УНИВЕРСИТЕТСКАЯ СЕРИЯ

Р. Я. Хамидуллин

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УЧЕБНОЕ ПОСОБИЕ

Рекомендовано Общественным советом содействия повышению качества высшего образования

Москва
2020

УДК 519.2
ББК 22.17
Х182тв

Серия удостоена диплома в номинации «Лучший издательский проект»
на IV Общероссийском конкурсе учебных изданий для высших учебных заведений
«Университетская книга – 2008»

Печатается по решению Ученого совета Университета «Синергия»

Ответственный редактор серии
член-корреспондент Российской академии образования,
доктор экономических наук, профессор **Ю. Б. Рубин**

Рецензент
заслуженный работник высшей школы РФ,
кандидат технических наук, профессор **Монсик В.Б.**

Хамидуллин Р. Я.
Х182тв Теория вероятностей и математическая статистика: учебное пособие / Р. Я. Хамидуллин. – М.: Московский финансово-промышленный университет «Синергия», 2020. –276 с. (Университетская серия).

ISBN 978-5-4257-0398-9

Учебное пособие предназначено для студентов вузов, аспирантов и преподавателей экономических и смежных специальностей, а также для слушателей заочного и вечернего обучения, может быть полезно лицам, применяющим вероятностные методы при решении практических задач.

УДК 519.2
ББК 22.17

ISBN 978-5-4257-0398-9

ПРЕДИСЛОВИЕ

Учебное пособие предназначено для подготовки специалистов в области управления, экономики, информационных технологий, преподавателей и научных работников, использующих вероятностные и статистические подходы в повседневной работе, и содержит теоретические основы и прикладные методы теории вероятностей и математической статистики.

В процессе подготовки данных специалистов (бакалавров, магистров) традиционно математико–статистические дисциплины вызывают сложности освоения у студентов. Данное пособие призвано помочь студентам, особенно обучающимся в системе вечернего и заочного образования, понять прикладной, практический смысл проблем, решаемых с помощью математического аппарата теории вероятностей и математической статистики.

Методы теории вероятностей широко применяются в различных отраслях науки и техники: в теории надежности, теории массового обслуживания, при экономическом обосновании различных хозяйственных процессов, в теоретической физике, геодезии, астрономии, теории ошибок, теории управления, теории связи и во многих других теоретических и прикладных науках. Теория вероятностей служит для обоснования математической статистики.

Математическая статистика – раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений с целью выявления статистических закономерностей. Методы математической статистики используются при планировании организации производства, анализе технологических процессов, для контроля качества продукции и многих других целей.

Учебное пособие состоит из двух разделов и четырнадцати глав, каждая глава содержит не менее четырех параграфов. Главы имеют сквозную нумерацию.

Формулы обозначены тремя цифрами: первая цифра – номер главы, вторая цифра – номер параграфа, третья цифра – порядковый номер формулы в параграфе.

В первом разделе рассматриваются основные положения теории вероятностей: основные понятия теории вероятностей, комбинаторные методы решения вероятностных задач, основные теоремы и формулы теории вероятностей, случайные величины, их законы распределения, системы случайных величин и предельные теоремы теории вероятностей.

Во втором разделе даются основы математической статистики: выборочный метод в математической статистике, оценивание законов распределения и моментных характеристик случайных величин по результатам испытаний (наблюдений). Учебное пособие содержит значительное число примеров, иллюстрирующих теоретические положения.

После каждой главы приведены контрольные вопросы и задания, их объем вполне достаточен для организации практических занятий по решению задач и организации самостоятельной работы студентов.

В основу учебного пособия положен курс лекций, которые читались автором многие годы в Московском финансово-промышленном университете «Синергия» и основные положения учебного пособия под его редакцией «Теория вероятностей и математическая статистика в примерах и задачах». Порядок изложения теоретического материала в учебном пособие построен таким образом, чтобы подготовить студента к самостоятельному решению задач по всем основным темам курса теории вероятностей и математической статистике.

При подготовке учебного пособия использовано большое количество литературы (список прилагается).

Автор выражает большую благодарность заслуженному работнику высшей школы РФ, профессору В. Б. Монсик за рецензирование рукописи и сделанные им замечания

Раздел I
ТЕОРИЯ ВЕРОЯТНОСТЕЙ

ГЛАВА 1
ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. Случайное событие. Классификация случайных событий

Первичными понятиями теории вероятностей являются такие понятия, как испытание и событие.

Случайным называется событие, которое может произойти или не произойти в результате некоторого опыта (испытания, эксперимента).

Случайные события принято обозначать прописными буквами латинского алфавита: $A, B, C, \ldots, X, Y, Z$.

Опыт (испытание, эксперимент) – это процесс, включающий определенные условия и приводящий к одному из возможных исходов. Исходом опыта может быть результат наблюдения или измерения.

Примеры случайных событий:

$A =$ {выпадение “решки”};

$B =$ {выпадение “орла”} – при бросании монеты;

$C =$ {появление “7” очков} – при бросании игральной кости;

$D =$ {в аудиторию вошла студентка};

$E =$ {в аудиторию вошла студентка старше 18 лет}.

Случайные события могут состоять из нескольких элементарных событий, подразделяющихся на:

- достоверные;
- невозможные;
- совместные;
- несовместные;
- единственно возможные;
- равновозможные;
- противоположные.

Событие, которое обязательно произойдет в результате опыта (испытания, эксперимента) называется **достоверным** (пример *A* и *B*). Достоверные события будем обозначать Ω.

Событие, которое не может произойти в результате данного опыта (испытания, эксперимента) называется **невозможным** (пример *C*).

Обозначение: Ø.

Несколько событий называются **совместными**, если в результате опыта (испытания, эксперимента) появление одного из них не исключает возможность появления других (пример *D* и *E*).

Несколько событий называются **несовместными** в данном опыте (испытании, эксперименте), если появление одного из них исключает возможность появления других (пример *A* и *B*).

События называются **единственно возможными**, если в результате опыта (испытания, эксперимента) хотя бы одно из них обязательно произойдет (пример *A* и *B*).

Несколько событий называются **равновозможными**, если в результате опыта (испытания, эксперимента) ни одно из них не имеет объективно большую возможность появления, чем другие (пример *A* и *B*).

Два единственно возможных и несовместных события называются **противоположными** (пример *A* и *B*).

Совокупность всех единственно возможных и несовместных событий называется **полной группой событий**.

События, которые образуют полную группу, несовместны и равновозможны, называют **случаями** (пример *A* и *B*).

1.2. Действия над случайными событиями

Операция умножения событий

Произведением или **пересечением** двух событий ***A*** и ***B*** называется событие ***C***, состоящее в совместном появлении события ***A*** и ***B***.

$$C=A\cdot B \tag{1.2.1}$$

Для иллюстрации действий над событиями будем использовать так называемые диаграммы Венна[1].

Операция умножения событий (рис. 1.2.1.).

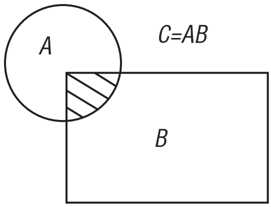


Рис. 1.2.1. Произведение двух событий

Операция умножения событий эквивалентна логической операции «и». В этом смысле формула (1.2.1) может быть представлена как:

$$C = A \text{ и } B \qquad (1.2.2)$$

Пример 1.

По некоторой цели производятся 3 выстрела.

Рассматриваются события $A1$, $A2$, $A3$ – попадание в цель первым, вторым и третьим выстрелами, V – попадание в цель всех трех выстрелов. Тогда очевидно:

$$V = A_1 \cdot A_2 \cdot A_3.$$

В данном примере рассматривается произведение трех событий – событие V. Иллюстрация дана на рис. 1.2.2.

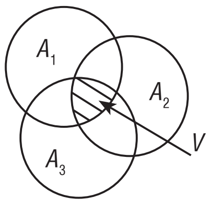


Рис. 1.2.2. Произведение трех событий

Тогда по примеру 1 в общем случае, произведение нескольких событий $A_1, A_2, A_3, \dots, A_n$ можно записать следующим образом:

$$V = \prod_{i=1}^{n} A_i \, . \qquad (1.2.3)$$

---

[1] Джон Венн (John Venn, 1834 1923) – английский математик-логик, построивший графический аппарат диаграмм. Основные труды: “Логика случая” (1866 г.), “Символическая логика” (1881 г.).

и заключается в появлении в данном опыте событий $A_1, A_2, A_3, \ldots, A_n$ вместе, то есть события A_1 и события A_2, ..., и события A_n.

Операция сложения событий

Суммой двух событий A и B называется событие D, состоящее в появлении события A или B, или обоих вместе.

Диаграммы Венна для несовместных и совместных событий будут различаться и представлены на рис. 1.2.3 и 1.2.4.

$$D=A+B \tag{1.2.4}$$

Операция сложения двух несовместных событий эквивалентна логической операции «или». В этом смысле формула (1.2.4) может быть пред-

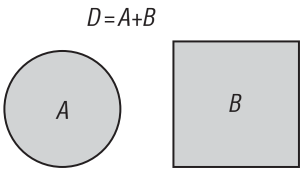


Рис. 1.2.3. Сумма двух несовместных событий

ставлена как:

$$D=A \text{ или } B \tag{1.2.5}$$

Сумма нескольких несовместных событий $A_1, A_2, A_3, \ldots, A_n$ может быть записана в виде:

$$D = \sum_{i=1}^{n} A_i. \tag{1.2.6}$$

Как видно из рис 1.2.4, сумма двух совместных событий представляет собой их объединение, которое состоит в появлении в данном опыте хотя

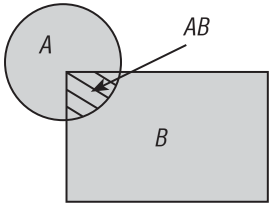


Рис. 1.2.4. Сумма двух совместных событий

бы одного из перечисленных событий – события A или события B или событий A и B вместе:

$$C = A \cup B. \tag{1.2.7}$$

Формула (1.2.7) эквивалентна логической записи:

$$C = A \text{ или } B \text{ или } AB. \tag{1.2.8}$$

Объединение нескольких совместных событий может быть представлено в виде:

$$C = \bigcup_{i=1}^{n} A_i. \tag{1.2.9}$$

Операция вычитания событий

Разностью двух событий ***A*** и ***B*** называется событие ***F***, которое состоит

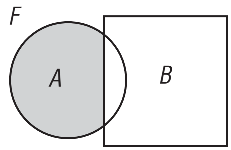


Рис. 1.2.5. Разность двух событий

из всех исходов, принадлежащих ***A***, исключая исходы, принадлежащие ***B***:

$$F = A - B. \tag{1.2.10}$$

На диаграмме разность событий представлена рис. 1.2.5.

Операция дополнения

Дополнением события A до события Ω называется событие $\overline{A}$, несовместное с A и образующее с ним полную группу событий (рис. 1.2.6). Дополнение A до Ω можно выразить в виде разности:

$$\overline{A} = \Omega - A. \tag{1.2.11}$$

События A и $\overline{A}$ являются противоположными событиями.

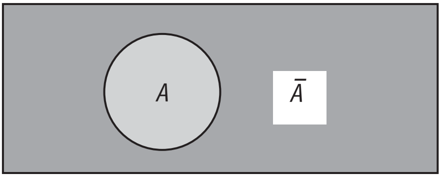


Рис. 1.2.6. Дополнение события A до полного пространства

Следовательно, для противоположных событий справедливы следующие соотношения:

$$A + \overline{A} = \Omega, \tag{1.2.12}$$

$$A \cdot \overline{A} = \emptyset. \tag{1.2.13}$$

1.3. Аксиоматическое определение вероятности события

Аксиоматическое определение вероятности было создано в 1933 году А. Н. Колмогоровым[2].

Вероятностью $P(A)$ называется числовая функция, определенная для всех событий A поля S (поля событий для данного эксперимента) и удовлетворяющая трем аксиомам вероятностей:

Аксиома 1 (*неотрицательности*): $P(A) \geq 0$. Вероятность неотрицательна.

Аксиома 2 (*нормировки*):. Вероятность достоверного события равна единице.

Аксиома 3 (*аддитивности*). Вероятность суммы несовместных событий A_1, A_2, A_3,..., A_n равна сумме вероятности этих событий:

$$P\sum_{i=1}^{n} A_i = \sum_{i=1}^{n} P(A_i), \quad A_i A_j = \emptyset, \quad i \neq j. \tag{1.3.1}$$

Смысл этих аксиом: вероятность есть неотрицательная, нормированная и аддитивная функция.

Аксиоматическое определение вероятности не дает способа конкретного вычисления вероятности, поэтому используются другие определения вероятности, на которых остановимся ниже.

[2] Андрей Николаевич Колмогоров (25 апреля 1903 — 20 октября 1987) — советский математик, один из крупнейших математиков XX века. Колмогоров — один из основоположников современной теории вероятностей.

1.4. Классическое определение вероятности

При классификации случайных событий мы использовали термин «возможность», то есть рассмотрели качественные признаки основных положений теории вероятностей. Для того, чтобы сравнивать события по степени возможности их появления, необходимо ввести количественную характеристику, то есть выразить численно.

Такой численной мерой объективной возможности наступления события является вероятность события.

Так называемый "классический" способ определения вероятности непосредственно используется для событий, называемых "случаями".

Случаями (шансами) называют несовместные, равновозможные события, образующие полную группу. Примерами событий, образующих группу случаев, являются:

- появление герба и цифры при одном бросании монеты;
- появление 1, 2, 3, 4, 5, 6 очков при однократном бросании игральной кости;
- появление различных шаров при вынимании наугад одинаковых на ощупь шаров из урны, содержащей несколько различных по цвету шаров.

Пусть события $A_1, A_2, A_3,..., A_n$ – несовместны, равновозможны и образуют полную группу – случаи, а события $A_1, A_2, A_3,..., A_m$ благоприятствуют появлению интересуемого нас события A ($m \leq n$). Тогда вероятность события A определяется по так называемой **классической формуле**:

$$P(A) = \frac{m}{n}. \quad (1.4.1)$$

где n – общее число случаев,
m – число случаев, благоприятствующих событию A.

Свойства вероятности события:

1. Вероятность достоверного события равна 1: $P(\Omega)=1$.
2. Вероятность невозможного события равна 0: $P(\varnothing)=0$.
3. Вероятность случайного события есть положительное число, заключенное между 0 и 1: $0<P<1$.

Таким образом, вероятность любого события: $0 \leq P \leq 1$.

Пример 2. Опыт заключается в однократном бросании игральной кости. Рассматриваются события:
A – появление четного числа очков;

B – появление не менее трех очков;

C – появление хотя бы одного очка.

Определим вероятности этих событий.

Решение.

Так как пространство исходов опыта – события $A_1, A_2, A_3, A_4, A_5, A_6$ – появление 1, 2, 3, 4, 5, 6 очков – образуют группу случаев, то искомые вероятности будем вычислять по формуле (1.4.1).

Событию A благоприятны исходы A_2, A_4, A_6, ($m_A = 3$) событию B благоприятны исходы A_3, A_4, A_5, A_6 ($m_B = 4$), событию C благоприятны все исходы $A_1, A_2, \ldots, A_6$ ($m_C = 6$). Вероятности этих событий соответственно равны:

$$P(A) = \frac{3}{6} = \frac{1}{2}, \qquad P(B) = \frac{4}{6} = \frac{2}{3}, \qquad P(C) = \frac{6}{6} = 1 = P(\Omega).$$

Пример 3. В урне находятся 12 шаров, пять из них красных. Найти вероятность, что наудачу вынутый шар – красный.

Решение.

A={наудачу вынутый из урны шар – красный}.

Всего шаров – 12, значит n=12 из них красных – 5, значит число благоприятствующих исходов m=5 .

Тогда по формуле (1.4.1):

$$P(A) = \frac{5}{12}.$$

1.5. Статистическое определение вероятности

Рассмотрим n одинаковых испытаний. Предположим, что результатом каждого испытания может быть появление или непоявление события A. Естественной характеристикой события A в этой последовательности испытаний является относительная частота (частость) его появления – отношение числа испытаний, в которых событие A появилось, к числу всех проведенных испытаний. Обозначая относительную частоту (частость) события A через $P^*(A)$, а число появлений события A в n испытаниях через m_A, получим:

$$P^*(A) = \frac{m_A}{n}. \tag{1.5.1}$$

Относительная частота невозможного события Ø:

$$P^*(\emptyset) = \frac{m_\emptyset}{n} = \frac{0}{n} = 0\,. \qquad (1.5.2)$$

Относительная частота достоверного события Ω:

$$P^*(\Omega) = \frac{m_\Omega}{n} = \frac{n}{n} = 1. \qquad (1.5.3)$$

Относительная частота случайного события A, которое может появиться или не появиться в результате испытания, очевидно, заключена между нулем и единицей. Поэтому относительная частота любого события представляет собой либо правильную дробь, либо равна нулю, либо равна единице:

$$0 \le P^*(A) \le 1 \ \text{ или в \% } \ 0\% \le P^*(A) \le 100\%. \qquad (1.5.4)$$

При многократном воспроизведении серии из n испытаний в одинаковых условиях изменяется число появлений m_A события A в каждой серии, следовательно, изменяется (является случайной) и относительная частота события (1.5.1).

Относительная частота события обладает замечательным свойством **устойчивости**, которое заключается в неограниченном приближении относительной частоты события к некоторой постоянной величине при неограниченном увеличении числа испытаний. Рассмотрим пример, характеризующий устойчивость частоты.

В таблице 1.5.1 приведены результаты испытаний известных ученых-статистиков Бюффона и Пирсона по наблюдениям за частотой появления герба при бросании монеты.

Таблица 1.5.1.

Экспериментатор	Число бросаний монеты	Число появлений герба	Частота
Бюффон (18 век)	4040	2048	0,5080
Пирсон (начало 20 века)	12000	6019	0,5016
Пирсон (начало 20 века)	24000	12012	0,5005

Результаты, приведенные в таблице, свидетельствуют о том, что относительная частота появлений герба неограниченно приближается к числу 0,5, которое назовем **статистической вероятностью события** G и будем обозначать $P(G)$. Таким образом, $P(G)$=0.50.

Мы ввели в рассмотрение статистическую вероятность – число, к которому неограниченно приближается относительная частота события при увеличении числа испытаний. Характер сходимости частоты и вероятности впервые изучен Я. Бернулли[3] и сформулирован в соответствующей теореме.

Величина X_n сходится по вероятности к величине a, если при сколь угодно малом $\varepsilon > 0$ вероятность неравенства $|X_n - a| < \varepsilon$ неограниченно приближается к единице при неограниченном увеличении числа n.

Применительно к частоте и вероятности это запишется в виде:

$$P(|P^*(A) - P(A)| < \varepsilon) > 1 - \delta, \quad \varepsilon > 0, \quad \delta > 0. \qquad (1.5.5)$$

Выражение (1.5.5) составляет содержание теоремы Я. Бернулли: частота события A при неограниченном увеличении числа испытаний сходится по вероятности к вероятности события A.

Следовательно, вероятность события $P(A)$ есть численная мера степени объективной возможности события в данном испытании, и с понятием "вероятность события" мы связываем определенный практический смысл: на основании накопленного опыта мы утверждаем, что наиболее вероятны те события, которые чаще происходят, а те события, которые редко происходят – наименее вероятны. Таким образом, понятие "вероятность события" связано с результатами опытов и, следовательно, с понятием "частота события".

1.6. Геометрическая вероятность

Одним из недостатков классического определения вероятности события, основанного на рассмотрении конечного числа равновозможных исходов испытания, является невозможность использования формулы (1.4.1) для случая бесконечного множества исходов испытания. Устранить этот недостаток можно используя так называемую геометрическую вероятность.

Общая задача, которая является моделью подхода к вычислению вероятности для такого случая, формулируется следующим образом. Рассмотрим на плоскости некоторую область G (рис.1.6.1) и область $g \subset G$. В область G наугад бросается точка x так, что она может равновероятно попасть в любую точку области, следовательно, вероятность попадания

[3] Якоб Бернулли (1654-1705) – швейцарский математик, профессор Базельского университета (с 1687). Фундаментальные достижения в теории вероятностей – разработка вероятностной модели независимых повторных испытаний (испытания Бернулли). Дал первое доказательство закона больших чисел. Основной труд по теории вероятностей "Искусство предположений" (опубликован в 1713 г).

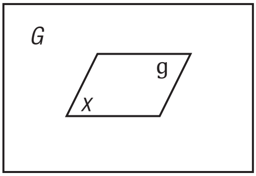


Рис. 1.6.1. Модель вычисления геометрической вероятности

в какую-либо часть области *G*, например, в область g, пропорциональна площади этой области и не зависит от её формы и расположения.

Поэтому, по определению, вероятность попадания точки *x* в область g при её бросании наугад в область *G* (событие *A*), естественно, будет равна:

$$P(A) = \frac{\text{mes } g}{\text{mes G}}, \tag{1.6.1}$$

где *mes* g – мера области $g \subset G$,
mes G – мера области *G*.

В качестве меры могут выступать длина, площадь, объем, угол и другое.

Рассмотрим несколько примеров вычисления геометрических вероятностей.

Пример 4. Два лица *A* и *B* условились о встрече в определенном месте в течение часа после 12 часов. Лицо, пришедшее первым – предположим *A*, ждет встречи в течение 20 минут. Когда они заканчиваются, лицо *A* уходит. Какова вероятность встречи *A* и *B*, если они приходят независимо друг от друга в случайное время после 12 часов до 13 часов?

Решение.

Рис. 1.6.2. Пространство исходов и область встречи *A* и *B*

Обозначим моменты прихода A через X, а B – через Y. Для встречи A и B необходимо и достаточно, чтобы было выполнено условие $|X - Y| \leq 20$.

Элементарное событие ω_1 (встреча A и B) будет характеризоваться двумя случайными параметрами X и Y и может быть представлено точкой с координатами (x, y) на плоскости $X0Y$. Построим на плоскости пространство элементарных исходов (событий) Ω. Это – квадрат (рис. 1.6.2) со стороной 60 минут. Условие встречи будет выполнено, если случайная точка (x, y) окажется в заштрихованной области g. Её площадь (мера области g) равна площади квадрата за вычетом площадей двух угловых треугольников:

$$S_g = S_G - 2\frac{1}{2}40^2,$$

$$S_g = 60^2 - 40^2 = 2000,$$

$$S_G = 60^2 = 3600,$$

$$P_{\text{вст.}} = \frac{2000}{3600} = \frac{5}{9} \quad .$$

Пример 5. В квадрат со стороной, равной a см, вписан круг. Бросается случайная точка. Какова вероятность того, что точка попадет внутрь круга.

Решение.

Вероятность попадания случайной точки в круг (событие) определяется как (см. рис. 1.6.3):

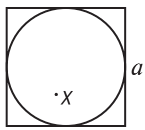


Рис. 1.6.3. Иллюстрация, к примеру 5

$$P(A) = \frac{S_{\text{кр.}}}{S_{\text{кв.}}}.$$

Найдем площадь круга и квадрата:

$$S_{\text{кр.}} = \pi\left(\frac{a}{2}\right)^2 = \frac{\pi a^2}{4},$$

$$S_{\text{кв.}} = a^2$$

тогда искомая вероятность будет равна:

$$P(A) = \frac{\pi a^2}{4a^2} = \frac{\pi}{4}.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что понимается под испытанием и событием в теории вероятностей?
2. Какова классификация событий в теории вероятностей?
3. Какое событие называется достоверным, какое невозможным и какое случайным?
4. Что является предметом изучения теории вероятностей?
5. На какие виды подразделяются случайные события?
6. Какие события называются равновозможными, какие несовместными, какие совместными и какие противоположными?
7. Что понимается под полной группой событий?
8. Что понимается под вероятностью события?
9. Что является единицей меры вероятности события?
10. Вероятность, равная нулю, соответствует какому событию?
11. В каких пределах находится вероятность случайного (возможного) события?
12. Какие события можно отнести к схеме случаев?
13. Какой случай называется благоприятствующим для события ?
14. По какой формуле определяется вероятность события , если испытание можно отнести к схеме случаев?
15. Что понимается под геометрической вероятностью на оси, на плоскости и в пространстве?
16. Что понимается под относительной частотой события ?
17. В чем отличие вероятности события от ее относительной частоты?
18. В чем состоит свойство устойчивости относительной частоты события ?

ЗАДАНИЯ

1.1. Из 1000 собранных на заводе телевизоров 5 штук бракованных. Эксперт проверяет один наугад выбранный телевизор из этой 1000. Найдите вероятность того, что проверяемый телевизор окажется бракованным.

1.2. В урне 9 красных, 6 желтых и 5 зеленых шаров. Из урны наугад достают один шар. Какова вероятность того, что этот шар окажется желтым?

1.3. На каждые 1000 электрических лампочек приходится 5 бракованных. Какова вероятность купить исправную лампочку?

1.4. В группе туристов 8 человек. С помощью жребия они выбирают шестерых человек, которые должны идти в село в магазин за продуктами. Какова вероятность того, что конкретно выбранный турист, входящий в состав группы, пойдет в магазин?

1.5. По железнодорожному полотну производится бомбометание. Длина полотна 3 км. Известно, что бомба попала на железнодорожное полотно. Какова вероятность попадания в железнодорожный состав длиною в 500 м находящийся на этом полотне?

1.6. В квадрат вписан равнобедренный треугольник так, что его основание совпадает со стороной квадрата. В квадрат случайным образом бросается точка. Найти вероятность того, что точка не попадет в треугольник?

1.7. На плоскость нанесена система параллельных линий, расположенных на расстоянии 3 см друг от друга. На плоскость случайным образом брошена монета диаметром 1 см. Какова вероятность того, что монета не пересечет ни одну из линий?

1.8. Перед началом первого тура чемпионата по шашкам участников разбивают на игровые пары случайным образом с помощью жребия. Всего в чемпионате участвует 26 шашистов, среди которых 3 участника из России, в том числе Петр Иванов. Найдите вероятность того, что в первом туре Петр Иванов будет играть с каким-либо шашистом из России?

1.9. При испытании партии приборов относительная частота годных приборов оказалась равной 0,9. Найти число годных приборов, если всего было проверено 200 приборов.

1.10. Два шахматиста играют одну партию. Событие А - выиграет первый игрок, В - выиграет второй игрок. Какое событие следует добавить к указанной совокупности, чтобы получилась полная группа событий?

1.11. События: А – хотя бы один из проверяемых приборов бракованный, - все приборы доброкачественные. Что означают события:

а) $A+B$?; б) AB?; в) $\overline{A}$; г) $\overline{B}$.

1.12. Из таблицы случайных чисел наугад выбраны два числа. События и соответственно, означают, что выбрано хотя бы одно простое число и хотя бы одно четное число. Что означают события AB?, $A+B$?.

1.13. В зоне поражения цели зенитной артиллерией, представляющей шар радиуса 100 м, оказалась цель, объем которой приближенно равен 1000 м^3. Какова вероятность поражения цели при одном выстреле?

ГЛАВА 2
КОМБИНАТОРИКА В ВЕРОЯТНОСТНЫХ ЗАДАЧАХ

2.1. Комбинаторный характер вероятностных задач

Комбинаторика – один из разделов дискретной математики, наука, изучающая комбинации и перестановки объектов произвольной природы, подчиненные определенным условиям.

С точки зрения теории множеств комбинаторика имеет дело со способами образования различных подмножеств конечных множеств, их объединениями и пересечениями, а также разрабатывает способы упорядочения этих подмножеств.

В методическом плане комбинаторика рассматривается как введение в теорию вероятностей, поскольку широко используется при решении вероятностных задач. Как правило, решение задач по классической схеме с конечным числом равновозможных исходов (1.4.1) сводится к использованию комбинаторных методов для подсчета общего числа случаев и числа случаев , благоприятствующих некоторому событию *A*:

$$P(A) = \frac{m}{n} \cdot$$

Кроме теории вероятностей, комбинаторика нашла применение во многих областях науки и техники, в частности, в теории информации, кодирования, статистической физике, теории графов, планировании экспериментов, теории программирования, генетике, биологии.

Рассмотрим пример, подтверждающий необходимость использования комбинаторных методов для их решения.

Пример 1. Игральная кость бросается дважды подряд. Вычисляется вероятность появления одного и того же числа очков при первом и втором бросаниях (событие *A*).

Решение.

Для определения общего числа случаев (исходов опыта) необходимо рассмотреть все возможные комбинации цифр, выпадающих на первой и второй костях, каждая из которых включает двухзначные числа, составленные из цифр от 1 до 6 на первой и второй позиции, то есть (1,1), (1,2), … , (3,6), … , (6,6). Число случаев, благоприятных событию *A*, равно числу двухзначных чисел, состоящих из одинаковых цифр, то есть (1,1), (2,2), ..., (6,6).

Способы подсчета чисел комбинаций в представленном примере, будет рассмотрен ниже.

2.2. Выборка из множества элементов

Изложение методов комбинаторики применительно к решению вероятностных задач ведется на основе понятия «выборка из множества элементов», что позволяет рассматривать все комбинаторные методы с единых позиций.

Если элементы выборки располагаются в порядке их выбора из множеств $X_1, X_2, \dots , X_k$ (в линейку), то такая выборка называется **упорядоченной**.

Если порядок элементов выборки не имеет значения (“куча”), то такая k^0 - выборка называется **неупорядоченной**.

Рассмотрим два основных правила подсчета чисел упорядоченных выборок (размещений) – правило произведения и правило суммы.

Пусть для образования упорядоченных выборок, обладающих заданным свойством, элемент $x_i^{(1)}$ из множества X_1 можно выбрать n_1 способами (по числу n_1 элементов этого множества), элемент из множества можно выбрать способами и т.д., наконец, элемент $x_j^{(2)}$ из множества X_2 можно выбрать n_2 способами и т.д., наконец, элемент $x_m^{(k)}$ из множества $X_k - n_k$ способами. Тогда число выборок n, обладающих заданным свойством, будет равно произведению чисел $n_1, n_2, \dots, n_k$:

$$n = n_1 n_2 \dots n_k = \prod_{i=1}^{k} n_i \,. \qquad (2.2.1)$$

Полученный результат представляет собой **правило произведения**.

Частный случай, если $X_1=X_2=\dots=X_k=X$ и $n_1=n_2=\dots=n_k=m$, то правило произведения (2.3.4) запишется в следующем виде:

$$n=m^k \qquad (2.2.2)$$

Пример 2. Определим возможное число n семизначных телефонных номеров.

Решение.

Искомое число семизначных телефонных номеров равно числу упорядоченных 7 - выборок из множества $X_1=\{1, 2, \dots , 9\}$, $n_1=9$ и шести одинаковых множеств $X_2=X_3=\ldots=X_7=\{0, 1, 2, \ldots, 9\}$, $n_2=10$. В соответствии с правилом произведения (2.3.4), (2.3.5) имеем:

$$n=n_1 n_2 \ldots n_7=9\cdot10\cdot10\cdot10\cdot10\cdot10\cdot10=9\cdot10^6.$$

Правило суммы используется в том случае, когда множество всевозможных выборок может быть представлено в виде некоторого числа непересекающихся различных классов $X_{j,j}=1 \ldots n$ (рис.2.2.1).

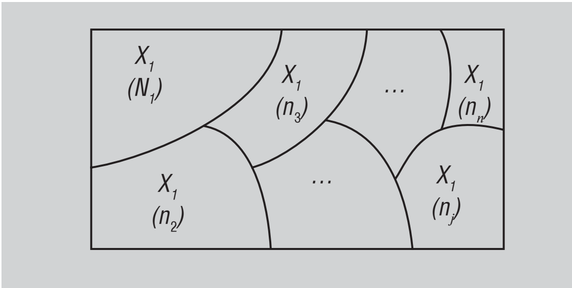


Рис. 2.2.1. Множество непересекающихся классов (подмножество) выборок

Общее число выборок в этом случае равно сумме чисел выборок в различных классах:

$$n = n_1 + n_2 + \ldots + n_n = \sum_{j=1}^{n} n_j. \tag{2.2.3}$$

Это и есть формула, определяющая **правило суммы**. Числа $n_1, n_2, \ldots, n_n$ выборок в различных классах определяются в соответствии с правилом произведения (2.2.1).

Пример 3. Определим, сколько различных сигналов можно составить из четырех различных сигнальных флагов на корабле.

Решение.

Множество вариантов составления различных сигналов можно разделить на 4 непересекающихся подмножества:

X_1 – множество сигналов из одного флага;

X_2 – множество сигналов из двух флагов;

X_3 – множество сигналов из трех флагов;

X_4 – множество сигналов из четырех флагов.

Общее число вариантов составления сигналов из 4-х флагов равно (2.2.3):

$$n=n_1+n_2+n_3+n_{4,}$$

где n_1= 4 варианта сигналов из одного флага;

n_2= 4·3 = 12 вариантов из двух флагов;

n_3= 4·3·2 = 24 варианта сигналов из трех флагов;

n_4= 4·3·2·1= 24 варианта сигналов из четырех флагов.

Таким образом, в результате получаем:

$$n= 4 + 12 + 24 + 24 = 64 \text{ (варианта сигналов).}$$

При подсчете чисел n_k использовалось правило произведения (2.2.1). Так, например, при подсчете n_3 в качестве первого флага для составления сигнала из 3-х флагов можно выбрать любой из 4-х флагов. Вторым флагом мог быть один из 3-х оставшихся, а третьим – один из двух оставшихся флагов. Отсюда и результат: n_3 = 4·3·2 = 24.

2.3. Размещения

Для вычисления суммы n – первых натуральных чисел существует очень удобная формула:

$$S = \frac{n(n+1)}{2}.$$

Для вычисления произведения n – первых натуральных чисел такой формулы нет, но зато эта величина получила специальное обозначение $n!$.

Читается как «эн» – факториал.

Примеры: 5! = 1·2·3·4·5 = 120, ..., 7! =1·2·3·4·5·6·7 = 120 = 5040.

Отметим, что принято считать 0!=1. Основанием этому утверждению служит свойство факториала: $n! = n\,(n-1)!$.

Приведенное выражение справедливо при n >1. Тогда естественно определить 0! так, чтобы было верно и при n=1. Тогда должно выполняться соотношение: 1!=1, но для этого необходимо считать, что 0!=1.

Размещением без повторений из *n* элементов по *m* элементов ($m<n$) называется упорядоченная m –выборка, элементы которой являются элементами одного и того же множества

$$X=\{x_1, x_2, \dots , x_n\}$$

и все различны.

Две такие выборки различны, если они отличаются друг от друга хотя бы одним элементом или порядком (расстановкой) одних и тех же элементов, то есть либо составом, либо порядком расположения, либо и тем, и другим.

Число размещений без повторений обозначается[4] A_n^m и определяется в соответствии с правилом произведения (2.2.1) при $n_1=n$, $n_2=n-1$, $n_3=n-2$, ... , $n_m=n-m+1$:

$$A_n^m = n(n-1)(n-2)\ldots(n-m+1). \qquad (2.3.1)$$

Для получения компактной формулы для A_n^m умножим и разделим выражение (2.3.1) на дополнение его до $n!$ – на величину $(n-m)!$:

$$(n-m)!=(n-m)(n-m-1)\ \ldots\ 3\cdot2\cdot1. \qquad (2.3.2)$$

Тогда окончательно получим:

$$A_n^m = \frac{n!}{(n-m)!}. \qquad (2.3.3)$$

Рассмотрим примеры определения числа размещений без повторений.

Пример 4. Состав и число размещений без повторений.

Из элементов множества $X=\{a, b, c, d, e\}$, $n=5$ составим размещения без повторений из 5 по 3 элемента и определим их число.

Решение. Из 5-элементного множества X можно образовать по формуле (2.3.3)

$$A_5^3 = \frac{5!}{2!} = \frac{2!\cdot3\cdot4\cdot5}{2!} = 3\cdot4\cdot5 = 60$$

размещений без повторений элементов, содержащих по 3 элемента.

Эти выборки могут отличаться друг от друга либо составом элементов, либо их порядком, но при этом повторение элементов в различных выборках не допускается. Перечислим их:

$$A=\{(a, b, c), (a, b, d), (a, b, e), (b, a, c), (b, c, a), \ldots, (e, d, c)\}.$$

Пример 5. Составление чисел.

Определим, сколько четырехзначных чисел можно составить из цифр 1, 2, 3, 4, 5, если повторения цифр в этих числах не допускается?

Решение. Количество четырехзначных чисел равно числу размещений без повторений из 5 элементов по 4 элемента:

$$A_5^4 = \frac{5!}{1!} = \frac{1!\cdot2\cdot3\cdot4\cdot5}{1!} = 2\cdot3\cdot4\cdot5 = 120 \text{ чисел}$$

Пример 6. Призеры хоккейного турнира.

Определим, сколько существует способов распределения золотых, серебряных и бронзовых медалей между призерами хоккейного турнира,

[4] буква А – начальная от arrangement с французского – размещение.

если предположить, что все 16 играющих команд потенциально равны по уровню игры и составу игроков?

Решение. Задача сводится к определению числа троек команд – 3 выборок из 16 элементов, отличающихся друг от друга либо составом команд, либо порядком распределения мест. Это – число размещений без повторений, которое равно

$$A_{16}^{3}=\frac{16!}{13!}=\frac{13!\cdot 14\cdot 15\cdot 16}{13!}=14\cdot 15\cdot 16=3360 \text{ способов}.$$

Размещения с повторениями

Размещением с повторениями из *n* элементов по *m* элементов называется упорядоченная m – выборка, составленная из элементов одного и того же n – множества

$$X=\{x_1, x_2, ..., x_n\}.$$

Элементы этой выборки могут совпадать. Две таких выборки различны, если они отличаются порядком неравных элементов или их составом.

