.

3. Получаем оценки вероятностей $p_1,\ldots,p_s$ по формуле (4.1).

4. Вычисляем значение статистики $\chi_n^2(p)$ по формуле (4.2).

5. По заданному уровню значимости q найдем C_q из условия $\chi^2_{(s-1)(k-1)}(C_q)=1-q$ (см. Приложение Б, таблица 3).

6. Гипотезу однородности принимаем, если $\chi_n^2(p)<C_q$ и отклоняем в противном случае.

На рисунке 4.1 приведен текст программы, реализующей проверку гипотезы однородности двух выборок по критерию χ^2 в среде Mathcad.

4.3 Задание к практическому занятию

а) Было проверено 2 партии теннисных мячей, произведенных на одном заводе. Первая партия состоит из N_1 штук, вторая - из N_2 штук. Каждый мяч был взвешен, веса мячей из первой партии приведены в файле homo-V-1.txt, второй партии - hom-V-2.txt (V - это номер вашего варианта). Проверить гипотезу однородности двух партий теннисных мячей с уровнем значимости q.

б) В файлах homog-V-1.txt, homog-V-2.txt, homog-V-3.txt

(V - это номер вашего варианта) находятся 3 независимые выборки, описывающих работу 3-х смен на заводе, изготавливающих одинаковые детали на одном и том же оборудовании. Элементы выборок - это количества бракованных деталей, произведенных каждым рабочим смены. В первой смене работало N_1 рабочих, во второй - N_2, в третьей - N_3. Проверить гипотезу однородности для этих выборок с уровнем значимости q.

$N1 := 200 \quad x1 := rpois(N1, 6)$

$N2 := 180 \quad x2 := rpois(N2, 6)$

Генерируем две выборки x1 и x2 из распределения Пуассона Π_6 объема 200 и 180 соответственно

$i := 0 \quad y_i := x1_0$

```
form1(x1, N1, r, z) := | for i ∈ 0..r
                       |   y_i ← z_i
                       | j ← r + 1
                       | flag ← 0
                       | for i ∈ 0..N1 − 1
                       |   | for k ∈ 0..j − 1
                       |   |   flag ← 1 if y_k = x1_i
                       |   | if flag = 0
                       |   |   | y_j ← x1_i
                       |   |   | j ← j + 1
                       |   | flag ← 0
                       | y
```

Функция формирования вектора y, компоненты которого - различные наблюдаемые значения в обеих сериях опытов

$z := form1(x1, N1, 0, y) \quad r := length(z)$

$y := form1(x2, N2, r - 1, z)$

Формируем вектор y, применив функцию form1(...) вначале к вектору x1, а затем к вектору x2

$y^T =$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0	6	3	7	8	10	11	5	4	9	1	2	12	13	14

$s := length(y) \quad s = 15$

Длина вектора y - число различных наблюдаемых значений

```
form2(x1, x2, N1, N2, y, s) := | for i ∈ 0..s − 1
                               |   | a_{i,0} ← 0
                               |   | a_{i,1} ← 0
                               | for i ∈ 0..s − 1
                               |   for j ∈ 0..N1 − 1
                               |     a_{i,0} ← a_{i,0} + 1 if y_i = x1_j
                               | for i ∈ 0..s − 1
                               |   for j ∈ 0..N2 − 1
                               |     a_{i,1} ← a_{i,1} + 1 if y_i = x2_j
                               | a
```

Функция для формирования матрицы ν, каждый элемент которой ν_{ij} -число реализаций исхода y_i в j-й серии опытов

$\nu := form2(x1, x2, N1, N2, y, s)$

Формируем матрицу ν

$\nu^T =$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	1	35	23	17	21	10	4	24	28	16	1	15	3	2	0
1	0	25	22	18	15	14	6	34	23	10	3	9	1	0	1

$i := 0..s - 1 \quad \mu_i := \nu_{i,0} + \nu_{i,1} \quad p_i := \dfrac{\mu_i}{N1 + N2}$

Формируем вектор p, каждый элемент которого p_i- оценка вероятности исхода y_i, найденная методом максимального правдоподобия

$p^T =$

	0	1	2	3	4	5	6	7
0	$2.632 \cdot 10^{-3}$	0.158	0.118	0.092	0.095	0.063	0.026	0.153

$$\chi := (N1 + N2)\left[\sum_{i=0}^{s-1}\left[\frac{(\nu_{i,0})^2}{N1 \cdot \mu_i} + \frac{(\nu_{i,1})^2}{N2 \cdot \mu_i}\right] - 1\right] \qquad \chi = 14.301$$

$\chi\_kp := qchisq(0.95, s - 1) \qquad \chi\_kp = 23.685$

Вычисляем значение статистики $\chi = \chi_{N1+N2}(p)$ и сравниваем ее значение с χ_кр. Гипотеза однородности принимается с уровнем значимости 0.05.

Рисунок 4.1 - Проверка гипотезы однородности с помощью критерия χ^2

Варианты заданий

1. а) $N_1 = 500, N_2 = 400, q = 0.02$; б) $N_1 = 200, N_2 = 180, N_3 = 190, q = 0.05$.
2. а) $N_1 = 300, N_2 = 250, q = 0.01)$; б) $N_1 = 200, N_2 = 220, N_3 = 195, q = 0.04$.
3. а) $N_1 = 350, N_2 = 250, q = 0.04$; б) $N_1 = 190, N_2 = 210, N_3 = 195, q = 0.03$.
4. а) $N_1 = 220, N_2 = 250, q = 0.05)$; б) $N_1 = 185, N_2 = 205, N_3 = 195, q = 0.02$.
5. а) $N_1 = 250, N_2 = 240, q = 0.03$; б) $N_1 = 210, N_2 = 205, N_3 = 200, q = 0.02$.
6. а) $N_1 = 350, N_2 = 400, q = 0.01$; б) $N_1 = 185, N_2 = 190, N_3 = 188, q = 0.05$.
7. а) $N_1 = 400, N_2 = 420, q = 0.04$; б) $N_1 = 220, N_2 = 215, N_3 = 218, q = 0.03$.
8. а) $N_1 = 300, N_2 = 310, q = 0.02$; б) $N_1 = 202, N_2 = 205, N_3 = 200, q = 0.04$.
9. а) $N_1 = 420, N_2 = 430, q = 0.05$; б) $N_1 = 198, N_2 = 202, N_3 = 200, q = 0.03$.
10. а) $N_1 = 500, N_2 = 490, q = 0.01$; б) $N_1 = 199, N_2 = 201, N_3 = 220, q = 0.05$.
11. а) $N_1 = 480, N_2 = 485, q = 0.02$; б) $N_1 = 181, N_2 = 183, N_3 = 187, q = 0.04$.
12. а) $N_1 = 450, N_2 = 420, q = 0.04$; б) $N_1 = 202, N_2 = 218, N_3 = 200, q = 0.02$.
13. а) $N_1 = 380, N_2 = 390, q = 0.03$; б) $N_1 = 150, N_2 = 200, N_3 = 210, q = 0.05$.
14. а) $N_1 = 390, N_2 = 360, q = 0.01$; б) $N_1 = 180, N_2 = 182, N_3 = 185, q = 0.03$.

15. а) $N_1 = 400, N_2 = 490, q = 0.02$; б) $N_1 = 179, N_2 = 180, N_3 = 185, q = 0.04$.

16. а) $N_1 = 350, N_2 = 500, q = 0.05$; б) $N_1 = 184, N_2 = 220, N_3 = 202, q = 0.02$.

17. а) $N_1 = 470, N_2 = 490, q = 0.01$; б) $N_1 = 199, N_2 = 213, N_3 = 200, q = 0.05$.

18. а) $N_1 = 360, N_2 = 300, q = 0.04$; б) $N_1 = 211, N_2 = 201, N_3 = 203, q = 0.03$.

19. а) $N_1 = 380, N_2 = 310, q = 0.02$; б) $N_1 = 214, N_2 = 200, N_3 = 198, q = 0.01$.

20. а) $N_1 = 410, N_2 = 390, q = 0.03$; б) $N_1 = 200, N_2 = 220, N_3 = 204, q = 0.05$.

Практическое занятие №5
Проверка гипотезы случайности

5.1 Построение критерия для проверки гипотезы случайности

В различных статистических задачах исходные данные $X=(X_1,\ldots,X_n)$ рассматривают как случайную выборку из некоторого распределения F, т.е. считают компоненты X_i вектора данных X независимыми и одинаково распределенными случайными величинами. Однако, иногда такое предположение нуждается в проверке.

Математически задачу можно сформулировать так: проверить гипотезу $H_0=\{F_x(x)=F(x_1)\cdot\ldots\cdot F(x_n), x=(x_1,\ldots,x_n)\}$, где F(x) - некоторая функция распределения, против альтернативной гипотезы $H_1 = \{H_0$ неверна$\}$.

Критерий для проверки этой гипотезы строится исходя из следующих соображений: если гипотеза случайности действительно имеет место, то компоненты вектора X "равноправны" и поэтому данные не должны быть ни в каком смысле упорядочены. Следовательно, критерий проверки гипотезы H_0 можно построить на основании статистик, измеряющих степень беспорядка исходных данных.

Одной из таких статистик является число инверсий в выборке. Говорят, что компонента X_i образует μ_i инверсий, если в вариационном ряду, построенном по выборке X левее X_i, стоит μ_i элементов выборки с большими номерами.

На рисунке 5.1 приведен текст программы, вычисляющей количество инверсий для любого элемента выборки по заданному вариационному ряду.

Общее число инверсий для выборки X можно найти по формуле:

$$T_n(X)=\mu_1+\ldots+\mu_{n-1}. \qquad (5.1)$$

$y :=$ C:\..\rand1.txt

Считываем из файла вариационный ряд. 1-я строка в y^T - это значение элементов выборки, а 2-я строка - это номера соответствующих элементов в исходной выборке $X=(X_0, \ldots, X_{N-1})$

$y^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.036	0.049	0.063	0.074	0.076	0.095	0.1	0.113	0.161	0.165
1	22	52	113	132	78	54	99	82	5	144

$N := \text{length}\left(y^{\langle 0\rangle}\right)$ $\quad N = 150$ $\quad$ Определяем объем исходной выборки

```
inv(k) := | num ← 0
          | for i ∈ 0..N − 1
          |   m ← i  if y[i,1] = k
          | for i ∈ 0..m − 1
          |   num ← num + 1  if y[i,1] > y[m,1]
          | num
```

Функция, вычисляющая количество инверсий, образованных k-м элементом исходной выборки

$inv(54) = 3$ $\quad$ Элемент исходной выборки X_{54} образовал 3 инверсии

Рисунок 5.1. Вычисление количества инверсий, образованных элементом X_k

Нормируем статистику T_n следующим образом:

$$T_n^*(x) = \left(T_n(X) - \frac{n(n-1)}{4}\right) \cdot \frac{6}{n^{3/2}}. \qquad (5.2)$$

Теорема 5.1 При $n \to \infty$ $T_n^* \to \xi$, где случайная величина ξ имеет стандартное нормальное распределение $N_{0,1}$.

По заданному уровню значимости q найдем C_q из условия $\Phi(-C_q) = q/2$ ($\Phi(x)$ — функция Лапласа). Построим критерий проверки гипотезы случайности.

$$\rho(X) = \begin{cases} H_0, & \text{если} T_n^* \le C_q, \\ H_1, & \text{в} \cdot \text{противном} \cdot \text{случае}. \end{cases}$$

Таким образом, получаем следующий алгоритм проверки гипотезы случайности:

1. По заданной выборке X составляем вариационный ряд.

2. Считаем значение статистик T_n и T_n^* по формулам (5.1), (5.2).

3. Для заданного уровня значимости q определяем C_q из условия $Ф(-C_q) = q/2$ (см. Приложение Б, таблица 1).

4. Если $T_n^* \le C_q$, то гипотезу случайности принимаем, в противном случае отклоняем.

5.2 Задание к практическому занятию

В файле rand-V.txt находится некоторая последовательность чисел. Можно ли считать эту последовательность случайной выборкой из некоторого распределения с уровнем значимости q = 0.0V (V - номер вашего варианта)?

Практическое занятие №6
Проверка гипотезы о независимости, вычисление коэффициента корреляции, построение уравнения линейной регрессии

6.1 Проверка гипотезы независимости с помощью критерия χ^2

Предположим, что в некотором эксперименте наблюдается случайная величина $\psi = (\xi, \eta)$ с неизвестной функцией распределения $F_\psi(x, y)$, и есть основание предполагать, что компоненты ξ и η независимы. В этом случае надо проверить гипотезу независимости $H_0 = \{F_\psi(x, y) = F_\xi(x) \cdot F_\eta(y)\}$, где $F_\xi(x)$ и $F_\eta(y)$ - некоторые одномерные функции распределения, против альтернативной гипотезы $H_1 = \{H_0$ неверна$\}$.

Итак, пусть имеется выборка $(X, Y) = ((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ из распределения случайной величины $\psi = (\xi, \eta)$. Простой критерий согласия для проверки гипотезы H_0 для этой выборки можно построить, основываясь на методике χ^2.

Как известно, эту методику применяют для дискретных моделей с конечным числом исходов, поэтому условимся считать, что случайная величина ξ принимает конечное число s различных значений, которые обозначим $u_1, u_2, ..., u_s$, а вторая компонента η - k значений $v_1, v_2, ..., v_k$.

Если исходная модель имеет другую структуру, то предварительно группируют возможные значения случайных величин отдельно по первой и второй компонентам: множество значений ξ разбивается на s интервалов $\Delta_1, \Delta_2, \ldots, \Delta_s$, множество значений η на к интервалов $\nabla_1, \nabla_2, \ldots, \nabla_k$, а само множество значений $\psi = (\xi, \eta)$ на $N = sk$ прямоугольников $\Delta i \times \nabla i$

Обозначим через v_{ij} число наблюдений пары (u_i, v_j) (или число элементов выборки, принадлежащих прямоугольнику $\Delta_i \times \nabla_j$, если данные группируются), так что $\sum_{i=1}^{s}\sum_{j=1}^{k} v_{ij} = n$. Результаты наблюдений удобно ***расположить в виде*** таблицы сопряженности двух признаков:

	η_j				
ξ_i	v_1	v_2	...	v_k	Сумма
u_1	v_{11}	v_{12}	...	v_{1k}	$v_{1\bullet}$
u_2	v_{21}	v_{22}	...	v_{2k}	$v_{2\bullet}$
...	...	...	...	...	...
u_s	v_{s1}	v_{s2}	...	v_{sk}	$v_{s\bullet}$
Сумма	$v_{\bullet 1}$	$v_{\bullet 2}$	...	$v_{\bullet k}$	n

Далее вычисляем значение статистики

$$\hat{X}_n^2 = n\left(\sum_{i,j}\frac{v_{ij}^2}{v_{i\bullet}v_{\bullet j}} - 1\right).$$

Теорема 6.1 ***В случае справедливости гипотезы H_0 при*** $n \to \infty$ $\hat{X}_n^2 \to \xi$***, где случайная величина*** ξ ***имеет распределение*** χ^2 ***с*** $(s-1)(k-1)$ ***степенями свободы.***

Построим критерий согласия для проверки гипотезы независимости***:***

$$\rho(X,Y) = \begin{cases} H_0, & \text{если } \hat{\chi}_T^2 < \chi^2_{1-q,(s-1)(k-1)}, \\ H_1, & \text{если } \hat{\chi}_T^2 \ge \chi^2_{1-q,(s-1)(k-1)} \end{cases}$$

На рисунке 6.1 приведен текст программы, реализующей проверку гипотезы независимости по критерию χ^2 в среде Mathcad.

$A :=$ C:\..\independent.txt

Считываем из файла матрицу А, 1-й столбец которой - значения случ. величины X, а 2-й - соответствующие значения случ. величины Y

$x := A^{\langle 0 \rangle}$ $\quad x^T =$

	0	1	2	3	4	5	6	7	8
0	0.342	0.33	0.339	0.371	3.333	3.315	0.1	0.249	2.406

$y := A^{\langle 1 \rangle}$ $\quad y^T =$

	0	1	2	3	4	5	6	7	8
0	2.836	0.868	0.724	1.209	3.032	1.073	2.323	5.851	2.279

$N := \mathrm{rows}(A) \qquad N = 200$

Определяем объем выборки (X,Y)

$$U := \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 7 \\ 1 & 2 & 3 & 4 & 5 & 7 & 14 \\ 77 & 61 & 23 & 12 & 9 & 12 & 6 \end{pmatrix}^T \qquad l\_X := \mathrm{rows}(U) \qquad l\_X = 7$$

$$V := \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 & 12 \\ 66 & 51 & 35 & 22 & 12 & 7 & 7 \end{pmatrix}^T \qquad l\_Y := \mathrm{rows}(V) \qquad l\_Y = 7$$

Составляем таблицы частот для каждого из признаков:
U - таблица частот для выборки X,
V - таблица частот для выборки Y

l_X и l_Y - число интервалов группировки в соответсвующих таблицах

$$I(i,j,x,y) := x \ge U_{i,0} \wedge x < U_{i,1} \wedge y \ge V_{j,0} \wedge y < V_{j,1}$$

$i := 0..l\_X - 1 \qquad j := 0..l\_Y - 1$

$$\nu_{i,j} := \sum_{k=0}^{N-1} I(i,j,x_k,y_k) \qquad \nu = \begin{pmatrix} 28 & 17 & 14 & 9 & 3 & 3 & 3 \\ 22 & 15 & 11 & 4 & 4 & 4 & 1 \\ 6 & 8 & 4 & 3 & 1 & 0 & 1 \\ 4 & 3 & 3 & 2 & 0 & 0 & 0 \\ 2 & 3 & 1 & 1 & 2 & 0 & 0 \\ 4 & 5 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 2 \end{pmatrix}$$

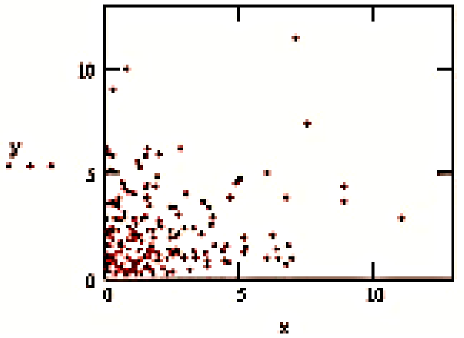
Составляем таблицу сопряженности 2-х признаков

$$\chi := N \cdot \left[\sum_{i=0}^{l\_X-1} \sum_{j=0}^{l\_Y-1} \frac{(\nu_{i,j})^2}{U_{i,2} \cdot V_{j,2}} - 1 \right] \qquad \chi = 40.315$$

Вычисляем значение статистики

$q := 0.05 \quad \chi\_kp := \mathrm{qchisq}[1 - q, (l\_X - 1) \cdot (l\_Y - 1)] \quad \chi\_kp = 50.998$

Т.к. χ<χ_kp, то гипотезу о независимости случайных величин X и Y принимаем с уровнем значимости q=0.05



Для наглядности элементы выборки (X,Y) нанесем на плоскость

Рисунок 6.1. Проверка гипотезы независимости с помощью критерия χ^2.

6.2 Выборочный коэффициент корреляции. Проверка гипотезы о значимости выборочного коэффициента корреляции

Как известно из курса теории вероятностей, коэффициент корреляции

$$r = \frac{M(\xi\eta) - M(\xi)M(\eta)}{\sigma(\xi)\sigma(\eta)}$$

характеризует наличие (или отсутствие) линейной зависимости между двумя случайными величинами ξ и η.

При $r \neq 0$ случайные величины ξ и η называются коррелированными, а при $r = 0$ - некоррелированными.

Необходимо помнить, что в общем случае некоррелированность случайных величин еще не означает их независимости.

Коэффициент корреляции удовлетворяет неравенству

$$-1 \leq r \leq 1,$$

и если $r = \pm 1$, то ξ и η связаны линейной функциональной зависимостью.

Пусть в результате эксперимента получена выборка $(X, Y) = ((X_1,Y_1), (X_2, Y_2), \ldots, (X_n,Y_n))$ из распределения случайной величины (ξ, η).

Исходя из определения коэффициента корреляции и точечных оценок для математического ожидания и среднеквадратического отклонения, дадим определение выборочного коэффициента корреляции r_B*:*

$$r_B = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_y - \overline{X} \cdot \overline{Y}}{s_0(X) s_0(Y)},$$

где $\overline{X}$ и $\overline{Y}$ - выборочные средние для выборок X и Y, а $s_0(X)$и $s_0(Y)$ - соответствующие несмещенные выборочные средние квадратические отклонения.

r_B является точечной оценкой коэффициента корреляции случайных величин ξ и η.

Разумеется, из того, что найденный по выборке $r_B \neq 0$ не следует, что коэффициент корреляции генеральной совокупности $r \neq 0$. В связи с этим проверяют гипотезу о равенстве нулю коэффициента корреляции генеральной совокупности (гипотезу о значимости выборочного коэффициента корреляции): $H_0 = \{r = 0\}$ против альтернативной гипотезы $H_1 = \{r \neq 0\}$.

Если гипотеза H_0 будет принята, что это будет означать, что выборочный коэффициент корреляции незначим, а величины ξ и η некоррелированы, если же будет принята H_1, то что выборочный коэффициент корреляции значим, а величины ξ и η коррелированы.

Предположим, что двумерная генеральная совокупность (ξ, η) распределена нормально. Тогда, если при проверки гипотезы о значимости выборочного коэффициента корреляции мы получим, что ξ и η некорре- лированы, то мы имеем право сделать вывод, что они независимы.

Вычисляем значение следующей статистики:

$$\hat{T}_n = \frac{r_B\sqrt{n-2}}{\sqrt{1-r_B}}.$$

Теорема 6.2. ***В случае справедливости гипотезы*** H_0 ***при*** $n \to \infty$ $\hat{T}_n \to \xi$, где ***случайная величина*** ξ ***имеет распределение Стьюдента с*** n - 2 ***степенями свободы.***

Построим критерий для проверки гипотезы о значимости выборочного коэффициента корреляции***:***

$$\rho(X,Y) = \begin{cases} H_0, & \text{если } \hat{T}_n < t_{1-q/2,n-2} \\ H_1, & \text{если } \hat{T}_n \geq t_{1-q/2,n-2} \end{cases},$$

где $t_{1-q/2,n-2}$ ***- квантиль распределения Стьюдента уровня 1 - q/2 (см. Приложение Б, таблица 2).***

6.3 Линейная регрессия

Пусть наблюдаемая случайная величина η зависит от случайной величины ξ. Обозначим через $f(x)$, функцию задающую зависимость среднего значения η от значений ξ

$$M(\eta/\xi = x) = f(x).$$

Уравнение y = f(x) называется уравнением регрессии.

Проведем n экспериментов, в результате которых случайная величина ξ примет последовательно значения $X_1, X_2, \ldots, X_n$, и получим соответствующие значения случайной величины η : $Y_1, Y_2, \ldots, Y_n$. Обозначим разницу между Y_i, и ее математическим ожиданием

$$\alpha_i = Y_i - M(\eta/\xi = X_i) = Y_i - f(X_i).$$

Обычно предполагают, что α_i - независимы и распределены нормально с параметрами 0, σ^2.

Требуется по значениям $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$ оценить как можно точнее функцию f (x). Сначала заранее определяют вид функции f (x). Будем предполагать, что f (x) - линейная функция

$$f(x) = ax + b.$$

Оценки неизвестных параметров a и b находят с помощью метода максимального правдоподобия или метода наименьших квадратов, суть которого мы рассмотрим несколько позже.

Эти оценки выглядят следующим образом:

$$a=\frac{\sigma(\eta)}{\sigma(\xi)}r,\, b=M(\eta)-rM(\xi)\frac{\sigma(\eta)}{\sigma(\xi)}.$$

Прямая

$$y=M(\eta)+r\frac{\sigma(\eta)}{\sigma(\xi)}(x-M(\xi))$$

называется ***прямой среднеквадратической регрессии*** η ***на*** ξ.

Величина $\Delta=\sigma^2(\eta)(1-r^2)$ называется остаточной дисперсией η на ξ**.** Она определяет величину ошибки приближенного равенства $\eta\approx a\xi+b$. Если $r=\pm 1$, то ошибка равна нулю, а величины η и ξ связаны линейной функциональной зависимостью.

Теперь, заменяя $M(\xi)$, $M(\eta), \sigma(\xi), \sigma(\eta)$ и r на их точечные оценки, получаем уравнение выборочной прямой среднеквадратической регрессии η на ξ:

$$y=\overline{Y}=r_B\frac{s_0(Y)}{s_0(X)}\left(x-\overline{X}\right).$$

Аналогично получается уравнение выборочной прямой среднеквадратической регрессии ξ на η:

$$x=\overline{X}+r_B\frac{s_0(X)}{s_0(Y)}\left(y-\overline{Y}\right).$$

6.4 Задание к практическому занятию

а) В файле ind-V.txt (V - это номер вашего варианта) в виде матрицы задана выборка (X, Y) из двухмерного распределения. Первый столбец матрицы - значения X, второй столбец - соответствующие значения Y. Проверить гипотезу о независимости случайных величин, представленных выборками X и Y с уровнем значимости q.

б) В файле cor-V.txt (V - это номер вашего варианта) находятся выборка (X, Y) из двумерного нормального распределения случайной величины (ξ, η). Первый столбец матрицы - значения X, второй столбец - соответствующие значения Y. Найти выборочный коэффициент корреляции. С уровнем значимости q проверить гипотезу о значимости выборочного коэффициента корреляции. Являются ли величины ξ и η независимыми?