Число размещений с повторениями обозначается $\hat{A}_n^m$ и определяется

по правилу произведения (2.2.1), (2.2.2) при:

$$\hat{A}_n^m=n^m. \tag{2.3.4}$$

Рассмотрим примеры определения числа размещений с повторениями.

Пример 7. Состав и число размещений с повторениями.

Из элементов множества $X=\{a, b, c, d, e\}$, $n=5$ составим размещения с повторениями (упорядоченные выборки) по 3 элемента в каждой и определим их число.

Решение. Из пяти элементов множества X можно образовать (2.3.4)

$$\hat{A}_5^3=5^3=125$$

размещений с повторениями по 3 элемента.

Эти выборки могут отличаться друг от друга либо составом элементов, либо их порядком:

$$\hat{A}=\{(a,\ a,\ a),\ \ (a,\ a,\ b),\ \ (a,\ a,\ c),\ \ (a,\ a,\ d),\ \ (a,\ a,\ e),$$
$$(a,\ b,\ a),\ \ (a,\ b,\ b),\ \ (b,\ b,\ b),\ \ \dots,\ \ (e,\ e,\ e)\}$$

Пример 8. Составление числа.

Определим, сколько четырехзначных чисел можно составить из множества цифр X={1, 2, 3, 4, 5, 6, 7, 8, 9}, если повторения цифр допустимы?

Решение. Количество четырехзначных чисел равно числу размещений с повторениями из 9-ти элементов по 4 элемента:

$$\hat{A}_9^4 = 9^4 = 6561.$$

2.4. Перестановки

Перестановками без повторений (или просто **перестановками**) называются размещения без повторений из всех элементов X, то есть из n элементов по n.

Их число обозначается P_n и определяется выражением (2.4.1):

$P_n = A_n^n = n(n-1)(n-2)\ldots 3 \cdot 2 \cdot 1 = n!$, то есть

$$P_n = n!. \qquad (2.4.1)$$

Две перестановки различны, если отличаются только порядком своих элементов, так как в их состав входит все n элементов исходного множества X. Рассмотрим примеры составления перестановок из n элементов множества X.

Пример 9. Состав и число перестановок. Из элементов множества $X=\{a, b, c, d, e\}$, n=5 составим всевозможные перестановки (размещения без повторений из пяти элементов по пять) и определим их число.

Решение. Число перестановок из пяти элементов a, b, c, d, e равно:

$$P_5 = 5! = 1\cdot 2\cdot 3\cdot 4\cdot 5 = 120.$$

Соответствующие перестановки представляются в виде:

$$\{P_5\} = \{(a,\ b,\ c,\ d,\ e),\ \ (b,\ a,\ c,\ d,\ e),\ \ (b,\ c,\ a,\ d,\ e),$$

$$(b,\ c,\ a,\ e,\ d),\ \ (c,\ a,\ b,\ d,\ e),\ \ \ldots,\ \ (e,\ d,\ c,\ b,\ a)\}\ .$$

Пример 10. Расстановка книг на полке.

Сколькими способами можно расставить 4 различные книги на книжной полке?

Решение. Число расстановок 4-х книг равно числу перестановок из 4-х элементов:

$$P_4 = 4! = 1\cdot 2\cdot 3\cdot 4 = 24 \text{ (способа)}.$$

Перестановки с повторениями

Перестановками с повторениями называются такие перестановки, в которых хотя бы два элемента одинаковы.

Пусть перестановка P_n состоит из n элементов множества $\hat{X}$, среди которых имеются одинаковые элементы, например, n_1 элементов A, n_2 элементов B, n_3 элементов $C, \ldots, n_m$ элементов L, так что сумма чисел элементов равна:

$$n_1+n_2+n_3+\ldots+n_m=n.$$

Одна из возможных перестановок из этих n элементов множества X, общее число которых равно $n!$, может иметь, например, следующий вид:

$$(P_n)_1 = \left(\overbrace{\underbrace{a, a, \ldots, a}_{n_1}, \underbrace{b, b, \ldots, b}_{n_2}, \underbrace{c, c, \ldots, c}_{n_3}, \ldots, \underbrace{z, z, \ldots, z}_{n_m}}^{n}\right). \qquad (2.4.2)$$

Перестановка $(P_n)_1$ – перестановка с повторениями. Число перестановок с повторениями обозначается символом:

$$\hat{P}(n_1,\ n_2,\ \ldots,\ n_m).$$

При его определении будем иметь в виду, что, переставляя n_1 одинаковых элементов $a, a, \ldots, a$, $n_1!$ способами, мы не получим новых перестановок, следовательно, общее число перестановок $n!$ следует уменьшить в $n_1!$ раз. Аналогично, переставляя n_2 одинаковых элементов $b, b, \ldots,$ b $n_2!$ способами, мы вновь не получим новых перестановок, следовательно, опять общее число перестановок $n!$ следует уменьшить в $n_2!$ раз. Рассуждая и дальше подобным образом, получим следующий результат:

$$\hat{P}(n_1,\quad n_2,\ \ldots,\ n_k) = \frac{n!}{n_1!\,n_2!\ldots n_m!}. \qquad (2.4.3)$$

Пример 11. Сколько различных “слов” можно составить из букв, образующих слово АВИАЦИЯ?

Решение. Слово АВИАЦИЯ содержит две буквы “А” (n_1=2), две буквы “И” (n_2=2), по одной “В”, “Ц”, “Я” ($n_3 = n_4 = n_5 =1$). Общее число букв n=7. Общее число перестановок из семи букв равно

$P_7 = 7! = 1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7 = 5040$.

Однако, учитывая то обстоятельство, что перестановки одинаковых букв “А” и “И” не дают новых “слов”, получим

$$\hat{P}(2,\ 2,\ 1,\ 1,\ 1) = \frac{7!}{2!\ 2!\ 1!\ 1!\ 1!} = \frac{5040}{4} = 1260.$$

2.5. Неупорядоченные выборки (сочетания)

Сочетания без повторений

Сочетанием без повторений или просто **сочетанием** *из n* элементов по m элементов ($m \le n$) называется неупорядоченная выборка, составленная из элементов одного и того же множества:

$$X=\{x_1, x_2, \ldots, x_n\}.$$

Элементы (состав) такой выборки все различны, а их порядок не имеет значения.

Два сочетания различны, если они отличаются хотя бы одним элементом.

Два сочетания равны, если они состоят из одних и тех же элементов.

Примеры сочетаний:

1. Выборка 5-ти чисел из 36 для заполнения карточки Спортлото (существенны сами числа, но не их порядок).
2. Выбор 22-х игроков в сборную команду по футболу из 40 кандидатов (существенны кандидатуры игроков, но не порядок их выбора).
3. Выбор 3-х членов ревизионной комиссии из 50 членов садового товарищества (существенны сами люди, а не порядок их выбора).

Число сочетаний без повторений (сочетаний) из n элементов по m элементов ($m \le n$) обозначается C_n^m.

Формула для подсчета числа сочетаний C_n^m из n элементов по m выводится из формулы (2.3.3) для числа A_n^m размещений из n элементов по m. Действительно, составим все сочетания из n по m элементов, число которых равно C_n^m, а затем произведем всевозможные перестановки из m элементов каждого сочетания, которых будет $m!$.

Получим множество упорядоченных выборок без повторений элементов – размещений из n элементов по m, различающихся либо самими элементами, то есть составом элементов, либо порядком элементов.

Таким образом, можно записать:
$A_n^m = C_n^m \cdot m!$, откуда с учетом выражения (2.3.3) для A_n^m получим:

$$C_n^m = \frac{A_n^m}{m!} = \frac{n!}{m!\,(n-m)!}. \tag{2.5.1}$$

Учитывая другую форму записи A_n^m (2.3.2), получим выражение для C_n^m более удобное для практических расчетов:

$$C_n^m = \frac{n(n-1)(n-2)\dots(n-k+1)}{m!}. \qquad (2.5.2)$$

Заметим, что полученная формула (2.5.2) для C_n^m совпадает с формулой для числа перестановок с повторениями (2.4.3) из n элементов, в состав которых входят m элементов одного типа и $(n - m)$ элементов другого типа (например, m "единиц" и $(n - m)$ "нулей"). Действительно,

$$P(m, n-m) = \frac{n!}{m!\,(n-m)!} = C_n^m. \qquad (2.5.3)$$

Рассмотрим примеры подсчета числа сочетаний.

Пример 12. Выбор шаров из урны.

Из урны, содержащей 10 одинаковых на ощупь шаров, среди которых могут быть шары разного цвета, вынимают наугад 4 шара. Сколько существует способов выбора 4 шаров из 10 различных шаров, если такие 4 выборки будут отличаться лишь составом, а не порядком (сочетания)?

Решение. Число способов выбора 4-х шаров из 10 различных шаров равно числу неупорядоченных выборок (сочетаний) из 10 шаров по 4 (2.5.1):

$$C_{10}^4 = \frac{10!}{4!\,6!} = \frac{6!\cdot 7\cdot 8\cdot 9\cdot 10}{6!\cdot 1\cdot 2\cdot 3\cdot 4} = 7\cdot 3\cdot 10 = 210 \text{ способов}.$$

Пример 13. Выбор членов комиссии.

В отделе 8 сотрудников, 5 из которых должны быть направлены в комиссию для проверки инвентарного учета. Сколькими способами можно выбрать комиссию?

Решение. Число способов выбора комиссии равно числу сочетаний (неупорядоченных выборок без повторений) из 8 по 5 элементов (2.5.1):

$$C_8^5 = \frac{8!}{3!\,5!} = \frac{5!\cdot 6\cdot 7\cdot 8}{6!\cdot 1\cdot 2\cdot 3} = 7\cdot 8 = 56 \text{ способов}.$$

Свойства сочетаний

Сочетания обладают рядом замечательных свойств, которые облегчают решение практических задач комбинаторики. Рассмотрим эти свойства, используя для доказательства формулу для числа сочетаний (2.5.1).

Свойство 1. Число сочетаний из n элементов по одному элементу равно n:

$$C_n^1 = n. \qquad (2.5.4)$$

Свойство 2. Число сочетаний из n элементов по n равно единице:

$$C_n^n = 1. \tag{2.5.5}$$

Свойство 3. Число сочетаний из n элементов по m равно числу сочетаний из n элементов по $(n - m)$ элементов:

$$C_n^m = C_n^{n-m}. \tag{2.5.6}$$

Например, $C_{10}^8 = C_{10}^2$, $C_{100}^{97} = C_{100}^3$ и т.д.

Свойство 4. Правило Паскаля.

Число сочетаний из n элементов по m равно сумме чисел сочетаний из n-1 элементов по m и по $(m - 1)$ элементов соответственно:

$$C_n^m = C_{n-1}^{m-1} + C_{n-1}^m, \quad m = 1, \ldots, n. \tag{2.5.7}$$

Выражение (2.5.7), известное под названием “правило Паскаля”, дает возможность легко определять числа C_n^m для различных значений n и m $(m = 1, \ldots, n)$ – так называемые биномиальные коэффициенты в разложении бинома Ньютона.

Сочетания с повторениями

Сочетанием с повторениями из элементов n типов по m элементов $(m > n)$ называется произвольная неупорядоченная m – выборка, составленная из элементов одного и того же множества

$$X = \{x_1, x_2, \ldots, x_m\}.$$

так, что в её состав входят m_1 элементов первого типа, m_2 элементов второго типа, ..., m_n элементов n -го типа и

$$m_1 + m_2 + \ldots + m_n = n. \tag{2.5.8}$$

Число сочетаний с повторениями обозначается $\hat{C}_n^m (m>n)$. Вывод формулы для $\hat{C}_n^m$ поясним на примере выбора 10 пирожных из 4 пирожных, например, эклеров, заварных, слоеных и фруктовых (m =10, n = 4). Представим один из возможных вариантов выбора 10 пирожных из четырёх типов в виде набора единиц (по числу пирожных) и нулей:

$$\left\{\underbrace{11}_{m_1}\,0\,\underbrace{1111}_{m_2}\,0\,\underbrace{11}_{m_3}\,0\,\underbrace{11}_{m_4}\right\}. \tag{2.5.9}$$

что означает выбор двух пирожных 1-го типа, четырех – 2-го, двух – 3-го типа и двух – 4-го типа.

Нули разделяют 4 группы единиц. Различным вариантам выбора будут соответствовать перестановки с повторениями (2.4.3) из 10 (m) единиц и 3 (n–1) нулей:

$$\hat{C}_4^{10} = \hat{P}(10{,}3) = \frac{13!}{3!\,10!} = 286$$

В общем случае неупорядочную выборку с повторениями можно записать по аналогии с перестановкой с повторениями в виде

$$\hat{P} = [m_1,\ m_2,\ \ldots,\ m_n], \tag{2.5.10}$$

где m_i (i=1...n) целые положительные числа, равные числам элементов множества X i -го типа так, что их сумма равна m_i (2.5.8). З аменяя каждое m_i группой из m_i единиц, а промежутки между ними заполняя нулями, получим выборку из (n+m – 1) элементов.
Число выборок равно $\hat{C}_n^m$ числу перестановок с повторениями (2.4.3) из m единиц и (n– 1) нулей:

$$\hat{C}_n^m = \hat{P}(n-1,\ m) = \frac{(n+m-1)!}{(n-1)!\ \ m!}. \tag{2.5.11}$$

Кроме того, формула (2.5.11) выражает число сочетаний без повторений (2.5.1) из (n+m– 1) элементов по m элементов:

$$\hat{C}_n^m = \frac{(n+m-1)!}{(n-1)!\,m!} = C_{n+m-1}^m.$$

Обобщим итоги типов комбинаций при выборке и результаты с целью удобства пользования сведем в таблицу (табл. 2.5.1 и табл. 2.5.2).

Схема подсчета числа комбинаций

Таблица 2.5.1

Комбинации без повторений			
Название комбинации	**Признак отличия**	**Пример**	**Формула подсчета числа комбинаций**
Размещения	Состав порядок	Размещения из 3-х элементов: a, b, c по 2: $A_3^2 = 2 \cdot 3 = 6$ $\{ab, bc, ca, ba, cb, ac\}$	$A_n^m = \dfrac{n!}{(n-m)!}$

Перестановки	Порядок	Перестановки из 3-х элементов: a, b, c: $P_n = 3! = 1 \cdot 2 \cdot 3 = 6$ $\{abc, bca, cba, cab, bac, acb\}$	$P_n = n!$
Сочетания	Состав	Сочетания из 3-х элементов: a, b, c по 2: $C_3^2 = 3$ $\{ab, ac, bc\}$	$C_n^m = \dfrac{n!}{m!\,(n-m)!}$

Таблица 2.5.2

Комбинации с повтореними			
Название комбинации	**Признак отличия**	**Пример**	**Формула подсчета числа комбинаций**
Размещения	Состав порядок	Размещения из 3-х элементов: a, b, c по 2 с повторениями: $\hat{A}_3^2 = 3^2 = 9$ $\{aa, ab, ba, bc, ac, ca, cc, bc, cb\}$	$\hat{A}_n^m = n^m$
Перестановки	Порядок	Перестановки из 4-х элементов: a, a, b, b: $\hat{P}(2,2) = \dfrac{4!}{2!\ 2!} = 2 \cdot 3 = 6$ $\{aabb, abba, bbaa, baab, abab, baba\}$	$\hat{P}(n_1,\ n_2,\ \dots,\ n_k) = \dfrac{n!}{n_1!\ n_2! \dots n_m!}$
Сочетания	Состав	Сочетания из 3-х элементов: a, b, c по 2 с повторениями: $\hat{C}_n^2 = \dfrac{4!}{2!\,2!} = 6$ $\{ab, bc, ca, aa, bb, cc\}$	$\hat{C}_n^m = \dfrac{(n+m-1)!}{(n-1)!\,m!} = C_{n+m-1}^m$

Примеры решения задач

Пример 14. В вазе с цветами 15 гвоздик: 5 белых и 10 красных. Из вазы наугад вынимают два цветка. Какова вероятность того, что эти цветки: а) оба белые; б) оба красные; в) разного цвета.

Решение:

а) Обозначим событие A={оба вынутых из вазы цветка белые}. Количество возможных способов взять 2 цветка из 15-ти равно C_{15}^{2}:

$$n = C_{15}^{2} = \frac{15!}{2! \cdot 13!} = \frac{14 \cdot 15}{1 \cdot 2} = 7 \cdot 15 = 105.$$

Количество возможных способов взять 2 белых цветка из 5-ти белых равно:

$$m = C_{5}^{2} = \frac{5!}{2! \cdot 3!} = \frac{4 \cdot 5}{1 \cdot 2} = 10.$$

Тогда по классической формуле (1.4.1) определения вероятности:

$$P(A) = \frac{m}{n} = \frac{10}{105} = 0{,}095$$

б) B={оба вынутых из вазы цветка красные}.

$$n = C_{15}^{2} = \frac{15!}{2! \cdot 13!} = \frac{14 \cdot 15}{1 \cdot 2} = 7 \cdot 15 = 105.$$

Количество возможных способов взять 2 красных цветка из 10-ти красных равно:

$$m = C_{10}^{2} = \frac{10!}{2! \cdot 8!} = \frac{9 \cdot 10}{1 \cdot 2} = 9 \cdot 5 = 45.$$

Тогда по формуле (1.4.1):

$$P(B) = \frac{m}{n} = \frac{45}{105} = 0{,}43.$$

в) C={цветы разного цвета, один белый, один красный}

$$n = C_{15}^{2} = \frac{15!}{2! \cdot 13!} = \frac{14 \cdot 15}{1 \cdot 2} = 7 \cdot 15 = 105.$$

Количество возможных способов взять один красный цветок из 10-ти красных «и» один белый цветок из 5-ти белых равно:

$$m = C_{10}^{1} \cdot C_{5}^{1} = 10 \cdot 5 = 50.$$

Тогда по формуле (1.4.1):

$$P(C) = \frac{50}{105} = 0{,}48.$$

Пример 15. Из 18 студентов, среди которых 5 отличников, формируются случайным образом две группы по 9 человек для сдачи зачета. Найти вероятности следующих событий: *A* – все отличники окажутся в какой-либо одной группе, *B* –два отличника окажутся в одной группе, а три – в другой группе.

Решение.

Общее число случаев *n* равно числу сочетаний из 18 студентов по 9:

$$n = C_{18}^{9}.$$

Число случаев *m*, благоприятствующих событию *A*, равно числу сочетаний из 5 отличников по 5 (все отличники – в одну группу), умноженному на число сочетаний из недостающих до 18 других 13 студентов по 4 (недостающих до состава группы из 9 человек), взятому дважды (отличники могут попасть в одну или другую группу):

$$m = 2C_5^5 C_{13}^4.$$

В итоге получаем:

$$P(A) = \frac{2 \cdot C_5^5 C_{13}^4}{C_{18}^9} = \frac{1}{34} = 0{,}03.$$

Число случаев *m* – благоприятствующих событию *B* равно:

$$m = C_5^2 \cdot C_{13}^7 + C_5^3 \cdot C_{13}^6,$$

тогда

$$P(B) = \frac{C_5^2 \cdot C_{13}^7 + C_5^3 \cdot C_{13}^6}{C_{18}^9} = \frac{12}{17} = 0{,}71.$$

Пример 1.6. В лифт десятиэтажного дома на первом этаже вошли 3 человека, каждый из которых с одинаковой вероятностью может выйти из лифта на любом этаже (со второго по десятый). Найти вероятность следующих событий:
A – все люди выйдут на 5-м этаже, *B* – все люди выйдут на одном и том же этаже, *C* – люди выйдут на разных этажах.

Решение.

Общее число случаев *n* равно числу размещений с повторениями из 9 по 3:

$$n = \hat{A}_9^3 = 9^3 = 729.$$

Число случаев, благоприятствующих событиям *A, B* и С, равны:

$m_A = 1$, $m_B = 9$, а m_C равно числу способов (сочетаний) распределить трех пассажиров лифта по девяти этажам (кроме первого):

$$m_C = C_9^3 = \frac{9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3} = 84.$$

Окончательно получим:

$$P(A) = \frac{1}{729} = 0{,}0014,$$

$$P(B) = \frac{9}{729} = 0{,}0123, \qquad P(C) = \frac{84}{729} = 0{,}1152.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что изучает комбинаторика?
2. Что вы понимаете под упорядоченной выборкой?
3. А что такое неупорядоченная выборка?
4. В чем суть правила произведения?
5. Что такое правило суммы в комбинаторике?
6. Что такое размещения без повторений? В чем особенность размещений без повторений?
7. Что такое размещения с повторениями? В чем их особенность?
8. Что такое перестановки без повторений? В чем особенность перестановок без повторений?
9. Что такое перестановки с повторениями? В чем их особенность?
10. Что такое сочетания без повторений? В чем особенность сочетаний без повторений?
11. Что такое сочетания с повторениями? В чем их особенность?
12. Какова роль комбинаторики в теории вероятностей?
13. Какие свойства сочетаний вы знаете?
14. Чему равно ?
15. Чему равно ?

ЗАДАНИЯ

2.1. Вычислите: а) $P_5, A_5^4, C_5^4, \hat{A}_5^4, \hat{C}_5^4$; б) $P_4, A_6^4, C_6^4, \hat{A}_6^4, \hat{C}_6^4$.

2.2. Вычислите: а) $P_6, A_7^3, C_7^3, \hat{A}_7^3, \hat{C}_7^3$; б) $P_8, A_8^5, C_8^5, \hat{A}_8^5, \hat{C}_8^5$.

2.3. Решить уравнение $C_n^5 = 2C_{n\text{-}1}^5$.

2.4. Возможно ли равенство P_n=36 $A_{n\text{-}1}^2$ и если да, то при каком n?

2.5. Решить уравнение: а)$xA_5^2 = C_5^4$; б)$xP_3 = C_6^4$.

2.6. Из 14 слов мужского рода, 8 женского и 10 среднего надо выбрать по одному слову каждого рода. Сколькими способами может быть сделан этот выбор?

2.7. Комиссия состоит из председателя, его заместителя и еще шести человек. Сколькими способами члены комиссии могут распределить между собой обязанности?

2.8. Сколькими способами можно указать на шахматной доске два квадрата: белый и черный? Если нет ограничений на цвет выбранных квадратов?

2.9. Сколькими способами можно расставить шесть книг разных авторов на полке в один ряд?

2.10. Сколькими способами можно рассадить 6 гостей за круглым столом?

2.11. Автомобильные номера состоят из одной, двух или трех букв и четырех цифр. Найти число таких номеров, если используются 28 букв русского алфавита.

2.12. У англичан принято давать детям несколько имен. Сколькими способами можно назвать ребенка, если общее число имен равно 300, а ему дают не более трех имен?

2.13. В букинистическом магазине лежат 6 экземпляров романа И. С. Тургенева «Рудин», 3 экземпляра его же романа «Дворянское гнездо» и 4 экземпляра романа «Отцы и дети». Кроме того, есть 5 томов, содержащих романы «Рудин» и «Дворянское гнездо», 7 томов, содержащих романы «Дворянское гнездо» и «Отцы и дети», и 3 тома, в которые входят «Рудин» и «Отцы и дети». Сколькими способами можно сделать покупку, содержащую по одному экземпляру каждого из этих романов?

2.14. Брокерская фирма предлагает акции различных компаний. Акции 10 из них продаются по наименьшей среди имеющихся акций цене и обладают одинаковой доходностью. Клиент собирается приобрести акции 3 таких компаний по 1 от каждой компании. Сколько существует способов выбора 3 таких акций из 10, если выбор осуществляется в случайном порядке?

2.15. Покупая карточку лотереи «Спортлото», игрок должен зачеркнуть 5 из 36 возможных чисел от 1 до 36. Если при розыгрыше тиража лотереи он угадает все 5 чисел, то имеет шанс выиграть значительную сумму денег. Сколько возможных комбинаций можно составить из 36 по 5, если порядок чисел безразличен?

2.16. В ящике находятся 3 синих и 4 красных карандаша. Наудачу одновременно извлекаются 3 карандаша. Найти:

а) вероятность того, что все 3 карандаша красные;

б) вероятность того, что красных карандашей больше, чем синих.

2.17. Решить уравнение: а) $xA_6^2 = C_6^4$; б) $xP_2 = C_6^4$.

2.18. Среди 6 телевизоров, находящихся в продаже, 4 требуют дополнительной настройки. Найти вероятность того, что из 2 купленных телевизоров 1 будет требовать дополнительной настройки.

2.19. На девяти карточках напечатана одна из следующих букв: О, 2 буквы В, Т, С, Н, Е, А, Р. Вынимая последовательно из ящика карточки наугад и выкладывая их в одну линию, можно получить слово «РАВЕНСТВО». Какова вероятность этого события?

2.20. В шахматном турнире участвуют 20 человек, которые будут по жребию распределены в две группы по 10 человек. Какое событие более вероятно: два наиболее сильных участника окажутся в разных группах; два наиболее сильных участника окажутся в одной группе?

2.21. На пяти одинаковых карточках написаны буквы Т, У, А, М, Б. Они тщательно перемешаны. Извлекаются наудачу поочередно:

а) по одной букве все пять;

б) три буквы и укладываются слева направо.

Какова вероятность того, что будет получено слово:

а) «ТУМБА»;

б) «БУМ»?

2.22. Что вероятнее выиграть у равносильного соперника по шахматам: четыре партии из шести или три партии из пяти?

2.23. В автобусе было 5 девушек и 10 юношей. На остановке вышли 4 человек. Найдите вероятность того, что среди них две девушки и двое юношей.

2.24. В ящике имеются 40 изделий первого сорта и 15 — высшего сорта. Из ящика наудачу берут одно за другим три изделия. Найти вероятность того, что все изделия окажутся высшего сорта

2.25. В кабинете работают 6 мужчин и 4 женщины. Для переезда наудачу отобраны 7 человек. Найти вероятность того, что среди отобранных лиц три женщины.

2.26. Для получения выигрыша в спортлото необходимо правильно указать 6 номеров из 49. Какова вероятность выигрыша при покупке одного билета?

2.27. В урне 4 белых и 6 черных шаров. Какова вероятность извлечения из урны двух белых шаров?

2.28. В партии из 50 деталей 6 бракованных. Из партии произвольно выбираются 4 детали. Найти вероятность того, что среди отобранных деталей буде ровно 2 бракованных.

2.29. В партии из деталей имеется стандартных. Наудачу отобраны деталей. Найти вероятность того, что среди отобранных деталей ровно стандартных.

2.30. В автобусе было 4 девушки и 6 юношей. На остановке вышли 6 человек. Найдите вероятность того, что среди них трое девушек и трое юношей.

ГЛАВА 3
ОСНОВНЫЕ ТЕОРЕМЫ И ФОРМУЛЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

3.1. Условные вероятности и условные относительные частоты события

Условной вероятностью будем называть вероятность события A, вычисленную при условии, что произошло другое событие B, связанное с A. Условная вероятность обозначается $P(A|B)$ и вычисляется по формуле:

$$P(A|B) = \frac{P(AB)}{P(B)}, \qquad P(B) \neq 0. \tag{3.1.1}$$

Соответственно условная вероятность $P(A|B)$ – по формуле:

$$P(B|A) = \frac{P(AB)}{P(A)}, \quad P(A) \neq 0. \tag{3.1.2}$$

Условная вероятность лежит в пределах $0 \leq P(A|B) \leq 1$.

Рассмотрим примеры определения условной вероятности.

Пример 1. Производится одновременное бросание трёх симметричных монет на поверхность стола.

Рассматриваются события: A – появление одного герба, B – появление нечетного числа гербов.

Определим условную вероятность $P(A|B)$.

Решение. Обозначим $Г$ и $Ц$ события, состоящие в появлении герба и цифры соответственно. Тогда полное пространство исходов опытов запишется в виде:

$$\Omega = \{ГГГ,\ ГГЦ,\ ГЦГ,\ ЦГГ,\ ГЦЦ,\ ЦГЦ,\ ЦЦГ,\ ЦЦЦ\}.$$

Каждый исход опыта представлен в виде трех букв, каждая из которых показывает исход опыта для первой, второй и третьей монеты соответственно.

Исходы опыта в совокупности составляют схему случаев, поэтому вероятности $P(AB)$ и $P(B)$ вычисляются по формуле (1.4.1):

$$P(B) = \frac{4}{8} = \frac{1}{2}, \qquad P(AB) = \frac{3}{8}.$$

Искомая вероятность $P(A|B)$ (3.1.1) равна:

$$P(A|B) = \frac{3 \cdot 2}{8 \cdot 1} = \frac{3}{4}.$$

Пример 2. Из урны, содержащей 2 белых и один черный шар, вынимают по одному шару. Рассматриваются события: A – появление белого шара, B – появление черного шара. Определим вероятности $P(A)$, $P(B)$, $P(A|B)$ и $P(B|A)$.

Решение. Исходы опыта образуют схему случаев. Поэтому искомые вероятности равны:

$$P(A) = \frac{2}{3}, \quad P(B) = \frac{1}{3}, \quad P(A|B) = \frac{2}{2} = 1, \quad P(B|A) = \frac{1}{2}.$$

Аналогично введем понятие условной относительной частоты.

Условной относительной частотой $P^*(A|B)$ события A при наличии другого события B, связанного с A, называется относительная частота события A, вычисленная не по совокупности всех n опытов, а лишь по совокупности только тех опытов, в которых вместе с A произошло событие B.

Формулы вычисления условных относительных частот:

$$P^*(A|B) = \frac{P^*(AB)}{P^*(B)}, \quad P^*(B) \neq 0. \qquad (3.1.3)$$

$$P^*(B|A) = \frac{P^*(AB)}{P^*(A)}, \quad P^*(A) \neq 0. \qquad (3.1.4)$$

3.2. Зависимые и независимые события

Событие A не зависит от события B, если появление в результате опыта события B не изменяет вероятности появления события A:

$$P(A|B) = P(A). \qquad (3.2.1)$$

Иначе говоря, если A не зависит от B, то условная вероятность $P(A|B)$ равна безусловной вероятности $P(A)$.

Преобразуем формулы (3.1.1) и (3.12):

$$P(AB) = P(A)P(B|A) = P(B)P(A|B). \qquad (3.2.2)$$

Подставим (3.2.1) в (3.2.2):

$$P(A)P(B|A) = P(B)P(A),$$

откуда следует:

$$P(B|A) = P(B), \tag{3.2.3}$$

то есть событие B не зависит от события A.

Следовательно, свойство независимости событий A и B взаимно, если не зависит от B, то и B не зависит от A:

$$\begin{cases} P(A|B) = P(A), \\ P(B|A) = P(B). \end{cases} \tag{3.2.4}$$

Понятие независимости событий играет значительную роль в теории вероятностей и её приложениях. В задачах практики для проверки независимости событий обычно пользуются интуитивными соображениями и редко обращаются к равенствам (3.2.4).

События A, B называются зависимыми, если появление одного из них в результате опыта изменяет вероятность появления другого события:

$$\begin{cases} P(A|B) \neq P(A), \\ P(B|A) \neq P(B). \end{cases} \tag{3.2.5}$$

Рассмотрим примеры независимых и зависимых событий.

Пример 3. Бросаются подряд две монеты. Появление герба на одной монете (событие A) и появление цифры на второй (событие B) – независимые события.

Пример 4. Вынимание шаров из урны с возвращением.

Из урны, содержащей одинаковые белые и черные шары (6 белых и 4 черных), вынимают один шар и кладут его обратно. Затем из урны вынимают еще один шар. Рассмотрим событие: A – появление белого шара при первом вынимании, B – появление белого шара при втором вынимании. Их вероятности, очевидно, равны:

$$P(A) = \frac{6}{10}, \qquad P(B|A) = \frac{6}{10} = P(B).$$

В данном примере события A и B независимы, так как возврат в урну вынутого шара восстанавливает первоначальный состав шаров – 6 белых и 4 черных.

Пример 5. Вынимание шаров из урны без возвращения.

При условиях примера 1.5.2 вынутый шар оказывается белым (событие A), но в урну не возвращается.

Вероятность события B – появление белого шара при втором вынимании (при условии, что вынутый первым шар не возвращен в урну) изменяется, так как белых шаров стало 5 вместо 6-ти:

$$P(B|A) = \frac{5}{9} \neq \frac{6}{10}.$$

Следовательно, события A и B в данном примере зависимы, так как появление A изменяет вероятность события B.

3.3. Теорема сложения вероятностей (относительных частот) несовместных событий

Теорема сложения вероятностей двух несовместных событий: вероятность суммы двух несовместных событий равна сумме вероятностей этих событий.

$$P(A + B) = P(A) + P(B). \quad AB = \emptyset. \qquad (3.3.1)$$

Доказательство проведем для схемы случаев (несовместных, равновозможных событий, образующих полную группу событий).

Пусть событию A благоприятствуют m элементарных исходов, а событию B – соответственно k исходов. Так как события A и B по условию теоремы несовместны, то событию $A + B$ благоприятствуют $m+k$ элементарных исходов из общего числа n исходов. Следовательно:

$$P(A + B) = \frac{m + k}{n} = \frac{m}{n} + \frac{k}{n} = P(A) + P(B),$$

где $P(A)$ – вероятность события A; $P(B)$ – вероятность события B.

Теорема доказана. Для случайных событий, которые не укладываются в схему случаев, теорема сложения вероятностей принимается за аксиому и называется законом сложения вероятностей событий.

Замечание. Теорему можно было строго доказать и для относительных частот суммы двух несовместных событий и на основании того, что при неограниченном возрастании числа опытов относительная частота события сходится по вероятности к вероятности события (см. п. 1.5 и 9.5.), и сделать вывод, что теорема сложения относительных частот справедлива и для вероятностей событий:

$$P^*(A + B) = P^*(A) + P^*(B). \qquad (3.3.2)$$

Теорема сложения вероятностей (относительных частот) справедлива для любого конечного числа несовместных событий : $A_1, A_2, \ldots, A_n$.

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \qquad A_iA_j = \emptyset, \;\; i \neq j, \qquad (3.3.3)$$

$$P^*\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P^*(A_i) \quad A_iA_j = \emptyset, \ i \neq j. \tag{3.3.4}$$

3.4. Теоремы умножения вероятностей и относительных частот

Во многих случаях вероятности появления одних событий зависят от того, произошло другое событие или нет. В этом случае говорят об условной вероятности события. Как было отмечено в пункте 3.1., вероятность события A, вычисленная при условии, что имело место другое событие B, называется условной вероятностью события A и обознается $P(A|B)$.

Теорема умножения вероятностей для двух событий: вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого события, вычисленную при условии, что произошло первое событие:

$$P(AB) = P(A)P(B|A), \tag{3.4.1}$$

$$P(AB) = P(B)P(A|B) \tag{3.4.2}$$

Доказательство. Предположим, что из n возможных элементарных исходов событию A благоприятствуют m исходов, из которых k благоприятствуют событию B, тогда вероятность события A будет равна:

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

Условная вероятность события B относительно события A будет равна:

$$P(B|A) = \frac{k}{m}.$$

Произведению событий A и B благоприятствуют только те исходы, которые благоприятствуют и событию A и событию B одновременно, то есть k исходов.

Следовательно, вероятность произведения событий A и B:

$$P(AB) = \frac{k}{n}.$$

Умножим числитель и знаменатель последней формулы на m получим:

$$P(AB) = \frac{k}{n}\frac{m}{m} = \frac{m}{n}\frac{k}{m} = P(A)P(B|A).$$

Таким же способом доказывается и формула:

$$P(AB) = P(B)P(A|B).$$

Доказанная теорема справедлива и для произведения относительных частот двух событий: относительная частота произведения двух событий равна произведению относительной частоты одного из них на условную относительную частоту другого события, вычисленную при условии, что произошло первое событие:

$$P^*(AB) = P^*(A) \cdot P^*(B|A), \tag{3.4.3}$$

$$P^*(AB) = P^*(B) \cdot P^*(A|B) \tag{3.4.4}$$

Распространим теорему умножения вероятностей событий на несколько событий $A_1, A_2, \ldots, A_n$:

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2A_3\ldots A_{n-1}). \tag{3.4.5}$$

Аналогично для относительных частот событий:

$$P^*(A_1A_2\ldots A_n) = P^*(A_1)P^*(A_2|A_1)P^*(A_3|A_1A_2)\ldots P^*(A_n|A_1A_2A_3\ldots A_{n-1}). \tag{3.4.6}$$

Порядок нумерации событий указывается произвольно.

Пример 6. В урне имеется 5 одинаковых на ощупь шаров, из них 2 белых и 3 черных шара. Из урны наугад вынимают один шар, а затем другой шар. Какова вероятность того, что оба шара будут черными?

Решение. Обозначим A – появление двух черных шаров, A_1 – появление черного шара при первом вынимании, A_2 – появление черного шара при втором вынимании шаров. Тогда можно записать выражение для события A как произведение событий A_1 и A_2: $A = A_1A_2$.

События A_1 и A_2 зависимы, поэтому по теореме умножения (3.4.1) получим:

$$P(A) = P(A_1)P(A_2|A_1) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10} = 0{,}30.$$

Событие A_2 зависит от A_1, так как появление A_1 изменяет условия опыта, (шар не возвращается) и, следовательно, изменяется вероятность события A_2.