На плоскости Oxy нанести элементы выборки (X, Y) и построить прямую среднеквадратической регрессии η на ξ, определить остаточную дисперсию η на ξ. Сделать вывод о правомерности описания зависимости $\eta(\xi)$ линейной функцией.

Варианты заданий

1. а) $q = 0.02$; б) $q = 0.05$.
2. а) $q = 0.01$; б) $q = 0.04$.
3. а) $q = 0.04$; б) $q = 0.03$.
4. а) $q = 0.05$; б) $q = 0.02$.
5. а) $q = 0.03$; б) $q = 0.02$.
6. а) $q = 0.01$; б) $q = 0.05$.
7. а) $q = 0.04$; б) $q = 0.03$.
8. а) $q = 0.02$; б) $q = 0.04$.
9. а) $q = 0.05$; б) $q = 0.03$.
10. а) $q = 0.01$; б) $q = 0.05$.
11. а) $q = 0.02$; б) $q = 0.04$.
12. а) $q = 0.04$; б) $q = 0.02$.
13. а) $q = 0.03$; б) $q = 0.05$.
14. а) $q = 0.01$; б) $q = 0.03$.
15. а) $q = 0.02$; б) $q = 0.04$.
16. а) $q = 0.05$; б) $q = 0.02$.
17. а) $q = 0.01$; б) $q = 0.05$.
18. а) $q = 0.04$; б) $q = 0.03$.
19. а) $q = 0.02$; б) $q = 0.02$.
20. а) $q = 0.03$; б) $q = 0.02$.

Практическое занятие №7
Дисперсионный анализ

Дисперсионный анализ - это статистический метод, предназначенный для выявления влияния отдельных факторов на результат эксперимента. Суть метода заключается в том, что общая вариация результирующего показателя расчленяется на части, соответствующие совместному и раздельному влиянию различных качественных факторов, и остаточную вариацию, аккумулирующую влияние неучтенных факторов. Статистическое изучение этих частей позволяет делать выводы о том, действительно ли тот или иной качественный фактор оказывает влияние на результирующий показатель.

Дисперсионный анализ основан на следующих допущениях:

1) наблюдения результирующего фактора ξ - это нормально распределенная случайная величина с центром распределения $M\xi = \phi(b_1, \ldots, b_m)$, где $b_1, \ldots, b_m$ - это m независимых управляющих качественных факторов;

2) дисперсия единичного наблюдения, обусловленная случайными ошибками, постоянна во всех опытах и не зависит от $b_1, \ldots, b_m$.

По числу факторов, влияние которых исследуется, различают однофакторный и многофакторный дисперсионный анализ.

7.1 Однофакторный дисперсионный анализ

Как следует из названия, данным методом исследуется влияние на результирующий признак одного качественного показателя.

Пусть в результате эксперимента получено r групп выборочных значений результирующего признака $X_{ij}\left(j = 1, \ldots, n_i, i = 1, \ldots, r\right)$, соответствующих r значениям качественного фактора; n_i - это количество наблюдений для i -го значения качественного фактора $\left(\sum_{i=1}^{r} n_i = n\right)$.

Пусть $a_i\ (i = 1, \ldots, r)$ - групповые средние, а $a = \frac{1}{r}\sum_{i=1}^{r} a_i$ - общее (генеральное) среднее.

Будем проверять гипотезу $H_0 = \{a_1 = \ldots = a_r = a\}$ о том, что качественный фактор не влияет на результирующий признак против альтернативной гипотезы $H_1 = \{H_0$ неверна$\}$.

Определим общее и групповые выборочные средние (соответственно $\overline{X}$ и $\overline{X}_i$):

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ji}, \overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ji}.$$

Как известно, выборочные групповые средние являются несмещенными и состоятельными оценками средних a_i.

Представим полную сумму квадратов отклонений результирующего признака от общего среднего в виду двух сумм квадратов отклонений:

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ji} - \overline{X}\right)^2 = \sum_{i=1}^{r} n_i\left(\overline{X}_i - \overline{X}\right)^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ji} - \overline{X}_i\right)^2 = Q_1 + Q_2$$

Сумма Q_1 представляет собой сумму квадратов отклонений групповых средних значений от общего среднего значения ("сумма квадратов между группами"), т.е. вариацию, обусловленную качественным фактором, а сумма Q_2 является суммой квадратов отклонения каждой величины от соответствующего группового среднего значения ("сумма квадратов внутри групп"), т.е. остаточную вариацию, обусловленную случайными отклонениями от групповых средних.

Теорема 7.1 В случае справедливости гипотезы H_0 величина

$$F = \frac{Q_1/(r-1)}{Q_2/(n-r)}$$

имеет распределение Фишера с $r-1, n-r$ степенями свободы.

Отсюда, для проверки гипотезы H_0 при уровне значимости q получаем следующий критерий:

$$\rho(X)=\begin{cases} H_0, & \text{если} \quad F \le F_{1-q,r-1,n-r,} \\ H_1, & \text{в} \cdot \text{противном} \cdot \text{случае} \end{cases}$$

На практике для вычисления сумм Q_1, Q_2, Q бывает удобнее пользоваться формулами

$$Q_1=\sum_{i=1}^{r}\frac{\left(\sum_{j=1}^{n_i}X_{ji}\right)^2}{n_i}-\frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{n_i}X_{ji}\right)^2}{n},$$

$$Q_2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}X_{ji}^2-\sum_{i=1}^{r}\frac{\left(\sum_{j=1}^{n_i}X_{ji}\right)^2}{n_i}, \quad Q=\sum_{i=1}^{r}\sum_{j=1}^{n_i}X_{ji}^2-\frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{n_i}X_{ji}\right)^2}{n}$$

Приведем пример. Предположим, на экспертную оценку отправлено 15 видов товара. Каждого вида товара оправлялось по 20 образцов. Оценив каждый образец, эксперт должен был дать среднюю оценку каждому виду товара. Экспертиза проводилась двумя экспертами. Необходимо выяснить, насколько субъективной была эта экспертиза. Экспертные оценки приведены в следующей таблице:

Средние оценки экспертов по каждому виду товара	Виды товара														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1-й эксперт	4	3,2	4,6	3,8	3,4	3,2	3,5	4,6	3,7	4,1	5	3,3	4	5	4,9
2-й эксперт	3,9	3,9	4,1	4,3	1,9	3,2	2,3	5	4,9	2,7	4,1	5	3,6	4	2,5

На рисунке 7.1 приведен текст программы в среде Mathcad, проверяющей гипотезу о том, что личность эксперта не влияет на оценку товаров. По результатам статистического анализа эта гипотеза была принята.

$n1 := 15 \qquad n2 := 15 \qquad n := n1 + n2$

Вводим количество наблюдений для каждого из значений качественного фактора и вычисляем общее число наблюдений

$x := (4\ \ 3.2\ \ 4.6\ \ 3.8\ \ 3.4\ \ 3.2\ \ 3.5\ \ 4.6\ \ 3.7\ \ 4.1\ \ 5\ \ 3.3\ \ 4\ \ 5\ \ 4.9)^T$

$y := (3.9\ \ 3.9\ \ 4.1\ \ 4.3\ \ 1.9\ \ 3.2\ \ 2.3\ \ 5\ \ 4.9\ \ 2.7\ \ 4.1\ \ 5\ \ 3.6\ \ 4\ \ 2.5)^T$

Вводим значения результирующего признака при каждом значении качественного фактора

$$Q1 := \frac{\left(\sum_{i=0}^{n1-1} x_i\right)^2}{n1} + \frac{\left(\sum_{i=0}^{n2-1} y_i\right)^2}{n2} - \frac{\left[\sum_{i=0}^{n1-1} x_i + \left(\sum_{i=0}^{n2-1} y_i\right)\right]^2}{n} \qquad Q1 = 0.8$$

Вычисляем "сумму квадратов между группами" - Q1, "сумму квадратов внутри групп" - Q2 и "общую сумму квадратов" - Q

$$Q2 := \sum_{i=0}^{n1-1} (x_i)^2 + \sum_{i=0}^{n2-1} (y_i)^2 - \left[\frac{\left(\sum_{i=0}^{n1-1} x_i\right)^2}{n1} + \frac{\left(\sum_{i=0}^{n2-1} y_i\right)^2}{n2}\right] \qquad Q2 = 19.613$$

$$Q := \sum_{i=0}^{n1-1} (x_i)^2 + \sum_{i=0}^{n2-1} (y_i)^2 - \frac{\left(\sum_{i=0}^{n1-1} x_i + \sum_{i=0}^{n2-1} y_i\right)^2}{n} \qquad Q = 20.414 \qquad Q1 + Q2 = Q = 1$$

$R1 := Q1 \qquad R2 := \dfrac{Q2}{n-3} \qquad R := \dfrac{Q}{n-1}$

Находим среднее квадратов - R1, R2, R3

$R1 = 0.8 \qquad R2 = 0.726 \qquad R = 0.704$

$F := \dfrac{R1}{R2} \qquad F = 1.102$

Вычисляем дисперсионное отношение F

$F\_кр := qF(0.95, 1, n-2) \qquad F\_кр = 4.196$

Т. к. F<F_кр, то делаем вывод о том, личность эксперта существенно не влияет на оценку товаров

Рисунок 7.1 - Проверка гипотезы об отсутствии влияния одного качественного фактора на результирующий показатель

7.2 Двухфакторный дисперсионный анализ

В данном случае исследуется наличие или отсутствие влияния на результирующий признак двух качественных показателей.

Пусть рассматривается два фактора - A и B. Фактор A может принимать r значений $(A = \{A_1, \dots, A_r\})$ а фактор B - s значений $(B = \{B_1, \dots, B_s\})$.

В результате эксперимента получены выборочные значения результирующего признака $Xj_{ik},\ j=1,\dots,n_{ik},\ i=1,\dots,r,\ k=1,\dots,s$; n_{ik} - это количество наблюдений при i-м значения качественного фактора A и k-м значении качественного фактора $B\left(\sum_{i=1}^{r}\sum_{k=1}^{s} n_{ik} = n\right)$.

По указанной выборке будем проверять справедливость следующих гипотез:

H_A - о том, что качественный фактор A не влияет на результирующий признак,

H_B - о том, что фактор B не влияет на результирующий признак,

H_{AB} - о том, что взаимодействие факторов A и B не влияет на результирующий признак.

Для этого вводим общее и групповое выборочные средние:

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{n_{il}} X_{jik},\ \overline{X}_{ik} = \frac{1}{n_{ik}}\sum_{j=1}^{n_{ik}} X_{jik}.$$

Вычисляем значения Q, Q_1, Q_2, Q_3, Q_4 по формулам:

$$Q = \sum_{k=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{n_{ik}}\left(X_{jik} - \overline{X}\right)^2, Q_1 = \sum_{k=1}^{s}\sum_{i=1}^{r} n_{ik}\left(\frac{1}{s}\sum_{k=1}^{s}\overline{X}_{ik} - \overline{X}\right)^2,$$

$$Q_2 = \sum_{k=1}^{s}\sum_{i=1}^{r} n_{ik}\left(\frac{1}{r}\sum_{i=1}^{r}\overline{X}_{ik} - \overline{X}\right)^2, \tag{7.1}$$

$$Q_3 = \sum_{k=1}^{s}\sum_{i=1}^{r} n_{ik}\left(\overline{X}_{ik} - \frac{1}{s}\sum_{k=1}^{s}\overline{X}_{ik} - \frac{1}{r}\sum_{i=1}^{r}\overline{X}_{ik} + \overline{X}\right)^2,$$

$$Q_4 = \sum_{k=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{n_{ik}}\left(X_{jik} - \overline{X}_{ik}\right)^2.$$

Проверку гипотез проводим по следующему критерию:

а) Если $\dfrac{Q_1/(r-1)}{Q_4(n-rs)} \ge F_{1-q,r-1,n-rs}$, то гипотеза H_A отвергается;

б) Если $\dfrac{Q_2/(s-1)}{Q_4(n-rs)} \ge F_{1-q,s-1,n-rs}$,то гипотеза H_B отвергается;

в) Если $\dfrac{Q_3/((r-1)(s-1))}{Q_4/(n-rs)} \ge F_{1-q,(r-1)(s-1),n-rs}$, то гипотеза H_{AB} отвергается (см. Приложение Б, таблица 4).

7.3 Задание к практическому занятию

Исследовать влияние на результирующий показатель: а) одного качественного фактора, б) двух качественных факторов. Уровень значимости для проверки гипотез взять равным $q = 0.05$.

Указание: Под буквой б) данные необходимо считывать из текстового файла. Данные в файле располагаются в виде матрицы следующим образом: в 1-м столбце данные соответствуют значениям факторов (A_1, B_1), во 2-м столбце - значениям (A_2, B_1), в 3-м - (A_1, B_2), в 4-м - (A_2, B_2).

Соответственно, чтобы можно было воспользоваться формулами (7.1), необходимо правильно считать данные из файла, как это сделано, например, в программе, текст которой приведен на рисунке 7.2.

Рисунок 7.2 - Пример считывания данных из файла

Если вы считали данные подобным образом, то в дальнейшем, чтобы обратиться к элементу Xj_{ik}, пользуйтесь записью $(X_{i,k})_j$.

7.4 Варианты заданий

1. а) Исследовать влияние посещения секций и кружков во внеклассное время на успеваемость школьников. Качественный фактор - количество часов, проводимых школьниками на дополнительных занятиях. Результирующий признак - средние баллы учеников по совокупности предметов за год. Согласно значениям качественного фактора ученики были поделены на 3 группы (по 16, 12 и 10 человек соответственно).
Данные о средних баллах школьников приведены в таблице:

б) Исследовать влияние на скорость прорастания семян томатов следующих факторов: температуры воздуха (фактор А) и влажности воздуха (фактор В). Значения фактора $A : A = A_1$ - температура воздуха ниже 22°, $A = A_2$ - температура воздуха выше 22°. Значения фактора $B : B = B_1$ - влажность воздуха ниже 90 процентов, $B = B_2$ - влажность воздуха выше 90 процентов. Время прорастания семян (в часах) при различных значениях факторов приведено в файле disp-1.txt.

Количество часов T, проводимое на дополнительных занятиях	Средний балл за год															
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}	X_{16i}
T=0	3,74	3,35	4,48	5,00	3,93	3,88	3,40	4,10	4,21	3,81	4,64	4,58	3,99	3,72	2,97	3,70
0<T≤3	4,36	4,06	3,52	4,44	3,22	4,27	4,42	3,89	3,68	3,95	4,17	4,29				
T>3	2,55	4,03	3,48	3,46	3,74	3,35	3,87	3,26	4,65	2,34						

2. а) Исследовать влияние поведения цены за барель нефти на курс акций некоторого предприятия. Результирующий признак - цена акции предприятия. Согласно значениям качественного фактора (поведения цены на нефть) данные были поделены на 3 группы (по 15, 14 и 13 значений соответственно). Данные о курсе акций предприятия приведены в следующей таблице:

Поведение цены на нефть	Курс акции														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
цена растет	106,12	97,22	99,55	100,07	99,13	98,98	98,13	102,12	99,55	105,00	100,18	101,60	96,35	99,35	100,60
цена стабильна	97,96	102,12	102,07	103,75	98,46	100,92	98,84	99,50	103,10	103,64	96,82	100,98	95,11	98,55	
цена падает	100,58	97,99	101,65	101,45	100,41	99,83	99,04	100,81	96,69	100,46	100,73	101,11	101,22		

б) Исследовать влияние на успеваемость студентов по физике следующих факторов: посещаемости лекционных занятий (фактор А) и активности при работе на практических занятиях (фактор В).Значения фактора $A : A = A_1$ - посещаемость ниже 60 процентов, $A = A_2$ - посещаемость ниже 60 процентов. Значения фактора $B : B = B_1$- низкая активность, $B = B_2$- высокая активность. Средние баллы студентов по физике за 4 семестра при различных значениях факторов приведены в файле disp-2.txt.

3. а) Исследовать влияние на урожайность огурцов уровня влажности воздуха в теплице. Результирующий признак - количество килограммов огурцов, собранных с одного куста за сезон. Согласно значениям качественного фактора данные были поделены на 3 группы (по 16, 14 и 15 значений соответственно). Данные об урожайности огурцов приведены в таблице

Влажность воздуха	Урожай с одного куста огурцов															
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}	X_{16i}
низкая	19,84	18,64	19,90	22,47	20,72	21,45	24,50	21,95	17,92	22,05	20,75	20,52	20,04	19,12	20,27	23,99
средняя	18,20	19,96	23,91	18,51	23,05	17,24	20,69	19,34	17,64	17,81	20,57	24,69	16,08	17,03		
высокая	20,05	20,74	21,82	20,39	18,33	22,59	18,53	21,82	19,74	23,85	20,90	18,71	17,68	18,22	21,25	

б) Исследовать влияние на размер выпускаемой на заводе детали следующих факторов: станка, на котором производится деталь (фактор А) и рабочего, изготавливающего деталь (фактор В). Значения фактора $A : A = A_1$ - 1-й станок, $A = A_2$ - 2-й станок. Значения фактора $B : B = B_1$- 1-й рабочий, $B = B_2$- 2-й рабочий. Размеры получаемых деталей (в миллиметрах) при различных значениях факторов приведены в файле disp-3.txt.

Сезон	Количество потребленной энергии														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
зима	255,95	295,30	324,45	292,69	278,12	291,27	290,55	305,73	330,64	269,81,	329,45	283,01	298,33	314,13	327,66
лето	319,14	274,65	325,24	288,73	309,48	326,89	320,48	314,63	293,57	314,86	275,53	329,28	287,35	304,11	
межсезонье	301,13	331,65	330,68	318,07	335,80	338,58	332,63	307,37	343,26	308,51	352,80	315,01	334,53	280,54	364,32

4. а) Исследовать влияние фактора сезонности на среднее за сезон количество потребляемой энергии семьей из 3-х человек.

Согласно значениям качественного фактора данные были поделены на 3 группы (по
15, 14 и 15 значений соответственно). Данные о количестве потребленной энергии (в кВт) приведены в таблице

б) Исследовать влияние на всхожесть семян следующих факторов: свежести семян (фактор А) и освещенности теплицы (фактор В). Значения фактора $A : A = A_1$- свежие семена (прошлого урожая), $A = A_2$ - не свежие семена. Значения фактора $B : B = B_1$- хорошая освещенность, $B = B_2$ - плохая освещенность. Доли взошедших семян из каждой пачки приведены в файле disp-4.txt.

5 а) Исследовать влияние уровня кислотности почвы на урожайность свеклы. Согласно значениям качественного фактора данные были поделены на 3 группы (по 13, 14 и 15 значений соответственно). Данные о среднем урожае свеклы с 1 Га земли для каждого хозяйства приведены в таблице

Вид почвы	Средняя урожайность свеклы в 1 Га (в Цт.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
кислая	546,10	594,95	570,20	605,79	565,66	601,77	592,77	598,43	576,24	600,90	579,93	603,81	572,39		
щелочная	592,60	615,74	589,76	583,63	584,38	619,75	581,59	577,47	611,24	586,72	609,74	611,95	609,35	602,63	607,07
нормальная	629,24	621,44	611,20	612,20	617,45	630,55	650,25	616,24	625,16	625,03	649,04	641,62	573,71	626,92	

б) Исследовать влияние на заболеваемость гриппом и ОРЗ следующих факторов: количества времени, выделяемых человеком на сон (фактор А) и регулярность занятий спортом (фактор В). Значения фактора $A : A = A_1$- сон менее 8 часов в сутки, $A = A_2$- сон более 8 часов. Значения фактора $B : B = B_1$- занятие спортом не реже 1 раза в неделю, $B = B_2$ - реже 1 раза в неделю. Исследования проводились в нескольких городах. Данные о том, сколько раз в среднем болеет за год житель каждого города приведены в файле disp-5.txt.

6 а) Исследовать влияние времени суток на количество вызовов скорой помощи. Анализ проводился в 15 городах с примерно одинаковой численностью населения в течение года. Данные о среднем количестве вызовов бригад скорой помощи в каждом городе приведены в таблице

Время суток	Среднее число вызовов скорой помощи за указанный период														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
8.00-16.00	62,98	89,72	86,03	95,21	91,80	88,17	73,01	81,72	85,98	81,28	91,88	91,29	70,46	49,99	85,75
16.00-24.00	94,45	95,84	91,08	82,14	97,79	100,08	96,74	103,86	105,55	112,58	81,53	125,17	110,20	99,24	95,50
0.00-8.00	132,64	133,18	132,43	141,26	135,21	127,13	133,41	134,68	146,26	133,27	133,86	145,24	132,56	155,29	137,19

б) Исследовать влияние на количество крупных ДТП (с участием более 2-х машин или с наличием пострадавших) следующий факторов: плотности транспортного потока (фактор А) и наличие гололеда (фактор В). Значения фактора $A : A = A_1$ - плотный транспортный поток, $A = A_2$ - свободное движение. Значения фактора $B : B = B_1$- наличие гололеда, $B = B_2$- отсутствие гололеда. Исследования проводились в нескольких городах. Данные о среднем количестве крупных ДТП в час приведены в файле disp-6.txt.

7 а) Исследовать влияние прослушиваемой водителем музыки на скорость движения автомобиля по незагруженной транспортом трассе. Согласно значениям качественного фактора данные были поделены на 3 группы (по 10, 14 и 15 значений соответственно). Данные о среднем скорости водителей приведены в таблице

Вид музыки	Средняя скорость автомобиля на трассе														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
классическая	74,75	82,83	88,62	82,04	84,69	93,39	90,19	92,05	76,44	76,95					
популярная	100,69	92,44	106,97	98,18	93,56	92,77	94,83	105,58	97,55	100,90	112,62	92,94	100,02	111,08	108,93
рок	104,81	105,32	107,65	109,92	107,38	98,16	88,85	119,11	114,72	93,44	106,44	121,97	101,80	84,69	90,89

б) Исследовать влияние на посещаемость человеком кинотеатров следующих факторов: возраста (фактор А) и активности человека как читателя художественной литературы (фактор В). Значения фактора $A : A = A_1$ - возраст от 18 до 30 лет, $A = A_2$- от 31 года. Значения фактора $B : B = B_1$- количество прочитанных за год книг 0-3, $B = B_2$ - более 3-х. Опрос проводился в нескольких кинотеатрах. Данные о том, сколько раз в год в среднем человек посещает кинотеатры, приведены в файле disp-7.txt.

8 а) Исследовать влияние уровня дохода семьи из 4-х человек на количество потребляемого за неделю хлеба. Согласно значениям качественного фактора данные были поделены на 3 группы (по 10, 15 и 11 значений соответственно). Данные о среднем потреблении хлеба (в кг) за неделю приведены в таблице

Уровень дохода семьи	Среднее количество потребляемого за неделю хлеба (в кг.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
низкий	2,79	3,44	2,72	2,55	2,57	3,55	2,49	2,37	3,31	2,63					
средний	3,06	3,02	2,81	2,67	3,05	2,98	2,43	3,30	3,26	3,42	3,31	3,21	2,78	2,86	3,24
высокий	3,30	3,08	2,79	2,82	2,97	3,33	3,88	2,93	3,18	3,18	3,75				

б) Исследовать влияние на объем входящего интернет-трафика следующих факторов: отношения пользователя к компьютерным играм (фактор А) и наличия в домашней сети пользователя бесплатных ресурсов (фактор В). Значения фактора $A: A = A_1$ - пользователь увлекается компьютерными играми, $A = A_2$ - не увлекается. Значения фактора $B: B = B_1$ - большой объем бесплатных ресурсов в домашней сети, $B = B_2$ - бесплатных ресурсов не имеется, либо их мало. Данные о среднем входящем интернет-трафике (в Мб.) приведены в файле disp-8.txt.

9 а) Исследовать влияние возраста покупателя на количество покупаемых в магазине глазированных сырков. Согласно значениям качественного фактора данные были поделены на 3 группы (по 12, 15 и 11 значений соответственно). Данные о среднем количестве покупаемых сырков за одно посещение магазина приведены в таблице

Возраст покупателя	Среднее количество покупаемых за 1 раз глазированных сырков														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
до 25	1,26	1,37	0,99	1,14	1,08	1,21	0,76	0,65	2,52	0,86	0,36	1,38			
25-45	3,40	3,07	2,50	2,54	2,56	3,14	2,80	3,65	2,31	3,19	3,22	3,88	2,68	3,28	3,77
старше 45	0,59	2,15	0,98	1,73	0,97	0,88	1,71	0,54	1,44	1,71	2,23				

б) Исследовать влияние на склонность к сердечно-сосудистым заболеваниям у людей старше 50 лет следующих факторов: пола пациента (фактор A) и курения (фактор B). Значения фактора $A : A = A_1$ - пациент - женщина, $A = A_2$ - пациент - мужчина. Значения фактора $B : B = B_1$ - пациент курит, $B = B_2$ - пациент не курит. Данные о среднем количестве инфарктов на 100 человек старше 50 лет приведены в файле disp-9.txt.

10 а) Исследовать влияние климата на количество потребляемых мясных продуктов. Анализ проводился в нескольких регионах. Согласно значениям качественного фактора данные были поделены на 3 группы (по 13, 15 и 9 значений соответственно). Данные о среднем количестве потребляемых в неделю мясных продуктов (в кг.) жителями каждого региона приведены в таблице

Климатические условия	Среднее количество мяса, употребляемого жителем региона за неделю (в кг.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
северные регионы	0,46	0,52	0,86	0,91	0,54	0,64	0,67	0,74	0,69	0,61	0,65	0,53	0,76		
средняя полоса	0,41	0,33	0,41	0,28	0,39	0,36	0,40	0,67	0,39	0,33	0,56	0,61	0,18	0,59	0,23
южные регионы	0,09	0,22	0,27	0,40	0,27	0,06	0,30	0,36	0,42						

б) Исследовать влияние на спрос на фотоуслуги следующих факторов: дня недели(фактор A) и времени года (фактор B). Значения фактора $A : A = A_1$ - рабочий день, $A = A_2$ - выходной. Значения фактора B : $B = B_1$- теплое время года, $B = B_2$- холодное время года. Данные о среднем количестве отпечатываемых за день фотографий по нескольким фотолабораториям приведены в файле disp-10.txt.

11 а) Исследовать влияние возраста на продолжительность разговоров по мобильному телефону. Согласно значениям качественного фактора данные были поделены на 3 группы (по 15, 15 и 14 значений соответственно). Данные о средней продолжительности телефонных разговоров в день (в минутах) приведены в таблице

Возраст абонента	Средняя продолжительность телефонных разговоров в день (в мин.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
16-25 лет	4,91	4,64	4,67	5,05	5,10	4,82	5,33	5,12	5,26	5,10	4,75	5,20	4,77	4,99	4,94
25-45 лет	3,78	4,14	4,02	4,10	4,21	3,99	4,11	4,30	3,67	3,90	4,04	3,95	4,04	4,13	4,05
45-60 лет	1,17	0,81	0,94	1,23	1,00	1,04	0,91	1,23	1,33	0,65	1,38	1,13	1,51	1,03	

б) Исследовать влияние на количество потребляемой человеком минеральной воды следующих факторов: образа жизни (фактор А) и пола (фактор В). Значения фактора $A: A = A_1$ - занимается спортом, $A = A_2$ - не занимается спортом. Значения фактора $B: B = B_1$ - мужчина, $B = B_2$ - женщина. Данные о среднем количестве выпиваемой за неделю минеральной воды (в литрах) приведены в файле disp-11.txt.

12 а) Исследовать объективность судей, оценивавших спортсменов на некотором соревновании. Перед судьями выступало 15 спортсменов, совершивших по 10 подходов. Данные о средних баллах спортсменов (по 6-ти бальной шкале) приведены в таблице

эксперт	Средние баллы спортсменов														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1-й	3,9	4,5	4,0	4,7	3,8	4,4	4,7	4,0	3,8	4,3	4,5	4,5	4,5	4,1	4,0
2-й	4,8	3,9	4,4	4,7	4,5	3,9	4,1	4,7	4,8	4,6	4,2	3,8	4,7	4,8	3,6
3-й	5,2	3,9	5,1	4,1	4,7	4,0	4,6	4,9	4,5	4,5	5,1	4,9	4,4	4,4	4,9

б) Исследовать влияние на заболеваемость ОРЗ следующих факторов: как человек провел летний отпуск (фактор А) и принимает ли он витаминные комплексы (фактор В). Значения фактора $A: A = A_1$ - совершил длительный выезд на природу/отдыхал на курорте, $A = A_2$ - провел отпуск дома. Значения фактора $B: B = B_1$ - принимает витаминные комплексы, $B = B_2$ - не принимает. Данные о среднем количестве перенесенных за год ОРЗ приведены в файле disp-12.txt.

13 а) Исследовать влияние "возраста" йогурта на содержание в нем молочнокислых бактерий. Согласно значениям качественного фактора данные были поделены на 3 группы (по 15, 13 и 13 значений соответственно). Данные о среднем содержании молочнокислых бактерий в 1 г. йогурта (в 10^6 КОЕ) приведены в таблице

Возраст йогурта	Содержание молочнокислых бактерий (в 10^6 КОЕ)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1/3 от срока годности	22,04	21,99	22,08	21,92	22,10	21,92	21,98	22,21	21,99	22,11	21,94	22,09	22,02	22,05	22,00
2/3 от срока годности	16,09	15,94	15,89	16,14	16,16	15,93	16,09	15,95	15,95	16,09	15,89	16,06	15,94		
конец срока годности	9,98	10,10	9,93	10,10	9,68	9,88	10,02	10,08	10,09	10,12	10,05	9,96	10,31		

б) Исследовать влияние на срок службы стиральных машин следующих факторов: жесткости воды (фактор А) и используемое число оборотов центрифуги (фактор В). Значения фактора $A : A = A_1$ - жесткая вода, $A = A_2$ - ph нейтральная или мягкая вода. Значения фактора $B : B = B_1$ - 1000 оборотов, $B = B_2$ - 800 оборотов. Данные о средней продолжительности работы стиральных машин некоторой марки (в часах) приведены в файле disp-13.txt.

14. а) Исследовать влияние вида применяемых удобрений на содержание калия в фасоли. Согласно значениям качественного фактора данные были поделены на 3 группы (по 12, 11 и 13 значений соответственно). Данные о среднем калия в фасоли (в мг. на 100 г.) приведены в таблице

Вид удобрения	Среднее содержание калия в фасоли (в мг. на 100 г.)												
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}
1	1062,1	1059,3	1059,7	1061,3	1060,3	1060,6	1058,3	1059,7	1059,1	1060,8	1060,8	1059,0	
2	1062,9	1061,9	1059,2	1060,7	1060,1	1061,1	1059,7	1060,1	1058,4	1061,5	1057,9		
3	1059,7	1059,2	1059,0	1058,8	1059,8	1058,5	1061,4	1058,5	1059,3	1059,7	1061,3	1060,2	1059,6

б) Исследовать влияние на количество детей у женщин следующих факторов: места проживания (фактор А) и наличия высшего образования (фактор В). Значения фактора $A : A = A_1$ - сельская местность, $A = A_2$ - город. Значения фактора $B : B = B_1$ - есть высшее образование, $B = B_2$ - высшего образования нет. Данные были собраны по 30 сельским пунктам и 30 городам. Данные о среднем количестве детей у женщин 40 лет приведены в файле disp-14.txt.

14 а) Исследовать влияние года обучения студента в ВУЗе на посещаемость им лекционных занятий. Согласно значениям качественного фактора данные были поделены на 3 группы - (по 15, 14 и 13 значений соответственно). Данные посещаемости лекционных занятий студентами каждой специальности (в процентах) приведены в таблице

Год обучения	Средняя посещаемость лекционных занятий (в %)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1-й	76,32	67,52	70,51	75,23	68,72	66,11	72,25	66,93	72,51	66,23	75,77	73,84	67,83	68,68	66,66
2-й – 3-й	63,06	61,53	58,18	61,09	61,24	63,21	62,58	55,92	56,45	58,92	66,67	63,11	63,02	62,15	
4-й – 5-й	51,82	49,86	48,87	51,81	54,51	58,69	48,14	52,42	53,48	53,23	57,99	54,53	55,78		

б) Исследовать влияние на курс акций некоторого предприятия следующих факторов: поведения курса акций предприятия А (фактор А) и поведения курса акций предприятия Б (фактор В). Значения фактора $A : A = A_1$ - курс растет, $A = A_2$ - курс падает. Значения фактора $B : B = B_1$- курс растет, $B = B_2$ - курс падает. Данные о цене акции изучаемого предприятия приведены в файле disp-15.txt.

15 а) Исследовать влияние сорта яблок на содержание в них железа. Согласно значениям качественного фактора данные были поделены на 3 группы - (по 15, 13 и 15 значений соответственно). Данные о содержание железа в яблоках (в мг. на 100 г.) приведены в таблице

Сорт яблок	Содержание железа в разных сортах яблок (в мг. на 100 г.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
сорт «А»	2,51	2,51	2,50	2,50	2,50	2,50	2,50	2,49	2,53	2,50	2,49	2,51	2,50	2,49	2,49
сорт «Б»	2,46	2,45	2,44	2,44	2,44	2,45	2,45	2,46	2,44	2,45	2,45	2,47	2,44		
сорт «С»	2,38	2,41	2,39	2,40	2,39	2,40	2,38	2,40	2,40	2,40	2,41	2,39	2,40	2,40	2,40

б) Исследовать влияние на урожайность пшеницы следующих факторов: количества осадков за сезон (фактор А) и давность содержания земли "под паром"(фактор В). Значения фактора $A : A = A_1$- количество осадков выше нормы, $A = A_2$ - количество осадком ниже нормы. Значения фактора $B : B = B_1$- в прошлом году земля находилась "под паром", $B = B_2$ - не находилась. Данные об урожайности пшеницы (в ц. с Га) приведены в файле disp-16.txt.

16 а) Исследовать влияние возраста покупателя на то, сколько денег он тратит на покупку мясной продукции в магазине. Согласно значениям качественного фактора данные были поделены на 3 группы (по 12, 15 и 13 значений соответственно). Данные о среднем количестве средств, потраченных покупателями магазина на мясную продукцию за 1 посещение, приведены в таблице

Возраст	Средние затраты покупателей на мясную продукцию за 1 посещение (в руб)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
18-25	159	158	139	146	166	150	147	145	154	155	149	151			
25-40	145	162	177	151	168	159	149	167	163	151	141	150	147	156	150
старше40	74	89	88	77	62	87	53	78	66	72	74	78	83		

б) Исследовать влияние на успеваемость школьников по биологии следующих факторов: участия в олимпиадах (фактор А) и прохождения летней практики на школьном приусадебном участке(фактор В). Значения фактора $A : A = A_1$ - принимал участие в олимпиадах, $A = A_2$ не принимал. Значения фактора $B : B = B_1$ - проходил практику на приусадебном участке, $B = B_2$ - не проходил. Данные средней оценке школьников за 2 года приведены в файле disp-17.txt.

17 а) Исследовать объективность экспертов, оценивавших качество "Докторской" колбасы от разных производителей. На экспертизу было представлено по 10 образцов колбасы от 15 производителей. Данные о средних баллах, выставленных экспертами каждому производителю (по 10-ти бальной шкале), приведены в таблице

Эксперт	Средние оценки, выставленные «Докторской» колбасе от 15 пользователей														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1-й	7,0	6,4	7,4	6,3	6,0	7,7	5,4	6,5	6,4	6,7	7,6	7,7	6,2	5,8	6,4
2-й	5,9	7,7	7,1	7,5	8,1	7,0	7,6	8,5	5,4	6,5	7,2	6,8	7,2	7,7	7,2
3-й	8,5	4,1	7,3	6,6	7,6	8,1	6,7	8,1	5,1	9,2	8,5	7,1	7,1	5,4	5,7

б) Исследовать влияние на формирование цены на товар "С" следующих факторов: изменение спроса на товар "А" (фактор А) и изменение спроса на товар "В" (фактор В). Значения фактора $A : A = A_1$ - спрос растет, $A = A_2$ - спрос падает. Значения фактора $B : B = B_1$ - спрос растет, $B = B_2$ - спрос падает. Данные о средней цене на товар "С" по городу (в руб.) приведены в файле disp-18.txt.

18 а) Исследовать влияние сорта томатов на содержание в нем витамина С. Согласно значениям качественного фактора данные были поделены на 3 группы (по 15, 14 и 13 значений соответственно). Данные о среднем содержании витамина С в томатах (в мг. на 100 г.) приведены в таблице

Сорт томатов	Содержание витамина С в томатах (в мг. на 100 г.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
сорт «А»	34,93	34,81	35,54	34,04	35,34	34,87	34,99	35,30	34,38	34,74	35,03	34,82	35,21	34,14	35,56
сорт «В»	34,06	35,42	34,73	35,72	34,60	35,60	35,35	35,51	34,90	35,58	35,00	35,66	34,79	34,76	
сорт «С»	35,34	35,06	34,31	34,98	35,17	35,73	35,13	34,73	35,10	35,41	35,26	34,31	35,14		

б) Исследовать влияние на склонность к инфекционным заболеваниям у детей следующих факторов: пола ребенка (фактор А) и употребления витаминных комплексов (фактор В). Значения фактора $A: A = A_1$ - девочка, $A = A_2$ - мальчик. Значения фактора $B: B = B_1$ - ребенок принимает витаминные комплексы, $B = B_2$ - не принимает. Данные о среднем количестве перенесенных ребенком инфекционных заболеваний за год приведены в файле disp-19.txt.

19 а) Исследовать влияние поведения атмосферного давления на количество вызовов скорой помощи. Анализ проводился в 15 городах с примерно одинаковой численностью населения в течение года. Данные
о среднем количестве вызовов бригад скорой помощи (в день) в каждом городе приведены в таблице

атм.давление	Среднее число вызовов скорой помощи в день														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
падает	97,2	97,9	95,5	91,1	98,9	100,0	98,4	101,9	102,8	106,3	90,8	112,6	105,1	99,6	97,8
стабильно	86,5	76,7	88,7	84,9	80,8	82,5	79,2	82,5	90,9	84,3	89,6	89,5	82,3	78,0	79,5
растет	102,2	94,6	98,5	95,8	95,7	100,5	101,6	97,6	106,4	96,1	106,3	98,0	95,6	97,7	102,7

б) Исследовать влияние на продолжительность исходящих звонков с сотового телефона следующих факторов: пола абонента(фактор А) и количество отправляемых абонентом смс в день (фактор В). Значения фактора $A: A = A_1$ - женский пол, $A = A_2$ - мужской пол. Значения фактора $B: B = B_1$ - абонент отправляет до 5 смс в день, $B = B_2$ - больше 5 смс в день. Данные о средней продолжительности одного разговора абонента по исходящей связи приведены в файле disp-20.txt.

Практическое занятие №8
Метод наименьших квадратов. Построение конкретных нелинейных моделей

8.1 Нелинейная регрессия

Здесь мы имели дело с частным случаем регрессии - линейной регрессией. Теперь рассмотрим общий случай и общую постановку задачи регрессионного анализа.

Пусть имеется выборка $(X, Y) = ((X_1, Y_1), (X_2, Y_2)$, $(X_2, Y_2),..., (X_n, Y_n))$ из распределения случайной величины $\psi = (\xi, \eta)$. И пусть известно, что случайные величины ξ и η зависимы. Важное прикладное значение имеет задача о представлении одной из этих величин как функции от другой.

Проведение регрессионного анализа можно разделить на три этапа: выбор формы зависимости (типа уравнения), вычисление параметров выбранного уравнения, оценка достоверности полученного уравнения.

Выбор вида уравнения регрессии производится на основании опыта предыдущих исследований, наблюдений расположения точек (X_i, Y_i) на плоскости и т.д.

Обозначим через $f(x, \theta)$ функцию задающую зависимость среднего значения η от значений ξ (здесь $\theta = (\theta_1,..., \theta_k)$ - вектор параметров):

$$M(\eta / \xi = x) = f(x, \theta).$$

Уравнение $y = f(x, \theta)$называется уравнением регрессии.

Для определения неизвестных параметров $\theta_1,..., \theta_k$ можно использовать метод наименьших квадратов.

Суть этого метода состоит в том, что наилучшим считается такое положение линии регрессии, при котором сумма квадратов отклонений значений $f(X_i, \theta)$ от соответствующих Y_i минимальна. Метод состоит в минимизации функции

$$Q(\theta)=\sum_{i=1}^{n}\left(y_i-f\left(x_i,\theta\right)\right)^2 .$$

Приведем пример построения нелинейной регрессии с использованием метода наименьших квадратов.

Пусть при проведении эксперимента получены следующие значения величин x и y:

x	6	6.1	6.3	6.5	6.7	7	7.5	8	8.2	8.5
y	4.5	4	3.5	3	2.5	2	1.5	1	0.7	0.5

Считая справедливой зависимость $y(x,D)=D_0e^{D_1x}$, находим неизвестные параметры D_0 и D_1 с помощью метода наименьших квадратов. В результате получаем следующее уравнение регрессии:

$$y=500.1e^{-0\ 79x} .$$

Текст программы, реализующей построение уравнения регрессии приведен на рисунке 8.1.

Рисунок 8.1 - Построение уравнения регрессии с помощью метода наименьших квадратов

В данной программе для минимизации функции $Q(D)$ используется встроенная функция Minerr(). Однако минимизацию можно провести известным методом исследования функции нескольких переменных на экстремум с помощью дифференциального исчисления.

8.2 Задание к практическому занятию

В файле regrV.txt (V - это номер вашего варианта) в виде матрицы задана выборка (X, Y). Первый столбец матрицы - значения X, второй столбец - соответствующие значения Y.

1. С помощью метода наименьших квадратов построить уравнения регрессии, считая справедливыми следующие формы зависимости y от x:

а) $y = a\sin(bx)$, б) $y = \log_a bx$, в) $y = a_0 + a_1x + a_2x^2$.

Поиск минимума функции $Q(D)$ проводить, исследуя эту функцию на экстремум с помощью частных производных.

2. На одном графике изобразить исходные данные и полученные линии регрессии. Сделать вывод о том, какая из функций наилучшим образом представляет зависимость y от x.

Приложение А

Некоторые параметрические семейства распределений

1. Равномерное распределение $U_{a,b}$. Функция плотности распределения и моменты распределения:

$$u_{a,b}(t)=\begin{cases}\dfrac{1}{b-a}, t\in[a,b],\\ 0, t\notin[a,b]\end{cases},\quad MX=\frac{b-a}{2}, DX=\frac{(b-a)^2}{12}.$$

2. Показательное распределение E_λ. Функция плотности распределения и моменты распределения:

$$e_\lambda(t)=\begin{cases}\lambda e^{-\lambda t}, t>0,\\ 0, t\le 0,\end{cases}\qquad MX=\frac{1}{\lambda}, DX=\frac{1}{\lambda^2}.$$

3. Гамма-распределение $\Gamma_{\alpha,\beta}$. Функция плотности распределения:

$$\gamma_{\alpha,\beta}(x)=\begin{cases}\dfrac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)}x^{\alpha}e^{-\frac{x}{\beta}}, & x>0\\ 0, & x\le 0\end{cases}, \alpha>-1, \beta>0.$$

Моменты распределения: $MX=(\alpha+1)\beta, DX=\beta^2(\alpha+1)$.

4. Распределение Пуассона $\Pi_\lambda(\lambda>0)$:

$$P(X=m)=\frac{\lambda^m e^{-\lambda}}{m!}(m=0,1,2\ldots), MX=\lambda, DX=\lambda$$

5. Геометрическое распределение G_p:

$$P(X=k)=(1-p)^k p, p\in(0,1), k=0,1,2,\dots$$
$$MX=\frac{1-p}{p}, DX=\frac{1-p}{p^2}.$$

6. Биномиальное распределение B_p^n:

$$P(X=k)=C_k^n p^k(1-p)^{n-k}(0\le k\le n),$$
$$MX=np, DX-np(1-p).$$

7. Нормальное распределение N_{a,σ^2}. Функция плотности распределения и моменты распределения:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}, MX=a, DX=\sigma^2$$

8. Бета-распределение $\beta_{m,n}$. Функция плотности распределения:

$$\beta_{m,n}(x)=\begin{cases}\frac{\Gamma(m+n)}{\Gamma(m)\cdot\Gamma(n)}x^{m-1}(1-x)^{n-1}, & x\in(0,1)\\ 0, & x\notin(0,1)\end{cases}$$

где m > 0, n > 0.

Моменты распределения: $MX=\frac{m}{m+n}, DX=\frac{mn}{(m+n)^2(m+n+1)}$.

9. Логарифмически нормальное (логнормальное) распределение. Функция плотности распределения:

$$l(x)=\begin{cases}\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-a)^2}{2\sigma^2}}, & x>0,\\ 0, & x\le 0.\end{cases}$$

Моменты распределения: $MX = e^{-\frac{\sigma^2}{2}+a}, DX = e^{\sigma^2+2a}\left(e^{\sigma^2} - 1\right)$.

10. Распределение χ_n^2 Функция плотности распределения:

$$h_n(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{при} \cdot x > 0, \\ 0 & \text{при} \cdot x \le 0. \end{cases}, \qquad n = 1,2,\ldots$$

Моменты распределения: MX = n, DX = 2n.

11. Распределение Стьюдента T_k. Функция плотности распределения:

$$t_k(x) = \frac{1}{\sqrt{\pi k}} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \frac{1}{\left(1 + x^2/k\right)^{(k+1)/2}}, k = 1,2,\ldots$$

$$MX = 0; DX = k\frac{\Gamma(3/2)\Gamma(k/2-1)}{\sqrt{\pi}\Gamma(k/2)}, k > 2.$$

12. Распределение Фишера $F_{k,m}$. Функция плотности распределения:

$$f_{k,m}(x) = \left(\frac{k}{m}\right)^{k/2} \frac{\Gamma((k+m)/2)}{\Gamma(k/2)\Gamma(m/2)} \frac{x^{k/2-1}}{(1 + kx/m)^{(k+m)/2}}, x > 0; k, m > 0.$$

$$MX = \frac{m}{m-2}, m > 2; DX = \frac{2m^2(k+m-2)}{k(m-2)^2(m-4)^2}, m > 4$$

.

13. Распределение Коши $K_{m,n}$. Функция плотности распределения:

14.

$$k_{m,n}(x) = \frac{1}{\pi}\frac{m}{m + (x-n)^2}, m > 0, -\infty < n < \infty.$$

MX и DX не существуют.

Приложение Б

Таблица Б.1 - Квантили стандартного нормального распределения T_ε

ε	T_ε	ε	T_ε
0.010	2.3263	0.250	0.6745
0.025	1.9600	0.300	0.5244
0.050	1.6449	0.350	0.3853
0.100	1.2816	0.400	0.2533
0.150	1.0364	0.450	0.1257
0.200	0.8416	0.500	0.0000

Таблица Б.2 - Квантили распределения Стьюдента $t_{p,k}$

k	p = 0.750	p = 0.900	p = 0.990	p = 0.999
1	1.000	3.078	31.821	318
2	0.816	1.886	6.965	22.3
3	0.765	1.638	4.541	102
4	0.741	1.533	3.747	7.173
5	0.727	1.476	3.365	5.893
6	0.718	1.440	3.143	5.208
7	0.711	1.415	2.998	4.785
8	0.706	1.397	2.896	4.501
9	0.703	1.383	2.821	4.297
10	0.700	1.372	2.764	4.144
11	0.697	1.363	2.718	4.025
12	0.695	1.356	2.681	3.930
13	0.694	1.350	2.650	3.852
14	0.692	1.345	2.624	3.787
15	0.691	1.341	2.602	3.733
20	0.687	1.325	2.528	3.552
30	0.683	1.310	2.457	3.385
40	0.681	1.303	2.423	3.307
60	0.679	1.296	2.390	3.232
80	0.677	1.289	2.358	3.160
∞	0.674	1.282	2.326	3.090

Таблица Б.3 - Квантили распределения $\chi^2_{\alpha,k}$

k	α = 0.010	α = 0.025	α = 0.05	α = 0.10
1	0.00016	0.00098	0.00393	0.01580
2	0.0201	0.05060	0.1030	0.2110
3	0.1150	0.2160	0.3520	0.5840
4	0.297	0.484	0.711	1.106
5	0.554	0.831	1.150	1.161
6	0.872	1.240	1.640	2.200
7	1.240	1.690	2.170	2.830
8	1.650	2.180	2.730	3.490
9	2.090	2.700	3.330	4.170
10	2.560	3.250	3.940	4.870
11	3.050	3.820	4.570	5.580
12	3.570	4.400	5.230	6.300
13	4.110	5.010	5.890	7.040
14	4.660	5.630	6.570	7.790
15	5.230	6.260	7.260	8.550
16	5.81	6.91	7.96	9.31
17	6.41	7.56	8.67	10.1
18	7.01	8.23	9.39	10.9
19	7.63	8.91	10.1	11.7
20	8.26	9.59	10.9	12.4
21	8.90	10.3	11.6	13.2
22	9.54	11.0	12.3	14.0
23	10.2	11.7	13.1	14.8
24	10.9	12.4	13.8	15.7
25	11.5	13.1	14.6	16.5
30	15.0	16.8	16.5	20.6
35	18.5	20.6	22.5	24.8
40	22.2	24.4	26.5	29.1
45	25.9	28.4	30.6	33.4
50	29.7	32.4	34.8	37.7
75	49.5	52.9	56.1	59.8
100	70.1	74.2	77.9	82.4

Таблица Б.4 - Квантили распределения Фишера $F_{p,k1,k2}$

k2	$k_1 = 1$	$k_1 = 2$	$k_1 = 1$	$k_1 = 2$	$k_1 = 1$	$k_1 = 2$
1	161.4	199.5	4052	4999.5	405300	500000
2	18.51	19.00	98.50	99.00	998.5	999
3	10.13	9.55	34.12	30.82	167.0	148.5
4	7.71	6.94	21.20	18.00	74.14	61.25
5	6.61	5.79	16.26	13.27	47.18	37.12
6	5.99	5.14	13.75	10.92	35.51	27.00
7	5.59	4.74	12.25	9.55	29.25	21.69
8	5.32	4.46	11.26	8.65	25.42	18.49
9	5.12	4.26	10.56	8.02	22.86	16.39
10	4.96	4.10	10.04	7.56	21.04	14.91
11	4.84	3.98	9.65	7.21	19.69	13.81
12	4.75	3.89	9.33	6.93	18.64	12.97
13	4.67	3.81	9.07	6.70	17.81	12.31
14	4.60	3.74	8.86	6.54	17.14	11.78
15	4.54	3.68	8.68	6.36	16.59	11.34
16	4.49	3.63	8.53	6.23	16.12	10.97
17	4.45	3.59	8.40	6.11	15.72	10.66
18	4.41	3.55	8.29	6.01	15.38	10.39
19	4.38	3.52	8.18	5.93	15.08	10.16
20	4.35	3.49	8.10	5.85	14.82	9.95
25	4.24	3.39	7.77	5.57	13.88	9.22
30	4.17	3.32	7.56	5.39	13.29	8.77
40	4.08	3.23	7.31	5.18	12.61	8.25
60	4.00	3.15	7.089	4.98	11.97	7.76
120	3.92	3.07	6.85	4.79	11.38	7.32
∞	3.84	3.00	6.63	4.61	10.83	6.91

Таблица Б.5 - Значения функции распределения Колмогорова K(t)

t	к (t)
1.36	0.9505
1.40	0.9603
1.45	0.9702
1.52	0.9803
1.63	0.9902

Практическое занятие №9
Непрерывные случайные величины

9.1 Краткие сведения из теории

Определение. Случайная величина ξ называется непрерывной, если ее значения целиком заполняют некоторый интервал (конечный или бесконечный).