Если события A и B независимы (3.2.4), то закон умножения вероятностей принимает следующий вид (3.4.1):

$$P(AB) = P(A) \cdot P(B), \tag{3.4.7}$$

теорема умножения относительных частот (3.4.3), для двух независимых событий запишется в виде:

$$P^*(AB) = P^*(A) \cdot P^*(B). \tag{3.4.8}$$

Пример 7. Рассмотрим условия первого примера, но после вынимания первого шара его возвращают обратно, восстанавливая при этом начальные условия опыта. Тем самым события A_1 и A_2 – появление черного шара при первом и втором выниманиях шаров из урны, содержащей 2 белых и 3 черных шара, – становятся независимыми.

Решение. По формуле (3.4.7) имеем:

$$P(A) = P(A_1A_2) = P(A_1)P(A_2) = \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{25} = 0{,}36.$$

Если события $A_1, A_2, \ldots, A_n$ – независимы (3.2.4), то выражения (3.4.5) и (3.4.6) для вероятности произведения и относительной частоты n событий запишутся в следующем виде:

$$P\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i); \tag{3.4.9}$$

$$P^*\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P^*(A_i). \tag{3.4.10}$$

Таким образом, вероятность (относительная частота) произведения n независимых событий равна произведению вероятностей (относительных частот) этих событий.

Пример 8. Три стрелка с разным уровнем подготовки независимо друг от друга стреляют по одному разу по мишеням. Вероятности попадания в мишень для стрелков соответственно равны $p_1 = 0{,}90$, $p_2 = 0{,}60$, $p_3 = 0{,}50$. Какова вероятность того, что все три стрелка попадут в мишень, сделав по одному выстрелу?

Решение. Обозначим A – три стрелка попали в мишень, A_1 – первый стрелок попал в мишень, A_2 – второй, A_3 – третий стрелок попали в мишень. Тогда событие A выразится через A_1, A_2, A_3 в виде произведения:

$$A = A_1 A_2 A_3.$$

Вероятность события в соответствии с (3.4.9) – события A_1, A_2, A_3 – независимы – равна:

$$P(A)=P(A_1)P(A_2)P(A_3)=p_1\,p_2\,p_3=0{,}90\cdot 0{,}60\cdot 0{,}50=0{,}27.$$

3.5. Теорема сложения вероятностей (относительных частот) совместных событий

Как указывалось выше, теорема сложения справедлива только для несовместных событий. В случае, когда события совместны, мы условились рассматривать вместо суммы событий их объединение.

Теорема: Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий, без вероятности их совместного появления:

$$P(A+B)=P(A)+P(B)-P(AB). \qquad (3.5.1)$$

Доказательство: Событие $A+B$ наступит, если наступит одно из трех несовместных событий $A\overline{B}$, $\overline{A}B$, AB. По теореме сложения вероятностей несовместных событий:

$$P(A+B)=P\left(A\overline{B}\right)+P\left(\overline{A}B\right)+P(AB). \qquad (3.5.2)$$

Событие A произойдет, если наступит одно из двух несовместных $A\overline{B}$, AB. Снова применим теорему сложения вероятностей несовместных событий, в результате получаем:

$$P(A)=P\left(A\overline{B}\right)+P(AB), \qquad (3.5.3)$$

откуда:

$$P\left(A\overline{B}\right)=P(A)-P(AB). \qquad (3.5.4)$$

Подобным образом для события B получаем:

(3.5.5)

Подставим формулы (3.5.4) и (3.5.5) в (3.5.2):

$$P(A+B)=P(A)-P(AB)+P(B)-P(AB)+P(AB)\Rightarrow$$

$$P(A+B)=P(A)+P(B)-P(AB) \qquad (3.5.6)$$

Теорема доказана.

Пример 9. Производится бросание двух монет. Рассматриваются события:
A – появление хотя бы одного герба;
B – появление хотя бы одной цифры;
AB – появление одного герба и одной цифры.
$C = A \cup B$.

Найти вероятность события C – появление любого исхода опыта.

Решение. Вероятности событий A, B и AB равны:

$$P(A) = \frac{3}{4}, \quad P(B) = \frac{3}{4}, \quad P(AB) = \frac{2}{4} = \frac{1}{2}.$$

Вероятность объединения событий A и B равна:

$$P(C) = P(A \cup B) = P(A) + P(B) - P(AB) = \frac{3}{4} + \frac{3}{4} - \frac{1}{2} = 1,$$

откуда следует, что событие C – достоверно, $C=\Omega$.

3.6. Следствия теоремы сложения

Теорема (закон) сложения вероятностей имеет два важных для практики решения вероятностных задач следствия, связанных с суммой несовместных событий (исходов опыта), образующих полную группу, иначе говоря, с полным пространством исходов опыта.

Следствие 1. Сумма вероятностей несовместных событий A_1, A_2, ..., An, образующих полную группу, равна единице:

$$\sum_{i=1}^{n} P(A_i) = 1, \quad A_i A_j = \emptyset, \quad i, j = 1 \ldots n; \quad i \neq j. \qquad (3.6.1)$$

Доказательство. Сумма событий A_1, A_2, ..., An, образующих полную группу (полное пространство исходов опыта), – событие достоверное, так как в результате опыта должно обязательно произойти хотя бы одно из них:

$$\sum_{i=1}^{n} A_i = \Omega,$$

а вероятность суммы и вероятность достоверного события, следовательно, равны единице:

$$P\left(\sum_{i=1}^{n} A_i\right) = P(\Omega) = 1.$$

С другой стороны, согласно закону сложения вероятностей (3.3.3), вероятность суммы несовместных событий равна сумме вероятностей этих событий:

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = 1.$$

Следствие 2. Сумма вероятностей противоположных событий равна единице:

$$P(A) + P(\overline{A}) = 1. \qquad (3.6.2)$$

Доказательство. Следствие является частным случаем доказанного выше первого следствия при n=2 (для двух событий). Два несовместных события, образующих полную группу, называются противоположными событиями. Для противоположных событий справедливы соотношения:

$$\begin{cases} A + \overline{A} = \Omega, \\ A \cdot \overline{A} = \emptyset. \end{cases} \qquad (3.6.3)$$

Вероятность их суммы равна единице:

$$P(A + \overline{A}) = P(\Omega) = 1. \qquad (3.6.4)$$

а по теореме сложения:

$$P(A + \overline{A}) = P(A) + P(\overline{A}). \qquad (3.6.5)$$

Совмещая (3.6.4) и (3.6.5), получаем доказательство следствия 2 (3.6.2).

Данное следствие весьма часто используется при решении вероятностных задач в такой форме:

$$P(A) = 1 - P(\overline{A}). \qquad (3.6.6)$$

Практически это означает, что вероятность события A можно вычислить через вероятность его противоположного события $\overline{A}$, которая в большинстве случаев решения задач определяется гораздо проще, чем вероятность события A.

Рассмотрим пример практического применения второго следствия теоремы сложения вероятностей.

Пример 10. Производится бросание 3-х монет. Какова вероятность того, что герб выпадет хотя бы один раз?

Решение. Рассмотрим пространство исходов этого опыта с точки зрения выпадения гербов на 3-х монетах. Очевидно, что число гербов может быть равно 0, 1, 2 или 3. Искомая вероятность события – "хотя бы один герб появился при трех бросаниях монеты" (событие A) – равна сумме вероятностей трех из четырех возможных исходов опыта A_1, A_2, A_3 (A_k– появление k гербов, k=1, 2, 3):

$$P(A) = P(A_1) + P(A_2) + P(A_3).$$

Противоположным событию A является событие A_0 – ни одного герба не выпало при бросании трех монет. Его вероятность равна:

$$P(A_0) = q^3 = 0{,}50^3 = 0{,}125,$$

где $q = 1 - p = 1 - 0{,}50 = 0{,}50$, $p = 0{,}50$ – вероятность появления герба при одном бросании монеты.

Таким образом, искомая вероятность $P(A)$ вычисляется, в конечном счете, по формуле:

$$P(A) = 1 - P(\bar{A}) = 1 - P(A_0) = 1 - (1 - p)^3,$$

$$P(A) = 1 - 0{,}125 = 0{,}875.$$

Преимущество решений примеров через противоположные события – очевидно.

3.7. Вероятность появления события хотя бы один раз в нескольких независимых опытах

Рассмотрим пример на случай независимых опытов, осуществляемых в изменяющихся или одинаковых условиях.

Пусть производится независимых опытов, условия которых изменяются от опыта к опыту. Например, стрелок, производящий стрельбу по мишени из спортивного пистолета, после каждого выстрела делает один шаг по направлению к мишени. Очевидно, вероятность попадания в мишень с каждым выстрелом растет.

Обозначим A_i– попадание в мишень при м выстреле, i=1...n, $\overline{A}_i$ – промах при м выстреле, – хотя бы одно попадание при выстрелах, – вероятность хотя бы одного попадания при выстрелах, – ни одного попадания в мишень при   выстрелах. Тогда событие может быть представлено в виде:

$$B = A_1\overline{A}_2\ldots\overline{A}_n + \overline{A}_1 A_2\overline{A}_3\ldots\overline{A}_n + \ldots + \overline{A}_1\overline{A}_2\ldots\overline{A}_{n-1}A_n + A_1A_2\overline{A}_3\ldots\overline{A}_n +$$

$$+ A_1\overline{A}_2A_3\overline{A}_4\ldots\overline{A}_n + \ldots + \overline{A}_1\overline{A}_2\ldots\overline{A}_{n-2}A_{n-1}A_n + \ldots + A_1A_2\ldots A_n \tag{3.7.1}$$

Противоположное событие $\overline{B}$ может быть представлено в виде:

$$\overline{B} = \overline{A}_1\overline{A}_2\ldots\overline{A}_n, \tag{3.7.2}$$

а его вероятность, очевидно, равна:

$$P(\overline{B}) = P\left(\prod_{i=1}^{n}(\overline{A}_i)\right) = \prod_{i=1}^{n}P(\overline{A}_i). \tag{3.7.3}$$

Обозначим вероятность появления события A_i в -м опыте – попадание i-м выстреле (i=1...n) – через $p_i = P(A_i)$, а вероятность промаха – появление события $\overline{A}$ – через:

$$q_i = 1 - p_i = P(\overline{A}_i) = 1 - P(A_i), \quad i = 1,2,\ldots,n\,.$$

Тогда выражение (3.7.3) можно представить в виде:

$$P(\overline{B}) = \prod_{i=1}^{n}q_i = \prod_{i=1}^{n}(1 - p_i). \tag{3.7.4}$$

Из (3.6.6) следует:

$$R_{1,n} = P(B) = 1 - P(\overline{B}), \tag{3.7.5}$$

откуда

$$R_{1,n} = 1 - \prod_{i=1}^{n}q_i = 1 - \prod_{i=1}^{n}(1 - p_i). \tag{3,7,6}$$

В случае, когда условия опытов не изменяются (например, стрельба по мишени производится с одной и той же позиции):

$$\begin{cases} P(A_i) = P(A) = p, \\ P(\bar{A}_i) = P(\bar{A}) = q = 1 - p. \end{cases} \tag{3.7.7}$$

Подставляя полученные результаты в (3.7.6), окончательно получим:

$$R_{1,n} = 1 - q^n = 1 - (1-p)^n. \qquad (3.7.8)$$

Формула (3.7.6) используется при вычислении вероятности $R_{1,n}$ – появление события хотя бы один раз в n независимых опытах, осуществлявшихся в изменяющихся от опыта к опыту условиях. Формула (3.7.8) дает возможность вычислять вероятность $R_{1,n}$ появления события хотя бы один раз в неизменяющихся (постоянных) условиях проведения n независимых опытов.

Пример 11. Трое рабочих при изготовлении некоторого изделия работают над ним последовательно и независимо один от другого, при этом качество изделия при передаче изделия следующему рабочему не проверяется. Первый рабочий допускает брак с вероятностью p_1=0,10, второй – с вероятностью p_2=0,08, третий – с вероятностью p_3=0,05. Какова вероятность того, что при изготовлении изделия будет допущен брак?

Решение. Обозначим A – допущен брак при изготовлении изделия тремя рабочими. Очевидно, событие произойдет, если брак допустит хотя бы один из рабочих. Тогда по формуле (3.7.6) получим:

$$P(A) = R_{1,3} = 1 - (1-p_1)(1-p_2)(1-p_3) = 1 - 0{,}90 \cdot 0{,}92 \cdot 0{,}95 = 0{,}21$$

Пример 12. При одном цикле обзора радиолокационной станции (РЛС) слежения за космическими объектами объект обнаруживается с вероятностью p=0,80. Обнаружение объекта в каждом цикле обзора РЛС происходит независимо от других циклов. За время нахождения объекта в зоне обзора РЛС она успевает сделать 5 циклов. Какова вероятность обнаружения объекта?

Решение. Обозначим B – обнаружение объекта за n циклов обзора РЛС (n=5). Очевидно, что при обнаружении объекта хотя бы в одном цикле обзора произойдет событие B. По формуле (3.7.8) вычислим искомую вероятность

$$P(B) = R_{1,5} = 1 - (1-p)^5 = 1 - 0{,}20^5 = 0{,}99968 \cdot$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какова формулировка теоремы сложения вероятностей для несовместных событий?
2. Как вычислить вероятность противоположного события?
3. Какая вероятность называется условной?
4. Какова формулировка теоремы умножения вероятностей?
5. Какие события называются независимыми?

6. Какие события называются зависимыми?
7. Как формулируется теорема умножения вероятностей для зависимых событий?
8. Чему равна сумма вероятностей противоположных событий?
9. Как формулируется теорема умножения вероятностей для независимых событий?
10. Как формулируется обобщенная теорема умножения вероятностей?

ЗАДАНИЯ

3.1. Для некоторой местности среднее число пасмурных дней в июле равно 6. Требуется найти вероятность того, что первого и второго июля будет ясная погода.

3.2. Вероятность того, что наудачу выбранное изделие из партии будет высшего качества равно 0,8. Найти вероятность того, что из трех взятых изделий два высшего сорта.

3.3. Для сигнализации об аварии установлены три независимо работающих устройства. Вероятности того, что устройство сработает при аварии, равны 0,95; 0,9; 0,85 соответственно. Какова вероятность того, что: а) при аварии сработает ровно два устройства; б) сработает хотя бы одно устройство.

3.4. Контроллер проверяет изделия на соответствие стандарту. Известно, что вероятность соответствия стандарту изделий равна 0,79. Какова вероятность того, что из двух проверенных изделий: а) оба будут стандартными; б) ровно одно будет стандартным; в) хотя бы одно будет стандартным.

3.5. Отдел технического контроля проверяет изделие на стандартность. Вероятность того, что изделие стандартно равно 0,9. Найти вероятность того, что из двух проверенных изделий только одно стандартное.

3.6. Два стрелка стреляют по мишени. Вероятность попадания в мишень для первого стрелка равна 0,7, а для второго стрелка равна 0,8. Найти вероятность того, что при одном залпе попадет только один стрелок.

3.7 Вероятность правильного оформления накладной при передаче продукции равна 0,85. Какова вероятность того, что из трех накладных: а) только две оформлены верно; б) хотя бы одна оформлена верно.

3.8. Вероятности своевременного выполнения задания тремя независимо работающими предприятиями соответственно равны 0,7; 0,8; 0,9. Найти вероятность своевременного выполнения задания: а) ровно одним предприятием из трех; б) хотя бы одним предприятием.

3.9. Устройство содержит два независимо работающих элемента. Вероятности отказов элементов равны 0,1 и 0,09 соответственно. Какова вероятность: а) отказа прибора, если для этого достаточно, чтобы отказал хотя бы один элемент; б) отказа ровно одного элемента из двух; в) отказа двух элементов.

3.10. Студент знает 20 из 25 вопросов программы. Найти вероятность того, что студент знает все три вопроса, предложенные экзаменатором.

3.11. В коробке содержится 6 одинаково занумерованных кубиков. Наудачу из коробки извлекают по одному кубику. Найти вероятность того, что номера извлеченных кубиков появятся в возрастающем порядке.

3.12. Для сигнализации об аварии установлены три независимо работающих устройства. Вероятности того, что устройство сработает при аварии, равны 0,8; 0,75; 0,65 соответственно. Какова вероятность того, что: а) при аварии сработает ровно два устройства; б) сработает хотя бы одно устройство.

3.13. При подготовке к экзамену студент успел выучить 25 вопросов из 30 предложенных преподавателем. В каждом билете по три вопроса. Найти вероятность того, что в билете, наудачу взятом студентом: а) студент знает все три вопроса; б) студент знает ответы только на два вопроса; в) студент знает ответ только на один вопрос; г) студент не знает ответов ни на один вопрос.

3.14. Контроллер проверяет изделия на соответствие стандарту. Известно, что вероятность соответствия стандарту изделий равна 0,85. Какова вероятность, что из двух проверенных изделий: а) оба будут стандартными; б) ровно одно будет стандартным; в) хотя бы одно будет стандартным.

3.15. Вероятность правильного оформления накладной при передаче продукции равна 0,8. Какова вероятность того, что из трех накладных: а) только две оформлены верно; б) хотя бы одна оформлена верно.

3.16. Вероятность того, что изготовленная на станке деталь стандартная равна 0,95. Из партии этих деталей взято три детали. Какова вероятность того, что: а) все детали стандартные; б) все детали бракованные; в) хотя бы одна деталь бракованная.

3.17. Магазин получил партию костюмов, где 90% первого сорта. Какова вероятность того, что из трех взятых наудачу костюмов два будут первого сорта?

3.18. Вероятности своевременного выполнения задания тремя независимо работающими предприятиями соответственно равны 0,7; 0,6; 0,5. Какова вероятность своевременного выполнения задания: а) ровно одним предприятием из трех; б) хотя бы одним предприятием.

3.19. Устройство содержит два независимо работающих элемента. Вероятности отказов элементов равны 0,04 и 0,02 соответственно. Какова вероятность: а) отказа прибора, если для этого достаточно, чтобы отказал хотя бы один элемент; б) отказа ровно одного элемента из двух; в) отказа двух элементов.

3.20. Для сигнализации об аварии установлены три независимо работающих устройства. Вероятности того, что устройство сработает при аварии, равны 0,75; 0,85; 0,9 соответственно. Какова вероятность того, что: а) при аварии сработает ровно два устройства; б) сработает хотя бы одно устройство.

3.21. Вероятность попадания цели в зону обзора РЛС равна 0,5. Вероятность обнаружения цели оператором РЛС в зоне обзора равна 0,9. Вероятность поражения цели при ее обнаружении равна 0,8. Найти вероятность поражения цели.

3.22. Из трех орудий произведен залп по цели. Вероятность попадания в цель первым орудием равна о,6, вторым орудием 0,5, третьим орудием 0,8. Цель может быть поражена только при двух попаданиях. Найти вероятность поражения цели.

3.23 Студент разыскивает нужную ему формулу в трех справочниках. Вероятности того, что формула содержится в первом, втором и третьем справочниках равны 0,6; 0,7 и 0,8. Найти вероятности того, что формула содержится: а) только в одном справочнике; б) только в двух справочниках; в) во всех трех справочниках.

3.24. Мастер обслуживает 4 станка, работающих независимо друг от друга. Вероятность того, что первый станок в течении смены потребует внимания рабочего, равна 0,3, второго – 0,6, третьего – 0,4 и четвертого – 0,25. Найти вероятность того, что в течении смены хотя бы один станок потребует внимания рабочего.

3.25. Вероятность своевременного выполнения студентом контрольной работы по каждой из трех дисциплин равна соответственно 0,6, 0,5 и 0,8. Найти вероятность своевременного выполнения студентом контрольной работы: а) по двум дисциплинам; б) хотя бы по двум дисциплинам.

3.26. Экспедиция издательства отправила газеты в три почтовых отделения. Вероятность своевременной доставки газет в первое отделение равна 0,95, во второе – 0,9, в третье 0,8. Найти вероятность следующих событий: а) только одно отделение получит газеты вовремя; б) хотя бы одно отделение получит газеты с опозданием

3.27. Прибор, работающий в течении времени t, состоит из трех узлов, каждый из которых независимо от других может за это время выйти из строя. Неисправность хотя бы одного узла выводит из строя прибор целиком. Вероятность безотказной работы в течении времени t первого узла равна 0,9, второго 0,95, третьего – 0,8. Найти вероятность того, что в течении времени t прибор выйдет из строя.

ГЛАВА 4
ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ. ФОРМУЛА БАЙЕСА

4.1. Формула полной вероятности

Формула полной вероятности является следствием теорем сложения и умножения вероятностей. Она позволяет определять вероятность некоторого события, которое может происходить в различных ситуациях с разной вероятностью, причем вероятности этих ситуаций можно оценить до опыта, а условные вероятности появления рассматриваемого события при каждой сложившейся ситуации должны быть известны.

С учетом вышеизложенного искомая вероятность определяется как средневзвешенная вероятность, а весами при этом являются вероятности всевозможных ситуаций, при которых данное событие может происходить.

Пусть требуется определить вероятность случайного события , которая зависит от условий опыта (ситуаций).

Об этих условиях всего можно сделать взаимоисключающих предположений (**гипотез**):

$$H_1,\ H_2,\ \ldots,\ H_n \quad \left(H_i \cdot H_j = \emptyset; \quad i,j = 1,2,\ldots,n, \quad i \neq j\right). \tag{4.1.1}$$

Каждая гипотеза H_i (i=1...n) представляет случайное событие, вероятность которого до опыта (априори[5]) оценивается некоторой вероятностью:

$$P(H_1),\ P(H_2),\ \ldots,\ P(H_n). \tag{4.1.2}$$

Сумма вероятностей гипотез, называемых априорными вероятностями, равна единице (следствие теоремы сложения (3.6.1)):

$$\sum_{i=1}^{n} P(H_i) = 1, \tag{4.1.3}$$

как сумма вероятностей несовместных событий (гипотез), образующих полную группу.

[5] A priori (греч.) – заранее, до опыта, без проверки, независимо от опыта.

Кроме того, предполагается, что условные вероятности $P(A|H_i)$, $i=1, 2, \ldots, n$ появления события A при каждой гипотезе H_i известны и равны:

$$P(A|H_1),\quad P(A|H_2),\quad \ldots,\quad P(A|Hn). \tag{4.1.4}$$

На основании сказанного выше, событие A может появиться только вместе с одной из гипотез, образуя сумму несовместных пересечений (произведений) с событиями H_i, $i=1, 2, \ldots, n$.

$$A = H_1A + H_2A + \ldots + H_nA = \sum_{i=1}^{n} H_iA. \tag{4.1.5}$$

Искомую вероятность $P(A)$ события A определим в соответствии с теоремой сложения вероятностей (3.3.3):

$$P(A) = P\left(\sum_{i=1}^{n} H_iA\right) = \sum_{i=1}^{n} P(H_iA), \tag{4.1.6}$$

а на основании теоремы умножения вероятностей (3.4.1) получим выражение:

$$P(H_iA) = P(H_i)P(A|H_i), \quad i = 1,2,\ldots,n \tag{4.1.7}$$

Подставляя (4.1.7) в формулу (4.1.6), окончательно получим:

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i). \tag{4.1.8}$$

Это и есть **формула полной вероятности**.

Формула (4.1.8) позволяет вычислять полную вероятность события A как сумму произведений априорных вероятностей – вероятностей гипотез $P(H_i)$ на условные вероятности $P(A|H_i)$ события A при каждой гипотезе H_i.

Формула (4.1.8) обычно применяется в тех случаях, когда об условиях опыта со случайным исходом можно сделать ряд взаимоисключающих предположений (гипотез) (4.1.1), оцениваемых до начала опыта априорными вероятностями (4.1.2), а вероятность события (случайного исхода опыта) изменяется в зависимости от условий опыта (4.1.4).

Пример 1. В данный магазин холодильники поставляются от трех производителей в соотношении 2:3:5. Вероятность того, что холодильник от первого поставщика проработает безотказно весь гарантийный срок равна 0,9, от второго – 0,8, от третьего – 0,7. Куплен один холодильник. Какова вероятность того, что данный холодильник проработает безотказно весь гарантийный срок?

Решение.

Начнем решение с обозначения интересуемого нас события или по-другому с ответа на поставленный вопрос:

A – {купленный холодильник проработает безотказно весь гарантийный срок}.

Мы не знаем, кто производитель холодильника, выдвинем гипотезы и найдем их вероятности из условия задачи:

H_1 – {холодильник от первого производителя};

H_2 – {холодильник от второго производителя};

H_3 – {холодильник от третьего производителя}.

Вероятности гипотез:

$$P(H_1) = \frac{2}{2+3+5} = 0{,}2, \quad P(H_2) = \frac{3}{2+3+5} = 0{,}3,$$

$$P(H_3) = \frac{5}{2+3+5} = 0{,}5.$$

$$\sum_{i=1}^{3} P(H_i) = 0{,}2 + 0{,}3 + 0{,}5 = 1.$$

Найдем условные вероятности: $P(A|H_1)=0{,}9$, $P(A|H_2)=0{,}8$, $P(A|H_3)=0{,}7$.

Вычислим вероятность интересуемого нас события по формуле (4.1.8):

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A|H_i) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) =$$

$$= 0{,}2 \cdot 0{,}9 + 0{,}3 \cdot 0{,}8 + 0{,}5 \cdot 0{,}7 = 0{,}77$$

Ответ: 0,77 или 77%

Пример 2. Детали поступают на конвейер с трех станков. Первый станок производит 25% всех деталей, второй 35% и третий 40% деталей. Первый станок выпускает 1% бракованных деталей, второй 3%, третий 5%. Определить вероятность того, что случайно выбранная с конвейера деталь окажется бракованной.

Решение.

Введем обозначения событий: A – деталь окажется бракованной; события H_1, H_2, H_3 – деталь изготовлена соответственно первым, вторым или третьим производителем.

По условию задачи:

$$P(H_1) = 0{,}25,\ P(H_2) = 0{,}35,\ P(H_3) = 0{,}40;$$

$$P(A|H_1) = 0{,}01,\ P(A|H_2) = 0{,}03,\ P(A|H_3) = 0{,}05$$

По формуле полной вероятности (4.1.8) находим:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) =$$

$$= 0{,}25 \cdot 0{,}01 + 0{,}35 \cdot 0{,}03 + 0{,}40 \cdot 0{,}05 = 0{,}033$$

4.2. Формула Байеса (теорема гипотез)

Формула Байеса[6] или теорема гипотез является следствием формулы полной вероятности (4.1.8) и теоремы умножения вероятностей (3.4.1). Она дает возможность пересчитывать априорные вероятности $P(H_i)$, $i=1, 2,..., n$, с учетом результата опыта, – определять вероятности $P(H_k|A)$, $k=1, 2,..., n$ называемые **апостериорными вероятностями** (a`posteriori – “после ” опыта, греч.).

Пусть об условиях опыта, в котором может произойти некоторое событие A, можно сделать ряд взаимоисключающих гипотез:

$$H_1,\quad H_2,\quad \ldots,\quad H_n;\quad H_iH_j = \emptyset;\quad i,j = 1,2,\ldots,n;\quad i \neq j\,, \qquad (4.2.1)$$

образующих полную группу событий:

$$\sum_{i=1}^{n} H_i = \Omega. \qquad (4.2.2)$$

Вероятности гипотез (априорные вероятности) оцениваются до опыта и известны:

$$P(H_1),\quad P(H_2),\quad \ldots,\quad P(H_n);\quad \sum_{i=1}^{n} P(H_i) = 1. \qquad (4.2.3)$$

[6] Названа по имени английского математика Томаса Байеса (T. Bayes) (1702-1761 г.г.). Работы Т. Байеса по теории вероятностей опубликованы в 1763 г. под названием “Опыт решения задач по теории вероятностей...”.

Предположим, что в результате опыта событие A произошло. Пересчитаем априорные вероятности (4.2.3) с учетом этого факта – определим апостериорные вероятности, которые являются условными вероятностями:

$$P(H_1|A),\quad P(H_2|A),\quad \dots\quad P(H_n|A). \tag{4.2.4}$$

Рассмотрим произведение H_iA, $i=1, 2,\dots, n$ и по теореме умножения вероятностей (3.4.1), (3.4.2) запишем его вероятность в двух формах:

$$(H_iA) = P(H_i)P(A|H_i) = P(A)P(H_i|A) \Rightarrow$$

$$P(A)P(H_i|A) = P(H_i)P(A|H_i)\qquad . \tag{4.2.5}$$

Разделим обе части равенства (4.2.5) на $P(A)\neq 0$ и получим выражение для $P(H_iA)$:

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)};\qquad i = 1,2,\dots,n. \tag{4.2.6}$$

Это и есть **формула Байеса**.

Вероятность $P(A)$ может определяться при необходимости по формуле полной вероятности (4.1.8):

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i) - \text{формула полной вероятности.}$$

Для иллюстрации рассмотрим пример предыдущего параграфа, для которых вычислим апостериорные вероятности по формуле Байеса.

Пример 3. Условия предыдущего примера. Купленный холодильник проработал безотказно весь гарантийный срок. Каким из производителей наиболее вероятно он выпущен?

Решение.

$$P(H_1|A) = \frac{0{,}2\cdot 0{,}9}{0{,}77} = \frac{18}{77},$$

$$P(H_2|A) = \frac{0{,}3\cdot 0{,}8}{0{,}77} = \frac{24}{77},$$

$$P(H_3|A) = \frac{0{,}5\cdot 0{,}7}{0{,}77} = \frac{35}{77}.$$

Ответ: наиболее вероятно третьим производителем, так как:

$$P(H_3|A) > P(H_2|A) > P(H_1|A).$$

Пример 4. В корзине находится один шар: или белый или черный – одинаковая вероятность. В корзину опускается белый шар. После перемешивания извлекаем один шар, который оказался белым. Какова вероятность того, что в корзине остался белый шар?

Решение. Пусть гипотеза H_1 в корзине исходно находится белый шар, гипотеза H_2 в корзине находится черный шар. Так как с равной вероятностью в корзине может находиться как белый, так и черный шар, то:

$$P(H_1) = P(H_2) = \frac{1}{2} = 0{,}5.$$

После того, как в корзину был опущен белый шар, вероятность вынуть белый шар (событие A) в предположении гипотезы H_1 есть: $P(A|H_1)=1$. Аналогично, вероятность вынуть белый шар в предположении гипотезы H_2: $P(A|H_1)=0{,}5$. Следовательно, по формуле полной вероятности (4.1.8):

$$P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) = 0{,}5 \cdot 1 + 0{,}5 \cdot 0{,}5 = 0{,}75 \cdot$$

Тогда вероятность, что в корзине остался белый шар (верна гипотеза):

$$P(H_1|A) = \frac{0{,}5 \cdot 1}{0{,}75} = \frac{2}{3}.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Следствием каких теорем является формула полной вероятности?
2. В каких случаях применяется формула полной вероятности?
3. Что представляет собой гипотеза в формуле полной вероятности?
4. Чему равна сумма вероятностей гипотез?
5. Влияет ли вероятность какой-нибудь из выдвинутых гипотез на вероятность интересуемого события A?
6. Следствием каких теорем является формула Байеса?
7. С какой целью применяют формулу Байеса?
8. Почему формулу Байеса называют теоремой гипотез?
9. Можете ли вы привести пример из жизни для использования формулы полной вероятности?

10. Можете ли вы привести пример из жизни для использования формулы Байеса?

ЗАДАНИЯ

4.1. Покупатель может приобрести товар в двух магазинах. Вероятность обращения в первый магазин равна 0,4, а во второй магазин – 0,6. Вероятность того, что к приходу покупателя в первый магазин есть нужный товар, равна 0,5, а для второго магазина – 0,3. Какова вероятность приобретения нужного товара покупателем?

4.2. Пассажир может приобрести билет в одной из двух касс. Вероятность обращения в первую кассу равна 0,4, а во вторую кассу – 0,6. Вероятность того, что к моменту прихода пассажира нужные ему билеты будут распроданы, равна 0,35 для первой кассы и 0,7 – для второй. Какова вероятность того, что пассажир приобрел билет во второй кассе?

4.3. Вероятность того, что деталь попадает первому контролеру, равна 0,7, а второму – 0,3. Вероятность того, что деталь признает стандартной первый контролер, равна 0,92, второй – 0,88. Найти вероятность того, что стандартную деталь проверил первый контролер.

4.4. Два контролера производят оценку качества выпускаемой продукции. Вероятность попадания продукции к первому контролеру равна 0,55, а ко второму – 0,45. Первый контролер выявляет дефект с вероятностью 0,8, а второй – с вероятностью 0,9. Найти вероятность того, что наудачу взятая продукция после контроля признана бездефектным?

4.5. Ваша фирма собирается приобрести ценный электронный прибор. Прибор может оказаться выпущенным одним из трех заводов, причем заранее не известно, каким именно. Обычно в продажу поступает 60% приборов с завода № 1, 30% с завода № 2 и 10% с завода № 3. Соответственно вероятность того, что прибор проработает весь гарантийный срок без поломки, для различных заводов составляет: для завода № 1 – 0,9, для завода № 2 – 0,8, для завода № 3 – 0,6. Какова вероятность того, что купленный фирмой прибор проработает весь гарантийный срок без поломки?

4.6. В магазин поступил одноименный товар, изготовленный двумя предприятиями. С первого предприятия поступило 150 единиц, из них 30 единиц первого сорта, а со второго предприятия поступило 200 единиц, из них 50 единиц первого сорта. Из общей партии наудачу берется один товар. Этот товар оказался первого сорта. Какова вероятность того, что взятый товар изготовлен на первом предприятии?

4.7. В данный район изделия поставляются тремя фирмами в соотношении 5:8:7. Среди продукции первой фирмы стандартные изделия составляют 90%, второй – 85%, третьей – 75%. Найти вероятность того, что: а) приобретенное изделие окажется нестандартным; б) приобретенное изделие оказалось стандартным. Какова вероятность того, что оно изготовлено третьей фирмой?

4.8. Два специалиста отдела технического контроля фабрики проверяют качество выпускаемой продукции. Вероятность попадания изделия к каждому из них одна и та же. Вероятность выявления дефекта первым специалистом равна 0,8, а вторым 0,9. Из массы проверенной продукции наудачу берется одно изделие. Оно оказалось с дефектом. Какова вероятность того, что ошибку допустил второй контролер?

4.9. Вероятности того, что во время работы цифровой электронной машины произойдет сбой в арифметическом устройстве, в оперативной памяти и в остальных устройствах, относятся как 3:2:5. Вероятности обнаружения сбоя в арифметическом устройстве, в оперативной памяти и в остальных устройствах соответственно равны 0,8; 0,9; 0,9. Найти вероятность того, что возникший в машине сбой будет обнаружен.

4.10. В двух одинаковых коробках находится одинаковое количество карандашей марки «Конструктор». Известно, что в первой коробке треть составляют карандаши твердости «ТМ», а во второй коробке такой твердости составляет четвертая часть. Наудачу берется коробка, и из нее наудачу извлекается карандаш. Он оказался твердости «ТМ». Какова вероятность того, что он извлечен из первой коробки?

4.11. Страховая компания разделяет застрахованных по классам риска: 1 класс – малый риск, 2 класс – средний, 3 – большой риск. Среди этих клиентов 50% первого класса риска, 30% – второго, 20% – третьего. Вероятность необходимости выплачивать страховое вознаграждение для первого класса риска составляет 0,01, второго – 0,03, третьего – 0,08. Какова вероятность того, что:

а) застрахованный получит вознаграждение за период страхования;

б) получивший денежное вознаграждение относится к группе малого риска?

4.12. Изделия были произведены с использованием двух технологических линий. На первой линии было произведено 3 изделия, на второй линии 2 изделия. Вероятность того, что изделие будет отличного качества при производстве на первой линии равна 0,85, на второй – 0,8. Какова вероятность того, что случайно выбранной изделие отличного качества произведено на первой линии?

4.13. Один из трех стрелков вызывается на линию огня и производит два выстрела. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,3, для второго – 0,5; для третьего – 0,8. Мишень не поражена. Найти вероятность того, что выстрелы произведены первым стрелком.

4.14. На складе имеется 24 комплектующих изделий от двух компаний поставщиков, из них 66 изделий от первой компании. Известно, что с вероятностью 0,8 среди поставок первой компании встречаются изделия, выполненные по новейшей технологии. Среди изделий второй компании такие встречаются с вероятностью 0,9. Какова вероятность того, что случайным образом выбранное изделие выполнено по новейшей технологии?

4.15. На трех станках-автоматах обрабатываются однотипные детали, поступающие после обработки на общий конвейер. Первый станок дает 2% брака, второй – 7%, третий – 10%. Производительность первого станка в 3 раза больше производительности второго, а третьего – в 2 раза меньше, чем второго.
а) Каков процент брака на конвейере?
б) Каковы доли деталей каждого станка среди бракованных деталей на конвейере?

4.16. Для проверки усвоения лекционного материала в студенческой группе был случайным образом выбран студент, и ему был предложен тест по теме лекции. В этой студенческой группе 6 отличников, 7 хороших студентов и три средних студента (по результатам прошедшей сессии). Было известно, что отличник справляется с тестом с вероятностью 0,85, хороший студент справляется с тестом с вероятностью 0,6, а средний студент справляется с тестом с вероятностью 0,3.
a) Вычислить априорную вероятность того, что был протестирован хороший студент.
в) Вычислить вероятность того, что студент не справился с тестом.
с) Вычислить вероятность того, что был выбран хороший студент, если известно, что студент с тестом не справился.

4.17. В первой коробке 20 радиоламп, из них стандартных 18; во второй 10, из них 9 стандартных. Из второй коробки переложили в первую одну наугад взятую лампу. Определить вероятность того, что затем наугад взятая лампа из первой коробки, является стандартной.