Функция

$$F_{\xi}(x) = P(\xi < x) \qquad (9.1)$$

называется функцией распределения случайной величины ξ.

Функция распределения $F_{\xi}(x)$ не убывает, непрерывна слева, также

$$F_{\xi}(-\infty) = 0; \quad F_{\xi}(+\infty) = 1 \qquad (9.2)$$

Для любого $x \in (-\infty;+\infty)$ имеем

$$0 \le F_{\xi}(x) \le 1 \qquad (9.3)$$

Если задана функция распределения $F_{\xi}(x)$, то для любого отрезка $[a; b]$ можно определить $P(\xi \in [a; b))$, а именно

$$P(a \le \xi < b) = F_{\xi}(b) - F_{\xi}(a). \qquad (9.4)$$

Плотностью распределения случайной величины ξ (плотностью вероятностей) называется функция

$$f_{\xi}(x) \frac{dF_{\xi}(x)}{dx}. \qquad (9.5)$$

Плотность распределения обладает свойствами:

$$f_{\xi}(x) \geq 0, \tag{9.6}$$

$$\int_{-\infty}^{+\infty} f_{\xi}(x)dx = 1. \tag{9.7}$$

Если известна плотность распределения случайной величины ξ то

$$P(a < \xi < b) = \int_{a}^{b} f_{\xi}(x)dx \tag{9.8}$$

З а м е ч а н и е. Для непрерывной случайной величины ξ справедливы соотношения

$$P(a \leq \xi \leq b) = P(a \leq \xi < b) = P(a < \xi \leq b) = P(a < \xi < b).$$

Формула (1.5) определяет плотность распределения тогда, когда известна функция распределения. Обратно:

$$F_{\xi}(x) = \int_{-\infty}^{x} f_{\xi}(u)du. \tag{9.9}$$

Эту формулу используют для нахождения функции распределения по заданной плотности.

Математическое ожидание непрерывной случайной величины вычисляется по формуле:

$$M\xi = \int_{-\infty}^{+\infty} x f_{\xi}(x)dx \tag{9.10}$$

Дисперсия определяется следующим равенством:

$$D\xi = M(\xi - M\xi)^2 = \int_{-\infty}^{+\infty} (x - M\xi)^2 f_\xi(x)dx \qquad (9.11)$$

Если распределение случайной величины ξ сосредоточено на отрезке $[a;b]$, то несобственные интегралы из формул (9.10)и (9.11) переходят в интегралы с конечными пределами, т. е.

$$M\xi = \int_a^b x f_\xi(x)dx \qquad (9.12)$$

$$D\xi = \int_a^b (x - M\xi)^2 f_\xi(x)dx \qquad (9.13)$$

З а м е ч а н и е. При решении задач на вычисление дисперсии иногда лучше использовать формулу (9.11), которая справедлива для любых случайных величин (как дискретных, так и непрерывных).

Начальный момент порядка определяется $k, k \in N$ соотношением

$$\alpha_k = M\xi^k = \int_{-\infty}^{+\infty} x^k f_\xi(x)dx \qquad (9.14)$$

Центральный момент порядка $k, k \in N$

$$\mu_k = M(\xi - M\xi)^k = \int_{-\infty}^{+\infty} (x - M\xi)^k f_\xi(x)dx \qquad (9.15)$$

Если значения случайной величины сосредоточены на $[a;b]$, то интегралы в (9.14) и (9.15) заменяются на интегралы по этому отрезку, т.е.

$$\alpha_k = \int_a^b x^k f_\xi(x)dx \quad , \qquad (9.16)$$

$$\mu_k = \int_a^b (x - M\xi)^k f_\xi(x)dx \,. \qquad (9.17)$$

Среднее квадратичное отклонение случайной величины ξ:

$$\sigma\xi = \sqrt{D\xi} \qquad (9.18)$$

9.2 Решение типовых задач

З а д а ч а 9.1. Случайная величина ξ имеет плотность распределения

$$f_\xi(x) = \begin{cases} a\sin x, x \in [0;\pi] \\ 0, x \notin [0,\pi] \end{cases}.$$

а) Найти параметр a.

б) Найти функцию распределения $f_\xi(x)$.

в) Вычислить $P(\xi) \in [0;\pi/4))$.

Р е ш е н и е. а) Для определения параметра a воспользуемся свойством (1.7) плотности, согласно которому

$$\int_0^\pi a\sin xdx = 1 \Leftrightarrow a\int_0^\pi \sin xdx = 1 \Rightarrow a = 1/2 \,.$$

Таким образом, случайная величина ξ имеет плотность распределения:

$$f_\xi(x)=\begin{cases}\frac{1}{2}\sin x, x\in[0,\pi]\\ 0, x\notin[0,\pi]\end{cases}.$$

б) Находим функцию распределения, используя соотношение (9.9). Для этого выделим 3 случая:

б$_1$) $x\le 0 \Rightarrow f_\xi(x)=0 \Rightarrow F_\xi(x)=\int\limits_{-\infty}^{x}0du=0$

б$_2$) $0<x\le\pi$. Тогда:

$$F_\xi(x)=\int\limits_{-\infty}^{x}f_\xi(u)du=\int\limits_{-\infty}^{0}f_\xi(u)du+\int\limits_{0}^{x}f_\xi(u)du=$$

$$=\int\limits_{0}^{x}f_\xi(u)du=\frac{1}{2}\int\limits_{0}^{x}\sin u du=-\frac{1}{2}\cos u\Big|_0^x=$$

$$=-\frac{1}{2}[\cos x-\cos 0]=\frac{1}{2}(1-\cos x)$$

б$_3$) $x>\pi$. Тогда:

$$F_\xi(x)=\int\limits_{-\infty}^{x}f_\xi(u)du=$$

$$=\int\limits_{-\infty}^{0}f_\xi(u)du+\int\limits_{0}^{\pi}f_\xi(u)du+\int\limits_{\pi}^{x}f_\xi(u)du=\int\limits_{0}^{\pi}f_\xi(u)du=$$

$$=\frac{1}{2}\int\limits_{0}^{\pi}\sin u du=-\frac{1}{2}\cos u\Big|_0^\pi=-\frac{1}{2}[\cos\pi-\cos 0]=1$$

$$F_\xi(x)=\begin{cases}0, & x\le 0\\ \frac{1}{2}(1-\cos x), & 0<x\le\pi\\ 1, & x>\pi\end{cases} \qquad (9.19)$$

в) Вычислим $P(\xi \in [0; \pi/4))$ по формуле (9.4):

$$P(\xi \in [0; \pi/4)) = F_\xi(\pi/4) - F_\xi(0) = \frac{1}{2}\left(1 - \cos\frac{\pi}{4}\right) -$$

$$-\frac{1}{2}(1 - \cos 0) = \frac{1}{2}\left(1 - \frac{\sqrt{2}}{2}\right) = \frac{2 - \sqrt{2}}{4} \approx 0{,}147.$$

Ответ. $a = 1/2$. Функция распределения имеет вид (9.19).

$$P(\xi \in [0; \pi/4)) \approx 0{,}147.$$

Задача 9.2. Задана функция распределения случайной величины ξ:

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ x/3, & 0 < x \le 3 \\ 1, & x > 3 \end{cases}$$

Найти:

а) Плотность распределения $f_\xi(x)$.

б) Математическое ожидание и дисперсию.

в) Графики функций $f_\xi(x)$ и $F_\xi(x)$.

Решение. а) Используя формулу (9.5), придем к плотности:

$$f_\xi(x) = \begin{cases} 1/3, & x \in (0,3] \\ 0, & x \notin (0,3] \end{cases} \tag{9.20}$$

б) $$M\xi = \int_0^3 x f_\xi(x)dx = \int_0^3 \frac{x}{3}dx = \frac{1}{3}\int_0^3 xdx = \frac{1}{3} \cdot \frac{x^2}{2}\Big|_0^3 = \frac{1}{3} \cdot \frac{3^2}{2} = \frac{3}{2}.$$

Для вычисления $D\xi$ применим формулу (9.11). Вычислим сначала $M\xi^2$:

$$M\xi^2 = \int_0^3 x^2 f_\xi(x)dx = \frac{1}{3}\int_0^3 x^2 dx = \frac{1}{3}\cdot\frac{x^3}{3}\Big|_0^3 = \frac{1}{3}\cdot\frac{3^3}{3} = \frac{3^3}{3^2} = 3.$$

$$\Rightarrow D\xi = 3-(3/2)^2 = 3/4.$$

в) Графики функций $f_\xi(x)$и $F_\xi(x)$ представлены на рисунке 9.1:

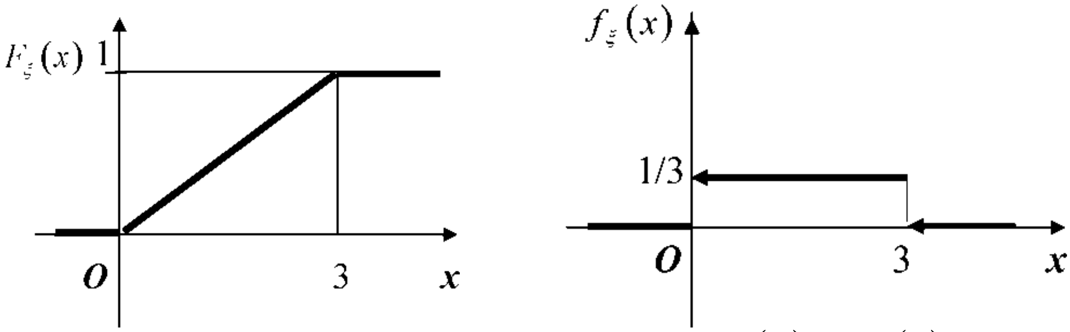


Рисунок 9.1 – Графики функций $f_\xi(x)$ и $F_\xi(x)$

Ответ. Плотность распределения задана формулой (9.20).

$M\xi$=3/2; $D\xi$=3/4.

Задача 9.3. Случайная величина ξ имеет плотность распределения

$$f_\xi(x) = \begin{cases} 0, & -\infty < x \le 1 \\ a/x^4, & 1 < x < +\infty \end{cases}.$$

Найти:

а) Значение параметра a.

б) Функцию распределения $F_\xi(x)$

в) Математическое ожидание и дисперсию.

г) $P(\xi \in [-2;2])$.

Решение. Для определения параметра a используем свойство (9.7) плотности распределения. Так как $f_\xi(x) = 0$ для $x \in (-\infty;1]$, то интегрирование будет по интервалу $(1;+\infty)$.

$$\int_1^{+\infty} f_\xi(x)dx = 1 \Leftrightarrow \int_1^{+\infty} \frac{a}{x^4} dx = 1$$

$$a\int_1^{+\infty} x^{-4}dx = 1 \Leftrightarrow a = \frac{1}{\int_1^{+\infty} x^{-4}dx} \Leftrightarrow a = 3.$$

Таким образом,

$$f_\xi(x) = \begin{cases} 0, & -\infty < x \le 1 \\ 3/x^4, & 1 < x < +\infty \end{cases}$$

б) Найдем функцию распределения величины ξ, используя соотношение (9.9).

б$_1$) Если $x \in (-\infty;1]$, то $f_\xi(x) = 0$. Поэтому $F_\xi(x) = 0$.

б$_2$) Пусть $x \in (1;\infty)$. Тогда

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(u)du = \int_{-\infty}^{1} f_\xi(u)du + \int_1^x f_\xi(u)du = \int_1^x f_\xi(u)du =$$

$$= \int_1^x \frac{3}{u^4} du = 3\int_1^x u^{-4}du = 3 \cdot \left.\frac{u^{-3}}{-3}\right|_1^x = -\left(x^{-3} - 1\right) = 1 - 1/x^3$$

Следовательно,

$$F_\xi(x) = \begin{cases} 0, & -\infty < x \le 1 \\ 1 - 1/x^3, & 1 < x < +\infty \end{cases} \tag{9.21}$$

в) Вычислим математическое ожидание и дисперсию.

$$M\xi=\int_{-\infty}^{\infty} x f_\xi(x)dx=\int_1^{\infty} x\cdot\frac{3}{x^4}dx=$$

$$=3\int_1^{\infty}\frac{dx}{x^3}=3\int_1^{\infty}x^{-3}dx=-\frac{3}{2}x^{-2}\Big|_1^{\infty}=3/2.$$

$$M\xi^2=3; D\xi=3-\left(\frac{3}{2}\right)^2=\frac{3}{4};$$

г) $P(\xi\in[-2;2])=P(-2\le\xi\le 2)=F_\xi(2)-F_\xi(-2)=\frac{7}{8}$

Ответ. a=3. Функция распределения задана формулой (9.21). $M\xi=\frac{3}{2}$; $D\xi=\frac{3}{4}$; $P(\xi\in[-2;2])=\frac{7}{8}$.

Задача 9.4. Случайная величина ξ имеет функцию распределения

$$F_\xi(x)=\begin{cases}0, & -\infty<x\le 1\\ \frac{1}{4}(x-1), & 1<x\le 5\\ 1, & x>5\end{cases}$$

Найти:

а) Плотность распределения $f_\xi(x)$.

б) $M\xi$ и $D\xi$.

в) $P(\xi\in(2;3))$.

Нарисовать графики функций $f_\xi(x)$. и $F_\xi(x)$.

Решение. а) По формуле (9.5) находим плотность:

$$f_\xi(x)=\begin{cases}\frac{1}{4}, x\in(1;5]\\ 0, x\notin(1;5]\end{cases} \tag{9.22}$$

б) Вычислим моменты $M\xi$ и $D\xi$

$$M\xi = \int_1^5 x f_\xi(x)dx = \int_1^5 x \cdot \frac{1}{4}dx = \frac{1}{4}\int_1^5 xdx =$$

$$= \frac{1}{4} \cdot \frac{x^2}{2}\Big|_1^5 = \frac{1}{8}(25-1) = \frac{24}{8} = 3$$

$D\xi = M\xi - (M\xi)^2$ Вычислим сначала $M\xi^2$

$$M\xi^2 = \int_1^5 x^2 f_\xi(x)dx = \frac{1}{4}\int_1^5 x^2 dx =$$

$$= \frac{1}{4} \cdot \frac{x^3}{3}\Big|_1^5 = \frac{1}{4}\left(\frac{125}{3} - \frac{1}{3}\right) = \frac{1}{4} \cdot \frac{124}{3} = \frac{31}{3}.$$

$$D\xi = \frac{31}{3} - 3^2 = \frac{31}{3} - 9 = \frac{31-27}{4} = \frac{4}{3}.$$

в) $P(\xi \in (2;3)) = P(2 < \xi < 3) = F_\xi(3) - F_\xi(2) =$

$$= \left[\frac{1}{4}(3-1)\right] - \left[\frac{1}{4}(2-1)\right] = \frac{1}{4}(2-1) = 1/4.$$

Графики функций $f_\xi(x)$.и $F_\xi(x)$ представлен на рисунке 9.2:

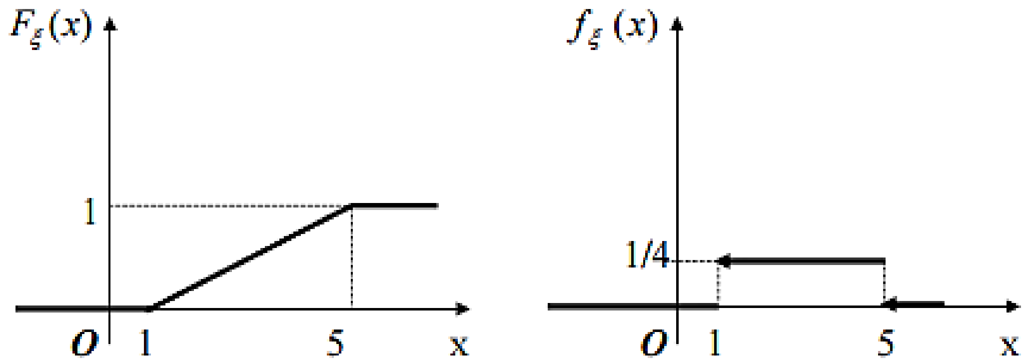


Рисунок 9.2 – Графики функций $f_\xi(x)$.и $F_\xi(x)$

Ответ. Плотность распределения задана формулой (9.22) $M\xi = 3$; $D\xi = \frac{4}{3}$; $P(\xi \in (2;3)) = \frac{1}{4}$ Графики представлены выше.

9.3 Задачи для упражнений

9.1 Случайная величина ξ имеет функцию распределения

$$F_\xi(x)=\begin{cases}0, & -\infty<x\le 0\\ \frac{1}{5}x, & 0<x\le 5\\ 1, & x>5\end{cases}$$

Найти:

а) Плотность распределения $f_\xi(x)$

б) Математическое ожидание, дисперсию и $M\xi^2$.

в) Вычислить вероятности: $P_1=P(\xi\in(0;3))$; $P_2=P(\xi\in(0;5))$; $P_3=P(\xi\in(0;6))$; $P_4=P(\xi\in(1;10))$. Построить графики функций $F_\xi(x)$ и $f_\xi(x)$.

9.2.Случайная величина ξ имеет функцию распределения

$$F_\xi(x)=\begin{cases}0, & -\infty<x\le 0\\ x^3+ax, & 0<x\le 1\\ 1, & x>1\end{cases}$$

Найти:

а) Значение параметра a .

б) Плотность распределения $f_\xi(x)$

в) Математическое ожидание и дисперсию.

г) $P(\xi\in(0,5;2))$

Построить графики функций $F_\xi(x)$. и $f_\xi(x)$

9.3.Задана плотность распределения случайной величины ξ

$$f_\xi(x)=\begin{cases}0, & x\le 0\\ cx^2e^{-\frac{x^2}{2a^2}} & x>0\end{cases}$$

где c некоторая константа, $a > 0$ - параметр.
Найти:

а) Значение константы c.

б) Математическое ожидание $M\xi$ и дисперсию $D\xi$.

9.4 Задана функция распределения непрерывной случайной величины ξ

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ \sin x, & 0 < x \le \pi/2 \\ 1, & x > \pi/2 \end{cases}$$

Найти:

а) Плотность распределения $f_\xi(x)$

б) Математическое ожидание и дисперсию.

в) $P(\xi \in [\pi/4; 2\pi/3])$

9.5 Задана плотность распределения случайной величины ξ.

$$f_\xi(x) = \begin{cases} \sin x, & x \in [0; \pi/2] \\ 0, & x \notin [0; \pi/2] \end{cases}$$

Найти:

а) Функцию распределения $F_\xi(x)$.

б) Математическое ожидание, дисперсию и среднее квадратичное отклонение.

в) $P(\xi \in [\pi/4; 2\pi/3])$

9.6 Случайная величина ξ задана плотностью распределения

$$f_\xi(x) = \begin{cases} x - \dfrac{1}{2}, & x \in (1; 2] \\ 0, & x \notin (1; 2] \end{cases}$$

Найти:

а) Функцию распределения величины ξ.

б) Математическое ожидание, дисперсию и среднее квадратичное отклонение.

в) $P(\xi \in (0;2))$.

Построить графики функций $F_\xi(x)$. и $f_\xi(x)$

9.7 Случайная величина ξ имеет плотность распределения:

$$f_\xi(x)\frac{2c}{1+x^2}$$

Найти:

а) Значение параметра c.

б) Функцию распределения $F_\xi(x)$.

в) Математическое ожидание.

г) $P(\xi \in [0;1])$

Построить графики функций $F_\xi(x)$. и $f_\xi(x)$.

9.8 Случайная величина ξ имеет плотность распределения

$$f_\xi(x) = \begin{cases} c\,\mathrm{arctg}\,x, & x \in [0;1] \\ 0, & x \notin [0;1] \end{cases}$$

Найти:

а) Значение параметра c.

б) Функцию распределения $F_\xi(x)$.

в) Математическое ожидание и дисперсию.

г) $P(\xi \in [0{,}5;2{,}5])$

Построить графики функций $F_\xi(x)$. и $f_\xi(x)$.

9.9 Задана плотность распределения случайной величины ξ

$$f_\xi(x) = \begin{cases} \dfrac{1}{\pi\sqrt{c^2 - x^2}}, & x \in (-c;c) \\ 0, & x \notin (-c;c) \end{cases}$$

Найти:

а) Математическое ожидание и дисперсию величины ξ

б) $P(\xi \in [0; c/2])$

9.10 Случайная величина ξ имеет плотность распределения

$$f_{\xi}(x) = \frac{1}{2} \cdot e^{-|x|}$$

Найти:

а) Функцию распределения $F_{\xi}(x)$.

б) Математическое ожидание и дисперсию.

в) $P(\xi \in [-1/2; 1/2])$

9.11 Задана функция распределения случайной величины ξ

$$F_{\xi}(x) = \begin{cases} 0, & x \le -1 \\ a + b\arcsin x, & -1 < x \le 1 \\ 1, & x > 1 \end{cases}$$

Найти:

а) Значения параметров a и b.

б) Плотность распределения $f_{\xi}(x)$

в) Математическое ожидание $M\xi$ и дисперсию $D\xi$

9.12 Задана плотность распределения случайной величины ξ

$$f_{\xi}(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{x^m}{m!} e^{-x}, & x > 0 \end{cases}$$

Найти:

а) Математическое ожидание $M\xi$ и дисперсию $D\xi$.

б) Среднее квадратичное отклонение $\sigma\xi$

9.13 Случайная величина ξ имеет функцию распределения

$$F_\xi(x)=\begin{cases} 0, & x \le 0 \\ x^2/16, & 0 < x \le 2 \\ x-7/4, & 2 < x \le 11/4 \\ 1, & x > 11/4 \end{cases}$$

Найти:

а) Плотность распределения $f_\xi(x)$

б) Математическое ожидание, дисперсию и среднее квадратичное отклонение величины ξ.

в) $P(\xi \in [1;3/2])$.

Построить графики функций $F_\xi(x)$ и $f_\xi(x)$.

9.14 Функция распределения случайной величины ξ имеет вид

$$F_\xi(x) = a + b \operatorname{arctg} x.$$

Найти:

а) Параметры a и b.

б) Плотность распределения $f_\xi(x)$.

Построить графики функций $F_\xi(x)$. и $f_\xi(x)$.

9.15 Случайная величина ξ имеет плотность распределения

$$f_\xi(x)=\begin{cases} 0, & x \le 1 \\ a/x^2, & x > 1 \end{cases}$$

Найти:

а) Параметр a.

б) Функцию распределения $F_\xi(x)$..

в) $P(\xi \in [3;4])$.

9.16 Случайная величина ξ имеет плотность распределения

$$f_{\xi}(x)=\begin{cases} -(3/4)x^2+(9/2)x-6, & x\in[2;4] \\ 0, & x\notin[2,4] \end{cases}$$

Найти:

а) Функцию распределения $F_{\xi}(x).$.

б) Математическое ожидание и дисперсию случайной величины ξ.

9.17 Задана плотность распределения некоторой случайной величины ξ

$$f_{\xi}(x)=\begin{cases} -(3/4)x^2+6x-45/4, & x\in[3;5] \\ 0, & x\notin[3;5] \end{cases}$$

Найти:

а) Функцию распределения $F_{\xi}(x).$.

б) Математическое ожидание и дисперсию ξ

9.18 Случайная величина ξ задана функцией распределения

$$F_{\xi}(x).=\begin{cases} 0, & x\le 2 \\ (x-2)^2, & 2<x\le 3 \\ 1, & x>3 \end{cases}$$

Найти:

а) Плотность распределения $f_{\xi}(x)$.

б) $M\xi$, $D\xi$ и $\sigma\xi$.

в) $P_1=P(\xi\in[1;2{,}5])$; $P_2=P(\xi\in[2{,}5;3{,}5])$.

Построить графики функций $F_{\xi}(x).$ и $f_{\xi}(x)$.

9.19 Дана функция

$$f_\xi(x)=\begin{cases} a\cdot\left(4x-x^3\right), & x\in(0;2] \\ 0, & x\notin(0;2] \end{cases}$$

При каком значении a $f_\xi(x)$ может быть плотностью распределения случайной величины ξ? Определить это значение a. Найти математическое ожидание и дисперсию случайной величины a.

9.20 Случайная величина ξ имеет плотность распределения

$$f_\xi(x)=\begin{cases} a\cdot x(3-x), & x\in[0;3] \\ 0, & x\notin[0;3] \end{cases}$$

Найти:

а) Параметр a.

б) Функцию распределения $F_\xi(x)$.

в) $P(\xi\in[1;2])$.

г) Математическое ожидание $M\xi$ и дисперсию $D\xi$.

Построить графики функций $F_\xi(x)$. и $f_\xi(x)$

9.21 Функция $f_\xi(x)=\dfrac{2a}{e^x+e^{-x}}$ является плотностью распределения случайной величины ξ. Найти:

а) Параметр a.

б) $P(\xi\in[0;2])$

в) $P(\xi<1)$

г) $P(\xi\geq 0)$

9.22 Случайная величина ξ подчинена закону распределения с плотностью

$$f_\xi(x)=\begin{cases} a\cdot\cos x, & x\in(0;\pi/2] \\ 0, & x\notin(0;\pi/2] \end{cases}$$

Найти:

а) Параметр a.

б) Функцию распределения $F_\xi(x)$.

в) Математическое ожидание $M\xi$ дисперсию $D\xi$ и среднее квадратичное отклонение $\sigma\xi$.

г) $P(\xi \in [0;\pi/4])$.

Построить графики функций $F_\xi(x)$ и $f_\xi(x)$.

9.23 Случайная величина ξ имеет функцию распределения

$$F_\xi(x) = \begin{cases} 0, & x \le -\dfrac{\pi}{4} \\ \dfrac{1}{2}(1+\sin 2x), & x \in \left(-\dfrac{\pi}{4}, \dfrac{\pi}{4}\right] \\ 1, & x > \dfrac{\pi}{4} \end{cases}$$

Найти:

а) Плотность распределения $f_\xi(x)$.

б) Математическое ожидание $M\xi$ и дисперсию $D\xi$.

в) $P(\xi \in [\pi/6;\pi/4])$

Практическое занятие №10
Классические распределения

10.1 Краткие теоретические сведения

Пусть испытание, в результате которого появляется событие A или $\overline{A}$, повторяется n раз; $P(A)=p, p(\overline{A})=1-p=q$. Число появлений события A в результате n повторений испытания является случайной величиной ξ со значениями $0,1,2,\ldots,m,\ldots n$ Известно, что

$$P(\xi=m)=P_n(m)=C_n^m p^m q^{n-m} \tag{10.1}$$

Закон распределения этой величины:

$$\xi:\begin{pmatrix} 0 & 1 & \ldots & m & \ldots & n \\ q^n & npq^{n-1} & \ldots & C_n^m p^m q^{n-m} & \ldots & p^n \end{pmatrix} \tag{10.2}$$

Определение. Случайная величина ξ, которая имеет закон распределения (10.2), называется биномиальной.

Распределение (10.2) называется биномиальным, так как вероятность (10.1) представляет собой общий член в разложении бинома $(q+p)^n$:

$$1=(q+p)^n=q^n+npq^{n-1}+\ldots+C_n^m p^m q^{n-m}+\ldots+nqp^{n-1}+p^n. \tag{10.3}$$

Если n велико, а $p=P(A)$ достаточно мала $(p\le 0{,}1)$, то вычисление $P_n(m)$ производят по следующей формуле:

$$P_n(m)\approx\frac{\lambda^m}{m!}e^{-\lambda} \tag{10.4}$$

где $np=\lambda$, $\lambda>0$ - параметр распределения.

Распределение вероятностей (10.4) обычно называется распределением Пуассона.

Случайная величина ξ имеет распределение Пуассона, если ее закон распределения характеризуется матрицей:

$$\xi:\begin{pmatrix} 0 & 1 & \ldots & m & \ldots \\ e^{-\lambda} & \lambda e^{-\lambda} & \ldots & \dfrac{\lambda^m}{m!}e^{-\lambda} & \ldots \end{pmatrix} \tag{10.5}$$

Легко проверить, что $\sum\limits_{m=0}^{\infty}\dfrac{\lambda^m}{m!}e^{-\lambda}=1$.

Пусть серия независимых испытаний прекращается с появлением события A.

Обозначим через ξ - число произведенных опытов до появления события A. Очевидно, ξ - случайная величина с возможными значениями $1,2,\ldots,m,\ldots,n,\ldots$

$$P(\xi=m)=q^{m-1}p$$

Распределение величины ξ:

$$\xi:\begin{pmatrix} 1 & 2 & \ldots & m & \ldots \\ p & qp & \ldots & q^{m-1}p & \ldots \end{pmatrix} \tag{10.6}$$

Вероятности из распределения (10.6) представляют собой бесконечно убывающую геометрическую прогрессию с первым членом p и знаменателем $q<1$. Следовательно,

$$\sum_{m=1}^{\infty}q^{m-1}p=\frac{p}{1-q}=1.$$

Распределение, которое характеризуется матрицей (10.6), называется геометрическим распределением.

Определение. Говорят, что случайная величина ξ имеет равномерное распределение на $[a, b]$, если:

$$f_\xi(x)\begin{cases}\dfrac{1}{b-a}, & x \in [a, b]\\ 0, & x \notin [a, b]\end{cases} \qquad (10.7)$$

$$M\xi = \frac{a+b}{2}; D\xi = \frac{(b-a)^2}{2} \qquad (10.8)$$

Говорят, что случайная величина ξ имеет нормальное распределение, если

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (10.9)$$

В формуле (10.9) a и $\sigma > 0$ являются параметрами распределения. $a = M\xi$ и $\sigma^2 = D\xi$

$$P(x_1 < \xi < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right) \qquad (10.10)$$

где $\Phi(x)$ является функцией Лапласа.

Для любого $\delta > 0$ имеем:

$$P(|\xi - a| < \delta) = 2\Phi(\delta / \sigma) = 2\Phi\left(\delta / \sqrt{D\xi}\right) \qquad (10.11)$$

Формула (10.11) выражает отклонение случайной величины ξ от ее математического ожидания на значение δ.

Говорят, что случайная величина ξ имеет показательное (экспоненциальное) распределение, если ее плотность распределения имеет вид:

$$f_{\xi}(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x}, & x\ge 0\end{cases} \qquad (10.12)$$

где $\lambda>0, \lambda$ - параметр распределения.

Математическое ожидание и дисперсия случайной величины ξ с плотностью (10.12) выражаются через параметр λ по формулам (10.13):

$$M\xi=\frac{1}{\lambda}; D\xi=\frac{1}{\lambda^2} \qquad (10.13)$$

10.2 Решение типовых задач

Задача 10.1. В партии из 20 деталей 5 имеют дефект. Из нее берут наудачу 5 деталей. Число деталей с дефектом среди выбранных, является случайной величиной ξ.

Найти:

а) Закон распределения величины ξ.

б) Функцию распределения $F_{\xi}(x)$.

в) Математическое ожидание $M\xi$ и дисперсию $D\xi$.

Решение. а) Очевидно, возможными значениями для величины ξ будут 0, 1, 2, 3, 4, 5. Соответствующие этим значениям вероятности находим по формуле Бернулли (10.1).

Событие А - « выбрали деталь с дефектом ».

$p=P(A)=5/20=0{,}25;$

$q=P(\overline{A})=1-p=0{,}75; n=5.$

$P(\xi=0)=C_5^0(0{,}25)^0\cdot(0{,}75)^5=1\cdot(0{,}75)^5=0{,}237.$

$P(\xi=1)=C_5^1(0{,}25)^1\cdot(0{,}75)^4=5\cdot0{,}25\cdot0{,}316=0{,}395.$

$P(\xi=2)=C_5^2(0{,}25)^2\cdot(0{,}75)^3=10\cdot0{,}0625\cdot0{,}422=0{,}264.$

$P(\xi=3)=C_5^3(0{,}25)^3\cdot(0{,}75)^2=0{,}1562\cdot0{,}562=0{,}088.$

$P(\xi=4)=C_5^4(0{,}25)^4\cdot0{,}75=5\cdot0{,}004\cdot0{,}75=0{,}015.$

$P(\xi=5)=C_5^5(0{,}25)^5=1\cdot0{,}001=0{,}001.$

Следовательно, случайная величина ξ имеет биномиальное распределение:

$$\xi : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0{,}237 & 0{,}395 & 0{,}264 & 0{,}088 & 0{,}015 & 0{,}001 \end{pmatrix} \tag{10.14}$$

б) Функцию распределения $F_\xi(x)$ найдем по формуле (10.4) используя закон распределения (10.14) величины ξ. Имеем

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ 0{,}237, & 0 < x \le 1 \\ 0{,}632, & 1 < x \le 2 \\ 0{,}896, & 2 < x \le 3 \\ 0{,}984, & 3 < x \le 4 \\ 0{,}999, & 4 < x \le 5 \\ 1, & x > 5 \end{cases} \tag{10.15}$$

в) Математическое ожидание биномиальной случайной величины можно вычислить по формуле:

$$M\xi = np \tag{10.16}$$

В нашем случае n=5; p=0,25, поэтому

$$M\xi = 5 \cdot 0{,}25 = 1{,}25$$

Дисперсия биноминального распределения

$$D\xi = npq \tag{10.17}$$

Следовательно,

$$D\xi = 5 \cdot 0{,}25 \cdot 0{,}75 \approx 0{,}94$$

Замечание. Такой же результат получится, если использовать формулы (10.5) и (10.6) для математического ожидания и дисперсии дискретной случайной величины.

Ответ. Закон распределения величины ξ имеет вид матрицы (10.14). Функция распределения является ступенчатой функцией и имеет вид (10.15). $M\xi = 1{,}25$, $D\xi \approx 0{,}94$.

Задача 10.2. В некотором радиоприемнике 1000 электроэлементов. Вероятность выхода из строя элемента в течение года равна 0,001. Найти:

а) закон распределения случайной величины ξ - числа элементов выходящих из строя в течение года, математическое ожидание $M\xi$ и дисперсию $D\xi$;

б) вероятность того что из строя выйдут два элемента.

в) вероятность выхода из строя по крайней мере двух элементов.

Решение. Очевидно, что мы имеем дело с распределением Пуассона. n=1000; $p = P(A) = 0{,}001$ р, где A - событие «элемент выходит из строя ». ξ - число элементов, которые выходят из строя в течение года.

а) возможными значениями числа элементов ξ, вышедших из строя в течение года будут 0, 1, 2, ... , 1000. Соответствующие этим значениям вероятности находим по формуле Пуассона (10.4),

$$\lambda = np = 1000 \cdot 0{,}001 = 1;\ m = 0;1;\ldots;1000,$$

$$P(\xi = m) = P_{1000}(m) = \frac{\lambda^m}{m!}e^{-\lambda} = \frac{1}{m!}e^{-1}$$

Итак, случайная величина ξ имеет распределение Пуассона (10.5)

$$\xi : \begin{pmatrix} 0 & 1 & 2 & \ldots & m & \ldots & 1000 \\ e^{-1} & e^{-1} & \frac{1}{2}e^{-1} & \ldots & \frac{1}{m!}e^{-1} & \ldots & \frac{1}{1000}e^{-1} \end{pmatrix} \qquad (10.18)$$

По формулам (10.16) и (10.17) находим:

$$M\xi = np = 1000 \cdot 0,001 = 1;$$

$$D\xi = npq = 1000 \cdot 0,001 \cdot \frac{999}{1000} = 0,99.$$

б) $P(\xi = 2) = P_{1000}(2) = \frac{\lambda^2}{2!} e^{-\lambda}$, где $\lambda = np = 1000 \cdot 0,001 = 1$, т.е. $P(\xi = 2) = (1/2) \cdot e^{-1} = \frac{1}{2e} \approx 0,184$.

в) $P(\xi \ge 2) = P(\xi = 2) + \ldots + P(\xi = 1000)$.

$$P(\xi \ge 2) = 1 - P(\xi < 2) = 1 - [P(\xi = 0) + P(\xi = 1)] = 1/2e \approx 0,264.$$

Ответ. а) Закон распределения величины ξ имеет вид (10. 18), $M\xi = 1$, $D\xi = 0,99$; б) $P(\xi = 2) \approx 0,184$; в) $P(\xi \ge 2) \approx 0,264$.

Задача 10.3. Стрелок стреляет по цели до первого попадания. Найти: а) закон распределения величины ξ - числа выстрелов и б) вероятность того, что он попадет четвертый раз, если вероятность попадания при одном выстреле равна 0,7.

Решение. А - « стрелок попадает в цель при одном выстреле ».

$$P(A) = p = 0,7,\ P(\overline{A}) = q = 1 - p = 0,3.$$

а) Очевидно, что число выстрелов ξ будет равняться $1, 2, \ldots, m, \ldots,$ и $P(\xi = m) = q^{m-1}p = (0,3)^{m-1} \cdot 0,7$. Распределение величины ξ является геометрическим и задается матрицей (10.6), т.е.

$$\xi : \begin{pmatrix} 1 & 2 & \ldots & m & \ldots \\ 0,7 & 0,3 \cdot 0,7 & \ldots & (0,3)^{m-1} \cdot 0,7 & \ldots \end{pmatrix} \qquad (10.19)$$

б) Пусть В - « стрелок попадет в цель четвертый раз ». Тогда:

$$B = \overline{A} \cap \overline{A} \cap \overline{A} \cap A.$$

$$P(B) = P(\overline{A}) \cdot P(\overline{A}) \cdot P(\overline{A}) \cdot P(A) = q^3 \cdot p =$$

$$= (0{,}3)^3 \cdot 0{,}7 = 0{,}027 \cdot 0{,}7 = 0{,}0189 \approx 0{,}019$$

Ответ. а) закон распределения величины имеет вид (10.19);

б) вероятность того, что стрелок попадет в цель четвертый раз, равна 0,019.

Задача 10.4. Найти математическое ожидание, дисперсию и среднее квадратичное отклонение случайной величины, равномерно распределенной на интервале [2; 8] .

Решение. Известно, что равномерное распределение на интервале [2; 8] согласно (10.7) будет иметь плотность распределения

$$f_\xi(x) = \begin{cases} 1/6, & x \in [2;8] \\ 0, & x \notin [2;8] \end{cases}$$

Тогда, по формулам (10.8), имеем

$$M\xi = \frac{a+b}{2} = \frac{2+8}{2} = \frac{10}{2} = 5,$$

$$D\xi = \frac{(b-a)^2}{12} = \frac{(8-2)^2}{12} = \frac{36}{12} = 3,$$

$$\sigma\xi = \sqrt{D\xi} = \sqrt{3} \approx 1{,}73$$

Ответ. $M\xi = 5$; $D\xi = 3$; $\sigma\xi \approx 1{,}73$.

Задача 10.5. Случайная величина ξ имеет нормальное распределение с параметрами $a = 2, \sigma = 1$. Найти $P(0 < \xi < 3)$.

Решение. Применим формулу (10.10) для $x_1 = 0, x_2 = 3, a = 2, \sigma = 1$. Тогда получим:

$$P\,(0<\xi<3)=\Phi\left(\frac{3-2}{1}\right)-\Phi\left(\frac{0-2}{1}\right)=\Phi(1)-\Phi(-2)=$$

$$=\Phi(1)+\Phi(2)=0{,}3413+0{,}4772=0{,}8185$$

Ответ. $P\,(0<\xi<3)\approx 0{,}82$

10.3 Задачи для упражнений

10.1 Механизм состоит из трех элементов, которые функционируют независимо. Вероятность того, что элемент выйдет из строя, равна 0,1. Найти распределение случайной величины ξ - числа элементов, выходящих из строя.

10.2 В некоторой партии содержатся 10% бракованных деталей. Наудачу берут 4 детали. Найти закон распределения биномиальной случайной величины ξ - числа бракованных деталей, находящихся среди выбранных.

10.3 Охотник, который имеет 4 патрона, стреляет до поражения цели (или до полного расхода патронов). Найти математическое ожидание и дисперсию числа использованных патронов, если вероятность поражения цели при одном выстреле равна 0,25.

10.4 В цель стреляют до получения двух попаданий. Найти математическое ожидание числа выстрелов, если вероятность попадания при одном выстреле равна 0,2.

10.5 В течение часа на коммутатор поступает в среднем 60 вызовов. Найти вероятность того, что в течение 30 секунд, сколько отсутствовала телефонистка, не будет ни одного вызова.

10.6 Текст, расположенный на 100 страницах, содержит 500 ошибок. Найти вероятность того, что одна страница содержит не менее трех ошибок.

10.7 Случайная величина ξ имеет нормальное распределение со средним значением $M\xi=40$ и дисперсией $D\xi=200$. Найти $P\,(30<\xi<80)$.

10.8 Рост взрослого человека является случайной величиной, имеющей нормальное распределение. Пусть средний рост равен 175 см, а среднее квадратичное отклонение - 6 см. Найти вероятность того, что хотя бы один из пяти человек, взятых наудачу, будет иметь рост между 170 см и 180 см.

10.9 Случайная величина ξ распределена нормально с плотностью $f_\xi(x)=\frac{1}{5\sqrt{2\pi}}e^{-\frac{(x-1)^2}{50}}$. Найти математическое ожидание, дисперсию и P (3<ξ<8).

10.10 Математическое ожидание и среднее квадратичное отклонение некоторой нормально распределенной случайной величины равны 10 и 2 соответственно. Найти вероятность того, что в результате опыта эта величина примет значения из интервала (12; 14).

10.11 Случайная величина ξ распределена нормально с математическим ожиданием Mξ=0 и Dξ=1. Найти вероятности: P (-0,5<ξ<-0,1) и P(1<ξ<2).

10.12 Ошибка, допущенная при измерении, является случайной величиной, которая распределена нормально. Найти $P(4\xi\in(-9;9))$, если a=0 и σ=3.

10.13 Найти число игральных костей, которые необходимо бросить для того, чтобы математическое ожидание числа костей, на которых выпало два очка, равнялось шести.

10.14 На стол высыпали 25 монет. Найти вероятность того, что число монет, упавших гербами вверх, заключено между 8 и 15, включая эти два крайних значения.

10.15 Непрерывная случайная величина ξ имеет показательное (экспоненциальное) распределение, а

$$F_\xi(x)=\begin{cases}1-e^{-0,6x}, & x\geq 0\\ 0, & x<0\end{cases}$$

Найти $P(\xi\in[2;5])$, $M\xi$ и $D\xi$.

10.16 Найти математическое ожидание, дисперсию и среднее квадратичное отклонение случайной величины ξ, имеющей плотность распределения

$$f_\xi(x)=\begin{cases}10e^{-10x}, & x\geq 0\\ 0, & x<0\end{cases}$$

10.17 Результаты измерения расстояния между двумя населенными пунктами подчинены нормальному закону распределения с параметрами a=16 км, σ=100 м. Найти вероятности того, что расстояние между этими пунктами:

а) не менее 15,8 км; б) не более 16,25 км;

в) от 15,75 км до 16,3 км.

10.18 Рост взрослой женщины является случайной величиной, распределенной по нормальному закону с параметрами a=164 см, σ =5,5 см. Найти плотность распределения и функцию распределения этой величины.

10.19 Рост взрослого мужчины является случайной величиной распределенной нормально. Пусть M ξ =170 см, D ξ =36 см. Вычислить вероятность того, что хотя бы один из наудачу выбранных четырех мужчин будет иметь рост от 168 см до 172 см.

10.20 Составить закон распределения случайной величины ξ — числа попаданий в мишень при четырех выстрелах, если вероятность попадания при одном выстреле равна 0,3. Вычислить математическое ожидание и дисперсию этой случайной величины.

10.21 Производится взвешивание некоторого вещества без систематических ошибок. Случайные ошибки взвешивания подчинены нормальному закону со среднем квадратичным отклонением σ =20 г. Найти вероятность того, что взвешивание будет произведено с ошибкой, не превосходящей по абсолютной величины 10 г.

10.22 Выпущено 150 почтовых голубей. Каждый из них возвращается с вероятностью p=0,75. Найти математическое ожидание и дисперсию числа вернувшихся голубей.

10.23 Случайная величина ξ имеет плотность распределения

$$f_{\xi}(x)=\begin{cases}3e^{-3x}, & x\ge 0\\ 0, & x<0\end{cases}$$

Найти:

а) математическое ожидание Mξ, дисперсию Dξ и среднее квадратичное отклонение $\sigma\xi$ случайной величины ξ

б) $p(\xi\in[1;2])$.

Практическое занятие №11
Двумерные случайные величины

11.1 Краткие сведения из теории

Результаты некоторых испытаний могут быть описаны более чем одним числом

Пример. Место падения снаряда определяется двумя числами: абсциссой и ординатой точки на плоскости.

Так как эти координаты изменяются от одного испытания к другому, то они являются случайными (заранее неизвестно в какое место попадет снаряд).

Говорят в этом случае, что задана двумерная случайная величина. Любая двумерная величина имеет две компоненты. Обозначим их через ξ и η. Двумерную случайную величину будем обозначать через ζ. Таким образом, $\zeta = (\xi; \eta)$.

Множество возможных значений

$$\{x; y\} = \{(x; y) : \xi = x\,; \eta = y\}$$

будет интерпретироваться как множество точек плоскости: x это абсцисса точки, а y ордината.

Также, как и для обычных случайных величин (одномерных), будем различать дискретные двумерные случайные величины и непрерывные двумерные случайные величины.

Законом распределения (распределением) дискретной двумерной случайной величины называется таблица, в которую занесены возможные значения:

$$\xi = x_i, i = 1,2,\ldots,m, \eta = y_j, j = 1,2,\ldots,n,$$

а также соответствующие этим значениям вероятности:

$$Pi_j = P\,(\xi = x_i,; \eta = y_j). \qquad (11.1)$$

Таким образом, распределение дискретной двумерной случайной величины можно изобразить в виде таблицы 11.1 с двойным входом.

Таблица 11.1 – Распределение двумерной случайной величины

$\xi \backslash \eta$	y_1	y_2	$\dots$	y_j	$\dots$	y_n	$P(\xi = x_i)$
x_1	p_{11}	p_{12}	$\dots$	p_{1j}	$\dots$	p_{1n}	p_1
x_2	p_{21}	p_{22}	$\dots$	p_{2j}	$\dots$	p_{2n}	p_2
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_i	p_{i1}	p_{i2}	$\dots$	p_{ij}	$\dots$	p_{in}	p_i
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_m	p_{m1}	p_{m2}	$\dots$	p_{mj}	$\dots$	p_{mn}	p_m
$P(\eta = y_i)$	q_1	q_2	$\dots$	q_j	$\dots$	q_n	

В этой таблице $x_1 < x_2 < \dots < x_m$, а также $y_1 < y_2 < \dots < y_n$. Имеют место очевидные соотношения:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1. \tag{11.2}$$

$$\sum_{j=1}^{n} p_{ij} = P(\xi = x_i) = p_i, i = 1, \dots, m. \tag{11.3}$$

$$\sum_{i=1}^{m} p_{ij} = P(\eta = y_j) = q_j, j = 1, \dots, n. \tag{11.4}$$

Закон распределения непрерывной двумерной случайной величины обычно задается функцией распределения или плотностью распределения.

Определение. Функцией распределения двумерной случайной величины $\zeta = (\xi; \eta)$ называется функция, обозначаемая через $F_\zeta(x; y)$

и выражаемая равенством:

$$F_{\zeta}(x;y)=P(\xi<x;\eta<y). \qquad (11.5)$$

Обозначим через $F_{\xi}(x)$ и $F_{\eta}(x)$ функции распределения компонент ξ и η:

$$F_{\xi}(x)=P(\xi<x);\;\; F_{\eta}(y)=P(\eta<y) \qquad (11.6)$$

Функции распределения (11.6) называются маргинальными функциями распределения. Известно, что

$$F_{\xi}(x)=F_{\zeta}(x;+\infty) \quad F_{\eta}(y)=F_{\zeta}(+\infty;y) \qquad (11.7)$$

Пусть $D=\{(x;y): a\le x\le b; c\le y\le d\}$. Тогда:

$$P(\zeta\in D)=F_{\zeta}(b;d)-F_{\zeta}(a;d)-F_{\zeta}(b;c)+F_{\zeta}(a;c) \qquad (11.8)$$

Плотностью распределения (плотностью вероятностей) случайной величины $\zeta=(\xi;\eta)$ называется функция:

$$f_{\zeta}(x;y)=\frac{\partial^2 F_{\zeta}(x;y)}{\partial x\partial y} \qquad (11.9)$$

Для любой области $D\in R^2$ имеем

$$P(\zeta=(\xi;\eta)\in D)=\iint\limits_{D} f_{\zeta}(x;y)dxdy \qquad (11.10)$$

Формула (11.9) выражает плотность распределения через функцию распределения двумерной случайной величины $\zeta=(\xi;\eta)$. Обратно,

$$F_{\zeta}(x;y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f_{\zeta}(u;v)dudv \qquad (11.11)$$

Плотность распределения $f_{\zeta}(x;y)$ обладает свойствами:

$$f_{\zeta}(x;y)\geq 0; \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f_{\zeta}(x;y)dxdy=1 \qquad (11.12)$$

Замечание. Очевидно, функция распределения (11.5) существует для любых двумерных случайных величин, как непрерывных, так и дискретных. Плотность распределения существует только для дифференцируемых функций распределения.