ГЛАВА 5
ПОВТОРНЫЕ ИСПЫТАНИЯ

5.1. Формула Бернулли

Повторение опытов связано с задачами, в которых осуществляется последовательность независимых опытов, в каждом из которых может произойти (или не произойти) некоторое событие A, вероятность которого известна. Задача заключается в определении вероятности появления события A ровно раз в независимых опытах, которая в дальнейшем будет обозначаться через $P_{m,n}$ (m=0...n).

Испытания со случайным исходом называются ***независимыми***, если вероятность исхода того или иного испытания не зависит от исходов других испытаний. В противном случае испытания будут ***зависимыми***.

Независимые испытания проводятся в одинаковых условиях, поэтому вероятность появления события A в каждом испытании одинакова и равна $P(A)=p$, а вероятность непоявления события A (появления противоположного события $\overline{A}$) равна $P(\overline{A})=1-p=q$. Такая последовательность испытаний носит название "**испытания Бернулли**". Требуется определить вероятность $P_{m,n}$ появления события A ровно m раз в n испытаниях (m = 0,1,...,n).

Примеры задач, связанных с испытаниями Бернулли:

1. Производится n бросаний симметричной монеты на гладкую поверхность стола. При каждом бросании герб (цифра) может появиться с одной и той же вероятностью $p=0{,}50$. Требуется определить вероятность $P_{m,n}$ появления герба (цифры) ровно m раз из n бросаний ($m=0,1,\ldots,n$).

2. Производится стрельба в тире по мишени n выстрелами с индивидуальным прицеливанием при каждом выстреле с одной дистанции. Вероятность попадания в "десятку" для данного стрелка оценивается величиной p. Требуется определить вероятность $P_{m,n}$ попадания в "десятку" m раз при n выстрелах ($m=0,1,\ldots,n$).

Докажем, что вероятность $P_{m,n}$ появления события A m раз в n независимых опытах определяется выражением (**формулой Бернулли**):

$$P_{m,n} = C_n^m p^m q^{n-m}, \quad m = 0,1,\ldots,n; \tag{5.1.1}$$

где C_n^m – число сочетаний из n элементов по m (2.5.1), (2.5.2),

$p = P(A)$– вероятность появления события в одном (любом) опыте (испытании),
$q = 1 - p$ – вероятность непоявления события A в одном (любом) опыте (испытании).

Для доказательства введем обозначение: обозначим через A_m событие, состоящее в появлении A ровно m раз в n независимых опытах:

$$P_n(A_m) = P_{m,n}.$$

Событие A_m при многократном воспроизведении серии из n опытов может появляться различными способами, при каждом из которых событие A должно появиться m раз, а остальные $(n - m)$ раз оно не должно появиться, должно появиться противоположное событие $\overline{A}$.

Обозначая факт появления события A единицей, а факт появления события $\overline{A}$ нулем, представим один из возможных способов появления события A_m в виде:

$$A_{m,j} = \{1,1,\ldots,1,0,0,\ldots,0\}, \qquad j = 1,2,\ldots,n. \tag{5.1.2}$$

Другие способы выражения A_m будут отличаться от (5.1.2) только порядком нулей и единиц при неизменном количестве тех и других (m единиц и $(n - m)$ нулей).

Число таких способов очевидно равно числу перестановок с повторениями из n элементов, среди которых m единиц и $(n - m)$ нулей (2.5.3), формально равному числу сочетаний из n по m:

$$N = P(m, n - m) = \frac{n!}{m!\,(n - m)!} = C_n^m. \tag{5.1.3}$$

Вероятность появления события A_m в любой комбинации m единиц и $n - m$ нулей очевидно равна:

$$P(A_{m,j}) = p^m q^{n-m}, \quad j = 1,2,\ldots,n. \tag{5.1.4}$$

Таким образом, с учетом (5.1.3) и (5.1.4) окончательно получим (5.1.1):

$$P_{m,n} = C_n^m p^m q^{n-m}, \quad m = 1,2,\ldots,n,$$

что и требовалось доказать. Число m_0 наступления события A в n независимых испытаниях называется наивероятнейшим, если вероятность $P_{m_0,n}$ по крайней мере не меньше вероятностей при любом m. Значение находят из следующей системы неравенств:

$$\begin{cases} P_{m_0,n} \geq P_{m_0+1,n}, \\ P_{m_0,n} \geq P_{m_0-1,n}. \end{cases}$$

Подставляя в последнее неравенство соотношения (5.1.1) и учитывая формулу для вычисления числа сочетаний, можно найти, что

$$np - q \le m_0 \le np + q. \qquad (5.1.5)$$

Заметим, что длина интервала равна $(np + q) - (np - q) = p + q = 1$, поэтому всегда существует целое число, удовлетворяющее условию (5.1.5), при этом если $np + q$ является целым числом, то наивероятнейших чисел два: $m_0 = np + q$ и $m_0 = np - q$.

Вероятности $P_{m,n}$ могут формально быть получены как коэффициенты в разложении бинома Ньютона $(q + px)^n$ по степеням независимой переменной x:

$$(q + px)^n = \sum_{m=0}^{n} C_n^m (px)^m q^{n-m} = \sum_{m=0}^{n} P_{m,n} x^m. \qquad (5.1.6)$$

Таким образом, функция

$$\varphi_n(x) = (q + px)^n \qquad (5.1.7)$$

независимой переменной x является “**производящей функцией**” для искомых вероятностей , которые называются **биноминальными коэффициентами**. Полагая в (5.1.6) x=1, получим:

$$\sum_{m=0}^{n} P_{m,n} = 1. \qquad (5.1.8)$$

Смысл этого соотношения: сумма вероятностей всевозможных исходов n опытов, состоящих в появлении события A ровно m раз (m=0, 1, 2, ..., n), равна единице.

Пример 1. Производится четыре независимых опыта, в каждом из которых с вероятностью $p = 0{,}60$ может произойти событие A. Определить вероятность появления события A не менее трех раз.

Решение. Обозначим B – событие, состоящее в появлении события A не менее трех раз в четырех независимых опытах. Тогда вероятность события B определяется как сумма вероятностей появления события A три или четыре раза:

$$P(B) = P_{3,4} + P_{4,4}.$$

По формуле (5.1.1) получим решение задачи:

$$P(B) = C_4^3 p^3 q + p^4 = 4 \cdot 0{,}60^3 \cdot 0{,}40 + 0{,}60^4 = 0{,}476.$$

Пример 2. Монета бросается пять раз. Какова вероятность того, что число выпавших гербов будет больше числа выпавших цифр?

Решение. Обозначим *A* искомое событие – число выпавших гербов больше числа цифр при пяти бросаниях монеты. Для выполнения события *A* необходимо, чтобы число гербов при пяти бросаниях монеты было 3,4 или 5 (при этом цифр будет соответственно 2,1 или 0). По формуле (5.1.1) получим:

$$P(A) = P_{3,5} + P_{4,5} + P_{5,6} = C_5^3 p^3 q^2 + C_5^4 p^4 q^4 + p^5 =$$

$$= 10 \cdot 0{,}50^3 \cdot 0{,}50^2 + 5 \cdot 0{,}50^4 \cdot 0{,}50 + 0{,}50^5 = 0{,}50.$$

5.2. Формула Пуассона

При больших вычисление по формуле Бернулли становится технически сложным, например, если требуется найти вероятность того, что «герб» выпадет 300 раз при 500 подбрасываниях монеты, то, согласно формуле (5.1.1), необходимо найти значение следующего выражения:

$$P_{300,500} = C_{500}^{300} \cdot (0{,}5)^{500}.$$

Поэтому при больших *n* используют формулы для приближенных вычислений.

Если при большом числе испытаний вероятность появления события *A* в одном опыте мала, а произведение $\lambda = np \leq 10$ сохраняет постоянное значение для разных серий опытов, то есть среднее число появлений события *A* в разных сериях испытаний остается неизменным, тогда используется **формула Пуассона**:

$$P_{m,n} \approx P_m(\lambda) = \frac{\lambda e^{-\lambda}}{m!}. \qquad (5.2.1)$$

Значения функции $P_m(\lambda)$ можно найти непосредственно по формуле (5.2.1) или по таблице 4, приведенной в приложении 4.

Пример 3. Завод отправил на базу 1000 бутылок минеральной воды в стеклянной таре. Вероятность того, что бутылка будет разбита в пути равна 0,002.

Какова вероятность того, что в пути будет разбито три бутылки?

Решение. По условию

$n = 1000$, $p = 0{,}002$, $m = 3$, $\lambda = np = 1000 \cdot 0{,}002 = 2$.

Следовательно, по формуле (5.2.1) искомая вероятность равна:

$$P_{3,1000} \approx \frac{2e^{-2}}{3!} = \frac{2}{1 \cdot 2 \cdot 3e^2} = \frac{2}{3e^2} \approx 0{,}18.$$

5.3. Локальная теорема Муавра-Лапласа

Как было отмечено, пользоваться формулой Бернулли при больших значениях n и m весьма неудобно, так как приходится обращаться с громоздкими вычислениями.

Локальная теорема Лапласа[7] даёт необходимую асимптотическую формулу, обеспечивающую возможность приближенно вычислить вероятность $P_{m,n}$ при достаточно больших n при $0<p<1$ и $npq\geq 20$.

Локальная теорема Муавра-Лапласа: если вероятность p наступления события A в каждом испытании постоянна и отлична от 0 и 1, то вероятность того, что событие A произойдет m раз в n независимых испытаниях $P_n(m)$ при достаточно большом числе n, то есть при n приближенно равна:

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi(x). \tag{5.3.1}$$

$$\text{где } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} - \text{функция Гаусса}, \qquad x = \frac{m - np}{\sqrt{npq}}.$$

Функция $\varphi(x)$ табулирована (приложение 1, табл. 1) и обладает следующими свойствами:

а) функция $\varphi(x)$ является четной, т.е. $\varphi(-x)=\varphi(x)$;

б) функция $\varphi(x)$ монотонно убывающая при положительных значениях x, причем при $x\to\infty$ $\varphi(x)\to 0$;

в) при $x\geq 4$ можно считать $\varphi(x)=0$. Уже при $x\geq 4$ $\varphi(x)\leq 0{,}0001$.

Пример 4. В некоторой местности из каждых 100 семей 80 имеют автомобили. Какова вероятность того, что из 400 семей у 300 имеются автомобили?

Решение. Вероятность того, что в семье имеется автомобиль, равна $p=80/100=0{,}8$. Так как $n=400$ достаточно велико (условие $npq=400\cdot 0{,}8\cdot 0{,}2=64\geq 20$ выполнено), то применим локальную формулу Муавра-Лапласа:

$$x = \frac{300 - 400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 0.2}} = -2{,}50.$$

[7] Лаплас Пьер Симеон (17491827) – выдающийся французский математик, физик, астроном, профессор Французской АН, председатель палаты мер и весов (1790), министр. Один из создателей математической теории вероятностей, для разработки которой ввел производящие функции и новое преобразование (Лапласа). Доказал биномиальный закон (1812), первые предельные теоремы теории вероятностей, развил теорию ошибок измерений и метод наименьших квадратов.

$\varphi(-2{,}50) = \varphi(2{,}50) = 0{,}0175$, получено из табл. 1, приложение 1.

$$P_{400}(300) \approx \frac{\varphi(-2{,}50)}{\sqrt{400 \cdot 0.8 \cdot 0.2}} = \frac{0{,}0175}{8} \approx 0.0022.$$

Пример 5. Найти вероятность появления события A ровно 80 раз в 400 независимых опытах в одинаковых условиях, если в одном опыте событие A появляется с вероятностью p=0,20.

Решение. По формуле (5.3.1) при n=400, m=80, p=0,20, (q=0,80) имеем:

$$x = \frac{80 - 400 \cdot 0{,}20}{\sqrt{400 \cdot 0{,}20 \cdot 0{,}80}} = 0,$$

$$\varphi(0) = 0{,}3989 \approx 0{,}40,$$

$$P_{80,400} \approx \frac{0{,}40}{\sqrt{400 \cdot 0{,}20 \cdot 0{,}80}} = 0{,}05,$$

5.4. Интегральная теорема Муавра–Лапласа

Если требуется найти вероятность того, что интересуемое нас событие включает определенный диапазон, связанный с громоздкими вычислениями, более рациональным является использование интегральной теоремы Муавра-Лапласа.

Интегральная теорема Муавра-Лапласа дает возможность определять вероятности событий вида:

$$P(m_1 \leq X \leq m_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{t^2}{2}} dt, \tag{5.4.1}$$

где
$$x_1 = \frac{m_1 - np}{\sqrt{npq}}, x_2 = \frac{m_2 - np}{\sqrt{npq}}. \tag{5.4.2}$$

Формулировка интегральной теоремы: вероятность $P(m_1 \leq X \leq \mathrm{m}_2)$ того, что событие A появится от m_1 до m_2 раз в n независимых опытах, если вероятность p появления события A в каждом опыте постоянна и равна

$0<p<1$, при $n\to\infty$ стремится к значению (5.4.1), где x_1, x_2 определяются по формулам (5.4.2) равна:

$$P(x_1 \le X \le x_2) = \Phi_0(x_2) - \Phi_0(x_1) \qquad (5.4.3)$$

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{t^2}{2}} dt, \qquad (5.4.4)$$

где $\Phi_0(x)$ – функция Лапласа (интеграл вероятности) – табулирована (см. приложение 3, табл. 3) и обладает следующими свойствами:

а) функция $\Phi_0(x)$ нечетная, то есть $\Phi_0(-x) = -\Phi_0(x)$;

б) функция $\Phi_0(x)$ монотонно возрастающая, причем при $x\to+\infty$ $\Phi_0(x)\to 0{,}5$ (практически можно считать, что уже при $x>4$ $\Phi_0(x)\approx 0{,}5$).

Интегральная формула, как и локальная, тем точнее, чем больше n. При условии $npq \ge 20$ интегральная формула (5.4.3) дает незначительную погрешность вычисления вероятностей.

Пример 6. Изготовленные детали проходят выборочную проверку в отделе технического контроля (ОТК). Вероятность того, что деталь пройдет проверку, равна 0,80. Какова вероятность того, что из 400 случайно выбранных для проверки деталей число деталей не прошедших ОТК, будет лежать в пределах от 70 до 100?

Решение. Вычислим искомую вероятность по формуле (5.4.1), а значения границ – по формулам (5.4.2) при n=400, p=0,20, q=0,80, m_1=70, m_2=100:

$$P(70 \le X \le 100) = \Phi_0(x_2) - \Phi_0(x_1),$$

где

$$x_1 = \frac{70 - 400 \cdot 0{,}20}{\sqrt{400 \cdot 0{,}20 \cdot 0{,}80}} = \frac{70 - 80}{\sqrt{64}} = -1{,}25;$$

$$x_2 = \frac{100 - 400 \cdot 0{,}20}{\sqrt{400 \cdot 0{,}20 \cdot 0{,}80}} = \frac{100 - 80}{\sqrt{64}} = 2{,}5;$$

$$\Phi_0(-1{,}25) = -\Phi_0(1{,}25) = -0{,}394;$$

$$\Phi_0(2{,}5) = 0{,}494$$

Окончательно получим:

$$P(70 \le X \le 100) = 0{,}494 + 0{,}394 = 0{,}888.$$

Пример 7. В некоторой местности из каждых 100 семей 80 имеют автомобили. Какова вероятность того, что от 300 до 360 (включительно) семей из 400 имеют автомобили.

Решение. Применим интегральную теорему Муавра-Лапласа ($npq = 64 > 20$).

$$x_1 = \frac{300 - 400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 02}} = -2.50; \qquad x_2 = \frac{360 - 400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 02}} = 5.0.$$

$$P_{400}(300 \leq X \leq 360) \approx \Phi_0(5{,}0) - \Phi_0(-2{,}50) =$$

$$= \Phi_0(5{,}0) + \Phi_0(2{,}50) \approx)0{,}5 + 0{,}4938 = 0{,}9938$$

5.5. Повторные испытания в изменяющихся условиях

Независимые опыты проводятся в изменяющихся условиях, поэтому вероятность появления события A в i-м опыте равна $P(A_i)=p_i$, а вероятность непоявления события A (появления противоположного события $\overline{A}$) равна $P(\overline{A}_i)=1-p_i=q_i$ ($i=1...n$). Требуется определить вероятность $P_{m,n}$ появления события A ровно m раз в n опытах.

Примеры задач, связанных с повторением опытов в изменяющихся условиях.

1. Производится стрельба по мишени из пистолета независимыми выстрелами. После каждого выстрела стрелок делает шаг по направлению к мишени. Вероятности попадания в мишень при каждом очередном выстреле растут: $p_1<p_2<p_3...<p_n$. Требуется определить вероятность $P_{m,n}$ попадания в мишень m раз при n выстрелах в изменяющихся условиях стрельбы ($m=0...n$).

2. Изделие подвергается контролю со стороны n независимо работающих контролеров. Вероятности прохождения контроля изделия у первого, второго,..., n-го контролера равны $p_1, p_2, ..., p_n$. Требуется определить вероятность $P_{m,n}$ того, что изделие прошло контроль успешно у m контролеров ($m=0...n$), но чаще вычисляется вероятность $P_{n,n}$ прохождения контроля у всех n контролеров.

Можно показать, что искомая вероятность $P_{m,n}$ будет равна коэффициенту при x^m в разложении производящей функции вида:

$$\varphi_n(x) = \prod_{i=1}^{n}(q_i + p_i x) = \sum_{m=0}^{n} P_{m,n} x^m. \tag{5.5.1}$$

по степеням независимой переменной x.

Действительно, если перемножить все n биномов, входящих в состав производящей функции (5.5.1):

$$\varphi_n(x) = (q_1 + p_1 x)(q_2 + px)\dots(q_n + p_n x), \qquad (5.5.2)$$

то каждый член результата будет обязательно содержать по одному элементу из каждого бинома. Поэтому коэффициент при x^m будет иметь в своем составе различные комбинации из m букв p с разными индексами, что соответствует m появлениям события A в n опытах и $(n-m)$ букв q с разными индексами, что соответствует $(n-m)$ появлениям события $\bar{A}$ (непоявлениям события A), например:

$$\left\{\underbrace{p_1,\ p_2,\ \dots,\ p_m}_{m},\ \underbrace{q_{m+1},\ q_{m+2},\ \dots,\ q_n}_{n-m}\right\}.$$

В частном случае при $p_1=p_2=\dots=p_n=p$, $(q_1=q_2=\dots=q_n=q)$ производящая функция (5.5.1) вырождается в бином Ньютона:

$$\varphi_n(x) = (q + px)^n = \sum_{m=0}^{n} P_{m,n} x^m . \qquad (5.5.3)$$

Пример 8. Производится три независимых опыта, условия которых изменяются от опыта к опыту. Вероятность появления события A в первом опыте равна $p_1=0{,}40$, во втором – $p_2=0{,}60$, в третьем – $p_3=0{,}80$. Определить вероятность $P_{m,3}$ появления события A ровно m раз ($m=0,1,2,3$) в трех независимых опытах.

Решение. Составим производящую функцию (5.5.1) и вычислим искомые вероятности $P_{m,3}$ как коэффициент при x^m в разложении этой функции по степеням независимой переменной x:

$$\varphi_3(x) = (0{,}60 + 0{,}40x)(0{,}40 + 0{,}60x)(0{,}20 + 0{,}80x) =$$

$$= 0{,}048 + 0{,}296x + 0{,}464x^2 + 0{,}192x^3$$

Искомые вероятности равны:

$$P_{0,3} = 0{,}048, \quad P_{1,3} = 0{,}296, \quad P_{2,3} = 0{,}464, \quad P_{3,3} = 0{,}192.$$

Правильность решения контролируется условием (5.1.8):

$$0{,}048 + 0{,}296 + 0{,}464 + 0{,}192 = 1{,}000 .$$

Пример 9. Четыре оператора радиолокационных станций независимо друг от друга анализируют воздушную обстановку в заданном секторе. Первый оператор распознает сигнал о воздушном объекте на фоне помех с вероятностью $p_1 = 0{,}90$, второй – с вероятностью $p_2 = 0{,}80$, третий – с вероятностью $p_3 = 0{,}70$, четвертый – с вероятностью $p_4 = 0{,}6$. Решение о появлении воздушного объекта в заданном секторе наблюдения принимается, если хотя бы три оператора распознают объект. Какова вероятность принятия решения о появлении объекта в секторе наблюдения четырех операторов?

Решение. Обозначим A – событие, состоящее в принятии решения о появлении воздушного объекта в заданном секторе наблюдения. Событие A произойдет в том случае, если 3 или 4 оператора распознают объект на фоне помех, следовательно, вероятность события A равна сумме вероятностей:

$$P(A) = P_{3,4} + P_{4,4}.$$

Вероятности $P_{3,4}$ и $P_{4,4}$ определим как коэффициенты при x^3 и x^4 в разложении производящей функции вида:

$$\varphi_4(x) = (q_1 + p_1 x)(q_2 + p_2 x)(q_3 + p_3 x)(q_4 + p_4 x) =$$

$$= (0{,}10 + 0{,}90x)(0{,}20 + 0{,}80x)(0{,}30 + 0{,}70x)(0{,}40 + 0{,}60x),$$

по степеням независимой переменной x:

$$\varphi_4(x) = 0{,}002 + 0{,}040x + 0{,}215x^2 + 0{,}440x^3 + 0{,}302x^4.$$

Проверка по (5.1.8):

$$0{,}002 + 0{,}040 + 0{,}215 + 0{,}440 + 0{,}302 = 1{,}000.$$

Вероятности $P_{m,4}$ равны:

$$P_{0,4} = 0{,}002,\quad P_{1,4} = 0{,}040,\quad P_{2,4} = 0{,}215,\quad P_{3,4} = 0{,}440,\quad P_{4,4} = 0{,}302.$$

Искомая вероятность равна:

$$P(A) = 0{,}440 + 0{,}302 = 0{,}742.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какой ряд испытаний называется независимым?
2. В чем состоит понятие «повторные независимые испытания»?
3. Что можно найти по формуле Бернулли для повторных независимых испытаний?
4. По какой формуле можно найти вероятность появления события от до в серии независимых испытаний?

5. По какой формуле определяется вероятность наступления события хотя бы один раз в серии независимых испытаний, в каждом из которых событие может произойти с вероятностью и не произойти с вероятностью?
6. Что понимается под наивероятнейшим числом появления события в серии независимых испытаний, в каждом из которых событие может произойти с вероятностью и не произойти с вероятностью?
7. В каких случаях пользуются формулой Пуассона?
8. В чем заключается локальная теорема Лапласа?
9. Каковы основные свойства нормированной функции Гаусса?
10. В чем заключается интегральная теорема Лапласа?
11. Каковы основные свойства функции Лапласа?

ЗАДАНИЯ

5.1. Вероятность поражения вирусным заболеванием куста земляники равна 0,15. Составить закон распределения числа кустов земляники, зараженных вирусом, из четырех посаженных кустов.

5.2. Завод отправляет в некоторый город 1500 автомобилей. Вероятность того, что в пути машина может получить повреждение, равна 0,002. Найти вероятность того, что в пути будет повреждено не более 4-х автомобилей

5.3. Вероятность попадания в мишень при одном выстреле для данного стрелка равна 0,7. Найти вероятность того, что при 200 выстрелах мишень будет поражена 160 раз.

5.4. Проверкой установлено, что цех в среднем выпускает 96% продукции высшего сорта. На базе приемщик проверяет 200 изделий этого цеха. Если среди них окажется более 10 изделий не высшего сорта, то вся партия бракуется, т.е. возвращается в цех. Какова вероятность того, что партия будет принята?

5.5. Вероятность рождения мальчика 0.515. В семье 6 детей. Найти вероятность того, что из них:

а) ровно три девочки;

б) не более трех девочек;

в) не менее двух, но не более четырех девочек.

5.6. Вероятность появления события в каждом их 2400 независимых испытаний постоянна и равна 0,6. Найти вероятность того, что это событие наступит ровно 1400 раз?

5.7. Сколько надо произвести независимых испытаний с вероятностью появления события в каждом испытании равной 0,4, чтобы наивероятнейшее число появлений события в этих испытаниях было равно 20?

5.8. Производство электронно-лучевых трубок для телевизоров дает в среднем 12% брака. Найдите вероятность наличия 215 годных трубок в партии из 250 штук.

5.9. Вероятность изготовления на автоматическом станке стандартной детали равна 0,8. Найти вероятности возможного числа появления бракованных деталей среди 5 отобранных.

5.10. Вероятность выигрыша в лотерею на один билет равна 0,3. Куплено 13 билетов. Найти наивероятнейшее число выигрышных билетов и соответствующую вероятность.

5.11. В новом микрорайоне поставлено 10000 кодовых замков на входных дверях домов. Вероятность выхода из строя одного замка в течение месяца равна 0,0002. Найти вероятность того, что за месяц откажут три замка.

5.12. На факультете насчитывается 1750 студентов. Какова вероятность того, что 1 января является днем рождения одновременно четырех студентов факультета?

5.13. При проведении маркетинговых исследований выявлено, что 30% опрошенных предпочитают использовать продукцию данной фирмы. Найти вероятности возможного числа пользователей продукцией фирмы в произвольно выбранной группе из пяти человек.

5.14. Из одного компьютера на другой необходимо передать файл объемом 8 000 символов. Вероятность ошибки при передаче символа равна 0,001. Найти вероятность того, что будет не менее двух ошибок при передаче файла.

5.15. В каждом из 700 независимых испытаний событие А происходит с постоянной вероятностью 0.35. Найти вероятность того, что событие А произойдет: а) 270 раз; б) не менее 230 и не более 270 раз; в) не менее 270 раз.

5.16. На телефонной станции неправильное соединение происходит с вероятностью 0,005. Найти вероятность того, что среди 400 соединений произойдет: а) одно неправильное соединение; б) меньше 3 неправильных соединений; в) больше 2 неправильных соединений.

5.17. Проводится 200 независимых опытов с вероятностью успеха в каждом 24%. Какова вероятность успешного проведения 50 опытов?

5.18. При пересадке саженцев в некотором хозяйстве в среднем погибает 30% саженцев. Каково наивероятнейшее число прижившихся саженцев из пяти пересаженных, и чему равна вероятность этого события?

5.19. На некотором производстве установлено, что в технологическом процессе в среднем 90% изделий являются стандартными. При выборочном контроле наудачу было взято 400 изделий. Каково наивероятнейшее число стандартных деталей среди отобранных, и чему равна его вероятность?

5.20. Вероятность того, что отклонение технических данных изделия в серийном производстве превышает норму от допустимого, равна 0,005. Найти вероятность того, что среди поступивших 1600 изделий с отклонениями от допустимых будет менее трех.

5.21. Вероятность выхода из строя за смену одного станка равна 0,1. Определить вероятность выхода из строя от 2 до 13 станков при наличии 100 станков.

5.22. Вероятность получения положительного результата в каждом из 2100 опытов равна 0,7. Найти вероятность того, что положительный результат будет получен в 1500 опытов.

5.23. Вероятность того, что при транспортировке фруктов процент порчи превысит норму, равна 0,03. Найти вероятность того, что из 100 прибывших вагонов процент порчи груза будет менее чем в 98 вагонах.

5.24. Мастер и ученик выпускают однотипные изделия, причем производительность мастера в два раза превышает производительности ученика. Найти вероятность того, что среди 450 изделий, случайно взятых из упаковочного контейнера, 300 окажутся изготовленными мастером.

5.26. Четыре покупателя приехали на оптовый склад. Вероятность того, что каждому из этих покупателей потребуется холодильник марки «Атлант», равна 0,4. Найти вероятность того, что холодильник потребуется: а) не менее чем двум покупателям; б) не более чем трем покупателям; в) всем четырем покупателям.

5.24. Для вычислительной лаборатории приобретено девять компьютеров, причем вероятность брака для одного компьютера равна 0,1. Какова вероятность, что придется заменить более двух компьютеров.

ГЛАВА 6
СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

6.1. Понятие случайной величины. Виды случайных величин

Одним из важнейших понятий теории вероятностей является понятие случайной величины.

Под **случайной величиной** понимается переменная, которая в результате испытания в зависимости от случая принимает одно из возможного множества своих значений, какое именно заранее не известно.

Примеры случайных величин:

1) число родившихся детей в течение суток в г. Москве;
2) количество бракованных изделий в данной партии;
3) число произведенных выстрелов до первого попадания;
4) дальность полета артиллерийского снаряда;
5) расход электроэнергии на предприятии за месяц.

Выделяют дискретные и непрерывные случайные величины.

Случайная величина называется **дискретной** (прерывной), если множество ее значений конечное или бесконечное, но счетное.

Под непрерывной случайной величиной будем понимать величину, бесконечное множество значений которой есть некоторый промежуток числовой оси.

Для дискретной случайной величины множество возможных значений случайной величины конечно или счетно, для непрерывной - бесконечно и несчетно.

Случайные величины обозначаются прописными буквами латинского алфавита $A, B, C, \ldots, X, Y, Z$, а их значения соответствующими индексированными строчными буквами $a_i, b_i, c_i, \ldots, x_i, y_i, z_i$.

Наиболее полным, исчерпывающим описанием случайной величины является ее закон распределения.

Закон распределения случайной величины — всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Для дискретной случайной величины закон распределения может быть задан в виде таблицы, аналитически (в виде формулы) и графически.

Простейшей формой задания закона распределения дискретной случайной величины является таблица (матрица), в которой в порядке возрастания перечислены возможные значения случайной величины и соответствующие им вероятности.

X:

x_i	x_1	x_2	...	x_{n-1}	x_n
p_i	p_1	p_2	...	p_{n-1}	p_n

Событие в ряде распределения несовместны и единственно возможны, так как образуют полную группу событий, поэтому для любой дискретной случайной величина суммы вероятностей ряда:

$$\sum_{i=1}^{n} p_i = p_1 + p_2 + \cdots + p_{n-1} + p_n = 1$$

Иллюстрацию дальнейших теоретических выкладок будем осуществлять на конкретном примере.

Пример 1. Пусть дан ряд распределения

X:

x_i	1	2	3	4	5
p_i	0,01	0,09	0,35	0,2	0,35

Ряд распределения может быть изображен графически, если по оси абсцисс откладывать значения случайной величины, а по оси ординат - соответствующие им вероятности.

Соединение полученных точек образуют ломаную, называемую многоугольником или полигоном распределения.

На рисунке 1 изображен многоугольник распределения для примера 1.

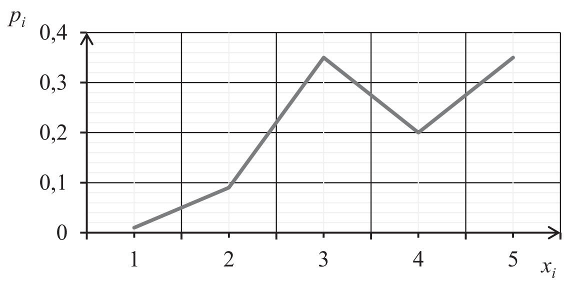


Рис. 1. Многоугольник распределения

6.2. Математическое ожидание дискретной случайной величины

Закон (ряд) распределения дискретной случайной величины дает исчерпывающую информацию о ней, так как позволяет вычислить вероятности любых событий, связанных со случайной величиной.

Однако такой закон (ряд) распределения бывает труднообозрим, не всегда удобным для анализа, а иногда и вовсе не нужным.

Рассмотрим следующий пример.

Пример 2. Известны количество выточенных деталей двумя токарями по дням недели за пять рабочих дней с вероятностью 0,2. Для первого: 3, 5, 12, 19, 21, для второго: 9, 11, 13, 14, 13. Предстоит сокращение штатов, перед руководством стоит вопрос, кому из них отдать предпочтение?

Решение.

Составим ряд распределения для случайных величин X (для первого токаря) и Y (для второго):

X:

x_i	3	5	12	19	21
p_i	0,2	0,2	0,2	0,2	0,2

Y:

y_i	9	11	13	14	13
p_i	0,2	0,2	0,2	0,2	0,2

Необходимо выяснить кто из них работает лучше, естественно работает лучше тот, кто в среднем за неделю вытачивает больше деталей.

Таким средним значением дискретной случайно величины является математическое ожидание.

Математическим ожиданием или **средним значением** (будем обозначать символами $M(X)$ или m_x) дискретной случайной величины называется сумма произведений возможных значений на соответствующие им вероятности, то есть:

$$M(X) = x_1p_2 + x_2p_2 + \dots + x_{n-1}p_{n-1} + x_np_n = \sum_{i=1}^{n} x_ip_i \qquad (6.2.1)$$

Пример 3. Найдем математические ожидания для примеров 1 и 2:

Пример 1: $M(X) = 1 \cdot 0{,}01 + 2 \cdot 0{,}09 + 3 \cdot 0{,}35 + 4 \cdot 0{,}2 + 5 \cdot 0{,}35 = 3{,}44$

Пример 2: $m_x = 3 \cdot 0{,}2 + 5 \cdot 0{,}2 + 12 \cdot 0{,}2 + 19 \cdot 0{,}2 + 21 \cdot 0{,}2 = 12$

$m_y = 9 \cdot 0{,}2 + 11 \cdot 0{,}2 + 13 \cdot 0{,}2 + 14 \cdot 0{,}2 + 13 \cdot 0{,}2 = 12$

Таким образом, пока нет ответа на вопрос, кому отдать предпочтение.

Математическое ожидание является обобщенной характеристикой, поэтому и не дает возможности однозначного принятия решений.

Рассмотрим основные свойства математического ожидания:

1. Математическое ожидание постоянной величины равно самой постоянной:
$$M(C) = C,\ \text{где}\ C = \text{const} \tag{6.2.2}$$
2. Постоянный множитель можно выносить за знак математического ожидания:
$$M(kX) = kM(X),\ \text{где}\ k = \text{const} \tag{6.2.3}$$
3. Математическое ожидание алгебраической суммы конечного числа случайных величин равно такой же сумме их математических ожиданий:
$$M(X \pm Y) = M(X) \pm M(Y) \tag{6.2.4}$$
4. Математическое ожидание произведения конечного числа независимых случайных величин равно произведению их математических ожиданий:
$$M(XY) = M(X)M(Y) \tag{6.2.5}$$
5. Если все значения случайной величины увеличить (уменьшить) на постоянную , то на эту же величину увеличится (уменьшится) математическое ожидание этой случайной величины:
$$M(X \pm C) = M(X) \pm C, \quad \text{где}\ C = \text{const}. \tag{6.2.6}$$

6.3. Дисперсия дискретной случайной величины

Мерой рассеивания значений случайной величины X, относительно её математического ожидания, является дисперсия[8] случайной величины X.

Дисперсией случайной величины X (будем обозначать $D(X)$ или d_x, D_x) называется математическое ожидание квадрата отклонения случайной величины X от ее математического ожидания:

$$D(X) = M\big(X - M(X)\big)^2 = \sum_{i=1}^{n} (x_i - M(X))^2 p_i \tag{6.3.1}$$

Пример 4. Найдем дисперсии для примеров 1 и 2:

[8] Дисперсия - (от лат. dispersio - рассеяние), в математической статистике наиболее употребительная мера рассеивания, отклонения случайных значений от среднего.

Пример 1:

$$D(X) = (1 - 3{,}44)^2 \cdot 0{,}01 + (2 - 3{,}44)^2 \cdot 0{,}09 + (3 - 3{,}44)^2 \cdot 0{,}35 + (4 - 3{,}44)^2 \cdot 0{,}2 + (5 - 3{,}44)^2 \cdot 0{,}35 = 1{,}2284$$

Пример 2:

$$D_x = (3 - 12)^2 \cdot 0{,}2 + (5 - 12)^2 \cdot 0{,}2 + (12 - 12)^2 \cdot 0{,}2 + (19 - 12)^2 \cdot 0{,}2 + (21 - 12)^2 \cdot 0{,}2 = 52.$$

$$D_y = (9 - 12)^2 \cdot 0{,}2 + (11 - 12)^2 \cdot 0{,}2 + (13 - 12)^2 \cdot 0{,}2 + (14 - 12)^2 \cdot 0{,}2 + (15 - 12)^2 \cdot 0{,}2 = 4{,}8$$

Так как дисперсия имеет размерность квадрата случайной величины, например для второго примера это (дет.)^2, вводят понятие среднего квадратического (стандартного) отклонения, как арифметический корень из дисперсии (будем обозначать σ(X) или σ_x):

$$\sigma(X) = \sqrt{D(X)} \text{ или } \sigma_{\text{x}} = \sqrt{D_x} \qquad (6.3.2)$$

Найдем средние квадратического отклонения для примеров 1 и 2:

Пример 1: $\sigma(X) = \sqrt{1{,}2284} = 1{,}1083.$

Пример 2: $\sigma_x = \sqrt{52} = 7{,}2.$ $\sigma_y = \sqrt{4{,}8} = 2{,}2.$

Таким образом, так как m_x=m_y, а σ_y=σ_x, предпочтение следует отдать второму токарю. При равных средних значениях, у него стабильность в работе, так как характеристика колеблемости – среднее квадратическое отклонение количества изготовленных деталеq – намного меньше, чем у первого токаря.