Если $f_{\xi}(x)$ и $f_{\eta}(y)$ являются плотностями распределения компонент ξ и η, то:

$$f_{\xi}(x)=\int\limits_{-\infty}^{\infty} f_{\zeta}(x;y)dy \qquad (3.13)$$

и

$$f_{\eta}(y)=\int\limits_{-\infty}^{\infty} f_{\zeta}(x;y)dx \qquad (3.14)$$

Пусть $\zeta=(\xi;\eta)$ имеет дискретное распределение. Обозначим через $p(x_i/y_j)$ условную вероятность

$$P(\xi=x_i/\eta=y_j).$$

Распределение

$$\begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ p(x_1/y_j) & p(x_2/y_j) & \ldots & p(x_m/y_j) \end{pmatrix}, j=1,\ldots,n \qquad (11.15)$$

называется распределением дискретной величины ξ при условии, что $\eta = y_j$ (условным распределением ξ).

Аналогично, распределение

$$\begin{pmatrix} y_1 & y_2 & \ldots & y_n \\ p(y_1/x_i) & p(y_2/x_i) & \ldots & p(y_n/x_i) \end{pmatrix}, j=1,\ldots,m \qquad (11.16)$$

является распределением дискретной величины ηпри условии, что $\xi = x_i$ (условным распределением η).

В силу теоремы умножения вероятностей, будем иметь:

$$p(y_j/x_i) = \frac{P(\xi = x_i; \eta = y_j)}{P(\xi = x_i)} = \frac{p_{ij}}{p_i}, i=1,\ldots,m, \qquad (11.17)$$

$$p(x_i/y_j) = \frac{P(\xi = x_i; \eta = y_j)}{P(\xi = y_j)} = \frac{p_{ij}}{q_j}, j=1,\ldots,n. \qquad (11.18)$$

Аналогично, если обозначить через $\varphi(x/y)$ и $\psi(y/x)$ условные плотности распределения непрерывных случайных величин ξ и η соответственно, то

$$\varphi(x/y) = \frac{f_\zeta(x;y)}{f_\eta(y)} \qquad (11.19)$$

$$\psi(y/x) = \frac{f_\zeta(x;y)}{f_\xi(x)} \qquad (11.20)$$

Из формул (11.19) и (11.20) следует:

$$f_{\zeta}(x;y)=f_{\xi}(x)\cdot\psi(y/x)=f_{\eta}(y)\cdot\varphi(x/y). \tag{3.21}$$

Используя формулы (11.13) и (11.14), из формул (11.19) и (11.20) выводим:

$$\varphi(x;y)=\frac{f_{\zeta}(x;y)}{\int\limits_{-\infty}^{\infty}f_{\zeta}(x;y)dx}, \tag{11.22}$$

$$\psi(y;x)=\frac{f_{\zeta}(x;y)}{\int\limits_{-\infty}^{\infty}f_{\zeta}(x;y)dy} \tag{11.23}$$

Условные плотности распределения $\varphi(x;y)$ и $\psi(y;x)$ обладают основными свойствами плотности распределения (11.12).

Условным математическим ожиданием дискретной случайной величины ξ при значении $\eta=y$ называется число

$$M(\xi/\eta=y)=\sum_{i=1}^{m}x_i p(x_i/y) \tag{11.24}$$

Аналогично определяется условное математическое ожидание дискретной величины η при значении $\xi=x$:

$$M(\eta/\xi=x)=\sum_{j=1}^{n}y_j p(y_j/x) \tag{11.25}$$

Для непрерывных распределений :

$$M(\xi/\eta=y)=\int\limits_{-\infty}^{\infty}x\varphi(x/y)dx, \tag{11.26}$$

$$M(\eta/\xi = x) = \int_{-\infty}^{\infty} y\psi(y/x)dy. \qquad (11.27)$$

Если x и y изменяются, то $M(\xi/\eta = y)$ и $M(\eta/\xi = x)$ будут функциями от y и x, таким образом $M(\xi/\eta = y) = g(y)$; $M(\eta/\xi = x) = h(x)$. Функция $h(x)$ называется функцией регрессии величины n на величину ξ, а функция $g(y)$ представляет собой функцию регрессии ξ на η.

Говорят, что случайная величина ξ не зависит от случайной величины η, если

$$\varphi(x/y) = f_\xi(x). \qquad (11.28)$$

Можно показать, что из (11.28) следует равенство

$$\psi(y/x) = f_\eta(x), \qquad (11.29)$$

т. е. независимость случайных величин взаимна. Для двух независимых случайных величин имеем:

$$f_\zeta(x;y) = f_\xi(x)\cdot f_\eta(y). \qquad (11.30)$$

Независимость случайных величин может быть выражена также в терминах функций распределения. Величины ξ и η независимы тогда и только тогда, когда

$$F_\zeta(x;y) = F_\xi(x)\cdot F_\eta(y). \qquad (11.31)$$

Двумерной случайной величине $\zeta = (\xi;\eta)$ можно сопоставить вектор математических ожиданий

$$\left(M\xi; M\eta\right) \quad (11.32)$$

и корреляционную матрицу

$$K = \begin{pmatrix} K_{\xi\xi} & K_{\xi\eta} \\ K_{\eta\xi} & K_{\eta\eta} \end{pmatrix}, \quad (11.33)$$

где

$$K_{\xi\eta} = M\left[\left(\xi - M\xi\right)\left(\eta - M\eta\right)\right] \quad (11.34)$$

является моментом корреляции величин ξ и η. Очевидно, что

$$K_{\xi\xi} = D\xi, K_{\eta\eta} = D\eta, K_{\xi\eta} = K_{\eta\xi}$$

Для момента корреляции справедлива формула:

$$K_{\xi\eta} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_i - M\xi\right)\left(y_i - M\eta\right)p_{ij}. \quad (11.35)$$

когда величины дискретны и формула

$$K_{\xi\eta} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x - M\xi\right)\left(y - M\eta\right)f_{\zeta}\left(x; y\right)dxdy, \quad (11.36)$$

если случайные величины ξ и η непрерывны. При вычислении момента корреляции часто используется формула

$$K_{\xi\eta} = M\left(\xi\eta\right) - M\xi \cdot M\eta, \quad (11.37)$$

которая следует из (11.34) после некоторых преобразований и использования свойств математического ожидания.

Определение. Коэффициентом корреляции случайных величин ξ и η называется число

$$k_{\xi\eta} = \frac{K_{\xi\eta}}{\sqrt{D\xi} \cdot \sqrt{D\eta}} \qquad (11.38)$$

Для любых случайных величин ξ и η имеем

$$\left| k_{\xi\eta} \le 1 \right|. \qquad (11.39)$$

Случайные величины ξ и η называются некоррелированными, если $k_{\xi\eta} = 0$.

Любые независимые величины также некоррелированы. Обратное утверждение неверно.

Проиллюстрируем применение формул (11.1) - (11.39) при решении типовых задач.

11.2 Решение типовых задач

Задача 11.1. Двумерная случайная величина $\zeta = (\xi; \eta)$ имеет распределение представленное в таблице 11.2:

Таблица 11.2 – Распределение двумерной случайной величины $\zeta = (\xi; \eta)$

η \ ξ	2,3	2,7
26	0,05	0,09
30	0,12	0,30
41	0,08	0,11
50	0,04	0,21

а) Найти законы распределения величин ξ и η.

б) Вычислить $M\xi$ и $M\eta$.

Решение. Случайная величина ξ принимает 4 значения:

$$x_1 = 26; x_2 = 30; x_3 = 41; x_4 = 50.$$

Вычислим по формулам (11.3) соответствующие им вероятности:

$$P(\xi = 26) = 0{,}05 + 0{,}09 = 0{,}14,$$
$$P(\xi = 30) = 0{,}12 + 0{,}30 = 0{,}42,$$
$$P(\xi = 41) = 0{,}08 + 0{,}11 = 0{,}19,$$
$$P(\xi = 50) = 0{,}04 + 0{,}21 = 0{,}25.$$

Таким образом, величина ξ имеет распределение

$$\xi : \begin{pmatrix} 26 & 30 & 41 & 50 \\ 0{,}14 & 0{,}42 & 0{,}19 & 0{,}25 \end{pmatrix} \qquad (11.40)$$

$$M\xi = 26 \cdot 0{,}14 + 30 \cdot 0{,}42 + 41 \cdot 0{,}19 + 50 \cdot 0{,}25 =$$
$$= 3{,}64 + 12{,}6 + 7{,}79 + 12{,}5 = 36{,}53$$

Случайная величина η принимает два значения: $y_1=2{,}3$ и $y_2=2{,}7$. Вычислим вероятности по формулам (3.4):

$$P(\eta = 2{,}3) = 0{,}05 + 0{,}12 + 0{,}08 + 0{,}04 = 0{,}29,$$
$$P(\eta = 2{,}7) = 0{,}09 + 0{,}30 + 0{,}11 + 0{,}21 = 0{,}71.$$

Следовательно, случайная величина η имеет закон распределения:

$$\eta : \begin{pmatrix} 2{,}3 & 2{,}7 \\ 0{,}29 & 0{,}71 \end{pmatrix} \qquad (11.41)$$

$$M\eta = 2{,}3 \cdot 0{,}29 + 2{,}7 \cdot 0{,}71 = 0{,}667 + 1{,}917 \approx 2{,}58.$$

Ответ. Законы распределения величин ξ и η представлены матрицами (11.40) и (11.41); $M\xi = 36{,}53$; $M\eta \approx 2{,}58$.

Задача 11.2. Двумерная случайная величина $\zeta = (\xi; \eta)$ имеет распределение, заданное в таблице 11.3.

Таблица 11.3 – Распределение двумерной случайной величины $\zeta = (\xi; \eta)$

η \ ξ	10	14	18
3	0,25	0,15	0,32
6	0,10	0,05	0,13

Найти:

а) Закон распределенияслучайной величины ξ, обусловленный значением $\eta = 10$.

б) Закон распределенияслучайной величины η, обусловленный значением $\xi = 6$.

Решение. а) Случайнаявеличина ξ принимает два значения: $x_1 = 3$ и $x_2 = 6$. Находим условные вероятности по формулам (11.18):

$$p(\xi = 3/\eta = 10) = \frac{P(\xi = 3; \eta = 10)}{P(\eta = 10)} = \frac{0{,}25}{0{,}25 + 0{,}10} = \frac{0{,}25}{0{,}35} = \frac{5}{7},$$

$$p(\xi = 6/\eta = 10) = \frac{P(\xi = 6; \eta = 10)}{P(\eta = 10)} = \frac{0{,}10}{0{,}35} = \frac{2}{7}$$

Условный закон распределения ξ, при значении $\eta = 10$ (формулы (11.15)) будет выглядеть так:

$$\xi/\eta = 10: \begin{pmatrix} 3 & 6 \\ 5/7 & 2/7 \end{pmatrix} \qquad (11.42)$$

б) Случайная величина η принимает три значения: $y_1 = 10$; $y_2 = 14$ и $y_3 = 18$. По формулам (11.17) находим условные вероятности:

$$p(\eta=10/\xi=6)=\frac{P(\eta=10/\xi=6)}{P(\xi=6)}=\frac{0,1}{0,1+0,05+0,13}=\frac{10}{28},$$

$$p(\eta=14/\xi=6)=\frac{P(\eta=14/\xi=6)}{P(\xi=6)}=\frac{0,05}{0,1+0,05+0,13}=\frac{5}{28},$$

$$p(\eta=18/\xi=6)=\frac{P(\eta=18/\xi=6)}{P(\xi=6)}=\frac{0,13}{0,1+0,05+0,13}=\frac{13}{28}.$$

Условный закон распределения величины η, при значении ξ=6 (формулы (11.16)), будет выглядеть так:

$$\eta/\xi=6:\begin{pmatrix} 10 & 14 & 18 \\ 10/28 & 5/28 & 13/28 \end{pmatrix} \qquad (11.43)$$

Ответ. Условные законы распределения случайных величин ξ и η представлены матрицами (11.42) и (11.43).

Задача 11.3. Двумерная случайная величина $\zeta=(\xi;\eta)$ имеет плотность распределения:

$$f_\zeta(x;y)=\frac{a}{1+x^2+x^2y^2+y^2}$$

Найти:

а) Значение параметра а.

б) Функцию распределения $F_\zeta(x;y)$.

в) Вероятность того, что величина ζ попадет в прямоугольник с вершинами в точках: О (0;0), А (0; 1),В ($\sqrt{3}$;1) и С ($\sqrt{3}$;0).

г) Показать, что случайные величины ξ и η независимы. Решение. а) Для определения параметра а воспользуемся свойством (11.12) плотности (второе соотношение). Имеем:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{a}{1+x^2+x^2y^2+y^2}dxdy=1$$

$$a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dxdy}{\left(1+x^2\right)^2\left(1+y^2\right)^2}=a\int_{-\infty}^{\infty}\frac{dx}{1+x^2}\int_{-\infty}^{\infty}\frac{dy}{1+y^2}=$$

$$=a\cdot \operatorname{arctg}x\Big|_{-\infty}^{\infty}\cdot \operatorname{arctg}xy\Big|_{-\infty}^{\infty}=a\pi^2=1\Rightarrow a=\frac{1}{\pi^2}$$

Следовательно, двумерная случайная величина $\zeta=(\xi;\eta)$ имеет плотность

$$f_\zeta(x;y)=\frac{1}{\pi^2\left(1+x^2\right)\left(1+y^2\right)}$$

б) Находим функцию распределения по формуле (11.11):

$$F_\zeta(x;y)=\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{dudv}{\pi^2\left(1+u^2\right)\left(1+v^2\right)}=$$

$$=\frac{1}{\pi^2}\operatorname{arctg}u\Big|_{-\infty}^{x}\cdot \operatorname{arctg}v\Big|_{-\infty}^{y}=\left(\frac{1}{\pi}\operatorname{arctg}x+\frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg}y+\frac{1}{2}\right)\quad(11.44)$$

в)Вычислим $P(\zeta\in D)$по формулам (3.10), где D прямоугольник с заданными вершинами (рисунок 11.1)

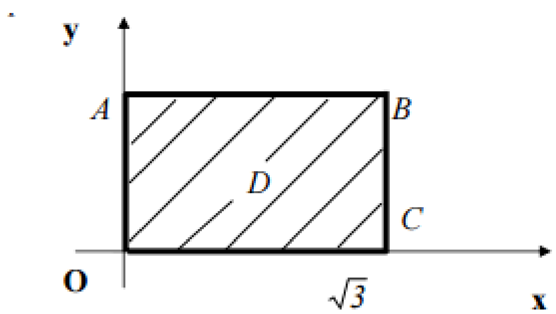


Рисунок 11.1 – Графическое изображение функции $P(\zeta\in D)$

Имеем

$$P(\zeta \in D) = P((\xi;\eta) \in D) = \iint_D f_\zeta(x;y) =$$

$$= \int_0^{\sqrt{3}}\int_0^1 \frac{dxdy}{\pi^2(1+x^2)(1+y^2)} = \frac{1}{\pi^2}\int_0^{\sqrt{3}} \frac{dx}{1+x^2}\int_0^1 \frac{dy}{1+y^2} =$$

$$= \frac{1}{\pi^2}\operatorname{arctg}x\Big|_0^{\sqrt{3}} \cdot \operatorname{arctg}y\Big|_0^1 = \frac{1}{\pi^2}\cdot\frac{\pi}{3}\cdot\frac{\pi}{4} = \frac{1}{12}$$

г) Плотность распределения $f_\zeta(x;y)$ может быть представлена так:

$$f_\zeta(x;y) = \frac{1}{\pi^2(1+x^2+x^2y^2+y^2)} =$$

$$= \frac{1}{\pi(1+x^2)}\cdot\frac{1}{\pi(1+y^2)} = f_\xi(x)\cdot f_\eta(y), \qquad (11.45)$$

где

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

откуда следует (формула (11.30)), что случайные величины ξ и η независимы.

Ответ. $a = 1/\pi^2$. Функция распределения задана формулой (11.44). $P(\zeta \in D) = 1/12$. Независимость следует из (11.45).

Задача 11.4. Двумерная случайная величина $\zeta = (\xi;\eta)$ распределена равномерно в круге радиуса r:

$$f_\zeta(x;y) = \begin{cases} \frac{1}{\pi r^2}, & x^2+y^2 \le r^2 \\ 0, & x^2+y^2 > r^2 \end{cases}$$

Найти:

а) Плотности распределения одномерных величин ξ и η.

б) Условные плотности распределения.

Решение. а) Вычислим $f_\xi(x)$ по формуле (11.13).

$$f_\xi(x)=\int_{-\infty}^{\infty} f_\zeta(x;y)dy=\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}\frac{dy}{\pi r^2}=\frac{2}{\pi r^2}\int_0^{\sqrt{r^2-x^2}}dy=\frac{2\sqrt{r^2-x^2}}{\pi r^2}$$

Очевидно, $|x|\le r$. Поэтому:

$$f_\xi(x)=\begin{cases}\dfrac{2\sqrt{r^2-x^2}}{\pi r^2}, & |x|\le r\\ 0, & |x|>r\end{cases}$$

Аналогично находим $f_\xi(y)$, используя формулу (11.14):

$$f_\xi(y)=\begin{cases}\dfrac{2\sqrt{r^2-y^2}}{\pi r^2}, & |y|\le r\\ 0, & |y|>r\end{cases}$$

б) Для определения условной плотности распределения $\varphi(x/y)$ используем формулу (11.19).

$$|x|<\sqrt{r^2-y^2}$$
$$\varphi(x/y)\,|x|>\sqrt{r^2-y^2},\ |y|<r$$

Аналогично,

$$\psi(y/x)=\begin{cases}\dfrac{1}{2\sqrt{r^2-x^2}}, & |y|<\sqrt{r^2-y^2}\\ 0, & |x|>\sqrt{r^2-y^2},\ |x|<r\end{cases}$$

Задача 11.5. Двумерная случайная величина $\zeta = (\xi; \eta)$ имеет плотность распределения:

$$f_\zeta(x;y) = \begin{cases} a\sin(x+y), & 0 \le x \le \pi/2; 0 \le y < \pi/2 \\ 0, & \text{вне} \cdot \text{квадрата} \end{cases}$$

а) Значение параметра a.

б) Математические ожидания и дисперсии случайных величин ξ и η.

в) Момент корреляции $K_{\xi\eta}$.

Решение. а) Используем второе из соотношений (11.12) для определения значения параметра a.

$$\int_0^{\pi/2}\int_0^{\pi/2} a\sin(x+y)dxdy = 1 \Leftrightarrow a = 1/2.$$

б) Определим, по формулам (11.13) и (11.14), $f_\xi(x)$ и $f_\eta(y)$:

$$f_\xi(x) = \int_{-\infty}^{\infty} f_\zeta(x;y)dy = \int_0^{\pi/2} \frac{1}{2}\sin(x+y)dy = -\frac{1}{2}\cos(x+y)\Big|_0^{\pi/2} =$$

$$= -\frac{1}{2}\left(\cos\left(x+\frac{\pi}{2}\right) - \cos x = \frac{1}{2}\left(\cos x - \cos\left(x+\frac{\pi}{2}\right) = \frac{\sqrt{2}}{2}\sin\left(x+\frac{\pi}{4}\right)\right.\right.$$

Аналогично находим $f_\eta(y) = \frac{\sqrt{2}}{2}\sin\left(y+\frac{\pi}{4}\right)$.Следовательно,

$$M\xi = \int_0^{\pi/2} xf_\xi(x)dx = \frac{\sqrt{2}}{2}\int_0^{\pi/2} x\sin\left(x+\frac{\pi}{4}\right)dx$$

Берем этот интеграл по частям:

$$\left.\begin{array}{c} x = u \\ dv = \sin\left(x + \dfrac{\pi}{4}\right) \end{array}\right| \begin{array}{c} dx = du \\ v = -\cos\left(x + \dfrac{\pi}{4}\right) \end{array}$$

Далее, для $M\xi$ получим:

$$M\xi = \frac{\sqrt{2}}{2}\left[-x\cos\left(x + \frac{\pi}{4}\right)\Big|_0^{\pi/2} + \int_0^{\pi/2}\cos\left(x + \frac{\pi}{4}\right)dx\right] =$$
$$= \frac{\sqrt{2}}{2}\left[-\frac{\pi}{2}\cos\left(x + \frac{\pi}{4}\right) + \sin\left(x + \frac{\pi}{4}\right)\Big|_0^{\pi/2}\right] = \frac{\pi}{4}$$

Аналогично, вычисляем $M\eta = \dfrac{\pi}{4}$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{\sqrt{2}}{2}\int_0^{\pi/2} x^2 \sin\left(x + \frac{\pi}{4}\right)dx - \frac{\pi^2}{16} = \frac{\pi^2}{16} + \frac{\pi}{2} - 2$$

в) Для вычисления момента корреляции $K_{\xi\eta}$ применим формулу (11.37), таким образом

$$K_{\xi\eta} = M(\xi\eta) - M\xi \cdot M\eta = \frac{1}{2}\int_0^{\pi/2}\int_0^{\pi/2} xy\sin(x + y)dxdy - \frac{\pi^2}{16} =$$
$$= \frac{1}{2}\int_0^{\pi/2}\int_0^{\pi/2} xy(\sin x\cos y + \cos x \sin y)dxdy - \frac{\pi^2}{16} =$$
$$\frac{1}{2}\int_0^{\pi/2} x\sin xdx \int_0^{\pi/2} y\cos ydy + \frac{1}{2}\int_0^{\pi/2} x\cos xdx \int_0^{\pi/2} y\sin ydy - \frac{\pi^2}{16} =$$
$$= \int_0^{\pi/2} x\sin xdx \int_0^{\pi/2} y\cos ydy - \frac{\pi^2}{16} = \frac{\pi}{2} - \frac{\pi^2}{16} - 1$$

Ответ.

а) a=1/2;

б) $M\eta = M\xi = \frac{\pi}{4}; D\eta = D\xi = \frac{\pi^2}{16} + \frac{\pi}{2} - 2;$

в) $K_{\xi\eta} = \frac{\pi}{2} - \frac{\pi^2}{16} - 1$

11.3 Задачи для упражнений

11.1 Случайная величина $\zeta = (\xi; \eta)$ распределена равномерно в треугольнике, образованном прямыми $y = x$, $y = 0$ и $x = 2$. Найти коэффициент корреляции между величинами ξ и η.

11.2 По цели стреляют два раза. Вероятность попадания в цель при одном выстреле равна p. Введем случайные величины: ξ - число попаданий; η - число промахов. Найти:

а) Закон распределения двумерной случайной величины $\zeta = (\xi; \eta)$

б) Математические ожидания $M\xi$ и $M\eta$; дисперсии $D\xi$ и $D\eta$; момент корреляции $K_{\xi\eta}$.

11.3 Случайные величины ξ и η. независимы и распределены равномерно на отрезках [-1; 1] и [0; 2] соответственно. Найти плотность распределения и функцию распределения двумерной случайной величины $\zeta = (\xi; \eta)$.

11.4 Двумерная случайная величина $\zeta = (\xi; \eta)$ имеет плотность распределения

$$f_\zeta(x; y) = \begin{cases} \cos x \cos y, 0 \le x \le \frac{\pi}{2}; 0 \le y \le \frac{\pi}{2} \\ 0, \text{вне} \cdot \text{квадрата} \end{cases}$$

Найти:

а) Функцию распределения $F_\zeta(x; y)$ величины $\zeta = (\xi; \eta)$.

б) Коэффициент корреляции $k_{\xi\eta}$.

11.5 Двумерная случайная величина задана таблицей 11.4.

Таблица 11.4 – Двумерная случайная величина

$\xi \backslash \eta$	-1	0	1
0	0,01	0,04	0,05
1	0,06	0,24	0,10
2	0,05	0,15	0,10
3	0,04	0,07	0,09

Найти:

а) Законы распределения величин ξ и η

б) Математические ожидания $M\xi$ и $M\eta$.

в) Дисперсии $D\xi$ и $D\eta$.

г) Коэффициент корреляции $k_{\xi\eta}$.

11.6 Двумерная случайная величина $\zeta = (\xi; \eta)$ имеет плотность распределения

$$f_{\zeta}(x; y) = \begin{cases} a\cos(x - y), 0 \le x \le \dfrac{\pi}{2}; 0 \le y \le \dfrac{\pi}{2} \\ 0, \text{вне} \cdot \text{квадрата} \end{cases}$$

Найти:

а) Значение параметра a.

б) Математические ожидания $M\xi$ и $M\eta$.

в) Дисперсии $D\xi$ и $D\eta$.

г) Момент $K_{\xi\eta}$ и коэффициент $k_{\xi\eta}$ корреляции.

11.7 Задана двумерная плотность распределения

$$f_{\zeta}(x; y) = \frac{a}{1 + \left(x^2 + y^2\right)^2}$$

Найти:

а) Значение параметра a.

б) Радиус круга с центром в начале координат, вероятность попасть в который, равна 0,5.

11.8 Плотность распределения двумерной случайной величины $\zeta = (\xi; \eta)$ задана функцией:

$$f_\zeta(x;y) = ae^{-(x+1)^2 - |y|}$$

Найти:

а) Значение параметра a.

б) Плотности распределения величин ξ и η.

в) Математические ожидания $M\xi$ и $M\eta$. Дисперсии $D\xi$ и $D\eta$.

г) Зависимы или нет величины ξ и η

11.9 Двумерная случайная величина $\zeta = (\xi; \eta)$ распределена равномерно в круге радиуса r=2:

$$f_\zeta(x;y) = \begin{cases} \frac{1}{4\pi}, & x^2 + y^2 \le 4 \\ 0, & x^2 + y^2 > 4 \end{cases}$$

Найти:

а) Математические ожидания $M\xi$ и $M\eta$.

б) Зависимы или нет величины ξ и η.

11.10 Монета бросается два раза. Пусть ξ - число появлений герба, а η - число появлений цифры. Найти закон распределения величины $\zeta = (\xi; \eta)$. Вычислить $M\xi$, $M\eta$, $D\xi$, $D\eta$ и $K_{\xi\eta}$.

11.11 Двумерная случайная величина $\zeta = (\xi; \eta)$ имеет плотность распределения

$$f_\zeta(x;y) = ae^{-\frac{(x+3)^2}{8} - \frac{(y-1)^2}{2}}.$$

Найти:

а) Значение параметра a.

б) $P(\xi < -3; \eta < 4)$.

11.12 Двумерная случайная величина $\zeta = (\xi; \eta)$ имеет независимые компоненты ξ и η с плотностями распределения.

$$f_\xi(x;y) = \begin{cases} 5e^{-5x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

$$f_\eta(x;y) = \begin{cases} 2e^{-2y}, & y \ge 0 \\ 0, & y < 0 \end{cases}$$

Найти:

а) Плотность распределения величины $\zeta = (\xi; \eta)$.

б) Функцию распределения величины $\zeta = (\xi; \eta)$.

в) Математические ожидания $M\xi$, $M\eta$. Дисперсии $D\xi$, $D\eta$.

г) Момент и коэффициент корреляции величин ξ и η.

11.13 Задана плотность распределения двумерной случайной величины $\zeta = (\xi; \eta)$:

$$f_\zeta(x;y) = \begin{cases} 4xye^{-x^2-y^2}, & x \ge 0, y \ge 0 \\ 0, & x < 0 \cdot \text{или} \cdot y < 0 \end{cases}$$

Найти:

а) Математические ожидания $M\xi$ и $M\eta$.

б) Дисперсии $D\xi$ и $D\eta$.

в) Момент и коэффициент корреляции между величинами ξ и η.

11.14 Двумерная дискретная случайная величина $\zeta = (\xi; \eta)$ имеет распределение представленное в таблице 11.5.

Таблица 11.5 - Двумерная дискретная случайная величина $\zeta = (\xi; \eta)$

η / ξ	-2	0	4
0	0,03	0,04	0,15
1	0,08	0,24	0,02
3	0,05	0,1	0,1
5	0,04	0,06	0,09

Найти:

а) Одномерные законы распределения величин ξ и η

б) Математические ожидания $M\xi$ и $M\eta$, дисперсии $D\xi$, $D\eta$ и коэффициент корреляции.

11.15 Случайная величина $\zeta = (\xi; \eta)$ задана таблицей 11.6.

Таблица 11.6 - Случайная величина $\zeta = (\xi; \eta)$

η / ξ	20	40	60
10	3λ	λ	0
20	2λ	4λ	2λ
30	λ	2λ	5λ

Найти:

а) Параметр λ;

б) $M\xi$ и $M\eta$;

в) $D\xi$ и $D\eta$;

г) $k_{\xi\eta}$

11.16 Дискретная двумерная случайная величина $\zeta = (\xi; \eta)$ имеет распределение представленное в таблице 11.7.

Таблица 11.7 – Дискретная двумерная случайная величина $\zeta = (\xi; \eta)$

η \ ξ	2	4	5	9
0,5	0,02	0,12	0,14	0,19
0,7	0,11	0,05	0,24	0,13

Найти:

а) Законы распределения случайных величин ξ и η.

б) Математические ожидания $M\xi$ и $M\eta$.

в) Дисперсии $D\xi$ и $D\eta$.

г) Условный закон распределения величины ξ при η=5.

д) Условный закон распределения величины η при ξ=0,7.

11.17 Задана дискретная двумерная случайная величина $\zeta = (\xi; \eta)$ (таблица 11.8).

Таблица 11.8 - Дискретная двумерная случайная величина $\zeta = (\xi; \eta)$

ξ \ η	0,4	0,8
2	0,15	0,05
5	0,30	0,12
8	0,35	0,03

Найти:

а) Законы распределения случайных величин ξ и η.

б) Условный закон распределения случайной величины ξ, если η=0,4.

в) Условный закон распределения η, если ξ= 5.

11.18 Двумерная случайная величина $\zeta = (\xi; \eta)$ задана таблицей 11.9.

Таблица 11.9 - Дискретная двумерная случайная величина $\zeta=(\xi;\eta)$

ξ \ η	1	2	3
2,5	0,12	0,15	0,2
3,5	0,15	0,21	0,19

Найти условные законы распределения случайной величины η.

11.19 Заданы законы распределения независимых случайных величин ξ и η.

$$\xi:\begin{pmatrix} -3 & 0 & 2 & 4 \\ 0{,}15 & 0{,}3 & 0{,}2 & 0{,}35 \end{pmatrix};\eta:\begin{pmatrix} 13 & 17 \\ 0{,}6 & 0{,}4 \end{pmatrix}$$

Найти:

а) Математические ожидания $M\xi$ и $M\eta$.

б) Дисперсии $D\xi$ и $D\eta$.

в) Закон распределения случайной величины $\zeta=(\xi;\eta)$

11.20 Плотность двумерной случайной величины $\zeta=(\xi;\eta)$ задана функцией

$$f_\zeta(x;y)=\begin{cases} x+y, & 0\le x\le 1; 0\le y\le 1 \\ 0, & \text{вне}\cdot\text{квадрата} \end{cases}$$

Найти коэффициент корреляции случайных величин ξ и η.

11.21 Задана двумерная дискретная случайная величина в таблице 11.10.

Таблица 11.9 - Дискретная двумерная случайная величина

η \ ξ	-4	0	2	5
10	0,05	0,15	0,17	0,13
20	0,12	0,08	0,24	0,06

Найти:

а) Дисперсии $D\xi$ и $D\eta$.

б) Коэффициент корреляции $k_{\xi\eta}$.

11.22 Случайная величина $\zeta = (\xi; \eta)$ имеет плотность распределения

$$f_\zeta(x; y) = ae^{-x^2-2xy-4y^2}$$

Найти:

а) Параметр a

б) Плотности распределения $f_\xi(x)$ и $f_\eta(y)$.

в) Условные плотности распределения $\varphi(x, y)$ и $\psi(y, x)$.

11.23 Задана плотность распределения двумерной случайной величины $\zeta = (\xi; \eta)$

$$f_\zeta(x; y) = \begin{cases} \frac{1}{4}\sin x \sin y, & 0 \le x \le \pi; 0 \le y \le \pi \\ 0, & \text{вне} \cdot \text{квадрата} \end{cases}$$

Найти:

а) Математические ожидания $M\xi$, $M\eta$, а также дисперсии $D\xi$ и $D\eta$.

б) Момент корреляции $K_{\xi\eta}$

Практическое занятие №12
Элементы математической статистики

4.1 Краткие сведения из теории

К основным задачам математической статистики относятся: указание или создание методов сбора и обработки статистических данных.

Статистической или генеральной совокупностью называется совокупность объектов, которую необходимо изучить.

Примеры. Множество людей некоторого города, множество студентов, множество электроламп, множество телевизоров, множество фирм, экономических агентов и т. д.

Число объектов генеральной совокупности называется объемом генеральной совокупности.

Обозначим это число через N .

Для изучения объектов генеральной совокупности производится выборка.

Выборочной совокупностью или просто выборкой называется совокупность (множество) случайно отобранных объектов.

Число объектов выборочной совокупности называется объемом выборки.

Обозначим это число через n. Обычно генеральную совокупность изучают относительно некоторого признака.

Признаком называется общее свойство, характерное для всех объектов совокупности.

Признаки бывают качественными и количественными.

Пример. Исследуется группа людей. Качественный признак: цвет волос, цвет глаз. Количественный признак: рост, возраст.

Пусть произвели выборку объема $n<N$. Исследуем ее объекты относительно некоторого количественного признака ξ, который принял значения $x_1, x_2, \ldots, x_m$; причем значение x_1 принял n_1 раз, значение x_2 - n_2 раза, ... , x_m принял n_m раз. Эти результаты заносятся в таблицу, состоящую из двух строк. В первую строку заносятся возможные значения признака ξ, записанные в возрастающем

порядке $x_1, x_2, \ldots, x_m$; а во вторую строку заносятся числа $n_1, n_2, \ldots, n_m$.

x_1	x_2	…	x_m
n_1	n_2	…	n_m

(12.1)

Абсолютной частотой значения x_i называется число n_i. Относительной частотой значения x_i называется отношение

$$f(x_i) = n_i/n. \qquad (12.2)$$

Иногда во вторую строку таблицы заносятся относительные частоты (12.3):

x_1	x_2	…	x_m
n_1 /n	n_2/n	…	n_m/n

(12.3)

Замечание. Чаще всего результаты выборки записываются в порядке их получения. Затем они представляются в виде (12.1), либо (12.3). Если результатов много, то они группируются по интервалам (см. задачу 12.1.).

Очевидно,

$$\sum_{i=1}^{m} n_i = n. \qquad (12.4)$$

Следовательно, с учетом (4.4),

$$f(x_1) + f(x_2) + \ldots + f(x_m) = n_1/n + n_2/n + \ldots n_m/n = 1$$

Признак или величина ξ может быть рассмотрена как случайная величина, чьими возможными значениями являются

$x_1, x_2, \ldots, x_m$, а частоты $f(x_i)$ могут быть интерпретированы как вероятности $p_i = P(\xi = x_i)$. Таким образом, (12.3) называется еще статистическим законом распределения или статистическим распределением признака (величины)

Функция распределения некоторого статистического распределения называется статистической функцией распределения или эмпирической функцией распределения.

Обозначим через $F_s(x)$ статистическую функцию распределения. Тогда

$$F_s(x) = \sum_{x_i < x} n_i / n \qquad (12.5)$$

Из (12.5) следует:

$$F_s(x) = \begin{cases} 0, & x \le x_1 \\ n_1 / n, & x_1 < x \le x_2 \\ (n_1 + n_2)/n, & x_2 < x \le x_3 \\ & \\ \sum\limits_{i=1}^{m-1} n_i / n, & x_{m-1} < x \le x_m \\ 1, & x > x_m \end{cases} \qquad (12.6)$$

График статистической функции распределения (12.6) - это график некоторой ступенчатой функции (рисунок 12.1), т.е.

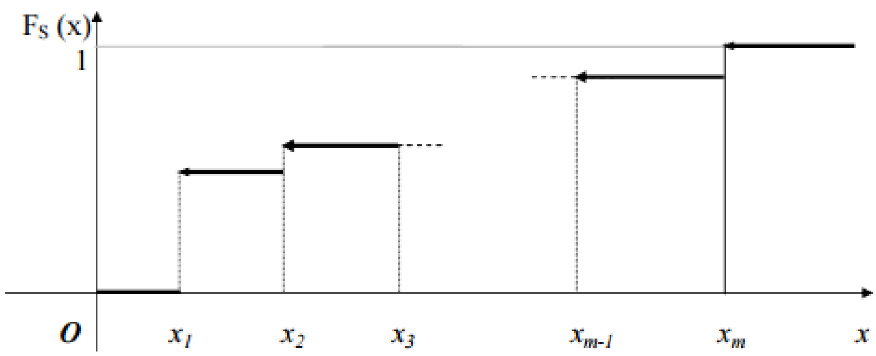


Рисунок 12.1 – График ступенчатой функции

Статистическое распределение (12.3) может быть представлено графически. В декартовой системе координат изображаем точки: $(x_1;n_1/n)$, $(x_2;n_2/n)$,..., $(x_m;n_m/n)$, которые затем соединяем между собой. Полученная линия называется полигоном частот (рисунок 12.2).

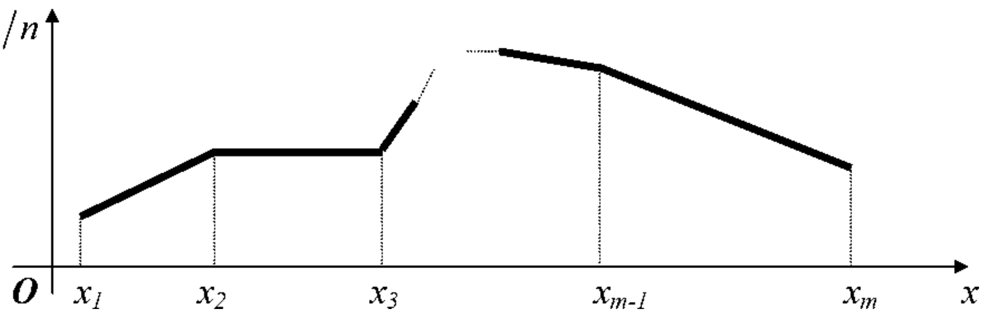


Рисунок 12.2 – Полигон частот

Зачастую накапливается много статистических данных и таблица значений получается громоздкой. Тогда эти данные группируются по интервалам и в таблицу переносят интервалы и частоты.

Пример. Для группы студентов измерили их рост. Полученные результаты занесены в таблицу 12.1.

Таблица 12.1 – Выборка «Рост студентов»

Рост	Число студентов
160 - 165	10
165 - 170	15
170 - 175	12
175 - 180	29
180 - 185	11
185 - 190	3

В этой таблице представлено интервальное распределение. Графически это распределение может быть изображено в виде гистограммы (рисунок 12.3): по оси абсцисс откладывают интервалы длины h и на них строят прямоугольники с высотами равными n_i/h.

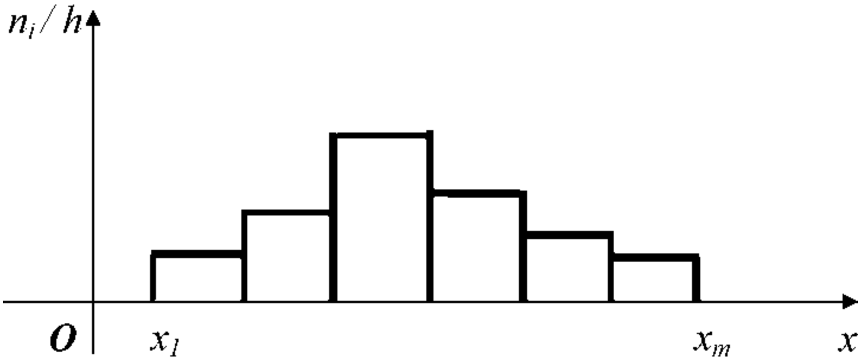


Рисунок 12.3 – Гистограмма

Выборочной средней распределения (12.3) называется число

$$M = \frac{1}{n}\sum_{i=1}^{m} x_i n_i \quad . \qquad (12.7)$$

Вычисления выборочной средней по формуле (4.7) довольно громоздки (особенно, если нет калькулятора), поэтому полезно знать следующую формулу:

$$M_s = \frac{k}{n}\sum_{i=1}^{m} \frac{x_i - c}{k} \cdot n_i + c \qquad (12.8)$$

где k и c - две специально подобранные константы, которые облегчают вычисления. Например, в формуле (4.8) в качестве c можно выбрать значение x_i, которому соответствует наибольшая частота n_i, а за k можно принять h.

Выборочной дисперсией называется число

$$D_s = \frac{1}{n}\sum_{i=1}^{m} (x_i^2 n_i - (M_s)^2 n_i \qquad (12.9)$$

После некоторых преобразований из (12.9) можно получить вычислительную формулу (12.10)

$$D_s = \frac{1}{n}\sum_{i=1}^{m}(x_i^2 n_i - (M_s)^2 \qquad (12.10)$$

Как и для выборочной средней здесь рекомендуется формула

$$D_s = \frac{k^2}{n}\sum_{i=1}^{m}\left(\frac{x_i - c}{k}\right)^2 \cdot n_i - (M_s - c)^2 \qquad (12.11)$$

которая, как правило, облегчает вычисления.

Средним квадратичным отклонением выборки называется число (12.12)

$$\sigma_s = \sqrt{D_s} \qquad (12.12)$$

Из некоторой генеральной совокупности делают выборку для того чтобы сделать определенные выводы относительно ее характеристик. Полученные результаты распространяются на все представители генеральной совокупности.

Выборочная средняя M_s, дисперсия D_s и среднее квадратичное отклонение σ_s могут быть вычислены по результатам выборки. Эти результаты позволяют делать некоторые выводы относительно всей совокупности, более точно, относительно некоторых ее характеристик.

Пусть M, D и σ - средняя, дисперсия и среднее квадратичное отклонение генеральной совокупности. Обычно эти параметры неизвестны. Их необходимо найти и оценить используя M_s, D_s и σ_s. Эти числа будут меняться от выборки к выборке, поэтому можно считать, что они случайны.

Из вышеизложенного следуют еще две задачи математической статистики:

1. Оценка параметров (получение некоторых приближенных значений параметров генеральной совокупности).
2. Проверка гипотез.

Пусть необходимо оценить параметр a закона распределения (например, нормальный закон содержит 2 параметра: a и σ. Для этого мы располагаем результатами выборки $x_1, x_2, \ldots, x_m$. Так как они изменяются от одной выборки к другой, то их можно считать случайными величинами.

Статистической оценкой $\tilde{a}$ неизвестного параметра a называется функция

$$\tilde{a} = f(x_1, x_2, \cdots, x_m) \tag{12.13}$$

Говорят, что оценка $\tilde{a}$ несмещенная или без систематической ошибки, если

$$M\tilde{a} = a. \tag{12.14}$$

Оценка $\tilde{a}$ называется эффективной, если

$$\lim_{n\to\infty} P(|\tilde{a} - a| < \varepsilon) = 1 \tag{12.15}$$

где s> 0 произвольное, сколь угодно малое число.

Например, оценкой (12.13) генеральной средней статистической совокупности является выборочная средняя M_s. Эта оценка несмещенная и эффективная, т.е. для нее выполняются соотношения (12.14) и (12.15).

Несмещенной оценкой для генеральной дисперсии является исправленная выборочная дисперсия (12.16)

$$D_s^* = \frac{1}{n-1}\sum_{i=1}^{m}(x_i - M_s)^2 \cdot n_i \tag{12.16}$$

Любая оценка вычисляется с некоторой точностью. Очевидно, что получим разные оценки для разных выборок. Таким образом, оценку $\tilde{a}$ параметра a можно рассматривать как случайную величину. Пусть Δ - точность оценки. Она достигается с некоторой вероятностью

$$P(|a - \tilde{a}|) \leq \Delta = \gamma. \qquad (12.17)$$

Соотношение (12.17) может быть записано в виде (12.18)

$$P(\tilde{a} - \Delta \leq a \leq \tilde{a} + \Delta) = \gamma \qquad (12.18)$$

Интервал $[\tilde{a} - \Delta; \tilde{a} + \Delta]$ называется доверительным интервалом, а γ - доверительной вероятностью (надежностью). Значения $\tilde{a} - \Delta$ и $\tilde{a} + \Delta$ называются доверительными границами.

Доверительный интервал для нормального распределения с параметрами a и σ задается соотношением:

$$P\left(M_s - U_\gamma \frac{\sigma}{\sqrt{n}} \leq a \leq M_s + U_\gamma \frac{\sigma}{\sqrt{n}}\right) = \gamma \qquad (12.19)$$

где

$$U_\gamma = \frac{\Delta}{\sigma}\sqrt{n},$$

$$\Phi(U_y) = \frac{\gamma}{2} \qquad (12.20)$$

С помощью (12.20), из таблиц находим аргумент U_γ. Соотношения (12.19) и (12.20) дают возможность найти доверительный интервал для a, когда параметр σ известен. Если σ не задано, то воспользуемся соотношением (12.21)

$$P\left(M_s - t_\gamma \sqrt{\frac{D_s^*}{n}} \leq a \leq M_s + t_\gamma \sqrt{\frac{D_s^*}{n}}\right) = \gamma \qquad (12.21)$$

для определения интервала $\Delta = t_\gamma \sqrt{D_s^*/n}$ параметра a.

Доверительный интервал для среднего квадратичного отклонения σ нормального распределения выражается неравенством

$$\sqrt{\frac{D_s^*}{U_2}} \le \sigma \le \sqrt{\frac{D_s^*}{U_1}} \tag{12.22}$$

где U_1 и U_2 находятся из таблиц распределения χ^2 (распределения Пирсона).

Иногда требуется установить и оценить зависимость между двумя или несколькими случайными величинами. Наиболее часто встречающаяся зависимость это линейная зависимость между выборочными средними.

Уравнение линии регрессии величины η на величину ξ имеет вид:

$$y_x - M_s\eta = k_{\xi\eta}^s \frac{\sigma_s\eta}{\sigma_s\xi}(x - M_s\xi) \,. \tag{12.23}$$

где y_x является условным средним величины η (для $\xi = x$). $M_s\xi$ и $M_s\eta$ - выборочные средние характеристик ξ и η соответственно; $\sigma_s\xi$ и $\sigma_s\eta$ - средние квадратичные отклонения; а $k_{\xi\eta}^s$ - выборочный коэффициент корреляции (12.24)

$$k_{\xi\eta}^s = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(x_i - M_s\xi)(y_j - M_s\eta)\cdot n_{ij}}{n\cdot\sigma_s\xi\cdot\sigma_s\eta} \tag{12.24}$$

Для облегчения процедуры счета, используется формула

$$k^s_{\xi\eta} = \frac{\dfrac{kl}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\dfrac{x_i - c_1}{k}\right)\left(\dfrac{y_j - c_2}{l}\right)n_{ij} - (M_s\xi - c_1)(M_s\eta - c_2)}{\sigma_s\xi \cdot \sigma_s\eta}, \quad (12.25)$$

где $M_s\xi$ и $M_s\eta$ вычисляются по формуле (12.8), а $D_s\xi$ и $D_s\eta$ - по формуле (12.11), где $c=c_1$ для ξ и $c=c_2$ для η.

Если между величинами ξ и η существует почти линейная зависимость, то линия регрессии величины ξ на величину η будет иметь уравнение:

$$x_y - M_s\xi = k^s_{\xi\eta}\frac{\sigma_s\xi}{\sigma_s\eta}(y - M_s\eta). \quad (12.26)$$

12.2 Решение типовых задач

Задача 12.1. Заданы результаты выборки:

135 133 124 132 104 152 134 130 129 120 122 124 117 123 123 129
121 122 125 131 147 124 137 112 126 128 111 129 115 147 131 132
137 119 125 120 129 125 123 127 132 118 133 132 132 134 131 120
135 132 125 132 108 114 121 133 133 135 131 125 114 115 122 131
125 132 120 126 115 117 118 118 132 134 127 127 124 135 128 127
115 144 129 120 137 127 125 116 132 120 117 127 118 109 127 122
120 135 116 118 133 136 125 126 119 126 129 127 129 124 127 132
126 131 127 130 126 134 135 127 124 123 123 130 132 143 122 129
120 134 108 132 121 111 123 140 137 120 125 131 118 120 120 136
129 127 116 138 128 133 122 131 128 140 138 134 120 126 109 137
111 115 117 130 113 126 115 124 125 118 115 128 123 129 128 120
115 134 118 135 134 123 117 120 124

Объем выборки n=185.

а) Найти относительные частоты.

б) Построить гистограмму и полигон заданного распределения.

в) Найти эмпирическую функцию распределения.

г) Вычислить: выборочную среднюю M_s, выборочную дисперсию D_s, среднее квадратичное отклонение σ_s выборки.

Решение. а) Используем статистические данные задачи для построения распределения выборки вида (12.1). Так как объем выборки n велик (185), то введем в рассмотрение интервалы длины n=5:

$$(100-105];(105-110];\ldots;(150-155].$$

После группировки данных по интервалам, приходим к следующему распределению выборки (таблица 12.2):

Таблица 12.2 – Распределение выборки

Интервал	105-110	105-110	110-115	115-120	120-125	125-130	130-135	135-140	140-145	145-150	150-155
n_i	1	4	15	31	35	40	42	12	2	2	1

Дальше построим таблицу вида (12.3), беря в качестве x_i середину каждого интервала. Приходим к таблице (12.3):

Таблица 12.3 – Результаты выборки

Xi	102,5	107,5	112,5	117,5	122,5	127,5	132,5	137,5	142,5	147,5	152,5
$\frac{n_1}{n}$	$\frac{1}{185}$	$\frac{4}{185}$	$\frac{15}{185}$	$\frac{31}{185}$	$\frac{35}{185}$	$\frac{40}{185}$	$\frac{42}{185}$	$\frac{12}{185}$	$\frac{2}{185}$	$\frac{2}{185}$	$\frac{1}{185}$

б) Для получения полигона распределения отложим по оси абсцисс значения x_i, а по оси ординат - частоты x_i.

Изобразим точки (x_i; x_i) (рисунок 12.4) и соединим их ломаной линией.