В экономике, финансах и менеджменте для анализа обычно используют так называемый коэффициент вариации, представляющий собой отношение среднего квадратического отклонения случайной величины к математическому ожиданию:

$$k_{var} = \frac{\sigma(X)}{M(X)} \text{ или } k_{var} = \frac{\sigma(X)}{M(X)} \cdot 100\%. \qquad (6.3.3)$$

Основные свойства дисперсии

1. Дисперсия постоянной величины равна нулю:

$$D(C) = 0, \quad \text{где } C = \text{const}. \tag{6.3.4}$$

2. Постоянный множитель можно выносить за знак дисперсии, возводя ее при этом в квадрат:

$$D(kX) = k^2 D(X), \text{ где } k = \text{const}. \tag{6.3.5}$$

3. Дисперсия алгебраической суммы конечного числа случайных величин равна сумме их дисперсий:

$$D(X \pm Y) = D(X) + D(Y). \tag{6.3.6}$$

4. Дисперсия случайной величины равна разности между математическим ожиданием квадрата случайной величины и квадратом ее математического ожидания:

$$D(X) = M(X^2) - M^2(X). \tag{6.3.7}$$

Математическое ожидание, дисперсия, среднее квадратическое отклонение и другие числа, призванные в сжатой форме отражать наиболее существенные черты распределения случайной величины, называются числовыми характеристиками.

6.4. Функция распределения

Функцией распределения случайной величины X $F(x)$ называется вероятность события ($X<x$) такого, что случайная величина X принимает значение, меньшее некоторого заданного (конкретного) значения :

$$F(x) = P(X < x). \tag{6.4.1}$$

На рисунке 6.4.1 представлена геометрическая интерпретация функции распределения $F(x)$, то есть случайная величина X принимает значения, лежащие левее точки x на оси абсцисс, следовательно, функция распределения непрерывна слева.

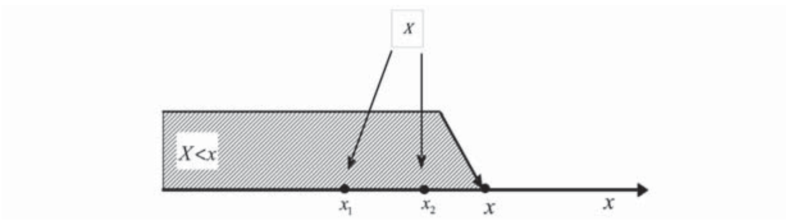


Рис. 6.4.1. К понятию функции распределения

Из определения функции распределения вытекают, очевидны следующие свойства функции распределения:

1. Вероятность попадания случайной величины X левее $(-\infty)$ – вероятность невозможного события – равна нулю $\Rightarrow F(-\infty)=P(X<-\infty)= P(\varnothing)=0$.

2. Вероятность попадания случайной величины X во всю числовую ось – вероятность достоверного события – равна единице $\Rightarrow F(+\infty)=P(X<+\infty)= P(\Omega)=1$.

3. При $x_2>x_1$, $F(x_2)\geq F(x_1)$ – функция распределения – неубывающая функция аргумента x.

4. Вероятность попадания случайной величины в промежуток $(x_1;x_2)$ равна приращению функции распределения на этом промежутке:

$$P(x_1 \leq X < x_2) = F(x_2) - F(x_1). \qquad (6.4.2)$$

Следовательно, выражение для нахождения функции распределения дискретной случайной величины можно представить следующим образом:

$$F(x) = p(X < x) = \begin{cases} 0, & \text{при } x \leq x_1 \\ p_1, & \text{при } x_1 < x \leq x_2 \\ p_1 + p_2, & \text{при} x_2 < x \leq x_3 \\ p_1 + p_2 + p_3, & \text{при } x_3 < x \leq x_4 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & \\ p_1 + p_2 + \ldots + p_{n-1}, & \text{при } x_{n-1} < x \leq x_n \\ 1, & \text{при } \ x > x_n. \end{cases} \qquad (6.4.3)$$

Пример 5. Пусть дан ряд распределения дискретной случайной величины:

X:	x_i	1	2	3	4	5
	p_i	0.2	0.1	0.3	0.15	0.25

Вычислим значения функции распределения $F(x)$ по формуле (6.4.3) и построим её график (рис. 6.4.2).

$$F(x) = P(x \leq 1) = 0;$$

$$F(x) = P(1 < x \leq 2) = 0{,}2;$$

$$F(x) = P(2 < x \leq 3) = 0{,}2 + 0{,}1 = 0{,}3;$$

$$F(x) = P(3 < \mathrm{x} \leq 4) = 0{,}2 + 0{,}1 + 0{,}3 = 0{,}6;$$

$$F(x) = P(4 < \mathrm{x} \leq 5) = 0{,}2 + 0{,}1 + 0{,}3 + 0{,}15 = 0{,}75;$$

$$F(x) = P(\mathrm{x} > 5) = 0{,}2 + 0{,}1 + 0{,}3 + 0{,}15 + 0{,}25 = 1.$$

Функция распределения

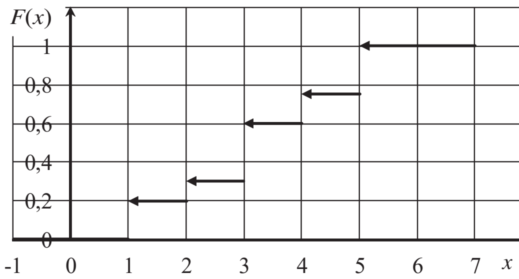


Рис. 6.4.2

Из определения анализа свойств $F(x)$ и примера 5 следует, что по мере увеличения значения аргумента растет значение функции распределения – суммируются вероятности p_i отдельных значений x_i случайной величины X. На этом основании функцию распределения $F(x)$ называют **интегральным законом распределения** случайной величины X.

Ясно, что в пределе при увеличении числа значений ($n \to \infty$), величины скачков стремятся к нулю – функция распределения $F(x)$ становится непрерывной функцией (рис. 6.4.3).

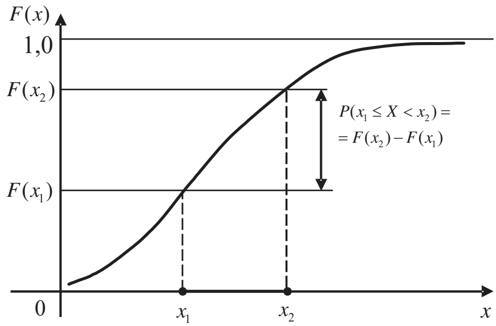


Рис. 6.4.3

В заключение заметим, что функция распределения универсальна, так как она существует для дискретных и непрерывных случайных величин – в первом случае она имеет вид скачкообразно изменяющейся разрывной функции со скачками, равными p_i в точках x_i, а во втором случае – $F(x)$ является непрерывной функцией. В обоих случаях изменяется в пределах от нуля до единицы.

6.5. Плотность вероятности

Плотность вероятности (плотность распределения) $f(x)$ случайной величины X имеет смысл только для непрерывных случайных величин и определяется как предел отношения вероятности попадания случайной величины X на отрезок к длине этого отрезка, когда последняя стремится к нулю:

$$f(x) = \lim_{\Delta x \to 0} \frac{P\big(X \in (x, x + \Delta x)\big)}{\Delta x}\,. \tag{6.5.1}$$

Выразим вероятность попадания случайной величины X на отрезок $[x, x+\Delta x]$ по формуле (6.4.2) как приращение $F(x)$ на этом отрезке:

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{dF(x)}{dx} = F'(x). \tag{6.5.2}$$

Плотность вероятности $f(x)$ равна производной функции распределения $F(x)$, поэтому получила название **дифференциальный закон распределения**.

Свойства плотности вероятности вытекают из её определения (6.5.1):

1. Плотность вероятности – неотрицательная функция аргумента x, график функции всегда расположен выше оси OX:

$$f(x) \geq 0. \tag{6.5.3}$$

Первое свойство следует из определения (6.5.1): в числителе стоит вероятность события, лежащая в пределах от нуля до единицы, а в знаменателе – длина отрезка – величина положительная.

2. Вероятность попадания непрерывной случайной величины в промежуток $[x_1. x_2)$ равна определенному интегралу от ее плотности вероятностей в этих пределах:

$$P(x_1 \leq X < x_2) = \int_{x_1}^{x_2} f(x)\,dx. \tag{6.5.4}$$

Вероятность попадания на отрезок геометрически иллюстрируется площадью под кривой распределения (рис. 6.5.1), ограниченной слева и справа ординатами в точках $x_1. x_2$ (заштрихована на рис. 6.5.1).

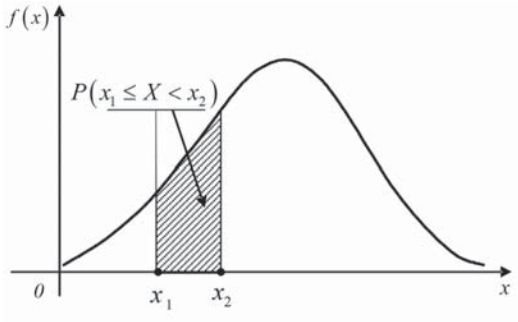


Рис. 6.5.1.

3. Функция распределения непрерывной случайной величины равна интегралу от плотности вероятностей в пределах от до:

$$F(x) = P(X < x) = P(-\infty < X < x) = \int_{-\infty}^{x} f(x)\,dx. \qquad (6.5.5)$$

Геометрическая интерпретация функции распределения $F(x)$, основанная на соотношении (6.5.5), показана на рис. 6.5.2 и представляет собой площадь под кривой распределения $f(x)$, лежащая левее ординаты в точке x (заштрихована на рис.).

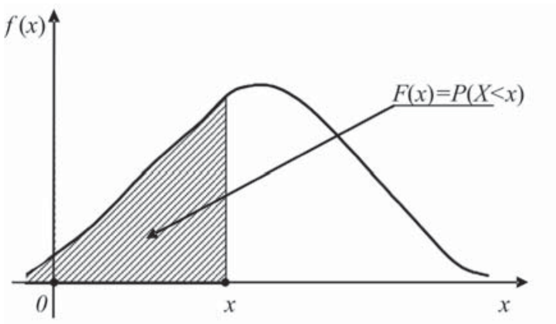


Рис. 6.5.2. Функция распределения

4. Интеграл в бесконечных пределах от плотности вероятности равен единице:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \qquad (6.5.6)$$

Очевидно, вероятность попадания случайной величины X в бесконечный отрезок (во всю числовую ось) есть вероятность достоверного события, равная единице:

$$P\big(X \in (-\infty, +\infty)\big) = \int_{-\infty}^{\infty} f(x)dx = P(\Omega) = 1.$$

Математическим ожиданием непрерывной случайной величины с плотностью вероятности $f(x)$, называется число:

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx. \qquad (6.5.7)$$

Свойства математического ожидания для дискретной случайной величины, остаются справедливыми и для непрерывных случайных величин, например:

$$M(CX) = \int_{-\infty}^{\infty} Cxf(x)dx = C\int_{-\infty}^{\infty} xf(x)dx = CM(X). \qquad (6.5.8)$$

Дисперсией непрерывной случайной величины с плотностью вероятности $f(x)$, называется число:

$$D(X) = \int_{-\infty}^{\infty} \big(x - M(X)\big)^2 f(x)dx. \qquad (6.5.9)$$

Формула для дисперсии может быть записана в более удобном виде для вычислений:

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \big(M(X)\big)^2. \qquad (6.5.10)$$

Свойства дисперсии для дискретной случайной величины, остаются справедливыми и для непрерывных случайных величин.

Среднее квадратическое или стандартное отклонение определяется из формулы:

$$\sigma(X) = \sqrt{D(X)}. \qquad (6.5.11)$$

Из свойств дисперсии вытекают соответствующие свойства средне1о квадратического отклонения:

1. $\sigma(C)=0$.
2. $\sigma(CX)=|C|\sigma(X)$.
3. $\sigma(C+X)=\sigma(X)$.

В завершении отметим, что для изучения свойств случайного явления, независящих от выбора масштаба измерения и положения центра группировки, заданную случайную величину X приводят к некоторому стандартному виду: ее центрируют, рассматривая разность (X-$M(X)$), геометрически это означает перенос начала координат в точку с абсциссой равной математическому ожиданию, впоследствии делят на среднее квадратическое отклонение $\sigma(X)$.

Полученную случайную величину:

$$Z=\frac{X-M(X)}{\sigma(X)}, \tag{6.5.12}$$

называют стандартной случайной величиной. Математическое ожидание стандартной случайной величины равно 0, а дисперсия 1. Покажем это:

$$M(Z)=M\left(\frac{X-M(X)}{\sigma(X)}\right)=\frac{1}{\sigma(X)}M\big(X-M(X)\big)=0,$$

$$D(Z)=\frac{1}{\sigma^2(X)}D\big(X-M(X)\big)=\frac{D(X)}{\sigma^2(X)}=1$$

Таким образом, Z-центрированная $M(Z)=0$ и нормированная ($D(Z)=1$) случайная величина.

6.6. Моменты случайных величин

Кроме математического ожидания и дисперсии случайной величины в теории вероятностей для описания некоторых свойств распределений случайных величин пользуются **моментными характеристиками** или **моментами**.

Начальным моментом -го порядка случайной величины X называется математическое ожидание k-й степени случайной величины X:

$$\alpha_k(X)=M(X^k). \tag{6.6.1}$$

Соответствующие формулы для вычисления начальных моментов дискретных и непрерывных случайных величин имеют вид:

$$\alpha_k(X)=\sum_{i=1}^{n}x_i^k\,p_i; \tag{6.6.2}$$

$$\alpha_k[X] = \int_{-\infty}^{\infty} x^k f(x)dx \cdot \tag{6.6.3}$$

Очевидно, что **математическое ожидание** случайной величины X не что иное, как первый начальный момент случайной величины:

$$\alpha_1(X) = M(X). \tag{6.6.4}$$

Отметим второй начальный момент, который будет использован ниже:

$$\alpha_2(X) = M(X^2). \tag{6.6.5}$$

Центральным моментом -го порядка случайной величины X называется математическое ожидание k-й степени центрированной случайной величины X, $\overset{o}{X}=X-m_x$:

$$\mu_k(X) = M\left(\overset{o}{X^k}\right). \tag{6.6.6}$$

Соответствующие формулы для вычисления центральных моментов дискретных и непрерывных случайных величин имеют вид:

$$\mu_k(X) = \sum_{i=1}^{n} (x_i - m_x)^k p_i; \tag{6.6.7}$$

$$\mu_k(X) = \int_{-\infty}^{\infty} (x - m_x)^k f(x)dx. \tag{6.6.8}$$

Первый центральный момент равен нулю:

$$\mu_1(X) = \int_{-\infty}^{\infty} (x - m_x) f(x)dx = \int_{-\infty}^{\infty} x f(x)dx - m_x \int_{-\infty}^{\infty} f(x)dx =$$

$$= m_x - m_x = 0.$$

Второй центральный момент, очевидно, **равен дисперсии** случайной величины X (4.3.6):

$$\mu_2(X) = M\left(\overset{o}{X^2}\right) = D(X). \tag{6.6.9}$$

В дальнейшем будем использовать третий и четвертый центральные моменты случайной величины X, которые записываются в виде:

$$\mu_3(X) = M\left(\overset{o}{X^3}\right), \quad \mu_4(X) = M\left(\overset{o}{X^4}\right). \tag{6.6.10}$$

Существует связь между начальными и центральными моментами случайной величины X. Получим, например, формулы связи центральных и начальных моментов различных порядков для непрерывной случайной величины X (аналогичный вывод будет и для дискретной случайной величины).

Рассмотрим второй центральный момент (дисперсию случайной величины):

$$\mu_2 = M\left(\overset{o}{X}{}^2\right) = D(X) = \int_{-\infty}^{\infty} (x - m_x)^2 f(x)dx = \tag{6.6.11}$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - 2m_x \int_{-\infty}^{\infty} xf(x)dx + m_x^2 \int_{-\infty}^{\infty} f(x)dx =$$

$$= \alpha_2(X) - 2m_x^2 + m_x^2 = \alpha_2(X) - m_x^2$$

Для третьего центрального момента, действуя аналогично, получим:

$$\mu_3(X) = M\left(\overset{o}{X}{}^3\right) = \int_{=\infty}^{\infty} (x - m_x)^3 f(x)dx =$$

$$= \int_{-\infty}^{\infty} x^3 f(x)dx$$

$$- 3m_x \int_{-\infty}^{\infty} x^2 f(x)dx + 3m_x^2 \int_{-\infty}^{\infty} xf(x)dx - m_x^3 \int_{-\infty}^{\infty} f(x)dx =$$

$$= \alpha_3(X) - 3m_x\alpha_2(X) + 2m_x^3$$

Выражения для центральных моментов порядков больше трех можно получить аналогично. В итоге получается сводка формул:

$$\begin{cases} \mu_1 = 0, \\ \mu_2 = \alpha_2 - m_x^2, \\ \mu_3 = \alpha_3 - 3m_x\alpha_2 + 2m_x^3, \\ \dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots \end{cases} \tag{6.6.12}$$

Формула, выражающая второй центральный момент через второй и первый начальные моменты, часто используется для вычисления дисперсии:

$$D(X) = \alpha_2(X) - m_x^2, \qquad (6.6.13)$$

которая для дискретных и непрерывных случайных величин принимает следующий вид:

$$D(X) = \sum_{i=1}^{n} x_i^2 p_i - m_x^2;$$

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - m_x^2. \qquad (6.6.14)$$

Вернемся к производящей функции, которая рассматривалась в пятой главе и использовалась для вычисления вероятностей событий. С помощью производящей можно находить и числовые характеристики случайных величин.

Основные числовые характеристики дискретных случайных величин с целыми неотрицательными значениями удобно находить с помощью производящих функций.

Пусть дискретная случайная величина X принимает значения 0, 1, 2, ..., k, ... с вероятностями $p_0, p_1, p_2, \ldots, p_k, \ldots$, тогда производящая функция для этой случайной величины имеет вид:

$$\varphi(z) = \sum_{k=0}^{\infty} p_k z^k = p_0 + p_1 z + p_2 z^2 + \cdots, \qquad (6.6.15)$$

где z – произвольный параметр, $0 < z \leq 1$.

Подчеркнем, что коэффициенты степенного ряда (6.6.15) являются вероятностями закона распределения дискретной случайной величины X.

Продифференцируем по z производящую функцию, в результате получим:

$$\varphi'(z) = \sum_{k=0}^{\infty} k p_k z^{k-1}. \qquad (6.6.16)$$

Тогда при $z=1$ получим начальный момент первого порядка, то есть математическое ожидание:

$$\varphi'(1) = \sum_{k=0}^{\infty} k p_k = \alpha_1(X) = M(X). \qquad (6.6.17)$$

Найдем вторую производную функции $\varphi(z)$ и при $z=1$ получим:

$$\varphi^{II}(z) = \sum_{k=0}^{\infty} k(k-1)p_k z^{k-2}. \qquad (6.6.18)$$

$$\varphi^{II}(1) = \sum_{k=0}^{\infty} k^2 p_k - \sum_{k=0}^{\infty} k p_k = \alpha_2(X) - \alpha_1(X), \qquad (6.6.19)$$

где $\alpha_2(X)$ и $\alpha_1(X)$ – начальные моменты соответственно второго и первого порядков.

Так как $\alpha_2(X)=M(X^2)$ найдем дисперсию $D(X)$:

$$D(X) = M(X^2) - \big(M(X)\big)^2 = \alpha_2(X) - \alpha_1^2(X) =$$

$$= \big(\alpha_2(X) - \alpha_1(X)\big) + \alpha_1(X) - \alpha_1^2(X).$$

Тогда:

$$D(X) = \varphi^{II}(1) + \varphi^{I}(1) - \big(\varphi^{I}(1)\big)^2. \qquad (6.6.20)$$

Третий центральный момент используется для описания **асимметрии** (скошенности) кривых распределения случайных величин, которая вычисляется по формуле:

$$S_k = \frac{\mu_3(X)}{\sigma^3(X)} \qquad (6.6.21)$$

Две асимметричные кривые распределения представлены на рис. 6.6.1.

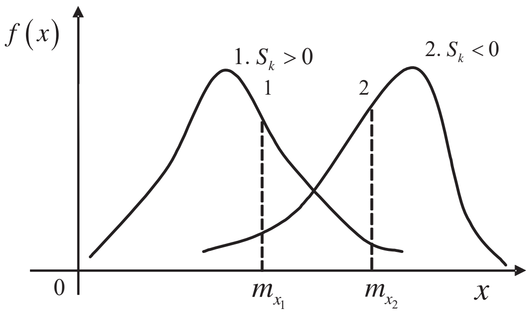


Рис. 6.6.1. Асимметрия (скошенность) кривых распределения

Четвертый центральный момент используется для описания крутости (остро или плосковершинности) кривых распределения, которая характеризуется так называемыми **эксцессом**. Эксцесс вычисляется по формуле:

$$E_x = \frac{\mu_4(X)}{\sigma^4(X)} - 3. \qquad (6.6.22)$$

Для нормального закона распределения, который является наиболее широко используемым на практике, величина $\mu_4/\sigma^4=3$. Поэтому для нормальной кривой распределения $E_x=0$. Все остальные кривые других распределений сравниваются с нормальной кривой (рис. 6.6.2).

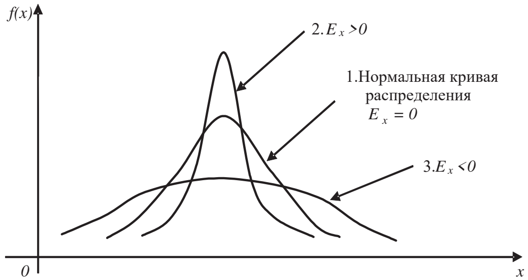


Рис. 6.6.2. Эксцесс (крутость) кривых распределения

При использовании различных методов математической статистики, особенно различных статистических критериев и методов построения интервальных оценок неизвестных параметров, широко используются понятия q– квантилей x_q и $100q$- процентных точек ω_q распределения $F(x)$.

Квантилем уровня q (или q - квантилем) непрерывной случайной величины X с функцией распределения $F(x)$ называется такое значение x_q этой случайной величины, для которого вероятность события ($X< x_q$) равна заданной величине q:

$$F(x_q) = P(X < x_q) = q. \qquad (6.6.23)$$

Частным случаем квантиля – 0,5-квантилем – является характеристика центра рассеивания значений случайной величины – медиана (M_e).

Квартилем уровня q называется -квантиль при фиксированных значениях q= 0,25; 0,50;...; 0,75, кратных четверти единицы. Квартиль $x_{0,5}$ является медианой (M_e).

Децилем уровня q называется -квантиль при фиксированных значениях q= 0,10; 0,20;...; 0,90, кратных десятой доле единицы.

Под $100q\%$-ной точкой случайной величины X понимают такое её значение ω_q, для которого вероятность события ($X \geq \omega_q$) равна q:

$$1 - F(\omega_q) = P(X \geq \omega_q) = q. \tag{6.6.24}$$

Процентилем уровня q называется -квантиль при фиксированных значениях q=0,01; 0,02; ...; 0,99, кратных сотой доле единице (проценту).

Из определений квантилей и процентных точек (6.6.23) и (6.6.24) вытекает их взаимное соотношение:

$$x_q = \omega_{1-q}. \tag{6.6.25}$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какая случайная величина называется дискретной?
2. Какая случайная величина называется непрерывной?
3. Что является законом распределения дискретной случайной величины?
4. Что является законом распределения непрерывной случайной величины?
5. В каком случае дискретная случайная величина называется полностью заданной с вероятностной точки зрения?
6. В каком случае непрерывная случайная величина называется полностью заданной с вероятностной точки зрения?
7. Что такое математическое ожидание дискретной случайной величины, что оно характеризует, и по какой формуле оно определяется?
8. Что такое математическое ожидание непрерывной случайной величины, что оно характеризует, и по какой формуле оно определяется?
9. Что понимается под дисперсией дискретной случайной величины, что она характеризует, и по какой формуле она определяется?
10. Что понимается под дисперсией непрерывной случайной величины, что она характеризует, и по какой формуле она определяется?
11. Что такое среднее квадратическое отклонение?
12. Что понимается под начальными моментами дискретной случайной величины?
13. Что понимается под начальными моментами непрерывной случайной величины?
14. Что понимается под центральным моментом дискретной случайной величины?
15. Что понимается под центральным моментом непрерывной случайной величины?
16. Какова связь между дисперсией и вторым начальным моментом?

17. Что понимается под медианой непрерывной случайной величины, и что характеризует медиана? Как найти значение медианы по заданной плотности распределения или функции распределения?
18. Что понимается под интегральной функцией распределения, и каковы ее свойства?
19. Что понимается под дифференциальной функцией распределения?
20. Каковы свойства дифференциальной функции распределения?

ЗАДАНИЯ

6.1. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины , зная закон ее распределения:

X:

x_i	0	1	2	3	4
p_i	0,2	0,4	0,3	0,08	0,02

6.2. Найти функцию распределения, числовые характеристики m_x, D_x, σ_x данной случайной величины и $P(0<X<1{,}5)$.

X:

x_i	2	5	6	8
p_i	0,1	0,3	0,2	0,4

6.3. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X, зная закон ее распределения:

X:

x_i	1	2	3	4
p_i	p	0,1	$2p$	0,4

6.4. Случайная величина задана плотностью распределения $f(x)=2x$ в интервале (0, 1); вне этого интервала $f(x)=0$. Найти m_x, D_x, σ_x.

6.5. Непрерывная случайная величина X подчинена закону распределения с плотностью:

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

Найти числовые характеристики m_x, D_x, σ_x данной случайной величины и $P(0<X<1{,}5)$.

6.6. Дискретная случайная величина задана рядом распределения:

X:

x_i	0	1	2	3	10
p_i	0,2	0,2	0,3	0,2	0,1

Требуется найти функцию распределения, математическое ожидание и дисперсию случайной величины X.

6.7. Найти функцию распределения, среднее квадратическое отклонение случайной величины X, заданной рядом распределения:

X:

x_i	1	2	3	4	5
p_i	0,1	0,2	0,3	0,2	0,2

6.8. При каждом независимом испытании вероятность пригодности детали равна 0,9. Найти ряд распределения вероятностей, математическое ожидание, дисперсию и среднее квадратическое отклонение числа пригодных деталей среди взятых наудачу 3 деталей.

6.9.Непрерывная случайная величина задана функцией распределения:

$$F(x) = \begin{cases} 0, & \text{при} \quad x \leq 0, \\ sin\, x & \text{при} \quad 0 < x \leq \dfrac{\pi}{2}, \\ 1 & \text{при} \quad x \leq \dfrac{\pi}{2}. \end{cases}$$

Требуется найти плотность ее распределения и вероятность попадания случайной величины X в интервал:

$$\left(\frac{\pi}{4}, \frac{\pi}{2}\right)$$

6.10. Случайная величина X задана плотностью распределения:

$$f(x) = \begin{cases} 0, & \text{при} \quad x < 0, \\ 3x^2 & \text{при} \quad 0 \leq x \leq 1, \\ 0 & \text{при} \quad x > 1. \end{cases}$$

Требуется найти функцию распределения случайной величины X, ее математическое ожидание, дисперсию и среднее квадратическое отклонение.

6.11. Непрерывная случайная величина подчинена закону распределения с плотностью:

$$f(x) = \begin{cases} 0 \quad \text{при } x < 1 \text{ и } x > 4; \\ a(4 - x) \text{ при } 1 \le x \le 4. \end{cases}$$

Найти коэффициент a, функцию распределения $F(x)$, $P(X>3{,}5)$ и $P(X\le3)$, $P(2<X<5)$, .

6.12. В лотерее выпущено 100 билетов. Разыгрывался один выигрыш в 50 у.е. и десять выигрышей по 10 у.е. Найти закон распределения величины X - стоимости возможного выигрыша - и построить график.

6.13. Два баскетболиста делают по одному броску мяча в корзину, причем вероятности попаданий равны 0,8 и 0,6 для каждого игрока. Соответственно, общее число попаданий – случайная величина X . Составьте закон распределения случайной величины X, найдите функцию распределения $F(x)$ и постройте график функции $F(x)$.

6.14. Стрелок, имея 3 патрона, стреляет в цель до первого попадания, или пока не израсходует все патроны. Вероятность попадания при каждом выстреле равна 0,3. Число израсходованных патронов – случайная величина X. Составьте ее закон распределения, найдите функцию распределения $F(x)$.

6.15. Дана дифференциальная функция распределения непрерывной случайной величины X:

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

Требуется найти интегральную функцию распределения $F(x)$, а также вероятность попадания случайной величины X в интервал:

$$\left(\frac{5}{4}, \frac{7}{4}\right)$$

по формулам функции и плотности распределения.

Сравнить полученные результаты.

ГЛАВА 7
ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

7.1. Биноминальное распределение

В пунктах 7.1, 7.2 и 7.3 рассмотрим законы распределения дискретных случайных величин.

Дискретная случайная величина X имеет биномиальный закон распределения, если она принимает целочисленные значения $\{0, 1, 2, \ldots, m, \ldots, n\}$ с вероятностями, определяемыми по формуле Бернулли (см. п. 5.1):

$$P(X = m) = P_n(m) = C_n^m p^m \tag{7.1.1}$$

где q=1-p, m=0...n.

Ряд распределения биномиального закона имеет вид:

X:	x_i	0	1	2	...	m	...	n
	p_i	q^n	$C_n^1 pq^{n-1}$	$C_n^2 p^2 q^{n-2}$	...	$C_n^m p^m q^{n-m}$	...	p^n

Многоугольники распределения для n=10 при различных значениях p представлены на рис. 7.1.1.

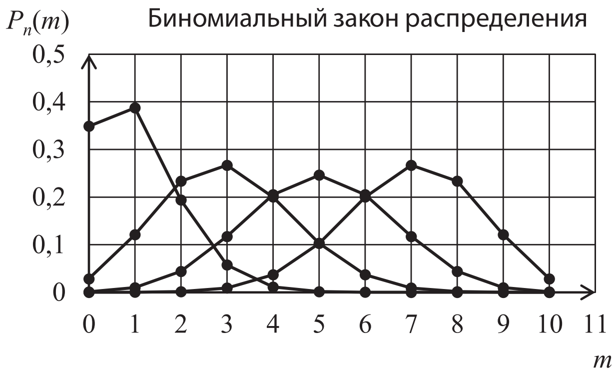


Рис. 7.1.1. Многоугольники биноминального распределения

Максимум вероятности $P_n(m)$ достигается при значениях m_0, удовлетворяющих условию:

$$np - q \leq m_0 \leq np + p \qquad (7.1.2)$$

такое значение m_0 называется наивероятнейшим значением (см. п. 5.1).

Математическое ожидание числа появлений события A в n независимых испытаниях, проводимых в одинаковых условиях, то есть распределенных по биномиальному закону, равно:

$$M(X) = np \qquad (7.1.3)$$

дисперсия равна:

$$D(X) = np(1-p) = npq \qquad (7.1.4)$$

среднее квадратическое отклонение равно:

$$\sigma(X) = \sqrt{npq} \qquad (7.1.5)$$

Доказательство: Дискретная случайная величина X_1 в одном испытании имеет только два возможных значения, например: x_1=1 и x_2=0, тогда:

X_1:

x_i	1	0
p_i	p	q

$$M(X_1) = 1 \cdot p + 0 \cdot q = p, \quad M(X_1^2) = 1^2 \cdot p + 0^2 \cdot q = p.$$

$$D(X_1) = M(X_1^2) - \left(M(X_1)\right)^2 = p - p^2 = p(1-p) = pq.$$

Далее на основании свойства математического ожидания (см. п. 6.2) для n независимых случайных величин $X_1, X_2, X_3, \ldots, X_n$, находим:

$$M(X_1 + X_2 + X_3 + \cdots + X_n) = \sum_{i=1}^{n} M(X_i) = \sum_{i=1}^{n} p = np.$$

Используя свойство дисперсии (см. п. 6.3), получим:

$$D(X_1 + X_2 + X_3 + \cdots + X_n) = \sum_{i=1}^{n} D(X_i) = \sum_{i=1}^{n} pq = npq.$$

Таким образом, справедливость формул (7.1.3) и (7.1.4) доказаны.

Биномиальный закон распределения широко используется в теории и практике статистического контроля качества продукции, при описании функционирования систем массового обслуживания, в теории стрельбы и в других областях.

Примеры случайных величин, распределенных по биноминальному закону:

1. Число гербов (цифр), выпавших при бросании n монет на гладкую поверхность;

2. Число попаданий в мишень при стрельбе n независимыми выстрелами в одинаковых условиях;

3. Число отказавших технических устройств из общего количества, подвергавшихся испытаниям в течение определенного времени в одинаковых условиях.

Пример 1. В данный магазин смартфоны поступают двух типов в соотношении 3:7. Куплено пять смартфонов. Составить закон распределения случайной величины X – числа смартфонов второго типа среди купленных. Найти математическое ожидание, дисперсию, и среднее квадратическое отклонение этой величины, вероятность того, что среди купленных смартфонов не мене трех будет второго типа.

Решение. Воспользуемся биномиальным законом распределения. Исходные данные согласно условию задачи:

$$n = 5 \quad p = \frac{7}{10} = 0{,}7, \quad q = 1 - p = 1 - 0{,}7 = 0{,}3$$

По формуле (7.1.1) находим вероятности:

$$P_5(0) = q^5 = 0{,}3^5 = 0{,}002,$$

$$P_5(1) = C_5^1 p q^4 = 5 \cdot 0{,}7 \cdot 0{,}3^4 = 0{,}028,$$

$$P_5(2) = C_5^2 p^2 q^3 = 10 \cdot 0{,}7^2 \cdot 0{,}3^3 = 0{,}133,$$

$$P_5(3) = C_5^3 p^3 q^2 = 10 \cdot 0{,}7^3 \cdot 0{,}3^2 = 0{,}309,$$

$$P_5(4) = C_5^4 p^4 q = 5 \cdot 0{,}7^4 \cdot 0{,}3 = 0{,}360,$$

$$P_5(5) = P^5 = 0{,}7^5 = 0{,}168$$

Убеждаемся, что:

$$\sum_{m=0}^{5} P_5(m) = 0{,}002 + 0{,}028 + 0{,}133 + 0{,}309 + 0{,}360 + 0{,}168 = 1{,}000$$

Ряд распределения с. в. X имеет следующий вид:

X:

m	0	1	2	3	4	5
$P_5(m)$	0,002	0,028	0,133	0,309	0,360	0,168

Математическое ожидание, дисперсию и среднее квадратическое отклонение находим по формулам (7.1.3, 7.1.4, 7.1.5):

$$m_x = np = 5 \cdot 0{,}7 = 3{,}5,$$

$$D_x = npq = 5 \cdot 0{,}7 \cdot 0.3 = 1{,}05,$$

$$\sigma_x = \sqrt{D_x} = 1{,}025$$

Вероятность того, что среди купленных смартфонов не мене трех будет второго типа определяется как:

$$P(X \geq 3) = P_5(3) + P_5(4) + P_5(5) = 0{,}309 + 0{,}360 + 0{,}168 = 0{,}837$$

7.2. Распределение Пуассона

Дискретная случайная величина X имеет распределение Пуассона[9], если она принимает целочисленные значения $\{0, 1, 2, ..., m, ...\}$ (бесконечное, но счетное множество значений) с вероятностями:

$$P(X = m) = P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}, \qquad (7.2.1)$$

где λ – параметр распределения.

Ряд распределения закона Пуассона имеет вид:

X:

x_i	0	1	2	...	m
p_i	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\lambda^2/2 \cdot e^{-\lambda}$	...	$\lambda^m/m! \cdot e^{-\lambda}$

---

[9] Пуассон Семеон Дени (1781...1840)- французский математик, механик, физик, профессор Парижского университета. В теории вероятностей сформулировал теорему Пуассона – частный случай закона больших чисел и одну из предельных теорем. Предложил распределение с. в., названное его именем – распределение Пуассона, одно из важнейших в теории вероятностей. Почетный член Петербургской АН (1826).

Многоугольники биномиального распределения для различных значений параметра представлены на рис. 7.2.1.

Наивероятнейшее значение m_0 определяется из условия:

$$\lambda\text{-}1 \leq m_0 \leq \lambda. \tag{7.2.2}$$

Если λ – целое число, то вероятность $P_n(m)$ достигает максимума при двух значениях m, равных $m=\lambda-1$ и $m=\lambda$, если $\lambda<1$, то $P_n(m)$ – максимальна при $m=0$, наконец, если λ не целое число ($\lambda>1$), то $P_n(m)$ – максимальна при $m=[\lambda]$ – целой части λ.

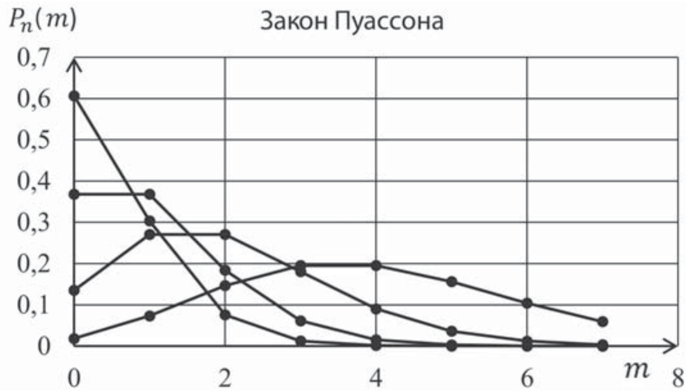


Рис. 7.2.1. Многоугольники распределения Пуассона

Примеры случайных величин, распределенных по закону Пуассона:

1. Число распадов радиоактивного вещества за определенное время;

2. Число элементарных космических частиц, попадающих на данную площадь за определенное время (в космический аппарат);

3. Число заявок на телефонные переговоры на телефонной станции за определенное время.

Математическое ожидание и дисперсия дискретной случайной величины, распределенной по закону Пуассона, совпадают и равны параметру закона распределения:

$$M(X)=\lambda, \quad D(X)=\lambda. \tag{7.2.3}$$

Доказательство начнем с математического ожидания:

$$M(X) = \sum_{k=1}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \left\{\begin{matrix} \dfrac{k}{(k-1)!\,k} = \dfrac{1}{(k-1)!} \\ m = k-1, 0! = 1, k = m+1 \end{matrix}\right\} \Rightarrow$$

$$= e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^{m+1}}{m!} = \lambda e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^{m}}{m!} \underset{\text{разложение в степенной ряд}}{=\!=\!=\!=\!=\!=\!=\!\Longrightarrow} \left\{e^x = \sum_{k=0}^{n}\frac{x^k}{k!}\right\} =$$

$$= \lambda e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^{m}}{m!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

Таким образом, $M(X)=\lambda$.

Найдем $M(X^2)$:

$$M(X^2) = \sum_{k=1}^{\infty}k^2\frac{\lambda^k e^{-\lambda}}{k!} = \left\{\underset{\text{умножим на}}{=\!=\!=\!=\!\Longrightarrow}\frac{\lambda}{\lambda}\right\} = \lambda\sum_{k=1}^{\infty}k\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!} = \left\{\underset{k+1-1}{=\!=\!\Longrightarrow}\right\} =$$

$$= \lambda\sum_{k=1}^{\infty}\big((k-1)+1\big)\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!} = \lambda\left[\sum_{k=1}^{\infty}(k-1)\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!} + \sum_{k=1}^{\infty}\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!}\right] =$$

$$= \{m = k-1\} = \lambda\left[\sum_{m=0}^{\infty}m\frac{\lambda^m e^{-\lambda}}{m!} + e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}\right] = \left\{\begin{matrix}\displaystyle\sum_{m=0}^{\infty}m\frac{\lambda^m e^{-\lambda}}{m!} = \lambda \\ \displaystyle\sum_{m=0}^{\infty}\frac{\lambda^m}{m!} = e^{\lambda}\end{matrix}\right\} =$$

$$= \lambda\big(\lambda + e^{\lambda}e^{-\lambda}\big) = \lambda(\lambda+1) = \lambda^2 + \lambda$$

Следовательно:

$$D(X)=M(X^2) - (M(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Таким образом, $D(X)=\lambda$ а среднее квадратическое отклонение равно: $\sigma(X)=\sqrt{\lambda}$.

Распределение Пуассона может быть получено из биномиального распределения путем предельного перехода при $n\to\infty$, $p\to 0$, при условии $np=$ const и этом смысле рассматривается как **закон редких явлений**.

Рассмотрим случайную величину X, распределенную по биномиальному закону с параметрами n, p и математическим ожиданием:

$$M(X)=np=\lambda,$$

и вероятностью появления события A в каждом опыте:

$$p = \frac{\lambda}{n}.$$

Вероятность события $A=\{(X=m)\}$ при n независимых опытах в одинаковых условиях определим по формуле (7.1.1):

$$P(A) = P(X = m) = C_n^m \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m}. \qquad (7.2.4)$$

Преобразуем это выражение к виду, удобному для предельного перехода при $n\to\infty$:

$$\lim_{n\to\infty} P(X = m) = \lim_{n\to\infty} \frac{n(n-1)\ldots(n-m+1)(n-m)!}{n^m(n-m)!} \cdot \frac{a^m}{m!} \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^m} =$$

$$= \lim_{n\to\infty} \frac{n(n-1)\ldots(n-m+1)}{n^m} \cdot \frac{\lambda^m}{m!} \cdot \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^m}$$

Очевидно, что знаменатель при $n\to\infty$ и конечном m стремится к единице, так как λ/n есть бесконечно малая величина $n\to\infty$:

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^m = \left\{\begin{matrix} \dfrac{\lambda}{n} \xrightarrow[\text{при } n\to\infty]{} 0 \\ m - \text{конечна} \end{matrix}\right\} = 1^m = 1$$

Остается рассмотреть предел числителя:

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = \langle 1^\infty \rangle = \left\{\begin{matrix} \dfrac{1}{t} = -\dfrac{\lambda}{n} \\ n = -\lambda t \\ n \longrightarrow \infty \Rightarrow \\ t \longrightarrow \infty \end{matrix}\right\} = \lim_{n\to\infty} \left(1 + \frac{1}{t}\right)^{-\lambda t} = \left[\lim_{t\to\infty} \left(1 + \frac{1}{t}\right)^t\right]^{-\lambda} =$$

$$= \left\{\begin{matrix} \lim\limits_{t\to\infty} \left(1 + \dfrac{1}{t}\right)^t = e \\ \text{второй замечательный предел} \end{matrix}\right\} = e^{-\lambda}$$

Следовательно:

$$P(X = m) = P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Таким образом, распределение Пуассона является предельным случаем для биномиального распределения.

Замечание:

На практике закон Пуассона используется, когда $n \to \infty$, $n \to 0$, при условии $0{,}1 \le np \le 10$.

Пример. Ткач обслуживает 1000 веретен. Вероятность обрыва нитки на одном из веретен в течении одной минуты равна 0,002. Найти вероятность того, что в течении одно минуты обрыв произойдет:

а) на 3 веретенах;

б) не менее трех;

в) хотя бы на одном.

Решение. Поскольку количество испытаний велико (n=1000), а вероятность отдельного испытания очень мала (p=0,002) то для вычисления искомой вероятности воспользуемся формулой Пуассона (7.2.1):

Параметр распределения тогда $\lambda = np = 0{,}002 \cdot 1000 = 2$, тогда:

а) $P(X = 3) = \frac{2^3}{3!}e^{-2} = \frac{8}{6}e^{-2} = \frac{4}{3e^2} = 0{,}1805$ или по табл. 2 приложения.

б) $P(X \ge 3) = 1 - \big(P(X = 0) + P(X = 1) + P(X = 2)\big) = 1 - (0{,}1353 + 0{,}2707 + 0{,}2707 = 0{,}3233.$

в) $P(X \ge 1) = 1 - P(X = 0) = 1 - 0{,}1353 = 0{,}8647.$

7.3. Геометрическое распределение

Пусть проводятся независимые испытания с двумя исходами, в каждом из которых вероятность появления события A равна p, $(0 < p < 1)$, а вероятность не появления события $\bar{A}$ соответственно $q = 1 - p$. Испытания заканчиваются, как только появится событие A. Пусть в первых $(m - 1)$ испытаниях событие A не появилось, в m-м испытании появилось. Тогда вероятность события (X=m) равна

$$P(A_m) = P(X = m) = q^{m-1}p. \tag{7.3.1}$$

Ряд распределения случайной величины X, имеющей геометрическое распределение имеет вид:

X:

x_i	1	2	2	...	m	...
p_i	p	qp	q^2p	...	pq^{m-1}	...

Проверка:

$$\sum_{k=0}^{\infty} pq^{m-1} = p\,\frac{1}{1-q} = \frac{p}{p} = 1.$$

Вероятности случайной величины X, вычисленные по формуле (7.3.1), представляют собой **геометрическую прогрессию** с первым членом равным p и знаменателем q, ($0<q<1$): p, qp, q^2p, ... , pq^{m-1} ... . Поэтому распределение (7.3.1) называют **геометрическим**.

Примеры геометрического распределения случайных величин:

1. Число бросаний монеты до первого появления герба;
2. Число попыток запустить двигатель самолета, отказавший в полете;
3. Число выстрелов по цели до первого попадания и т.д.

Найдем математическое ожидание и дисперсию геометрического распределения.

Воспользуемся производящей функцией, согласно (6.6.15), она имеет вид:

$$\varphi(z) = \sum_{k=0}^{\infty} q^{m-1}pz^m = \sum_{k=0}^{\infty} q^{m-1}pzz^{m-1} = pz\sum_{k=0}^{\infty}(qz)^{m-1} = pz\,\frac{1}{1-qz} \Rightarrow$$

$$\varphi(z) = \frac{pz}{1-qz} \qquad (7.3.2)$$

Найдем первую и вторую производные функции (7.3.2):

$$\varphi^{I}(z) = \left(\frac{pz}{1-qz}\right)^{I} = \frac{p(1-qz) - pz(-q)}{(1-qz)^2} = \frac{p - qzp + qzp}{(1-qz)^2} = \frac{p}{(1-qz)^2}.$$

$$\varphi^{II}(z) = \left(\frac{p}{(1-qz)^2}\right)^{I} = -p\,\frac{-2q(1-qz)}{(1-qz)^4} = \frac{2pq}{(1-qz)^3}$$

Следовательно, по формулам (6.6.17) и (6.6.20) (см. п. 6.6):

$$M(X) = \varphi^{I}(1) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

$$X) = \varphi^{II}(1) + \varphi^{I}(1) - \left(\varphi^{I}(1)\right)^2 = \frac{2pq}{(1-q)^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{2pq}{p^3} + \frac{1}{p} - \frac{1}{p^2}$$

$$= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q + p - 1}{p^2} = \frac{2q - (1-p)}{p^2} = \frac{q}{p^2}$$

Подведем итог:

$$M(X) = \frac{1}{p},$$

$$D(X) = \frac{q}{p^2},$$

$$\sigma(X) = \frac{\sqrt{q}}{p} \qquad (7.3.5)$$

Пример 3. Вероятность попадания в цель при одном выстреле для стрелка равна 0,2.

Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X – числа выстрелов по цели до первого попадания.

Решение Случайная величина X имеет геометрическое распределение с параметром p=0,2. По формулам (7.3.3), (7.3.4), (7.3.5):

$$M(X) = \frac{1}{0{,}2} = 5.$$

$$D(X) = \frac{0{,}8}{(0{,}2)^2} = 20.$$

$$\sigma(X) = \sqrt{20} = 2\sqrt{5}$$

В последующих параграфах рассмотрим законы распределения непрерывных случайных величин.

7.4. Равномерное распределение

Непрерывная случайная величина X имеет равномерный закон распределения на отрезке $[a,b]$, если ее плотность вероятности $f(x)$ постоянна на этом отрезке и равна нулю вне ее, то есть:

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & \text{при } a\le x\le b,\\ 0, & \text{при } x<0,\ x>b.\end{cases} \qquad (7.4.1)$$

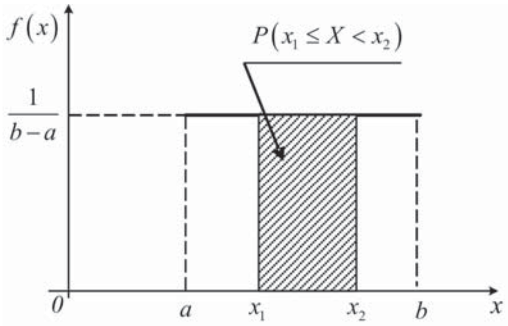


Рис. 7.4.1. Плотность вероятности равномерного распределения

Найдем функцию распределения $F(x)$ по формуле (6.5.5):
При $x<a$ функция распределения $F(x)=0$.
При $a \le x \le b$ по формуле:

$$F(x)=\int_a^x \frac{1}{b-a}dx=\frac{x}{b-a}\Big|_a^x=\frac{x-a}{b-a}$$

При $x > b$ ясно, что:

$$F(x)=\int_a^b \frac{1}{b-a}dx=\frac{x}{b-a}\Big|_a^b=\frac{b-a}{b-a}=1.$$

Таким образом:

$$F(x)=\begin{cases}0 & \text{при } x<0,\\ \dfrac{x-a}{b-a} & \text{при } a\le x\le b\\ 1 & \text{при } x>b.\end{cases} \qquad (7.4.2)$$

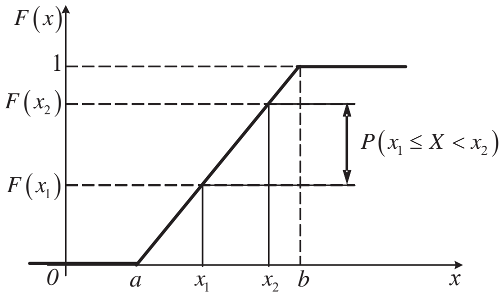


Рис. 7.4.2. Функция распределения равномерного распределения

Найдем математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X, распределенной равномерно по формулам (6.5.7), (6.5.9) и (6.5.11):

$$M(X) = \int_{-\infty}^{a} x \cdot 0dx = \int_{a}^{b} \frac{x}{b-a} dx + \int_{b}^{\infty} x \cdot 0dx = \frac{1}{b-a} \int_{a}^{b} xdx = \frac{1}{b-a} \frac{x^2}{2} \bigg|_{a}^{b} =$$

$$= \frac{1}{2(b-a)}(b^2 - a^2) = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.$$

$$D(X) = \left(\int_{-\infty}^{a} \left(x - \frac{a+b}{2} \right)^2 \cdot 0dx \right.$$

$$\left. = \int_{a}^{b} \left(x - \frac{a+b}{2} \right)^2 \left(\frac{1}{b-a} \right) dx + \int_{b}^{\infty} \left(x - \frac{a+b}{2} \right)^2 \cdot 0dx \right) =$$

$$= \frac{1}{b-a} \int_{a}^{b} \left(x - \frac{a+b}{2} \right)^2 dx = \frac{1}{(b-a)} \frac{1}{3} \left(x - \frac{a+b}{2} \right)^3 \bigg|_{a}^{b} =$$

$$= \frac{1}{3(b-a)} \left(\left(\frac{2b-a-d}{8} \right)^3 - \left(\frac{2a-a-b}{8} \right)^3 \right) =$$

$$= \frac{1}{3(b-a)}\left(\frac{(b-a)^3}{8} - \frac{(a-b)^3}{8}\right) = \frac{1}{3(b-a)}\left(\frac{(b-a)^3}{8} + \frac{(b-a)^3}{8}\right) =$$

$$= \frac{1}{3(b-a)}\left(\frac{(b-a)^3}{8} + \frac{(b-a)^3}{8}\right) = \frac{1}{3(b-a)}\frac{2(b-a)^3}{8} = \frac{(b-a)^2}{12}$$

Таким образом, для равномерно распределенной непрерывной случайной величине имеем:

$$M(X) = \frac{a+b}{2}, \quad (7.4.3)$$

$$D(X) = \frac{(b-a)^2}{12}, \quad (7.4.4)$$

$$\sigma(X) = \frac{b-a}{2\sqrt{3}}. \quad (7.4.5)$$

Равномерные распределения случайных величин применяются в основном при имитационном моделировании ремонтных работ, загрузки оборудования, динамики расхода средств, в управлении запасами.

Пример 4. Поезда метрополитена идут по расписанию с интервалом две минуты. Пассажир выходит на платформу в случайный момент времени. Написать выражение для плотности вероятности, функции распределения случайной величины X – времени ожидания поезда. Найти математическое ожидание и среднее квадратическое отклонение этой величины. Какова вероятность того, что ждать пассажиру придется не больше полминуты?

Решение. Случайная величина X – время ожидания поезда подчинена равномерному закону распределения, где a=0, b=2, тогда по формулам (7.4.1), (7.4.2), (7.4.3), (7.4.4), (7.4.5):

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

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

$$M(X)=\frac{0+2}{2}=1$$

$$D(X)=\frac{(2-0)^2}{12}=\frac{4}{12}=\frac{1}{3}$$

$$\sigma(X)=\frac{2-0}{2\sqrt{3}}=\frac{1}{\sqrt{3}}$$

Вероятность того, что ждать придется не более полминуты по формуле (6.4.2):

$$P\left(0\le X\le\frac{1}{2}\right)=F\left(\frac{1}{2}\right)-F(0)=\frac{1}{2}\cdot\frac{1}{2}-\frac{1}{2}\cdot 0=\frac{1}{4}=0{,}25$$

7.5. Показательный закон распределения

Непрерывная случайная величина X имеет показательный (экспоненциальный) закон распределения, если ее плотность вероятности имеет вид:

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & \text{при } x\ge 0,\\ 0, & \text{при } x<0,\end{cases} \tag{7.5.1}$$

где $\lambda>0$ – параметр закона распределения.

Функция распределения показательного закона распределения имеет вид:

$$F(x)=\begin{cases}0, \text{ при } x\le 0\\ 1-e^{-\lambda x}, \text{ при } x>0\end{cases} \tag{7.5.2}$$

Покажем это по формуле (6.5.5):

$$F(x)=\int_{-\infty}^{0}0\cdot dt+\int_{0}^{x}f(t)dt=\int_{0}^{x}\lambda e^{-\lambda t}dt=-\frac{\lambda}{\lambda}\int_{0}^{x}e^{-\lambda t}d(-\lambda t)=-e^{-\lambda t}\Big|_{0}^{x}=$$

$$-e^{-\lambda t}\Big|_{0}^{x}=-\left(e^{-\lambda x}-e^{-\lambda\cdot 0}\right)=-\left(e^{-\lambda x}-1\right)=1-e^{-\lambda x}$$

Графики плотности и функции показательного распределения представлены на рис. 7.5.1, 7.5.2.

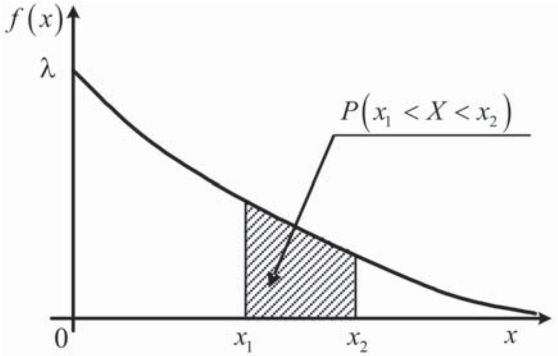


Рис. 7.5.1. Плотность вероятности показательного закона

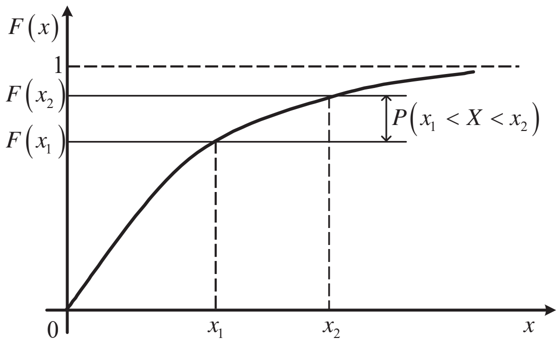


Рис. 7.5.2. Функция распределения показательного закона

Вероятность попадания случайной величины X, распределенной по показательному закону на интервал (x_1, x_2), можно определить как приращение функции распределения на (x_1, x_2) (рис. 7.5.2) формула (6.4.2):

$$P(x_1 < X < x_2)) = F(x_2) - F(x_1) = e^{-\lambda x_2} - e^{-\lambda x_1} \,. \quad (7.5.3)$$

На графике плотности вероятности (рис. 7.5.1) вероятность попадания на отрезок выражается в виде площади под кривой $f(x)$, ограниченной ординатами точек x_1 и x_2.

По формулам (6.5.7), (6.5.9) и (6.5.11) найдем числовые характеристики непрерывной случайной величины X распределенной по показательному закону:

$$M(X)=\int\limits_0^\infty xf(x)dx=\lambda\int\limits_0^\infty xe^{-\lambda x}dx=\lim_{b\to\infty}\int\limits_0^b xe^{-\lambda x}dx=$$

$$=\{\text{интегрируем по частям}\}=\left\{\begin{array}{c} u=x \\ du=dx \\ dv=e^{-\lambda x} \\ v=-\frac{\lambda}{\lambda}\int e^{-\lambda x}d(-\lambda x)=-e^{-\lambda x}\end{array}\right\}=$$

$$=\lim_{b\to\infty}\left(-xe^{-\lambda x}\Big|_0^b+\int\limits_0^b e^{-\lambda x}dx\right)=\lim_{b\to\infty}\left(-xe^{-\lambda x}\Big|_0^b-\frac{1}{\lambda}e^{-\lambda x}\Big|_0^b\right)=$$

$$=\lim_{b\to\infty}\left(-xe^{-\lambda x}\Big|_0^b-\frac{1}{\lambda}e^{-\lambda x}\Big|_0^b\right)=0-\frac{1}{\lambda}(0-1)=\frac{1}{\lambda}.$$

$$D(X)=\int\limits_0^\infty x^2f(x)dx-\big(M(X)\big)^2=\lambda\int\limits_0^\infty x^2e^{-\lambda x}dx-\frac{1}{\lambda^2}=$$

$$=\lambda\lim_{b\to\infty}\left(\int\limits_0^b x^2e^{-\lambda x}dx\right)-\frac{1}{\lambda^2}=\{\text{дважды интегрируем по частям}\}=$$

$$=\lambda\lim_{b\to\infty}\left(\left(-\frac{x^2}{\lambda}e^{-\lambda x}+\frac{2}{\lambda}\left(-\frac{\mathrm{x}}{\lambda}e^{-\lambda x}-\frac{1}{\lambda^2}e^{-\lambda x}\right)\right)\Bigg|_0^b\right)-\frac{1}{\lambda^2}=$$

$$=\lambda\left(0+\frac{2}{\lambda}\left(0+0-\frac{1}{\lambda^2}(0-1)\right)\right)-\frac{1}{\lambda^2}=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$$

Следовательно, математическое ожидание непрерывной случайной величины *X*, распределенной по показательному закону равно:

$$M(X) = \frac{1}{\lambda} \ , \qquad (7.5.4)$$

дисперсия:

$$D(X) = \frac{1}{\lambda^2} \, , \qquad (7.5.5)$$

среднее квадратическое отклонение:

$$\sigma(X) = \frac{1}{\sqrt{\lambda^2}} = \frac{1}{\lambda} \ . \qquad (7.5.6)$$

Таким образом, математическое ожидание и среднее квадратическое отклонение обратно пропорциональны параметру распределения λ, а дисперсия квадрату параметра распределения λ^2.

Как видно из графика плотности вероятности, показательное распределение имеет вид плавно снижающейся экспоненты и применяется в основном в организации производства и в планировании обслуживания коммуникаций. Например, считается, что по показательному закону распределены длительности выполнения некоторых технологических операций, интервалы времени между последовательными отказами оборудования, время прибытия клиентов и транспорта и т.д.

Пример 6. Установлено, что время ремонта телевизоров есть случайная величина *X*, распределена по показательному закону. Определить вероятность того, что на ремонт телевизора потребуется не менее 20 дней, если среднее время ремонта телевизоров составляет 15 дней. Найти плотность вероятности, функцию распределения и среднее квадратическое отклонение случайной величины *X*

Решение. Дано $M(X) = 15$.

По формуле (7.5.4):

$$M(X) = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{M(X)} = \frac{1}{15}.$$

Напишем выражения для плотности вероятности и функции распределения по формулам (7.5.1) и (7.5.2):

$$f(x) = \begin{cases} \frac{1}{15} e^{-\frac{1}{15}x}, \text{ при } x \geq 0, \\ 0, \qquad \text{при } x < 0. \end{cases}$$

$$F(x) = \begin{cases} 0, \text{ при } x \leq 0; \\ 1 - e^{-\frac{1}{15}x}, \text{ при } x > 0. \end{cases}$$

Согласно формуле (7.5.6) среднее квадратическое отклонение равно математическому ожиданию:

$$\sigma(X) = M(X) = \frac{1}{15}\cdot$$

Вероятность того, что на ремонт телевизора потребуется не менее 20 дней найдем с помощью функции распределения через вероятность противоположного события

$$P(X \geq 20) = 1 - P(X < 20) = 1 - F(20) = 1 - \left(1 - e^{-\frac{20}{15}}\right) = e^{-\frac{20}{15}} = 0{,}264$$

7.6. Нормальный закон распределения

Нормальный закон распределения (закон Гаусса[10]) наиболее часто встречается на практике и играет исключительную роль в теории вероятностей. Основная особенность закона Гаусса состоит в том, что он является предельным законом, к которому приближаются, при определенных условиях, остальные законы распределения.

Непрерывная случайная величина X распределена по нормальному закону с параметрами a и $\sigma>0$, если ее плотность вероятности имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in R\,. \tag{7.6.1}$$

Большинство случайных величин, используемых в задачах практики, как, например, ошибки измерений, ошибки изготовления технических устройств, ошибки стрельбы и многие другие, могут быть представлены как суммы большого числа элементарных ошибок, возникающих независимо от других под действием различных факторов. Особенности распределений этих случайных ошибок в сумме взаимно нивелируются так, что сумма подчиняется нормальному закону распределения, а также по нормальному закону могут распределяться:

- количество работающих на предприятии;
- процент брака;
- точности изготовления деталей и сборочных единиц;

[10] Карл Фридрих Гаусс (1777 ... 1855)- выдающийся немецкий математик, астроном, геодезист. Профессор, директор обсерватории Гёттингенского университета (1807...1855). Получил блестящие результаты в высшей алгебре, теории чисел, дифференциальной геометрии, геодезии, небесной механике, теоретической астрономии, теории электричества и магнетизма. Разработал метод наименьших квадратов (1820), закон распределения ошибок при точных наблюдениях в геодезии, астрономии (закон Гаусса).

- параметры технологического процесса изготовления;
- рост человека, случайно выбранного из большой группы;
- величины окружающей среды в производстве (температура, влажность, электромагнитное поле, напряжение, сила тока) и другие.

Нормальное (или гауссовское) распределение с параметрами a и σ обозначается: $N\in(a, \sigma)$.

Плотность вероятности зависит от двух параметров a, σ (представлена на рис. 7.6.1) и называется кривой распределения (кривой) Гаусса. Кривая Гаусса имеет специфический симметричный холмообразный вид. Максимальное значение плотности вероятности достигается при $x=a$ (рис. 7.6.1).

$$f_{max}(a) = \frac{1}{\sigma\sqrt{2\pi}} \tag{7.6.2}$$

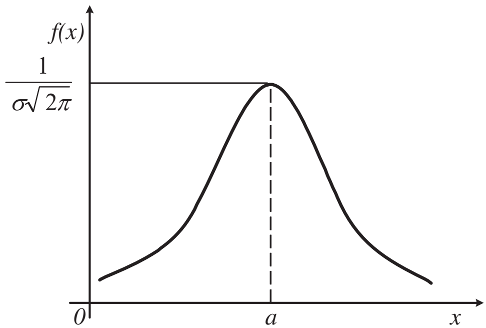


Рис. 7.6.1. Нормальная плотность вероятности (кривая Гаусса)

Покажем, что выполняется свойство (6.5.6) плотности вероятности:

$$\int_{-\infty}^{\infty} f(x) = 1.$$

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-a}{\sqrt{2}\sigma}\right)^2} dx = \left\{ \begin{array}{c} z = \dfrac{x-a}{\sqrt{2}\sigma} \\ x = \sqrt{2}\sigma z + a \\ dx = \sqrt{2}\sigma dz \end{array} \right\} =$$

$$= \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x-a}{\sqrt{2}\sigma}\right)^2} d\left(\frac{x-a}{\sqrt{2}\sigma}\right) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2} dz = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1$$

Замечание. Здесь воспользовались интегралом Эйлера-Пуассона:

$$\int_{-\infty}^{\infty} e^{-z^2} dz = \sqrt{\pi}.$$

Из интеграла Эйлера-Пуассона следует:

$$\int_{0}^{\infty} e^{-z^2} dt = \frac{\sqrt{\pi}}{2}, \quad \int_{0}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{\frac{\pi}{2}} = \int_{-\infty}^{0} e^{-\frac{z^2}{2}} dz.$$

Функция распределения непрерывной случайной величины (рис. 7.6.2) $N \in (a, \sigma)$ задается выражением:

$$F(x) = \int_{-\infty}^{x} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(z-a)^2}{2\sigma^2}} dz$$

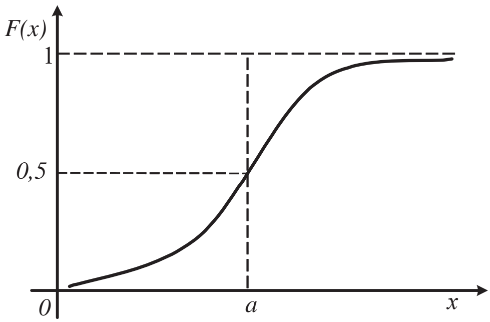


Рис. 7.6.2. Нормальная функция распределения.

Если $a = 0$ и $\sigma = 1$, то нормальное распределение называется стандартным и плотность вероятности данной величины имеет вид:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad (7.6.3)$$

а функция распределения (приложение 2, табл.2) случайной величины $N \sim N(0,1)$ выражается как:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz. \qquad (7.6.4)$$

Функция $\varphi(x)$ приведена в табличной форме в приложении 1 (табл.1) и обладает следующими свойствами:

1. $f(x)$ – четная, то есть $f(-x)= f(x)$;
2. $f(x)$ – симметрична относительно 0Y;
3. $f(x) \approx 0$, если $x > 0$;
4. $\varphi_{max}(x) = \varphi(0) = \frac{1}{\sqrt{2\pi}}$.

Заметим, формула (7.6.4) называется функцией Лапласа и связана с нормированной функцией Лапласа:

$$\Phi(x) = 0{,}5 + \Phi_0(x). \quad (7.6.5)$$

Напомним нормированная функция Лапласа определяется из выражения:

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz. \quad (7.6.6)$$

Функция Лапласа приведена в табличной форме в приложении (приложение 2, табл. 2). Очевидны следующие свойства функции Лапласа:

1. $\Phi_0(0) = 0$;

2. $\Phi_0(+\infty) = 0{,}50$;

3. $\Phi_0(-z) = -\Phi_0(z)$.

Найдем математическое ожидание, дисперсию, среднее квадратическое отклонение нормально распределенной случайной величины и установим связь данных числовых характеристик с параметрами закона $N \in (a, \sigma)$:

$$M[X] = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left\{ \begin{matrix} z = \frac{x-a}{\sqrt{2}\sigma} \\ x = \sqrt{2}\sigma z + a \\ dx = \sqrt{2}\sigma dz \end{matrix} \right\} =$$

$$\frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma z + a) e^{-z^2} dz = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} x e^{-z^2} dz + \frac{a}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2} dz = \frac{a}{\sqrt{\pi}} \cdot \sqrt{\pi} = a.$$

(7.6.7)

Первый интеграл (7.6.7) – интеграл от нечетной функции в симметричных пределах – равно нулю. Второй интеграл – интеграл Эйлера-Пуассона равный $\sqrt{\pi}$ (см. выше).

Таким образом, получили, что параметр равен математическому ожиданию:

$$M(X) = a. \tag{7.6.8}$$

Найдем дисперсию:

$$D[X] = \int_{-\infty}^{\infty} (x-a)^2 f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \{\text{ см. выше}\} =$$

$$= \frac{\sqrt{2}\sigma 2\sigma^2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2} dz = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2} dz = \left\{ \begin{matrix} u = z \\ du = dz \\ dv = ze^{-z^2} dz \\ v = -\frac{1}{2} e^{-z^2} \end{matrix} \right\} =$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \left(-\frac{1}{2} z e^{-z^2} \Big|_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} e^{-z^2} dz \right) = \frac{2\sigma^2}{\sqrt{\pi}} \left((0-0) + \frac{1}{2} \frac{2\sigma^2}{\sqrt{\pi}} \right) = \frac{2\sigma^2\sqrt{\pi}}{2\sqrt{\pi}} = \sigma^2$$

Таким образом:

$$D(X) = \sigma^2, \tag{7.6.9}$$

что означает – второй параметр нормального разложения есть не что иное, как среднее квадратическое отклонение случайной величины X:

$$\sigma(X) = \sqrt{D(X)} = \sigma. \tag{7.6.10}$$

Для оценки моментов нормального распределения, которые будут использоваться ниже, воспользуемся известными результатами. Нечетные центральные моменты нормальной с. в. X все равны нулю:

$$\mu_1 = \mu_3 = \mu_5 = \ldots = 0, \tag{7.6.11}$$

а четные центральные моменты определяются из рекуррентного соотношения:

$$\mu_k = (k-1)\sigma^2 \mu_{k-2} \ (k = 2,4,6,\ldots) \tag{7.6.12}$$

Из (7.6.12) следует:

$$\mu_2 = \sigma^2, \quad \mu_4 = 3\sigma^4, \qquad \mu_6 = 15\sigma^6 \text{ и т. д.}$$

Также очевидно, что асимметрия (скошенность (6.6.21)) и эксцесс (6.6.22) нормального распределения равны нулю, в этом смысле кривая Гаусса является эталоном:

$$S_k = \frac{\mu_3}{\sigma^3} = 0, \quad E_x = \frac{\mu_4}{\sigma^4} - 3 = 0.$$

В точке $x = a$ (точке перегиба) кривая $F(x)$ изменяет темп возрастания (вторая производная становится отрицательной) и далее стремится к единице. На оси ординат координатой точкой перегиба является значение 0,5.

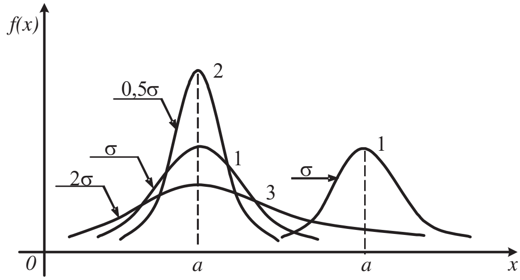


Рис. 7.6.3. Влияние параметров на кривую Гаусса

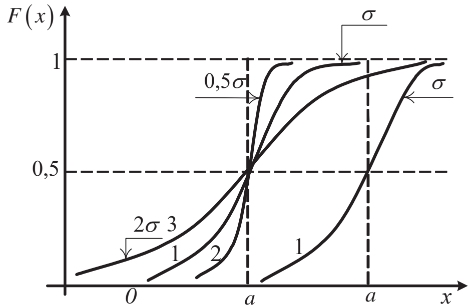


Рис. 7.6.4. Влияние параметров на нормальную функцию распределения

При изменении параметра a кривые $f(x)$ и $F(x)$ смещаются вдоль оси абсцисс, не изменяя своей конфигурации (рис. 7.6.3, 7.6.4).

При изменении параметра σ кривые $f(x)$ и $F(x)$ изменяют свою форму. Кривая Гаусса при уменьшении σ (уменьшении рассеивания случайной величины) вытягивается вверх (см. рис. 7.6.3), а при увеличении σ (увеличении рассеивания случайной величины X) сплющивается, прижимаясь к оси абсцисс, но так, что площадь под кривой Гаусса и осью абсцисс всегда остается равной единице. Это показано на рис. 7.6.3. Кривая нормальной функции распределения (рис. 7.6.4) при уменьшении становится более крутой, а при увеличении σ – более пологой по сравнению с опорной кривой (1).

В качестве вывода сформулируем основные свойства плотности вероятности $f(x)$ нормального закона распределения:

1. $f(x) > 0$, при $x \in R$; график функции расположен выше оси 0X;
2. Ось 0X является асимптотой графика функции $f(x)$, так как $\lim\limits_{n \to \pm\infty} f(x) = 0$;
3. Функция $f(x)$ унимодальная имеет один максимум при $x = a$, равный:

$$f(a) = \frac{1}{\sigma\sqrt{2\pi}};$$

4. График функции $f(x)$ симметричен относительно прямой $x = a$;
5. Точки перегиба графика функции $f(x)$:

$$G_1\left(a - \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right), \quad G_2\left(a + \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right).$$

Определим вероятность попадания нормальной с. в. X на отрезок $[x_1,\ x_2]$. Для этого используем формулу (6.5.4):

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-a)^2}{2\sigma^2}} dx. \qquad (7.6.13)$$

Сделаем замену переменной:

$$z = \frac{x - a}{\sigma} \Rightarrow$$

$x = \sigma z + a$, $dx = \sigma dz$, а пределы интегрирования для новой переменной z равны:

$$z_1 = \frac{x_1 - a}{\sigma}, \quad z_2 = \frac{x_2 - a}{\sigma}.$$

Подставляя новую переменную вместо , получим:

$$P(x_1 \le X \le x_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{Z^2}{2}} dz. \qquad (7.6.14)$$

$$P(x_1 \le X \le x_2) = \frac{1}{\sqrt{2\pi}} \int_{0}^{z_1} e^{-\frac{Z^2}{2}} dz - \frac{1}{\sqrt{2\pi}} \int_{0}^{z_2} e^{-\frac{Z^2}{2}} dz. \qquad (7.6.15)$$

Используя функцию Лапласа (7.6.6) (7.6.15) запишем в виде:

$$P(x_1 \le X \le x_2) = \Phi_0(z_2) - \Phi_0(z_1). \qquad (7.6.16)$$

Если вернуть исходные переменные, то есть подставив в (7.6.16) вместо z_1 и z_2 их выражения через x_1, x_2 и а, σ окончательно получим:

$$P(x_1 \leq X \leq x_2) = \Phi_0\left(\frac{x_2 - a}{\sigma}\right) - \Phi_0\left(\frac{x_1 - a}{\sigma}\right). \quad (7.6.17)$$

Через функцию Лапласа выражается и функция распределения $F(x)$ нормально распределенной случайной величины X:

$$F(x) = \frac{1}{2} + \Phi_0\left(\frac{x - a}{\sigma}\right). \quad (7.6.18)$$

Если имеется функция Лапласа $\Phi_0(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz,$$

$$F(x) = \Phi\left(\frac{x - a}{\sigma}\right), \quad (7.6.19)$$

что непосредственно вытекает из (7.6.5) и (7.6.18), тогда равенство (7.6.17) можно переписать и в виде:

$$P(x_1 \leq X \leq x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right). \quad (7.6.20)$$

Вероятность попадания нормальной случайной величины X с математическим ожиданием $M(X)=m_x$ и средним квадратическим отклонение $\sigma(X)=\sigma x$ на отрезок, симметричный относительно математического ожидания, длина которого $2l$, определяется по формулам (7.6.17) или (7.6.20) в зависимости от используемых табличных функций. Вычислим вероятность попадания случайной величины X в интервал $(m_x - l,\ m_x + l)$ по формуле (7.6.17):

$$P\big((m_x - l) \leq X \leq (m_x - l)\big) = \Phi_0\left(\frac{m_x + l - m_x}{\sigma}\right) - \Phi_0\left(\frac{m_x - l - m_x}{\sigma}\right) =$$

$$P(|X - m_x| < l = 2\Phi_0\left(\frac{l}{\sigma}\right) = 2\Phi_0\left(\frac{l}{\sigma}\right) - 1 \quad (7.6.21)$$

В равенстве (7.6.20) возьмем последовательно: $l=\sigma$, $l=2\sigma$, $l=3\sigma$

$$P(|X - m_x| < \sigma) = 2\Phi_0\left(\frac{\sigma}{\sigma}\right) = 2\Phi_0(1) = 2 \cdot 0{,}3413 = 0{,}683,$$

$$P(|X - m_x| < 2\sigma) = 2\Phi_0\left(\frac{2\sigma}{\sigma}\right) = 2\Phi_0(2) = 2 \cdot 0{,}4772 = 0{,}954,$$

$$P(|X - m_x| < 3\sigma) = 2\Phi_0\left(\frac{3\sigma}{\sigma}\right) = 2\Phi_0(3) = 2 \cdot 0{,}4987 = 0{,}997.$$

Таким образом, можно утверждать, что с вероятностью, близкой к единице (0, 997), все значения случайной величины X, распределенной по нормальному закону, укладываются на отрезке $[m_x \pm 3\sigma]$. Данная закономерность называется **“правилом трех сигм”**.