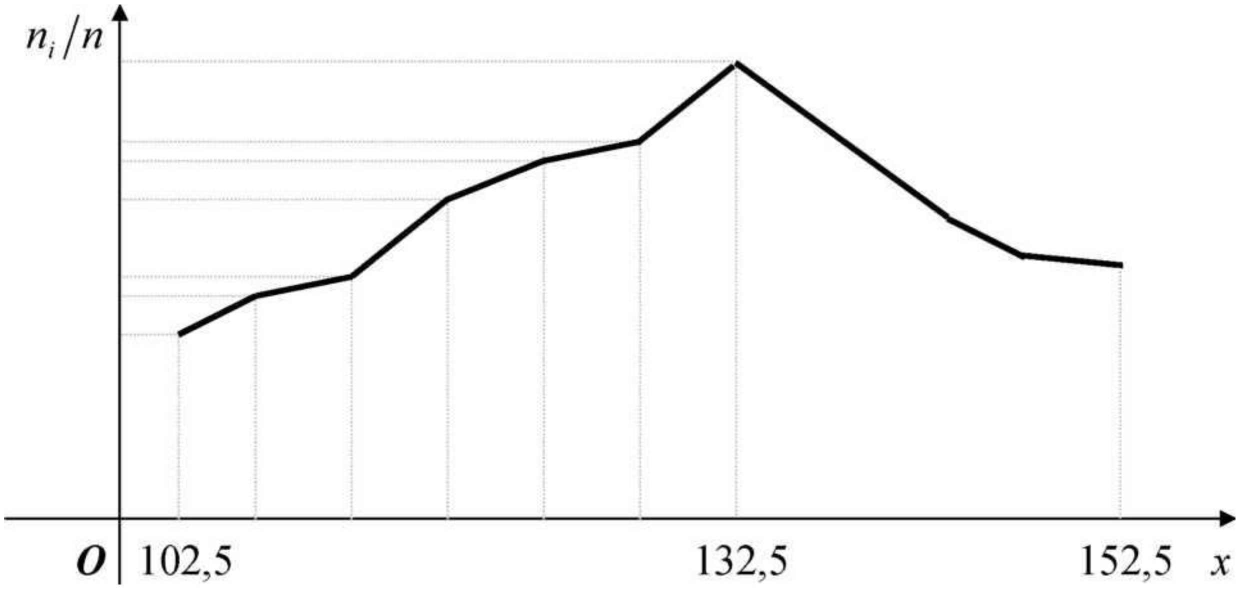


Рисунок 12.4 – График по координатам (x_i; x_i)

Построим теперь гистограмму (рисунок 12.5). По оси абсцисс отложим точки $100, 105, \ldots, 155$. На каждом интервале $[100;105]$, $[105;110], \ldots, [150;155]$ построим прямоугольники, у которых одна сторона совпадает с интервалом, а другая равна n_i/h, гдеh длина интервала. В нашем случае $h=5$.

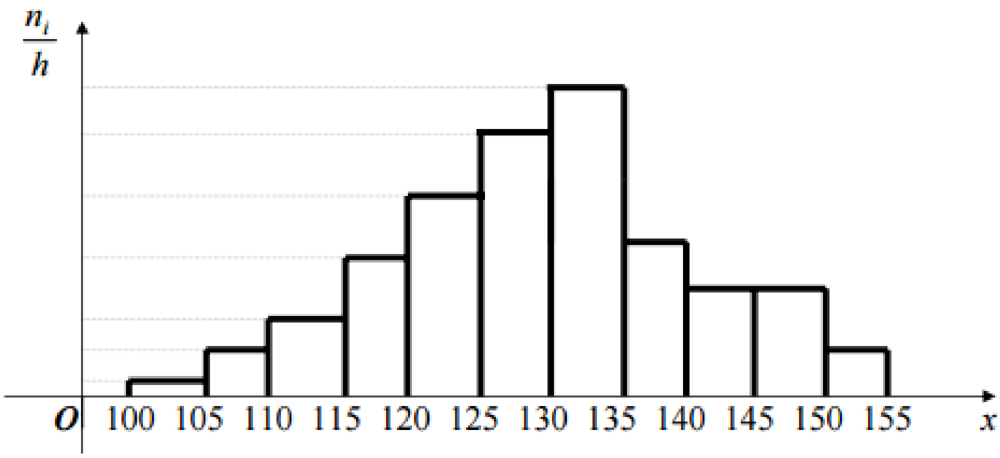


Рисунок 12.5 - Гистограмма

в) Для нахождения эмпирической функции распределения воспользуемся формулой (12.6). В качестве значений x_i выберем середины интервалов. Следовательно,

$$F_s(x)=\begin{cases}0, & x\le 102.5\\ 1/185, & 102{,}5<x\le 107{,}5\\ 5/185, & 107{,}5<x\le 112{,}5\\ 20/185, & 112{,}5<x\le 117{,}5\\ 51/185, & 117{,}5<x\le 122{,}5\\ 86/185, & 122{,}5<x\le 127{,}5\\ 126/185, & 127{,}5<x\le 132{,}5\\ 168/185, & 132{,}5<x\le 137{,}5\\ 180/185, & 137{,}5<x\le 142{,}5\\ 182/185, & 142{,}5<x\le 147{,}5\\ 184/185, & 147{,}5<x\le 152{,}5\\ 1, & x>125{,}5\end{cases}$$

г) Вычислим M_s, D_s и σ_s. Для облегчения вычислений используем формулы (12.8), (12.11) и (12.12). Оформим данные в таблице 12.4. Последний столбец является контрольным.

Таблица 12.4 – Полученные данные $c=127{,}5; k=5$

Серед. интер.	n_i	$x_i - c$	$\frac{x_i - c}{k}$	$\frac{x_i - c}{k}n_i$	$\frac{x_i - c}{k}+1$	$\left(\frac{x_i - c}{k}\right)^2$	$\left(\frac{x_i - c}{k}\right)^2 n_i$	$\left(\frac{x_i - c}{k}+1\right)^2 n$
102,5	1	-25	-5	-5	-4	25	25	16
107,5	4	-20	-4	-16	-3	16	64	36
112,5	15	-15	-3	-45	-2	9	135	60
117,5	31	-10	-2	-62	-1	4	124	31
122,5	35	-5	-1	-35	0	1	35	0
127,5	40	0	0	0	1	0	0	40
132,5	42	5	1	42	2	1	42	168
137,5	12	10	2	24	3	4	48	108
142,5	2	15	3	6	4	9	18	32
147,5	2	20	4	8	5	16	32	50
152,5	1	25	5	5	6	25	25	36
$\sum$	185			-78			548	577

Для проверки правильности счета используем равенство:

$$\sum n_i (u_i + 1)^2 = \sum n_i \cdot u_i^2 + 2\sum n_i u_i + n \qquad (12.27)$$

где $u_i = \dfrac{x_i - c}{k}$ Так как $577 = 548 - 2 \cdot 78 + 185$, то делаем вывод, что вычисления в таблице сделаны правильно. Теперь вычислим M_s и D_s, используя данные таблицы.

$$M_s = \frac{-78}{185} \cdot 5 + 127{,}5 = -2{,}1 + 127{,}5 = 125{,}4$$

$$D_s = \frac{1}{185} \cdot 548 \cdot 5^2 - (125{,}4 - 127{,}5)^2 \approx$$

$$\approx 2{,}96 \cdot 25 - (-2{,}1)^2 \approx 74 - 4{,}41 \approx 69{,}9$$

$$\sigma_s = \sqrt{D_s} \approx \sqrt{69{,}9} \approx 8{,}34$$

Задача 12.2. Случайная величина ξ распределена нормально со средним квадратичным отклонением $\sigma = 3$. Найти доверительный интервал для математического ожидания $M\xi$, если выборочная средняя $M_s = 4{,}1$; объем выборки n=36. Доверительная вероятность $\gamma = 0{,}95$.

Решение. Составим уравнение (12.20).

$$\Phi(U_\gamma) = \gamma/2 \Leftrightarrow \Phi(U_\gamma) = 0{,}95/2 = 0{,}475.$$

Из таблицы находим $U_\gamma \approx 1{,}96$. Следовательно,

$$\Delta = \frac{U_\gamma \cdot \sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 3}{\sqrt{36}} = 0{,}98 \Rightarrow \Delta = 0{,}98$$

Таким образом, доверительный интервал для $M\xi$ будет

$$M_s - 0{,}98 < M\xi < M_s + 0{,}98$$

или

$$4{,}1-0{,}98<M\xi<4{,}1+0{,}98.$$

$$3{,}12<M\xi<5{,}08.$$

Итак, с вероятностью 0,95 можно утверждать, что неизвестный параметр $M\xi$ нормального распределения ($\sigma=3$) будет принадлежать интервалу $(3{,}12;5{,}08)$.

Ответ. $P(M\xi\in(3{,}12;5{,}08))=0{,}95$.

Задача 12.3. Найти уравнения линий регрессии ξ на η и η на ξ по данным корреляционной таблицы 12.5:

Таблица 12.5 – Корреляционная таблица

ξ \ η	100	120	140	160	180	n_ξ
5	2	3				5
10	1	4				5
15		3	5			8
20			10	1		11
25			8			8
30				6		6
35				1	4	5
40				1	1	2
n_η	3	10	23	9	5	n=50

Решение. Из таблицы следует, что случайная величина ξ принимает значения:

$$x_1=5;x_2=10;x_3=15;x_4=20;$$

$$x_5=25;x_6=30;x_7=35;x_8=40$$

а величина η принимает значения:

$$y_1=100;y_2=120;y_3=140;y_4=160;y_5=180.$$

Возьмем в качестве c_1 значение 20, а в качестве c_2 значение 140. $c_1 = 20; c_2 = 140; k = 5; l = 10$.

Переходим к значениям:

$$u_i = \frac{x_i - c_1}{k} = \frac{x_i - 20}{5} \quad \nu_j = \frac{y_j - c_2}{l} = \frac{y_j - 140}{10}$$

i = 1,2,3,4,5,6,7,8; j = 1,2,3,4,5, которые заносим в таблицу 12.6.

Таблица 12.6 – Полученные значения

u_i \ v_j	-4	-2	0	2	4	n_ξ
-3	2	3				5
-2	1	4				5
-1		3	5			8
0			10	1		11
1			8			8
2				6		6
3				1	4	5
4				1	1	2
n_η	3	10	23	9	5	n= 50

Далее вычислим $M_s\xi$ и $D_s\xi$. Для этого используем таблицу 12.7:

Таблица 12.7 – Рассчитанные значения

u_i	n_i	$u_i \cdot n_i$	$u_i + 1$	u_i^2	$u_i^2 \cdot n_i$	$(u_i+1)^2$	$(u_i+1)^2 \cdot n_i$
-3	5	-15	-2	9	45	4	20
-2	5	-10	-1	4	20	1	5
-1	8	-8	0	1	8	0	0
0	11	0	1	0	0	1	11
1	8	8	2	1	8	4	32
2	6	12	3	4	24	9	54
3	5	15	4	9	45	16	80
4	2	8	5	16	32	25	50
$\sum$	50	10			182		252

Проверим вычисления с помощью соотношения (12.27).

$$252 = 182 + 2 \cdot 10 + 50 \Leftrightarrow 252 = 252.$$

Из (12.8) получим

$$M_s\xi = \frac{10 \cdot 5}{50} + 20 = 1 + 20 = 21 \Rightarrow M_s\xi = 21$$

Формула (12.10) даёт

$$D_s\xi = \frac{25}{50} \cdot 182 - (21 - 20)^2 = \frac{182}{2} - 1^2 = 91 - 1 = 90$$

Следовательно, из (12.12)

$$\sigma_s\xi = \sqrt{D_s\xi} = \sqrt{90} \approx 9{,}49$$

Запишем полученные значения в таблицу 12.8.

Таблица 12.8 – Полученные значения

v_j	n_j	$v_j \cdot n_j$	$v_j + 1$	v_j^2	$v_j^2 \cdot n_j$	$(v_j + 1)^2$	$(v_j + 1)^2$
-4	3	-12	-3	16	48	9	27
-2	10	-20	-1	4	40	1	10
0	23	0	1	0	0	1	23
2	9	18	3	4	36	9	81
4	5	20	5	16	80	25	125
Σ	50	6			204		266

Проверяем вычисления по (12.27):

$$266 = 204 + 2 \cdot 6 + 50 \Leftrightarrow 266 = 266.$$

По формулам (12.8) и (12.11), где $c = c_2 = 140$ получим,

$$M_s\eta = \frac{1}{50}\cdot 6\cdot 10 + 140 = \frac{6}{5} + 140 = 1{,}2 + 140 = 141{,}2$$

$$D_s\eta = \frac{100}{50}\cdot 204 - (141{,}2 - 140)^2 = 2\cdot 204 + (1{,}2)^2 =$$
$$= 408 - 1{,}44 \approx 406{,}56$$

$$\sigma_s\eta = \sqrt{D_s\eta} = \sqrt{406{,}56} \approx 20{,}16\,.$$

Теперь перейдем к вычислению выборочного коэффициента корреляции $k^s_{\xi\eta}$. Для этого составим новую корреляционную таблицу 12.9, аналогичную заданной в начале. В нее впишем значения u_i и v_j:

Таблица 12.9 – Корреляционная таблица

u_i \ v_j	-4	-2	0	2	4	$v = \sum n_{ij} v_j$	$u_i v$
-3	${}_{-6}2^{-8}$	${}_{-9}3^{-6}$				-14	42
-2	${}_{-2}1^{-4}$	${}_{-8}4^{-8}$				-12	24
-1		${}_{-3}4^{-6}$	${}_{-5}5^{0}$			-6	6
0			10	1^{2}		2	0
1			${}_{8}8^{0}$			0	0
2				${}_{12}6^{12}$		12	24
3				${}_{3}1^{2}$	${}_{12}4^{16}$	18	54
4				${}_{4}1^{2}$	${}_{4}1^{4}$	6	24
$u = \sum n_{ij} u_j$	-8	-20	3	19	16		
uv_j	32	40	0	38	64		$\Sigma = 174$

В каждую клетку этой таблицы вписаны: в центре n_{ij}; в правом верхнем углу $n_{ij}v_j$; а в левом нижнем $n_{ij}u_i$. Вычислим по формуле (12.25), используя данные последней таблицы. Имеем:

$$k^s_{\xi\eta} = \frac{\frac{5\cdot 10}{50}\cdot 174-(21-20)(141{,}2-140)}{9{,}49\cdot 20{,}16} =$$
$$= \frac{174-1\cdot 1{,}2}{191{,}32} = \frac{172{,}8}{191{,}32} \approx 0{,}9 \Rightarrow k^s_{\xi\eta} \approx 0{,}9$$

Коэффициент корреляции близок к 1,следовательно, линиями регрессии будут прямые, заданные уравнениями (12.23) и (12.26). Для (12.23) находим:

$$y_x - 141{,}2 = 0{,}9\cdot\frac{20{,}16}{9{,}49}(x-21)$$

или

$$y_x = 141{,}2 = 1{,}91(x-21) \Leftrightarrow y_x - 141{,}2 = 1{,}91x - 40{,}1.$$

Таким образом, линия регрессии η на величину ξ будет иметь уравнение $y_x = 1{,}91x + 101{,}1$.

Для определения регрессии ξ на η используем уравнение (12.26). Подставляя сюда данные, будем иметь:

$$x_y - 21 = 0{,}9\cdot\frac{9{,}49}{20{,}16}(y-141{,}2)$$
$$x_y - 21 = 0{,}42y - 59{,}8$$
$$x_y = 0{,}42y - 38{,}8$$

12.3 Задачи для упражнений

12.1 Построить полигон частот по данному распределению выборки ξ:

x_i	1	4	5	7
n_i	20	10	14	6

12.2 Построить полигон относительных частот по данному распределению выборки:

x_i	2	4	5	7	10
n_i/n	0,15	0,2	0,1	0,1	0,45

12.3 В таблице представлены результаты некоторой выборки:

56 76 65 66 76 62 89 48 62 50 47 80 67 87 78 55 67 51 73 75
61 88 46 57 65 60 72 28 75 51 69 68 65 34 77 63 57 61 42 85
49 41 62 63 80 62 65 75 56 66 92 60 43 52 80 68 70 76 62 55
42 87 81 67 65 81 90 38 58 60 79 79 50 64 70 58 77 73 54 58
77 86 52 61 42 70 93 54 65 51 53 64 65 76 88 59 62 67 62 90
88 69 61 81 65 72 58 68 94 54 58 58 81 57 70 71 78 52 93 89
57 68 70 58 72 57 62 63 87 61 91 57 57 66 68 40 63 86 48 75
66 83 64 55 75 65 67 54 70 44 51 86 67 58 73 71 46 86 68 79
50 58 66 69 61 64 78 78 60 46 71 71 74 79 65 61 62 84 53 67 83
43 64 67 50 60 83 61 83 67 67 58 46 73 58 47 76 81 72 66 83
73 71 70 60 68 52 51 63 63 75 61 80 51 63 62 46 48 53 59

Объем выборки равен 220.

а) Сгруппировать данные по интервалам длины h = 5.

б) Расписать распределение выборки по интервалам.

в) Построить гистограмму распределения.

г) Составить эмпирическую функцию распределения.

д) Вычислить выборочную среднюю, дисперсию и среднее квадратичное отклонение.

12.4. В таблице приведены результаты измерения роста 100 студентов:

Рост	154-158	158-162	162-166	166-170	170-174	174-178	178-182
Число студентов	10	14	22	27	17	7	3

Вычислить выборочную среднюю и дисперсию.

12.5 Найти дисперсию выборки:

x_i	2502	2804	2903	3028
n_i	8	30	60	2

12.6 В таблице приведены результаты некоторого опроса:

71 62 43 80 70 44 42 25 48 55 58 44 14 55 56
49 54 63 60 57 70 52 74 65 61 60 72 69 68 47
30 62 81 56 55 38 68 55 74 50 29 35 55 52 27
58 50 62 80 49 68 68 81 66 64 41 45 48 68 79
56 82 76 84 47 44 72 58 58 80 61 55 66 36 69
44 88 88 73 39 70 70 35 51 69 50 59 35 43 71
54 65 85 63 59 52 88 64 60 61 31 64 48 49 50
41 62 42 76 81 76 70 76 75 53 66 87 74 61 68
13 44 61 53 46 69 71 58 63 73 56 65 53 77 39
83 45 55 77 61 42 72 49 52 67 62 68 72 46 76
67 53 70 76 56 62 38 59 53 50 76 52 73 34 51
60 61 58 72 49

Объем выборки n=170.

а) Сгруппировать данные по интервалам длины h=5.

б) Построить гистограмму распределения абсолютных и относительных частот.

в) Найти эмпирическую функцию распределения и построить ее график.

г) Вычислить выборочную среднюю, дисперсию и среднее квадратичное отклонение.

12.7 Найти доверительный интервал для неизвестного параметра a нормального распределения, если известен второй параметр σ, выборочная средняя M_s и $\gamma = 0{,}99$:

а) $\sigma = 4; M_s = 10{,}2; n = 16$.

б) $\sigma = 5; M_s = 16{,}8; n = 25$.

12.8 Найти уравнения линий регрессии признака ξ на η и η на ξ по данным, представленным в таблице:

ξ \ η	14-19	19-24	24-29	29-34	34-39	n_ξ
16,5-18,5	2	2				4
18,5-20,5	3	4				7
20,5-22,5		15	6	2		23
22,5-24,5		11	22	14	2	49
24,5-26,5			18	29	10	57
26,5-28,5				3	7	10
n_η	5	32	46	48	19	n=150

12.9. Найти уравнения линий регрессии по данным выборки:

ξ \ η	20-25	25-30	30-35	35-40	40-45	45-50	n_ξ
16 - 18	2	3					5
18 - 20	3	5	1				9
20 - 22		9	25	3			37
22 - 24			13	4	5	2	24
24 - 26					3	2	5
n_η	5	17	39	7	8	4	n=80

12.10. Найти уравнения линий регрессии по данным таблицы:

ξ \ η	125	150	175	200	225	250	n_ξ
18		1					1
23	1	2	3				6
28		5	2	1			8
33			12	8			20
38				7	3		10
43					3	1	4
48						1	1
n_η	1	8	17	16	6	2	n=50

12.11. Найти уравнения линий регрессии по данным таблицы:

ξ \ η	16-21	21-26	26-31	31-36	36-41	41-46	n_ξ
150	3	7	2				12
200		2	8	6			16
250		3	50	4			57
300			2	6			8
350				2	1	1	4
400					2	1	3
n_η	3	12	62	18	3	2	n=100

12.12. Найти уравнения линий регрессии по данным таблицы:

ξ \ η	10-20	20-30	30-40	40-50	50-60	60-70	n_ξ
14 - 20	3						3
20 - 26	7	2	3				12
26 - 32	2	8	50				60
32 - 38		6	4	2	2		14
38 - 44				6	3	1	10
44 - 50					4	2	6
n_η	12	16	57	8	9	3	n=105

12.13 Найти выборочные уравнения прямых линий регрессии η на ξ и ξ на η по данным, приведенным в корреляционной таблице

ξ \ η	20	30	40	50	60	n_ξ
15	7	5				12
25	20	23				43
35		30	47	2		79
45		10	11	20	6	47
55			9	7	3	19
n_η	27	68	67	29	9	200

12.14 Найти уравнения прямых линий регрессии по данным таблицы

$\xi \backslash \eta$	15	20	25	30	35	n_ξ
40	5	7				12
50		4	16	23		43
60		8	20	32	27	87
70			11	29	2	42
80				9	7	16
n_η	5	19	47	93	36	200

12.15 По данным таблицы составить уравнения прямых линий регрессии

$\xi \backslash \eta$	3	9	15	21	27	33	n_ξ
25				1		1	2
35			1	5	4	5	15
45			2	18	10	2	32
55		6	14	2	2		24
65		6	3				9
75	4	8					12
85	6						6
n_η	10	20	20	26	16	8	100

Таблица 12.10 - Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

x	0	1	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9	10	11
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3947	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685

Продолжение таблицы 12.10

1	2	3	4	5	6	7	8	9	10	11
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
4,0	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001

Таблица 12.11 - Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$

x	0	1	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9	10	11
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4996	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4983	4983	4984	4984	4985	4985	4986	4986
3,0	4987	4987	4987	4988	4988	4989	4989	4989	4990	4990
3,1	4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
3,2	4993	4993	4994	4994	4994	4994	4994	4995	4995	4995
3,3	4995	4995	4995	4996	4996	4996	4996	4996	4996	4997
3,4	4997	4997	4997	4997	4997	4997	4997	4997	4997	4998
3,5	4998	4998	4998	4998	4998	4998	4998	4998	4998	4998

Продолжение таблицы 12.11

1	2	3	4	5	6	7	8	9	10	11
3,6	4998	4998	4999	4999	4999	4999	4999	4999	4999	4999
3,7	4999	4999	4999	4999	4999	4999	4999	4999	4999	4999
3,8	4999	4999	4999	4999	4999	4999	4999	4999	4999	4999
3,9	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
4,0	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000

Таблица 12.12 - Таблица значений функции $P(m) = \dfrac{\lambda^m}{m!} e^{-\lambda}$

λ	m						
	0	1	2	3	4	5	6
0,1	0,9048	0,0905	0,0045	0,0002			
0,2	0,8187	0,1637	0,0164	0,0011	0,0001		
0,3	0,7408	0,2222	0,0333	0,0033	0,0002		
0,4	0,6703	0,2681	0,0536	0,0072	0,0007	0,0001	
0,5	0,6065	0,3033	0,0758	0,0126	0,0016	0,0002	
0,6	0,5488	0,3293	0,0988	0,0198	0,0030	0,0004	
0,7	0,4966	0,3476	0,1217	0,0284	0,0050	0,0007	0,0001
0,8	0,4493	0,3595	0,1438	0,0383	0,0077	0,0012	0,0002
0,9	0,4066	0,3659	0,1647	0,0494	0,0111	0,0020	0,0003
1,0	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031	0,0005
1,1	0,3329	0,3662	0,2014	0,0738	0,0203	0,0045	0,0008
1,2	0,3012	0,3614	0,2169	0,0867	0,0260	0,0062	0,0012
1,3	0,2725	0,3543	0,2303	0,0998	0,0324	0,0084	0,0018
1,4	0,2466	0,03452	0,2417	0,1128	0,0395	0,0111	0,0026
1,5	0,2231	0,3347	0,2510	0,1255	0,0471	0,0141	0,0035
1,6	0,2019	0,3230	0,2584	0,1378	0,0551	0,0176	0,0047
1,7	0,1827	0,3106	0,2640	0,1496	0,0636	0,0216	0,0061
1,8	0,1653	0,2975	0,2678	0,1607	0,0723	0,0260	0,0078
1,9	0,1496	0,2842	0,2700	0,1710	0,0812	0,0309	0,0098
2,0	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0120
3,0	0,0498	0,1494	0,2240	0,2240	0,1680	0,1008	0,0504
4,0	0,0183	0,0733	0,1465	0,1954	0,1954	0,1563	0,1042
5,0	0,0067	0,0337	0,0842	0,1404	0,1755	0,1755	0,1462
6,0	0,0025	0,0149	0,0446	0,0892	0,1338	0,1606	0,1606
7,0	0,0009	0,00064	0,0223	0,0521	0,0912	0,1277	0,1490
8,0	0,0003	0,0027	0,0107	0,0286	0,0572	0,0916	0,1221
9,0	0,0001	0,0011	0,0050	0,0150	0,0337	0,0607	0,0911
10,0	0,0000	0,0004	0,0023	0,0076	0,0189	0,0378	0,0631

Библиографический список

1. Васильев, А.В. Mathcad 13 на примерах / А.В. Васильев. - СПб.: БХВ - Петербург, 2006. - 528с.
2. Кирьянов, Д.В. Самоучитель Mathcad 11 / Д.В. Кирьянов. - СПб.: БХВ-Петербург, 2003. - 560 с.
3. Новикова, Н.М. Компьютерный практикум по теории вероятности в среде MATHCAD: учебно-методическое пособие для вузов / Н.М. Новикова. – Воронеж: Издательско-полиграфический центр ВГУ, 2008. – 23 с.
4. Плис, А.И. Mathcad 2000. Математический практикум для экономистов и инженеров: Учеб. Пособие / А.И. Плис, Н.А. Сливина. - М.: Финансы и статистика, 2000. - 656с.