Пример 7.

Самолет производит посадку на аэродром в сложных метеоусловиях. Ширина взлетно-посадочной полосы (ВПП) l=60 м.

Система посадки обеспечивает процесс приземления самолета на ВПП с точностью в боковом направлении σ_z=12 м без систематической ошибки (m_z=0).

Какова вероятность успешной посадки самолета в пределах ВПП?

Решение.

Обозначим через A – успешную посадку самолета. Событие A происходит, если система обеспечивает выполнение условия , где - боковое уклонение самолета от осевой линии ВПП. Таким образом, искомая вероятность равна (7.6.21):

$$P(A) = P(|Z| < 30м) = 2\Phi_0\left(\frac{30}{12}\right) = 2\Phi_0(2{,}5) = 0{,}988.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. В каких случаях применяется биномиальный закон распределения?
2. Чему равны математическое ожидание, дисперсия и среднее квадратическое отклонение дискретной случайной величины , распределенной по биномиальному закону?
3. Как по-другому называют закон распределения Пуассона и почему?
4. В каких случаях используется закон распределения Пуассона?
5. Что такое геометрическое распределение, и откуда такое название?
6. В чем суть равномерного закона распределения?
7. Можете ли вы назвать управленческие и экономические процессы, имеющие равномерное распределение?
8. Почему закон распределения назвали показательным?
9. В чем суть нормального закона распределения, и почему его так назвали?
10. Что такое правило «трех сигм»?

ЗАДАНИЯ

7.1. Вероятность поражения вирусным заболеванием куста малины равна 0,15. Составить закон распределения числа кустов малины, зараженных вирусом, из четырех посаженных кустов.

7.2. Случайная величина распределена равномерно на отрезке [-3; 2]. Требуется:

1) записать функцию плотности вероятностей случайной величины;
2) построить график функции плотности;
3) найти математическое ожидание и дисперсию;
4) найти вероятность $P(-2 \leq X \leq 1)$.

7.3. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,1. Составить закон распределения числа отказавших элементов в одном опыте, построить многоугольник распределения. Найти функцию распределения и построить ее график. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение дискретной случайной величины.

7.4. Среднее время безотказной работы прибора равно 80 ч. Полагая, что время безотказной работы прибора имеет показательный закон распределения, найти: а) выражение его плотности вероятности и функции распределения; б) вероятность того, что в течение 100 ч. прибор не выйдет из строя.

7.5. Фирма снабжает своей продукцией пять магазинов. От каждого магазина может поступить заявка на очередной день с вероятностью 0,4 независимо от заявок других магазинов.

а) Какова вероятность того, что поступит не более двух заявок?

б) Какова вероятность, что количество поступивших заявок будет лежать в пределах от двух до четырех?

в) Найти наивероятнейшее число заявок в день и вероятность получения именно такого числа заявок?

7.6. Аппаратура содержит 2000 одинаково надежных элементов, вероятность отказа для каждого из которых равна 0,0005. Какова вероятность отказа аппаратуры, если он наступает при отказе хотя бы одного элемента?

7.7. По каналу связи передаются пять сообщений. Каждое сообщение, независимо от других может быть искажено с вероятностью 0,3. Составить ряд распределения числа искажений. Найти среднее значение числа искажений, среднее квадратическое отклонение от среднего значения. Найти вероятность того, что искажений будет не менее двух.

7.8. Текущая цена акции может быть смоделирована с помощью нормального закона распределения с математическим ожиданием 15 у.е. и средним квадратическим отклонением 0,2 у.е. Найти вероятность того, что цена акции не выше 15,4 у.е. и не ниже 15,3 у.е.

7.9. В шкафу находится 9 приборов, из них 5 исправных. Из шкафа наугад берется 4 прибора. Вычислить математическое ожидание числа исправных приборов и построить ряд распределения.

7.10. Непрерывная СВ X распределена нормально и задана плотностью распределения

$$f(x) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(x+2)^2}{32}}.$$

Найти: m_x, D_x, σ_x данной случайной величины и $P(-4 < X < 22)$, построить график функции $f(x)$.

7.11. Дискретная случайная величина X, принимающая четыре возможных значения, распределена по биномиальному закону. Найти дисперсию этой случайной величины, если $M(x)=3,2$.

7.12. Среднее время безотказной работы утюга (наработки на отказ) равно 12 месяцев. Какова вероятность того, что утюг не перегорит за 24 месяца?

7.13. Среднее время безотказной работы телевизора (наработки на отказ) равно 32 месяцев. Какова вероятность того, что телевизор проработает без отказа от 40 до 48 месяцев?

7.14. Непрерывная случайная величина X распределена равномерно и задана плотностью распределения:

$$f(x) = \begin{cases} C, \text{при } x \in [2; 6]; \\ 0, \text{при} x < 2\ , x > 6. \end{cases}$$

Найти значение C, ..., m_x, D_x, σ_x данной случайной величины и $P(3<X<5)$.

7.15. Непрерывная СВ X распределена нормально и задана плотностью распределения:

$$f(x) = \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{(x-10)^2}{8}} .$$

Найти m_x, D_x, σ_x данной случайной величины и $P(12<X<14)$, построить график функции $f(x)$.

7.16. На АТС поступило 1000 звонков от абонентов. Вероятность неправильного соединения равна 0,005. Какова вероятность, что произошло 8 неправильных соединений? Найти наивероятнейшее число неправильных соединений и соответствующую этому вероятность.

7.17. По многолетним статистическим данным известно, что вероятность рождения мальчика равна 0,6. Составить закон распределения случайной величины X – числа мальчиков в семье из 4 детей. Найти m_x и σ_x этой случайной величины.

7.18. Охотник стреляет по дичи до первого попадания, но успевает делать не более четырех выстрелов. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение числа промахов, если вероятность попадания в цель при одном выстреле равна 0,7.

7.19. Из 15 сбербанков 7 расположены за чертой города. Для обследования случайным образом отобрано 6 сбербанков. Какова вероятность того, что среди отобранных сбербанков только один окажется за чертой города?

7.20. Известно, что 15% открывающихся малых предприятий прекращают свою деятельность в течение года. Какова вероятность того, что из пяти малых предприятий не более двух прекратят свою деятельность в течение года? Найти наивероятнейшее число малых предприятий, которые прекратят свою деятельность, и соответствующую этому вероятность.

7.21. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 5 мин. Найти вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее трех минут. Чему равно среднее время ожидания автобуса?

7.22. Минутная стрелка электрических часов перемещается скачком в конце каждой минуты. Найти вероятность того, что в данное мгновение часы покажут время, которое отличается от истинного не более, чем на 20 с.

7.23. Корректура в 500 страниц содержит 500 опечаток. Найти вероятность того, что на странице не меньше трех опечаток.

7.24. Случайная величина X распределена по нормальному закону с математическим ожиданием m_x=4 и дисперсией σ_x=2. Записать ее плотность и найти $P(X \leq 5)$, $P(1 \leq X \leq 5)$.

7.25. Вероятность отказа в работе телефонной станции при каждом вызове равна 0,0012. Поступило 2000 вызовов. Определить вероятность трехразового отказа.

7.26. Продано 1200 лотерейных билетов. Вероятность выиграть квартиру равна 0,005. Какова вероятность того, что два человека получат ключи от квартиры?

7.27. Среднее время безотказной работы батарейки (наработки на отказ) равно 700 часов. Найти вероятность того, что батарейка проработает от 750 до 800 часов.

7.28. В банке оператор тратит на обслуживание одного клиента в среднем 20 минут. Какова вероятность того, что: а) за один час оператор обслужит два клиента; б) менее двух клиентов?

7.29. Среднее время безотказной работы стиральной машины (наработки на отказ) равно 24 месяца. Найти вероятность того, что стиральная машина проработает без отказа от 30 до 36 месяцев.

7.30. Случайная величина имеет показательной распределение с параметром $\lambda=2$ Найти вероятность попадания этой случайной величины в промежуток (-1; 3). Построить график плотности этого распределения и указать на нем фигуру, соответствующую найденной вероятности. Найти математическое ожидание X и показать его на графике. Найти дисперсию и среднее квадратическое отклонение этой случайной величины.

7.31. Непрерывная случайная величина имеет плотность распределения следующего вида: $f(x)=a\cdot\sin x$, $x\in(0,\pi)$; $f(x)=0$, $x\notin(0,\pi)$. Найти неизвестный параметр распределения. Вычислить математическое ожидание, дисперсию и среднее квадратическое отклонение. Показать на графике плотности значения математического ожидания и среднего квадратического отклонения. Найти вероятность попадания значений случайной величины в интервал $[0;\ \pi/4]$, показать на графике эту вероятность.

7.32. На заводе 1000 станков, каждый из которых выходит из строя в течение часа с вероятностью 0,0025. Какова вероятность того, что за смену выйдет из строя не больше 4 станков и ровно 4 станка?

ГЛАВА 8
СИСТЕМЫ СЛУЧАЙНЫХ ВЕЛИЧИН

8.1. Основные понятия

В практической деятельности чаще всего результат испытания зависит ни от одной случайной величины, а от двух и более. Так, например, попадание в цель, лежащую на плоскости, характеризуется двумя координатами – абсциссой и ординатой, а попадание в пространственную цель характеризуется тремя координатами – абсциссой, ординатой и аппликатой. Совокупность двух и более случайных величин, исследуемых в испытаниях, называется системой случайных величин.

Систему случайных величин $X, Y, Z, \ldots, W$ обозначают символически так: $(X, Y, Z, \ldots, W)$.

Систему случайных величин $(X, Y, Z, \ldots, W)$ иначе называют n-мерным случайным вектором, который может быть охарактеризован своей n-мерной функцией распределения:

$$F(x, y, z, \ldots, w) = P(X < x, Y < y, Z < z, \ldots, W < w) \qquad (8.1.1)$$

Функцию $F(x, y, z, \ldots, w)$ также часто называют распределением вектора $(X, Y, Z, \ldots, W)$ или совместным распределением величин $X, Y, Z, \ldots, W$.

8.2. Система двух случайных величин

Системой двух случайных величин называется совокупность двух случайных величин, например, X и Y, рассматриваемых совместно. Двумерная случайная величина (X, Y) геометрически интерпретируется как случайная точка (X, Y) на плоскости $0xy$ или как радиус вектор этой точки (X, Y) (рис. 8.2.1).

Таким образом, под функцией $F(x, y)$ распределения системы двух случайных величин (X, Y) понимается вероятность совместного выполнение двух неравенств:

$$F(x, y) = P(X < x, Y < y). \qquad (8.2.1)$$

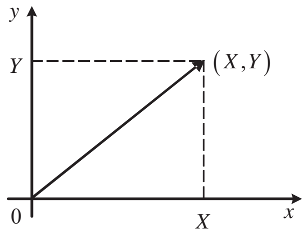


Рис. 8.2.1

Геометрически функция распределения (8.2.2) представляет вероятность попадания случайной точки (X, Y) в бесконечный квадрат с вершиной в точке (x, y), лежащей левее и ниже ее (рис. 8.2.2).

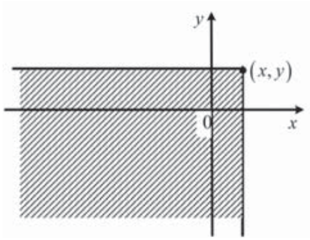


Рис. 8.2.2

Свойства функции распределения двумерной случайной величины

Из определения функции распределения системы двух случайных величин (X, Y) видно, что:

1. Функция распределения $F(x, y)$ – неубывающая функция своих аргументов $X=x, Y=y$ и $0\leq F(x, y)\leq 1$;
2. Если любой из двух или оба аргумента $F(x, y)$ положить равными $(-\infty)$, то функция распределения равна нулю:

$$F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0 \qquad (8.2.2)$$

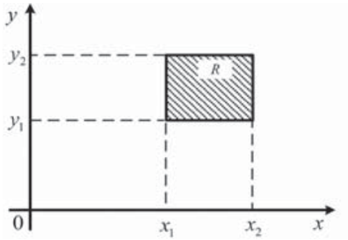
3. При обращении одного из аргументов $F(x, y)$ x или y в $(+\infty)$, функции распределения будут зависеть только от другого аргумента и являются функциями распределения случайных величин X и Y, рассматриваемых отдельно:

$$F(x, +\infty)=F_1(x), \qquad (8.2.3)$$
$$F(+\infty, y)=F_2(y); \qquad (8.2.4)$$

4. При $x=y=+\infty$ функция распределения $F(x, y)$ равна единице:

$$F(+\infty, +\infty) = 1; \tag{8.2.5}$$

5. Вероятность попадания случайной точки (X, Y) в прямоугольник R можно вычислить как (рис. 9.1.3):



$$P\big((X,Y) \in R\big) = P(x_1 \le X < x_2,\ y_1 \le Y < y_2) =$$
$$= F(x_2, y_2) + F(x_1, y_1) - F(x_2, y_1) - F(x_1, y_2) \quad . \tag{8.2.6}$$

8.3. Плотность вероятности системы двух случайных величин

Двумерная плотность вероятности $f(x, y)$ выражается через вторую смешанную частную производную двумерной функции распределения $F(x, y)$:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = F''_{xy}(x,y), \tag{8.3.1}$$

при условии, $F(x, y)$ что есть дважды непрерывно дифференцируемая функция.

Свойства:

1. $f(x, y) \ge 0$

2. Вероятность попадания случайной точки в некоторую область D на плоскости $0xy$ можно вычислить следующим образом:

$$P\big((X,Y) \in D\big) = \iint\limits_D f(x,y)dxdy; \tag{8.3.2}$$

3. Ясно, что:

$$F(x,y) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x,y)dxdy = 1; \tag{8.3.3}$$

4. Кроме того:

$$F_1(x) = F(x, +\infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x, y) dx dy, \quad (8.3.4)$$

$$F_2(y) = F(+\infty, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f(x, y) dx dy. \quad (8.3.5)$$

При этом:

$$f_1(x) = \frac{\partial F_1(x)}{\partial x} = \int_{-\infty}^{\infty} f(x, y) dy, \quad (8.3.6)$$

$$f_2(y) = \frac{\partial 2(y)}{\partial x} = \int_{-\infty}^{\infty} f(x, y) dx, \quad (8.3.7)$$

являются плотностями вероятности непрерывных случайных величин X и Y, рассматриваемых отдельно.

Условный закон распределения случайной величины X из системы (X, Y) есть закон распределения X, вычисленный при условии, что другая величина Y приняла определенное фиксированное значение. Обозначают через:

$F(x|y)$ и $F(y|x)$ – условные функции распределения;

$f(x|y)$ и $f(y|x)$ – условные плотности распределения.

При этом справедливо соотношение:

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

Случайные величины X и Y называются независимыми, если условный закон распределения одной из них не зависит от того, какое значение примет другая. В противном случае X и Y называются зависимыми.

Понятно, что случайные величины X и Y независимы только в том случае, когда:

$$f(x|y) = f_1(x) \text{ и } f(y|x) = f_2(y) \Rightarrow f(x, y) = f_1(x) f_2(y) \quad (8.3.8)$$

Аналогично, $F(x, y) = F_1(x)\ F_2(y)$. (8.3.9)

8.4. Числовые характеристики системы двух случайных величин

Если в системе (X, Y) случайная величина X рассматривается отдельно от составляющей Y, тогда для дискретной X математическое ожидание и дисперсия будут определяться следующим образом:

$$m_x = M(X) = \sum_i \sum_j x_i p_{ij}, \tag{8.4.1}$$

$$D_x = D(X) = \sum_i \sum_j (x_i - m_x)^2 p_{ij}, \tag{8.4.2}$$

где $p_{ij}=P(X=x_i,\ Y=y_j)$ есть вероятность того, что случайная величина X примет значение x_i и одновременно случайная величина Y примет значение y_j.

Для непрерывной случайной величины X числовые характеристики будут определяться через соответствующие интегралы:

$$m_x = \int_{-\infty}^{\infty} x f_1(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy, \tag{8.4.3}$$

$$D_x = \int_{-\infty}^{\infty} (x - m_x)^2 f_1(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)^2 f(x,y) dx dy, \tag{8.4.4}$$

где $f(x, y)$ – совместная плотность распределения системы (X, Y).

Среднее квадратическое отклонение:

$$\sigma_x = \sigma(X) = \sqrt{D_x}. \tag{8.4.5}$$

Аналогично, если в системе (X, Y) случайная величина Y рассматривается отдельно от составляющей X. Тогда для дискретной Y:

$$m_y = M(Y) = \sum_i \sum_j y_j p_{ij}, \tag{8.4.6}$$

$$D_y = D(Y) = \sum_i \sum_j \left(y_j - m_y\right)^2 p_{ij}. \tag{8.4.7}$$

Для непрерывной Y:

$$m_y = \int_{-\infty}^{\infty} y f_2(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy, \tag{8.4.8}$$

$$D_y = \int_{-\infty}^{\infty} (y - m_y)^2 f_2(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - m_y)^2 f(x,y) dx dy, \tag{8.4.9}$$

где $f(x, y)$ – совместная плотность распределения системы (X, Y).

Среднее квадратическое отклонение:

$$\sigma_y = \sigma(Y) = \sqrt{D_y}\,. \tag{8.4.10}$$

Одной из основных характеристик системы случайных величин (X, Y) является второй смешанный центральный момент $\mu_{1,1} = M\left(\overset{o}{X}\overset{o}{Y}\right)$, называемый корреляционным моментом случайных величин X и Y:

$$K_{x,y} = M\left(\overset{o}{X}\overset{o}{Y}\right) = M\left((X - m_x)(Y - m_y)\right). \tag{8.4.11}$$

Для дискретных и непрерывных случайных величин X и Y соответственно будем иметь:

$$K_{x,y} = \sum_i \sum_j (x_i - m_x)(y_j - m_y) p_{ij}. \tag{8.4.12}$$

$$K_{x,y} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} (x - m_x)(y - m_y) f(x,y) dy. \tag{8.4.13}$$

Корреляционный момент характеризует, помимо рассеивания случайных величин X и Y еще и связь между ними. Например, для независимых случайных величин K_{xy}=0.

Безразмерной величиной, характеризующей связь между случайными величинами X и Y, является коэффициент корреляции:

$$r_{x,y} = \frac{K_{x,y}}{\sigma_x \cdot \sigma_x}\,, \tag{8.4.14}$$

где σ_x, σ_x— средние квадратические отклонения величин X и Y.

Коэффициент корреляции $r_{x,y}$ характеризует только линейную зависимость между случайными величинами X и Y.

Следует отметить, что из независимости случайных величин X и Y следует равенство $r_{xy}=0$ или $K_{x,y}=0$. Обратное утверждение не всегда имеет место, то есть из нек,оррелированности случайных величин X и Y еще не следует их независимость: отсутствие линейной зависимости не исключает наличие нелинейной зависимости.

Если случайные величины X и Y связаны тесной линейной зависимостью:

$$Y=aX+b,$$

то $r_{x,y}=\pm 1$, причем знак плюс или минус совпадает со знаком a.

В общем случае, когда X и Y связаны произвольной вероятностной зависимостью: $-1< r_{x,y}<1$.

В случае, когда $r_{x,y}>0$ говорят о положительной корреляции, а при $r_{x,y}<0$ отрицательной корреляции.

Свойства числовых характеристик системы двух случайных величин и

1. Математическое ожидание суммы двух случайных величин равно сумме их математических ожиданий:

$$M(X+Y)=M(X)+M(Y); \tag{8.4.15}$$

2. Математическое ожидание линейной функции:

$$\sum_{i=1}^{n}(a_iX_i+b_i), \tag{8.4.16}$$

где a_i и b_i действительные числа равно:

$$M\left(\sum_{i=1}^{n}a_iX_i+b_i\right)=\sum_{i=1}^{n}a_i\cdot M(X_i)+b_i; \tag{8.4.17}$$

3. Дисперсия суммы двух случайных величин равна сумме их дисперсий плюс удвоенный корреляционный момент:

$$D(X+Y)=D(X)+D(Y)+2K_{x,y}; \tag{8.4.18}$$

4. Дисперсия линейной функции:

$$\sum_{i=1}^{n}(a_iX_i+b_i), \tag{8.4.19}$$

где a_i и b_i действительные числа равна:

$$D\left(\sum_{i=1}^{n} a_i X_i + b_i\right) = \sum_{i=1}^{n} a_i^2 \cdot D(X_i) + 2\sum_{i<j} a_i\, a_j K_{x,y}. \qquad (8.4.20)$$

8.5. Равномерный и нормальный законы распределения двумерной случайной величины

Равномерный закон распределения

Двумерная случайная величина (X, Y) называется равномерно распределенной в некоторой области G на плоскости $0xy$ с конечной площадью $S(G)$, если ее плотность вероятности $f(x, y)$ определяется формулой:

$$f(x,y) = \begin{cases} \dfrac{1}{S(G)}, \text{если } (x,y) \in G, \\ 0, \quad \text{если } (x,y) \notin G. \end{cases} \qquad (8.5.1)$$

Вероятность попадания случайной точки (X, Y) в заданную область g, целиком лежащую в области G и имеющую площадь $S(g)$, определяется по формуле:

$$P\big((x,y) \in g\big) = \frac{S(g)}{S(G)}. \qquad (8.5.2)$$

Нормальный закон распределения

Двумерная случайная величина (X, Y) называется нормально распределенной на плоскости $0xy$, если ее плотность вероятности $f(x, y)$ определяется формулой:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}}} e^{-\frac{1}{2\left(1-r_{xy}^2\right)}\left(\frac{(x-m_x)^2}{\sigma_x^2}+\frac{\left(y-m_y\right)^2}{\sigma_y^2}-2r_{xy}\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y}\right)}. \qquad (8.5.3)$$

где m_x и m_y – математические ожидания составляющих случайных величин X и Y;

σ_x и σ_y – средние квадратические отклонения X и Y, причем считается, что $\sigma_x\sigma_y \neq 0$ и r_{xy} коэффициент корреляции, при этом $r_{xy}{}^2 \neq 1$.

Если X и Y независимы, то $r_{xy}=0$ и формула плотности вероятности $f(x, y)$ принимает вид:

$$(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e^{-\left(\frac{(x-m_x)^2}{\sigma_x^2}+\frac{\left(y-m_y\right)^2}{\sigma_y^2}\right)}, \tag{8.5.4}$$

или $f(x,y)=f_1(x)\,f_2(y)$, где

$$f_1(x)=\frac{1}{2\pi\sigma_x}e^{-\frac{(x-m_x)^2}{\sigma_x^2}}$$ - плотность составляющей X,

$$f_2(y)=\frac{1}{2\pi\sigma_y}e^{-\frac{\left(x-m_y\right)^2}{\sigma_y^2}}$$ - плотность составляющей Y.

В этом случае вероятность попадания случайной точки (X, Y) в прямоугольник G:$\{x_1\le x\le x_2,\ y_1\le y\le y_2\}$ со сторонами, параллельными осям координат, равна:

$$P\big((X,Y)\in G\big)=\left(\Phi\left(\frac{x_2-m_x}{\sigma_x}\right)-\Phi\left(\frac{x_1-m_x}{\sigma_x}\right)\right)\left(\Phi\left(\frac{y_2-m_y}{\sigma_y}\right)-\Phi\left(\frac{y_1-m_y}{\sigma_y}\right)\right).$$

Пример 1. Два стрелка независимо один от другого производят по одному выстрелу, каждый по своей мишени. Рассматриваются случайные величины X - число попаданий первого стрелка, Y - число попаданий второго стрелка. Вероятность промаха для первого стрелка равна u, для второго стрелка v. Найти функцию распределения $F(x, y)$ системы случайных величин (X, Y).

Решение.

По условию случайны величины X и Y независимы, поэтому $F(x,y)=F_1(x)F_2(y)$. Запишем законы распределения для X и Y:

X:

x_i	0	1
p_i	u	$1-u$

Y:

y_i	0	1
p_i	v	$1-v$

Тогда функции распределения $F_1(x)$ и $F_2(y)$ строятся так:

$$F_1(x)=\begin{cases}0, & x\le 0,\\ u, & 0<x\le 1,\\ 1, & x>1,\end{cases}$$

$$F_2(y)=\begin{cases}0, & y\le 0,\\ v, & 0<y\le 1,\\ 1, & y>1.\end{cases}$$

Построим таблицу значений $F(x, y)$:

x \ y	y≤0	1	y>1
x≤0	0	0	0
0<x≤1	0	uv	u
x>1	0	v	1

Пример 2. Система двух случайных величин (X, Y) подчинена закону распределения с плотностью:

$$f(x,y)=\frac{2}{\pi^2(1+x^2)(4+y^2)}.$$

Найти функцию распределения $F(x, y)$ и вероятность попадания случайной точки (X, Y) в квадрат $R{:}\{0\le x\le 1, 0\le y\le 1\}$.

Решение.

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{2}{\pi^2(1+x^2)(4+y^2)}dxdy=\frac{2}{\pi^2}\int_{-\infty}^{x}\frac{dx}{1+x^2}\int_{-\infty}^{y}\frac{dx}{4+y^2}=$$

$$=\frac{2}{\pi^2}arctgx|_{-\infty}^{x}\frac{1}{2}arctg\frac{y}{2}\Big|_{-\infty}^{y}=\frac{1}{\pi^2}\left(arctgx+\frac{\pi}{2}\right)\left(arctg\frac{y}{2}+\frac{\pi}{2}\right)$$

Найдем вероятность попадания случайной точки (X, Y) в квадрат R рис. 9.5.1:

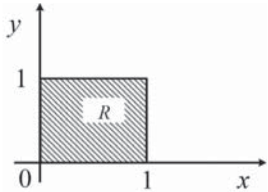


Рис. 9.5.1

$$P\big((X,Y) \in R\big) = \int_0^1 \int_0^1 \frac{2}{\pi^2(1+x^2)(4+y^2)} = \frac{2}{\pi^2} arctgx|_0^1 \frac{1}{2} arctg\frac{y}{2}\Big|_0^1 =$$

$$= \frac{1}{\pi^2}\left(\frac{\pi}{4} - 0\right)\left(arctg\frac{1}{2} - 0\right) = \frac{1}{4\pi} arctg\frac{1}{2} \approx 0{,}03$$

Пример 3. Система двух случайных величин $(X,\ Y)$ подчинена закону равномерной с плотности в квадрате R:$\{0 \le x \le v1, 1 \le y \le 2\}$. Найти значение плотности $F(x,\ y)$, m_x, m_y, D_x, D_y, K_{xy}.

Решение.

По условию рис. 9.1.5

$$f(x,y) = \begin{cases} C, & (X,Y)\ \in R, \\ 0, & (X,Y) \notin R. \end{cases}$$

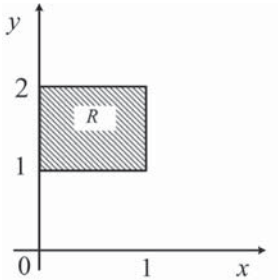


Рис. 9.5.2

По свойствам:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1 \Rightarrow \int_R \int Cdxdy = 1 \Rightarrow \iint_R Cdxdy = 1 \Rightarrow$$

$$= C \iint_R dxdy = 1 \Rightarrow C \cdot S(R) = 1 \Rightarrow C \cdot 1 = 1 \Rightarrow C = 1$$

Следовательно:

$$f(x,y) = \begin{cases} 1, & (X,Y)\ \in R, \\ 0, & (X,Y) \notin R. \end{cases}$$

Тогда для $(X, Y) \in R$ будет:

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy = \int_{0}^{x} dx \int_{1}^{y} dy = x(y-1).$$

Таким образом:

$$F(x,y) = \begin{cases} 0, & \text{при } x \le 0, \quad y \ge 1, \\ x(y-1), & \text{при } (x,y) \in \mathrm{R}, \\ 1, & \text{при } x > 1, \quad y > 2. \end{cases}$$

Найдем числовые характеристики и коэффициент корреляции:

$$m_x = \int_{0}^{1} xdx \int_{1}^{2} dy = \left.\frac{x^2}{2}\right|_{0}^{1} y|_1^2 = \frac{1}{2}\cdot 1 = \frac{1}{2}$$

$$m_y = \int_{0}^{1} dx \int_{1}^{2} ydy = x|_0^1 \left.\frac{y^2}{2}\right|_{1}^{2} = 1\cdot\frac{1}{2}(4-1) = \frac{3}{2}$$

$$D_x = \int_{0}^{1}\left(x-\frac{1}{2}\right)^2 dx \int_{1}^{2} dy = \left.\frac{1}{3}\left(x-\frac{1}{2}\right)^3\right|_{0}^{1} y|_1^2 = \frac{1}{3}\left(1-\frac{1}{2}\right)^3\cdot 1 = 1\cdot\frac{2}{24} = \frac{1}{12}$$

$$D_y = \int_{0}^{1} dx \int_{1}^{2}\left(y-\frac{3}{2}\right)^2 dy = x|_0^1 \left.\frac{1}{3}\left(y-\frac{3}{2}\right)^3\right|_{1}^{2} = \frac{1}{3}\left(1-\frac{3}{2}\right)^3\cdot 1 = 1\cdot\frac{2}{24} = \frac{1}{12}$$

$$K_{xy} = \int_{0}^{1}\left(x-\frac{1}{2}\right)dx \int_{1}^{2}\left(y-\frac{3}{2}\right)dy = \left.\frac{1}{2}\left(x-\frac{1}{2}\right)^2\right|_{0}^{1} \left.\frac{1}{2}\left(y-\frac{3}{2}\right)^2\right|_{1}^{2} = 0\cdot 0 = 0$$

Пример 4. Дана функция распределения двумерной случайной величины (X, Y):

$$F(x,y) = \begin{cases} (1-e^{-3x})(1-e^{-5y}), & \text{при } x \ge 0, \quad y \ge 0, \\ 0, & \text{при } x < 0, \quad y < 0. \end{cases}$$

Найти плотность вероятности этой системы.

Решение.

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}=0$$

при $x<0$ и $y<0$.

При $x>0$ и $y>0$:

$$\frac{\partial F(x,y)}{\partial x}=3e^{-3x}(1-e^{-5y}),\quad \frac{\partial^2 F(x,y)}{\partial x\partial y}=3e^{-3x}5e^{-5y}=15e^{-(3x+5y)}$$

$$f(x,y)=\begin{cases}15e^{-(3x+5y)}, & \text{при } x\ge 0, \quad y\ge 0,\\ 0, & \text{при } x<0, \quad y<0.\end{cases}$$

Пример 5. Дискретная двумерная величина (X, Y) задана таблицей

X \ Y	0,2	0,3
2	0,1	0,2
3	0,2	0,15
4	0,3	0,05

Найти:

а) безусловные законы распределения составляющих X и Y;

б) условный закон распределения составляющей X при условии, что составляющая Y приняла значение y_1=0,2.

Решение.

а) Первое, сложим в таблице вероятности в строках, а потом в столбцах, и запишем безусловные законы распределения соответственно для X и Y:

X:

x_i	2	3	4
p_i	0,3	0,35	0,35

Y:

y_i	0,2	0,3
p_i	0,6	0,4

б) Условные вероятности значений X, при условии, что $Y=y_1=0{,}2$, вычисляются следующим образом:

$$p(2|0{,}2)=\frac{p(2|0{,}2)}{p(0{,}2)}=\frac{0{,}1}{0{,}6}=\frac{1}{6},$$

$$p(3|0{,}2)=\frac{p(3|0{,}2)}{p(0{,}2)}=\frac{0{,}2}{0{,}6}=\frac{1}{3},$$

$$p(4|0{,}2)=\frac{p(4|0{,}2)}{p(0{,}2)}=\frac{0{,}3}{0{,}6}=\frac{1}{2}$$

Тогда требуемый условный закон распределения X имеет вид:

X	2	3	4	4
$p(X\|Y=0{,}2)$	1/6	1/3	1/2	1

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что понимается под системой случайных величин?
2. Что понимается под функцией распределения системы двух случайных величин, и каков ее геометрический смысл?
3. Каковы свойства функции распределения системы двух случайных величин?
4. Что понимается под плотностью распределения $f(x, y)$ системы двух случайных величин (X, Y) и по какой формуле она определяется?
5. По какой формуле функция распределения $F(x, y)$ выражается через плотность распределения $f(x, y)$?
6. Каковы свойства плотности распределения системы двух случайных величин?
7. По каким формулам функции и плотности распределения отдельных случайных величин X и Y выражаются чрез плотность распределения системы этих величин?
8. Что называется условным законом распределения случайной величины X, входящей в систему (X, Y)?

9. В чем заключается теорема умножения законов распределения системы двух случайных величин (X, Y)?
10. Какие случайные величины X и Y, входящие в систему случайных величин (X, Y), называются независимыми?
11. Какие существуют моменты системы двух случайных величин, и что под ними понимается?
12. Чему равны первые начальные моменты и вторые центральные моменты?
13. Что называется корреляционным моментом случайных величин X и Y?
14. Что понимается под корреляционным моментом системы случайных величин X и Y?
15. Какова связь между некоррелируемостью и независимостью случайных величин?

ЗАДАНИЯ

8.1. Два экипажа независимо друг от друга выполняют пуски ракет по своей неподвижной наземной цели. Первый экипаж пускает две ракеты с вероятностью попадания 0,6 в каждом пуске, второй пускает одну ракету с вероятностью попадания 0,8. Рассматриваются случайные величины: X – число попаданий первого экипажа и Y – число попаданий второго экипажа. Построить функцию распределения $F(x, y)$ системы случайных величин (X, Y).

8.2. Дана дискретная двумерная случайная величина (X, Y).

X	**Y**		
	2	5	8
0,4	0,15	0,3	0,35
0,8	0,05	0,12	0,03

Найти:

а) условный закон распределения составляющей X, при условии Y=0,4;

б) условный закон распределения Y при условии X=5 .

8.3. Дана дискретная двумерная случайная величина (X, Y).

X	Y		
	1	4	7
0,3	0,12	0,35	0,30
0,5	0,08	0,10	0,05

Найти:

а) условный закон распределения составляющей X, при условии $Y = 0{,}5$;

б) условный закон распределения Y, при условии $X = 4$.

8.4. Двумерная дискретная случайная величина задана табличным законом распределения.

X	Y		
	0	3	5
1	0	0,05	0,1
5	0,1	0,1	0,15
12	0,1	0,15	0,25

Построить линии регрессии X на Y.

8.5. Система двух независимых случайных величин (X, Y) распределена по нормальному закону параметрами: $m_x = 3$, $m_y = -2$, $\sigma_x = 2$, $\sigma_y = 4$.

Найти:

а) плотности вероятности $f_1(x, y)$ и $f_1(x, y)$ случайных величин X и Y в отдельности;

б) плотность вероятности $f(x, y)$ системы (X, Y);

в) вероятность события $\{1 \leq x \leq 5, -6 \leq y \leq 2\}$.

8.6 Двумерная случайная величина задана плотностью распределения:

$$f(x,y) = \begin{cases} \dfrac{1}{\pi r^2}, & \text{при} \quad x^2 + y^2 \leq r^2, \\ 0, & \text{при} \quad x^2 + y^2 > r^2. \end{cases}$$

Найти условные законы распределения вероятностей составляющих.

8.7 Двумерная дискретная случайная величина задана табличным законом распределения.

X	Y		
	0	3	5
2	0	0,04	0,16
6	0,1	0,13	0,12
14	0,14	0,10	0,21

Построить линии регрессии Y на X.

8.8. Плотность распределения двумерной непрерывной случайной величины (X, Y) задана формулой:

$$f(x,y) = \frac{3}{\pi^2(9 + x^2)(1 + y^2)}$$

Найти функцию распределения $F(x, y)$ и вероятность попадания случайной величины (X, Y) в прямоугольник

$$G\colon \{-\sqrt{3} \leq x \leq \sqrt{3},\ -1 \leq y \leq 1\}.$$

8.9. Плотность распределения двумерной непрерывной случайной величины (X, Y):

$$f(x,y) = \begin{cases} 3(x+y), & \text{при} \quad (x,y) \in D; \\ 0, & \text{при} \quad (x,y) \notin D, \end{cases}$$

$$\text{где} \;\; D = \{(x,y): x \geq 0, y \geq 0, x+y \leq 1\}$$

где $D = \{(x, y): x \geq 0,\ y \geq 0,\ x + y \leq 1\}$.

Найти функцию регрессии X на Y.

8.10. Система случайных величин (X, Y) равномерно распределена в прямоугольнике Ω с вершинами (0, 0), (2, 0), (2, 3), (0, 3). Требуется:

а) Найти плотность вероятности $f(x, y)$ системы (X, Y);

б) Найти функцию распределения $F(x, y)$ системы (X, Y);

в) Найти плотности вероятности $f_1(x, y)$ и $f_2(x, y)$, составляющих случайных величин X и Y в отдельности;

г) Найди вероятность попадания случайной точки (X, Y) в прямоугольник:

$$G: \begin{cases} 0 \leq x \leq 1, \\ 0 \leq y \leq 2. \end{cases}$$

д) Узнать являются ли X и Y зависимыми.

8.11. Дискретная двумерная случайная величина задана матрицей распределения.

X	Y		
	0	1	3
2	0,1	0,15	0,1
4	0,2	0,3	p_{23}

Найти неизвестное значение вероятности p_{13}, условные законы распределения составляющих, линию регрессии Y на X.

8.12. Система случайных величин (X, Y) равномерно распределена в треугольнике Ω с вершинами (0, 0), (4, 0), (0, 4). Требуется:

а) Найти плотность вероятности $f(x, y)$ системы (X, Y);

б) Найти функцию распределения $F(x, y)$ системы (X, Y);

в) Найти плотности вероятности $f_1(x, y)$ и $f_2(x, y)$ составляющих случайных величин X и Y в отдельности;

г) Найти функции распределения и , составляющих случайных величин и в отдельности;

д) Найти вероятность попадания случайной точки (X, Y) в треугольник G с вершинами (0, 0), (2, 0), (1, 3);

е) Узнать являются ли X и Y зависимыми.

ГЛАВА 9
ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

9.1. Основные понятия

Теория вероятностей изучает закономерности, свойственные массовым случайным явлениям. Предельные теоремы теории вероятностей устанавливают зависимость между случайностью и необходимостью. Изучение закономерностей, проявляющихся в массовых случайных явлениях, позволяет научно предсказывать результаты будущих испытаний.

Предельные теоремы теории вероятностей делятся на две группы, одна из которых получила название закона больших чисел (ЗБЧ), а другая - центральной предельной теоремы (ЦПТ).

Первая группа теорем устанавливает устойчивость средних значений: при большом числе испытаний их средний результат перестает быть случайным и может быть предсказан с достаточной точностью.

Вторая группа теорем устанавливает условия, при которых закон распределения суммы большого числа случайных величин неограниченно приближается к нормальному.

9.2. Неравенство Чебышева[11]

В начале рассмотрим неравенство Чебышева, которое можно использовать:

- для грубой оценки вероятностей событий, связанных со случайными величинами, распределение которых неизвестно;
- для доказательства ряда предельных теорем. Неравенство Чебышева является для этих теорем леммой, при определении «сходимости по вероятности» последовательности случайных величин к некоторой неслучайной величине.

[11] Чебышев Пафнутий Львович (1821...1894)- русский математик и механик, основоположник петербургской математической школы. В теории вероятностей доказал центральную предельную теорему, предложил простое и общее доказательство закона больших чисел (1867). От Чебышева и его учеников ведут своё начало важнейшие направления русской и советской математики.

Для любой случайной величины X с законом распределения $f(x; m_x, \sigma_x)$ справедливо утверждение: каково бы ни было положительное число ε, вероятность того, что случайная величина X отклонится от своего математического ожидания – m_x не меньше, чем на величину ε, ограничена сверху величиной D_x / ε^2:

$$P(|X - m_x| \geq \varepsilon) \leq \frac{D_x}{\varepsilon^2} \tag{9.2.1}$$

Доказательство:

1. Рассмотрим событие:

$$A = (|X - m_x| \geq \varepsilon), \quad (9.2.2)$$

которое имеет смысл попадания случайной величины X за пределы отрезка $(m_x\text{-}\varepsilon, m_x+\varepsilon)$ на числовой оси OX (рис. 9.2.1).

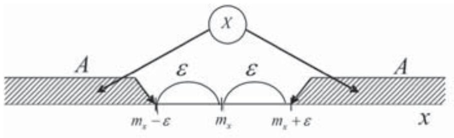


Рис. 9.2.1. Иллюстрация события

Вероятность попадания случайной величины X в эту область равна:

$$P(A) = 1 - \int_{m_x-\varepsilon}^{m_x+\varepsilon} f(x)dx = \int_{-\infty}^{m_x-\varepsilon} f(x)dx + \int_{m_x-\varepsilon}^{\infty} f(x)dx. \tag{9.2.3}$$

2. Дисперсия случайной величины определяется формулой:

$$D_x = \int_{-\infty}^{\infty} (x - m_x)^2 f(x)dx = \int_{-\infty}^{\infty} |x - m_x|^2 f(x)dx \tag{9.2.4}$$

Интеграл (9.2.4) не увеличится, если в подынтегральной функции заменить область интегрирования $(-\infty, \infty)$ на область, в которой $(X-m_x)^2 \geq \varepsilon^2$:

$$D_x \geq \int_{-\infty}^{m_x-\varepsilon} |x - m_x|^2 f(x)dx + \int_{m_x+\varepsilon}^{\infty} |x - m_x|^2 f(x)dx. \tag{9.2.5}$$

Замечание. В области $(-\infty, m_x-\varepsilon)$, $(m_x+\varepsilon, \infty)$ справедливо неравенство $|X-m_x|\geq \varepsilon$, следовательно, там же справедливо неравенство $(X-m_x)^2\geq \varepsilon^2$ (см. рис. 9.2.1).

3. Аналогично интеграл (9.2.5) не увеличится, если величину заменить на не превосходящую её величину :

$$D_x \geq \varepsilon^2 \int_{-\infty}^{m_x-\varepsilon} |x-m_x|^2 f(x)dx + \varepsilon^2 \int_{m_x+\varepsilon}^{\infty} |x-m_x|^2 f(x)dx = \qquad (9.2.6)$$

$$= \varepsilon^2 \left(\int_{-\infty}^{m_x-\varepsilon} |x-m_x|^2 f(x)dx + \int_{m_x+\varepsilon}^{\infty} |x-m_x|^2 f(x)dx \right)$$

4. С учетом (9.2.3) полученное выражение (9.2.6) можно представить в виде:

$$D_x \geq \varepsilon^2 P(A) = \varepsilon^2 P(|X-m_x| \geq \varepsilon) \qquad (9.2.7)$$

Разделив на $\varepsilon^2>0$ левую и правую части неравенства (9.2.7), получим доказательство леммы – неравенства Чебышева (9.2.1):

$$P(|X-m_x| \geq \varepsilon) \leq \frac{D_x}{\varepsilon^2}.$$

Отметим, что события $|X-m_x|\geq \varepsilon)$ и $|X-m_x|< \varepsilon)$ противоположны, неравенство Чебышева можно записать в следующем виде:

$$P(|X-m_x| < \varepsilon) \geq 1 - \frac{D_x}{\varepsilon^2}. \qquad (9.2.8)$$

Применим неравенство Чебышева в форме (9.2.8) для ряда случайных величин:

- для случайной величины $X=m$, имеющей биноминальный закон распределения с математическим ожиданием $m_x=np$ и дисперсией $D_x=npq$:

$$P(|m-np| < \varepsilon) \geq 1 - \frac{npq}{\varepsilon^2}; \qquad (9.2.9)$$

- для относительной частоты m/n события в n независимых испытаниях, в каждом из которых оно может произойти с одной и той же вероятностью:

$$m_x = M\left(\frac{m}{n}\right) = p,$$

и имеющей дисперсию:

$$D\left(\frac{m}{n}\right) = \frac{pq}{n},$$

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{pq}{n\varepsilon^2}. \qquad (9.2.10)$$

Рассмотрим применение неравенства Чебышева на примерах, иллюстрирующих “правило трех сигм”.

Пример 1. Оценим сверху вероятность того, что случайная величина X с законом распределения $f(x; m_x, \sigma_x)$ отклонится от своего математического ожидания m_x не меньше, чем на три значения среднего квадратического отклонения ($3\sigma_x$). Для этого положим в неравенстве Чебышева (9.2.1) $\varepsilon = 3\sigma_x$:

$$P(|X - m_x| \geq 3\sigma_x) \leq \frac{\sigma_x^2}{(3\sigma_x)^2} = \frac{1}{9} \approx 0{,}111.$$

Оценивая результат, видим, что вероятность попадания значений случайной величины X за пределы отрезка ($m_x \pm 3\sigma_x$) не превышает величины 1/9 (11,1%). Можно показать, что для большинства случайных величин, встречающихся на практике, ошибка “правила трех сигм” гораздо меньше этого значения. Покажем это на конкретных примерах.

Пример 2. Оценим вероятность ошибки “правила трех сигм” для показательного распределения случайной величины X:

$$F(x) = 1 - e^{-\lambda x}, \qquad \lambda > 0, x > 0,$$

характеристики которого равны:

$$m_x = \frac{1}{\lambda}, \quad \sigma_x = \frac{1}{\lambda}.$$

Вероятность отклонения значений случайной величины X на величину более трех значений среднего квадратического отклонения при $x > 0$ равна:

$$P(X \geq m_x + 3\sigma_x) = P\left(X \geq \frac{4}{\lambda}\right) = 1 - \left(X < \frac{4}{\lambda}\right) =$$

$$= 1 - F\left(\frac{4}{\lambda}\right) = 1 - (1 - e^{-\lambda\frac{4}{\lambda}}) = e^{-4} \approx 0{,}018 < 0{,}111$$

Пример 3. Оценим вероятность ошибки для нормального распределения случайной величины X:

$$P(|X - m_x| \geq 3\sigma_x) = 1 - P(|X - m_x| < 3\sigma_x) =$$

$$= 1 - 2\Phi_0(3) = 1 - 2 \cdot 0{,}4987 = 0{,}0026 << 0{,}111.$$

Пример 4. Длина изготавливаемых деталей является случайной величиной, среднее значение которой 60 мм. Среднеквадратичное отклонение этой величины равно 0,1 мм. Оценить вероятность того, что отклонение длины изготовленной детали от ее среднего значения по абсолютной величине не превзойдет 0,5 мм.

Решение. Для оценки вероятности используем неравенство Чебышева (9.2.8):

$$P(|X - 60| < 0{,}5) \geq 1 - \frac{(0{,}1)^2}{(0{,}5)^2} = 0{,}96.$$

Пример 5. Среднесуточное потребление электроэнергии в населенном пункте равно 21 000 кВт/ч, а среднеквадратичное отклонение – 300 кВт/ч. Какого потребления электроэнергии в этом населенном пункте можно ожидать в ближайшие сутки с вероятностью, не меньшей 0,95?

Решение. Воспользуемся неравенством Чебышева (9.2.8):

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

Подставим в правую часть неравенства вместо D_x величину $D_x = (300)^2 = 90000$, сделаем ее большей или равной 0,95:

$$1 - \frac{90000}{\varepsilon^2} \geq 0{,}95 \Rightarrow \frac{90000}{\varepsilon^2} \leq 1 - 0{,}95 \Rightarrow \varepsilon^2 \geq \frac{90000}{0{,}05} \Rightarrow \varepsilon^2 \geq 1800000 \Rightarrow$$

$$\varepsilon \geq 1341{,}6$$

Следовательно, в этом населенном пункте можно ожидать, с вероятностью не меньшей 0,95, потребление электроэнергии $X \in [19658{,}4;\ 22341{,}6]$.

Ответ: от 19658,4 до 22341,6.

Как было отмечено в начале параграфа доказательство предельных теорем, (о которых речь пойдет ниже) составляющих закон больших чисел, основывается на неравенстве Чебышева и определении “сходимости по вероятности” последовательности случайных величин к некоторой неслучайной величине. Выясним смысл этого понятия.

Последовательность $\{X_n\}$ случайных величин $X_1, X_2, \ldots, X_n$ **сходится по вероятности к неслучайной величине** a, если при неограниченном увеличении числа n вероятность события $A=(|\{X_n\}-a|<\varepsilon)$, где ε – произвольное положительное малое число, стремится к единице.

Говоря иными словами (как при доказательстве теорем), для любых произвольных малых заданных чисел $\varepsilon > 0$ и $\delta > 0$ существует такое число N, что для всех $n>N$ справедливо неравенство:

$$P(|\{X_n\}-a|<\varepsilon) > 1-\delta, \qquad (9.2.11)$$

которое может быть записано в таком виде:

$$\lim_{n\to\infty} P(|\{X_n\}-a|<\varepsilon) = 1. \qquad (9.2.12)$$

Сходимость по вероятности $\{X_n\}$ к a при $n\to\infty$ в дальнейшем будем записывать в следующем обобщенном виде:

$$\{X_n\} \xrightarrow[n\to\infty]{\mathcal{P}} a. \qquad (9.2.13)$$

9.3. Неравенство Маркова (лемма Чебышева)

Теорема. Если случайная величина X принимает только неотрицательные значения и имеет математическое ожидание, то для любого положительного числа A верно неравенство:

$$P(X > A) \leq \frac{M(X)}{A}. \qquad (9.3.1)$$

Приведем доказательство для дискретной случайной величины X. Расположим ее значения в порядке возрастания (рис. 9.3.1), из которых часть значений $x_1, x_2, \ldots, x_k$ будут не больше числа A, а другая часть - $x_{k+1}, \ldots, x_n$ будут больше A, математически это выражается следующим образом:

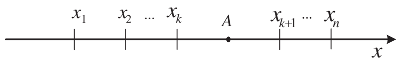


Рис. 9.3.1

Запишем выражение для математического ожидания $M(X)$:
$x_1p_1 + x_2p_2 + \ldots + x_kp_k + x_{k+1}p_{k+1} + \ldots + x_np_n = M(X)$, где $p_1, p_2, \ldots, p_n$ – вероятности того, что случайная величина X примет соответствующие значения: x_1, $x_2, \ldots, x_n$.

Отбрасывая первые k неотрицательных слагаемых, получим: $x_{k+1}p_{k+1}+$ $+\ldots+x_n p_n \leq M(X)$.

Заменяя в неравенстве значения $x_{k+1}, \ldots, x_n$ меньшим числом A, получим более сильное неравенство: $A(p_{k+1}+\ldots+p_n) \leq M(X)$ или

$$p_{k+1}+\ldots+p_n \leq \frac{M(X)}{A}.$$

Сумма вероятностей в левой части полученного неравенства представляет собой сумму вероятностей событий $X=x_{k+1}, \ldots, X=x_n$, то есть вероятность события $X>A$, поэтому:

$$P(X>A) \leq \frac{M(X)}{A}.$$

События $X>A$ и $X \leq A$ являются противоположными, то вероятность события можем запасать как $P(X \leq A)=1-P(X>A)$, тогда:

$$(X \leq A) \geq 1-\frac{M(X)}{A}. \tag{9.3.2}$$

Неравенство Маркова применимо к любым неотрицательным случайным величинам.

Пример 6. Оценить вероятность того, что в течение ближайшего дня потребность в воде в населенном пункте превысит 150 000 л, если среднесуточная потребность в ней составляет 50 000 л.

Решение. Используя неравенство Маркова (9,3.1) получим:

$$P(X>150000) \leq \frac{50000}{150000} \leq \frac{1}{3}.$$

Пример 7. Среднее число солнечных дней в году для данной местности равно 90. Оценить вероятность того, что в течение года в этой местности будет не более 240 солнечных дней.

Решение. Используя неравенство Маркова (9,3.1) получим:

$$(P \leq 240) \geq 1-\frac{90}{240} \geq 0{,}625.$$

9.4. Теорема Чебышева

Теорема Чебышева устанавливает связь между средним арифметическим Y_n наблюденных значений $X_1, X_2, \ldots, X_n$ случайной величины X:

$$Y_n=\frac{1}{n}\sum_{i=1}^{n} X_i, \tag{9.4.1}$$

и её математическим ожиданием – $M(X)=m_x$, при следующих условиях:

– испытания, в результате которых наблюдают значения $X_1, X_2, ..., X_n$ случайной величины X, независимы;

– условия испытаний не изменяются от испытания к испытанию, то есть постоянны.

Теорема: при достаточно большом числе независимых испытаний, проводимых в одинаковых условиях, среднее арифметическое наблюдённых значений с. в. X сходится по вероятности к её математическому ожиданию m_x:

$$Y_n \xrightarrow[n\to\infty]{\mathcal{P}} m_x .$$

Доказательство. Пусть производится n независимых испытаний в одинаковых условиях, в результате которых наблюдаются значения $X_1, X_2, ..., X_n$ случайной величины X с законом распределения $f(x; m_x, \sigma_x)$. Следовательно, результат наблюдения X_i ($i=1...n$) имеет характеристики $M(X_i)=m_x$, $D(X_i)=D_x$. Иначе говоря, результаты наблюдений одинаково распределены.

Найдем вероятностные характеристики среднего арифметического наблюденных значений Y_n:

$$M(Y_n) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n} n m_x \cdot \qquad (9.4.2)$$

$$D(Y_n) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2} n D_x = \frac{D_x}{n} . \qquad (9.4.3)$$

Применим к Y_n неравенство Чебышева (9.2.1):

$$P(|Y_n - M[Y_n]| \geq \varepsilon) \leq \frac{D[Y_n]}{\varepsilon^2} . \qquad (9.4.4)$$

С учетом (9.4.2) и (9.4.3) получим:

$$P(|Y_n - m_x| \geq \varepsilon) \leq \frac{D_x}{n\varepsilon^2} . \qquad (9.4.5)$$

Для любого малого $\varepsilon > 0$ всегда можно выбрать такое число N, что как только $n > N$, правая часть (9.4.5) станет не больше произвольного малого числа $\delta > 0$:

$$\frac{D_x}{n\varepsilon^2} \le \delta. \tag{9.4.6}$$

и неравенство (9.4.5) запишется в виде:

$$P(A) = P(|Y_n - m_x| \ge \varepsilon) \le \delta. \tag{9.4.7}$$

Переходя к противоположному событию:

$$\overline{A} = (|Y_n - m_x| < \varepsilon).$$

Окончательно получим:

$$P(|Y_n - m_x| < \varepsilon) > 1 - \delta. \tag{9.4.8}$$

Ясно, что:

$$\lim_{n\to\infty}\left(\delta \ge \frac{D_x}{n\varepsilon^2}\right) = 0 \Longrightarrow 1-\delta=1.$$

Таким образом, (9.4.8) среднее арифметическое Y_n сходится по вероятности к математическому ожиданию случайной величины :

$$\lim_{n\to\infty} P(|Y_n - m_x| < \varepsilon) = 1, \text{ или }. \tag{9.4.9}$$

$$Y_n \xrightarrow[n\to\infty]{\mathcal{P}} m_x. \tag{9.4.10}$$

Если рассмотреть величину $Z_n = Y_n - m_n$, то она, как очевидно, сходится по вероятности к нулю:

$$Z_n \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{9.4.11}$$

9.5. Теоремы Бернулли и Пуассона

Первой и наиболее простой формой закона больших чисел является теорема Бернулли. Теорема Бернулли теоретически обосновывает свойство устойчивости относительной частоты $P^*(A)$.

Теорема. Если вероятность появления события A в одном испытании равна p, число наступления этого события в n независимых испытаниях равно m_A, то для любого числа $\varepsilon > 0$ имеет место равенство:

$$\lim_{n\to\infty} P\left(\left|\frac{m_A}{n} - p\right| < \varepsilon\right) = 1. \tag{9.5.1}$$

Следовательно, относительная частота $P^*(A)$ события A сходится по вероятности к вероятности p события A:

$$P^*(A) \xrightarrow[n\to\infty]{\mathcal{P}} P(A) . \tag{9.5.2}$$

Пусть производится n независимых испытаний, в каждом из которых вероятность появления события A равна p. Другими словами, пусть имеет место схема Бернулли.

Введем случайные величины $X_1, X_2, \ldots, X_n$ следующим образом: $X_i = 1$, если в i-м испытании появилось событие A, а если не появилось, то $X_i = 0$. Тогда число успешных появлений m_A можно выразить как:

$$m_A = \sum_{i=1}^{n} X_i .$$

Математическое ожидание и дисперсия случайной величины X_i соответственно равны $M(X_i) = 1 \cdot p + 0 \cdot (1 - p) = p$,

$$D(X_i) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p) = pq.$$

Закон распределения случайной величины X_i при любом i имеет вид:

X_i	0	1
P	$1-p$	p

Следовательно, случайные величины X_i независимы их дисперсии ограничены одним и тем же числом 1/4, так как:

$$p(1-p) = p - p^2 = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \le \frac{1}{4}.$$

Поэтому к этим случайным величинам можно применить теорему Чебышева (9.4.9):

$$\lim_{n\to\infty} P(|Y_n - m_x| < \varepsilon) = 1.$$

Здесь $Y_n = m_A/\text{n}$, $m_x = p$, тогда:

$$\lim_{n\to\infty} P\left(\left|\frac{m_A}{n} - p\right| < \varepsilon\right) = 1,$$

Или:

$$P^*(A) \xrightarrow[n\to\infty]{\mathcal{P}} P(A)$$

Что и требовалось доказать.

Таким образом, теорема Бернулли является простейшим частным случаем теоремы Чебышева.

Пример 8. Из 1000 изделий, отправляемых в сборочный цех, было подвергнуто обследованию 200, отобранных случайным образом изделий. Среди них оказалось 25 бракованных. Приняв долю бракованных изделий среди отобранных за вероятность изготовления бракованного изделия, оценить вероятность того, что во всей партии бракованных изделий окажется не менее 10% и не более 15%.

Решение.

Определим вероятность изготовления бракованного изделия:

$$p = \frac{25}{200} = 0{,}125\cdot$$

Число испытаний m =1000.

Наибольшее отклонение относительной частоты появлений бракованных изделий от вероятности по абсолютной величине равно:

$$\left|\frac{m}{n} - p\right| = 0{,}025.$$

Используя формулу (9.2.10), находим искомую вероятность:

$$P\left(\left|\frac{m}{n} - p\right| < 0{,}025\right) \geq 1 - \frac{0{,}125 \cdot 0{,}875}{1000 \cdot (0{,}025)^2} > 0{,}825.$$

Теорема Пуассона

Теорема Чебышева и Бернулли доказаны для случая достаточно больших испытаний, проводимых в одинаковых условиях. Однако замечено, что если условия опыта меняются, то свойство устойчивости относительной частоты появления события A сохраняется. Это обстоятельство доказано Пуассоном.

Теорема Пуассона.

При неограниченном увеличении числа независимых испытаний, проводимых в переменных условиях, относительная частота появления события A сходится по вероятности к среднему арифметическому вероятностей появления данного события в каждом из опытов, то есть:

$$\lim_{n\to\infty}\left(\left|\frac{m_A}{n} - \frac{1}{n}\cdot\sum_{i=1}^{n} p_i\right| < \varepsilon\right) = 1, \quad 0 < \varepsilon \approx 0. \qquad (9.5.3)$$

Обозначим среднеее арифметическое вероятностей появления данного события в каждом из опытов:

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

Тогда выражение теоремы Пуассона можем представить следующим образом:

$$P^*(A) \xrightarrow[n\to\infty]{\mathcal{P}} \overline{p}.$$

Нетрудно убедиться, что теорема Пуассона является частным случаем теоремы Чебышева.

9.6. Центральная предельная теорема (Теорема Ляпунова)

Рассмотренные теоремы закона больших чисел касаются вопросов приближения некоторых случайных величин к определённым предельным значениям, независимо от их закона распределения. В теории вероятностей, как уже отмечалось, существует другая группа теорем, касающихся предельных законов распределения суммы случайных величин. Общее название этой группы теорем – центральная предельная терема. Различные её формы различаются условиями, накладываемыми на сумму составляющих случайных величин. Впервые одна из форм центральной предельной теоремы была доказана выдающимся русским математиком А. М. Ляпуновым в 1900 году с использованием специально разработанного им метода характеристических функций.

Теорема Ляпунова. Закон распределения суммы независимых случайных величин $X_1, X_2, \ldots, X_n$ приближается к нормальному закону распределения при неограниченном увеличении n (то есть, при $n \to \infty$), если выполняются следующие условия:

а) все X_i, $i=1, 2, \ldots, n$ имеют конечные математические ожидания и дисперсии: $M(X_i)=m_i$; $M((X_i-m_i)^2)=D(X_i)$; абсолютный центральный момент третьего порядка $M(|X_i-m_i|^3)=C_i$,
где $\varepsilon>0$, $i=1, 2, \ldots n$;

б) ни одна из случайных величин по степени своего влияния на всю сумму случайных величин не отличается от остальных (то есть, влияние каждой из случайных величин на всю сумму ничтожно мало). Другими словами, выполняется условие (9.6.1).

Обозначим:

$$C_\Sigma = \sum_{i=1}^{n} C_i,$$

$$D_\Sigma = \left(\sum_{i=1}^{n} D(X_i)\right)^{\frac{3}{2}},$$

$$\lim_{n\to\infty} \frac{C_\Sigma}{D_\Sigma} = 0. \qquad (9.6.1)$$

Тогда:

$$\lim_{n\to\infty} P(Y<y) = \frac{1}{\sqrt{2\pi}\cdot\sigma_y}\int_{-\infty}^{y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}\,dy. \qquad (9.6.2)$$

где:

$$Y = \sum_{i=1}^{n} X_i,\quad m_y = \sum_{i=1}^{n} M(X_i),\quad \sigma_y^2 = \sum_{i=1}^{n} D(X_i)$$

Следует отметить, что центральная предельная теорема справедлива не только для непрерывных, но и для дискретных случайных величин. Практическое значение теоремы Ляпунова огромно. Опыт показывает, что закон распределения суммы независимых случайных величин, сравнимых по своему рассеиванию, достаточно быстро приближается к нормальному. Уже при числе слагаемых порядка десяти закон распределения суммы можно заменить на нормальный (в частности, примером такой суммы может быть среднее арифметическое наблюдаемых значений случайных величин то есть:

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Частным случаем центральной предельной теоремы является теорема Лапласа. Далее, вероятность того, что Y заключено в интервале [α;β] можно вычислить по формуле:

$$P(a < Y < b) \approx \frac{1}{\sqrt{2\pi}\cdot\sigma_y}\int_{\alpha}^{\beta} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}\,dy. \qquad (9.6.3)$$

Используя функцию Лапласа, последнюю формулу можно записать в удобном для расчётов виде:

$$P(\alpha \le Y \le \beta) \approx \Phi_0(z_2) - \Phi_0(z_1). \qquad (9.6.4)$$

где:

$$z_1 = \frac{\alpha - M(Y)}{\sigma_y}, \quad z_2 = \frac{\beta - M(Y)}{\sigma_y}.$$

Пример 9. Независимые случайные величины *Xi* равномерно распределены на отрезке [0,1]. Найти закон распределения случайной величины:

$$Y = \sum_{i=1}^{50} X_i,$$

а также вероятность попадания случайной величины Y в интервал (30, 60).

Решение. Условие центральной предельной теоремы соблюдаются, поэтому случайная величина имеет приближенно плотность распределения:

$$f_Y(y) \approx \frac{1}{\sqrt{2\pi}\cdot\sigma_y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}.$$

По формулам вычисления числовых характеристик для равномерного закона распределения получаем:

$$M(X_i) = \frac{1-0}{2} = \frac{1}{1}, \qquad D(X_i) = \frac{(1-0)^2}{12} = \frac{1}{12}, \qquad \sigma(X_i) = \frac{1}{2\sqrt{3}} \quad .$$

Тогда:

$$m_y = M\left(\sum_{i=1}^{50} X_i\right) = 50 \cdot \frac{1}{2} = 25.$$

$$\sigma_y^2 = D\left(\sum_{i=1}^{50} X_i\right) = 50 \cdot \frac{1}{12} = \frac{25}{6}.$$

$$\sigma_y = \sqrt{\frac{25}{6}} = \frac{5}{\sqrt{6}} = \frac{5 \cdot \sqrt{6}}{6}.$$

Следовательно:

$$f_Y(y) \approx \frac{\sqrt{6}}{5\sqrt{2\pi}} e^{-\frac{6(y-25)^2}{50}}.$$

Используя формулу (9.6.4):

$$P(30 < Y < 60) \approx \Phi_0\left(\frac{60-25}{\frac{5\cdot\sqrt{6}}{6}}\right) - \Phi_0\left(\frac{30-25}{\frac{5\cdot\sqrt{6}}{6}}\right) =$$

$$= \Phi_0\left(7\sqrt{6}\right) - \Phi_0\left(\sqrt{6}\right) = (17{,}15) - \Phi_0(2{,}45) = 1 - 0{,}9853 = 0{,}0147\,.$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что называют законом больших чисел?
2. Какой смысл имеет это название?
3. В чем суть неравенства Чебышева?
4. Какова формулировка теоремы Чебышева?
5. В чем смысл теоремы Бернулли?
6. Какова роль предельных теорем в теории вероятностей?
7. Какой из законов распределения фигурирует в качестве предельного закона?

8. В чем состоит центральная предельная теорема Ляпунова?
9. В чем состоит центральная предельная теорема Ляпунова?
10. Как можно истолковать теорему Лапласа в качестве предельной теоремы теории вероятностей??

ЗАДАНИЯ

9.1. Длина изготовляемых изделий представляет случайную величину, среднее значение которой (математическое ожидание) равно 80 см. Дисперсия этой величины равна 0,025. Используя неравенство Чебышева, оценить вероятность того, что: а) отклонение длины изготовленного изделия от ее среднего значения по абсолютной величине не превзойдет 0,3; б) длина изделия выразится числом, заключенным между 79.6 и 80,4 см.

9.2. Среднее значение расхода воды в населенном пункте составляет 60000 л. в день. Оценить вероятность того, что в этом населенном пункте расход воды не будет превышать 140000 л. в день.

9.3. Средний вес клубня картофеля равен 150 г. Оценить вероятность того, что наудачу взятый клубень картофеля весит не более 500 г.?

9.4. Устройство состоит из 12 независимо работающих элементов. Вероятность отказа каждого элемента за время равна 0,05. Используя неравенство Чебышева, оценить вероятность того, что абсолютная величина разности между числом отказавших элементов и средним числом (математическим ожиданием) отказов за время t окажется меньше трех.

9.5. В осветительную сеть параллельно включено 18 ламп. Вероятность того, что за время лампа будет включена, равна 0,85. Пользуясь неравенством Чебышева, оценить вероятность того, что абсолютная величина разности между числом включенных ламп и средним числом (математическим ожиданием) включенных ламп за время окажется: а) меньше трех; б) не меньше трех.

9.6. Длина изготавливаемых деталей является случайной величиной, среднее значение которой 50 мм. Среднее квадратическое отклонение этой величины равно 0,2 мм. Оценить вероятность того, что отклонение длины изготовленной детали от ее среднего значения по абсолютной величине не превзойдет 0,4 мм.

9.7. Известно, что дисперсия каждой из данных независимых случайных величин не превышает 5. Определить число таких величин, при котором вероятность отклонения средней арифметической случайных величин от средней арифметической их математических ожиданий не более чем на 0,2 превысит 0,95.

9.8. Дисперсия каждой из 900 независимых случайных величин не превышает 10. Какой должна быть верхняя граница абсолютной величины отклонения средней арифметической случайных величин от средней арифметической их математических ожиданий, чтобы вероятность такого отклонения превышала 0,99?

9.9. За значение некоторой величины принимают среднеарифметическое достаточно большого числа ее измерений. Предполагая, что среднее квадратическое отклонение возможных результатов каждого измерения не превосходит 5 мм, оценить вероятность того, что при 1000 измерений неизвестной величины отклонение принятого значения от истинного по абсолютной величине не превзойдет 0,5 мм.

9.10. Выборочным путем требуется определить средний рост мужчин двадцатилетнего возраста. Какое количество мужчин, отобранных случайным образом, нужно измерить, чтобы с вероятностью, превышающей 0,98, можно было утверждать, что средний рост у отобранной группы будет отличаться от среднего роста всех двадцатилетних мужчин по абсолютной величине не более чем на один см. Известно, что среднеквадратичное отклонение роста для каждого мужчины из отобранной группы не превышает 5 см.

9.11. Произведено 650 независимых испытаний, в 260 из них вероятность появления события была равна 0,6, в 2400 – 0,5 и в 150 – 0,4. Оценить снизу вероятность того, что отклонение от средней вероятности не превысит по абсолютной величине 0,05.

Раздел 2
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ГЛАВА 10
ВВЕДЕНИЕ В МАТЕМАТИЧЕСКУЮ СТАТИСТИКУ

10.1. Предмет и задачи математической статистики

Математическая статистика – это наука, занимающаяся методами сбора, обработки и анализа данных, полученных в результате наблюдений за случайными явлениями.

Все методы математической статистики основаны на теории вероятностей. Однако, в силу специфичности решаемых задач, математическая статистика выделяется из теории вероятностей в самостоятельную область. Если в теории вероятностей считается заданной модель явления и производится расчет возможного реального течения этого явления, то в математической статистике подбирается подходящая теоретико-вероятностная модель, исходя из статистических данных.

Основными задачами математической статистики являются:

- определение закона распределения случайной величины или системы случайных величин;
- определение неизвестных параметров распределения;
- проверка правдоподобия гипотез.

При проверке гипотез широкое применение находит ряд теоретических законов распределения. Наиболее важным из них является нормальное распределение. С ним связаны распределения: *Хи* – квадрат распределение, T – распределение Стьюдента, F – распределение Фишера-Снедекора, а также интеграл вероятностей. Особенностью этих законов распределения является то, что они, как правило, описывают распределения некоторых функций - результатов наблюдений, так называемых “статистик”: $\theta=\theta(X_1, X_2, \ldots, X_n)$. Нормальное распределение (см. п. 7.6) занимает важное место в статистике, поскольку очень многие эмпирические распределения жизненных явлений приближаются к нему.

Нормальное распределение является наиболее важным в связи с центральной предельной теоремой теории вероятностей: распределение суммы независимых случайных величин стремится к нормальному с увеличением их количества при произвольном законе распределения отдельных слагаемых, если слагаемые обладают конечной дисперсией.

Так как реальные физические, химические, биологические, экономические явления часто представляют собой результат суммарного воздействия многих факторов, то в таких случаях нормальное распределение является хорошим приближением наблюдаемых значений. Функция плотности нормального распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}, \qquad x \epsilon R. \tag{10.1.1}$$

унимодальная, симметричная, аргумент x может принимать любые действительные значения. На рис. 10.1.1 представлены графики плотности нормального распределения: сплошная линия: $\bar{x}$=0, σ=1 штрихпунктирная линия $\bar{x}$ =0, σ=1,5.

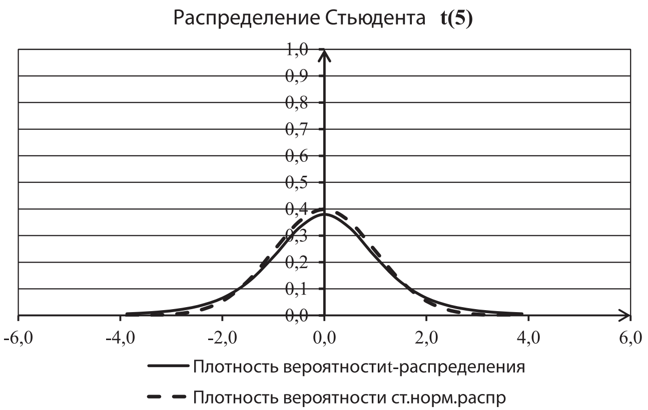


Рис. 10.1.1

10.2. *Хи* – квадрат распределение

Распределению хи-квадрат (x^2-распределение) с k степенями свободы соответствует распределение суммы $x^2=\Sigma u_i^2$ квадратов n стандартизованных случайных величин u_i, каждая из которых распределена по нормальному закону, причем k из них независимы, $n>k$. Функция плотности распределения x^2 с k степенями свободы:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\left(\frac{k}{2} - 1\right)} e^{-\frac{x}{2}}, & \text{при } x > 0, \\ 0 & \text{при } x \leq 0, \end{cases} \qquad (10.2.1)$$

где $x=x^2$

$$\Gamma(z) \int\limits_0 x^{z-1}\, e^{-x} dx \text{ — гамма функция.} \qquad (10.2.2)$$

Число степеней свободы k определяет количество независимых слагаемых в выражении для x^2. Функция плотности, при k равном одному или двум, — монотонная, а при $k > 2$ — унимодальная, несимметричная.

Математическое ожидание и дисперсия величины x^2 равны соответственно k и $2k$. Распределение x^2 является частным случаем более общего гамма-распределения, а величина, равная корню квадратному из *Хи* - квадрат с двумя степенями свободы, подчиняется закону Рэлея.

С увеличением числа свободы ($k > 30$) распределение x^2 приближается к нормальному с математическим ожиданием k и дисперсией $2k$. В таких случаях критическое значение $x^2(k,a) \gg (u_{1-a}(k,2k))$, где $u_{1-a}(k,2k)$ — квантиль нормального распределения. Погрешность аппроксимации не превышает нескольких процентов (рис. 10.2.1).

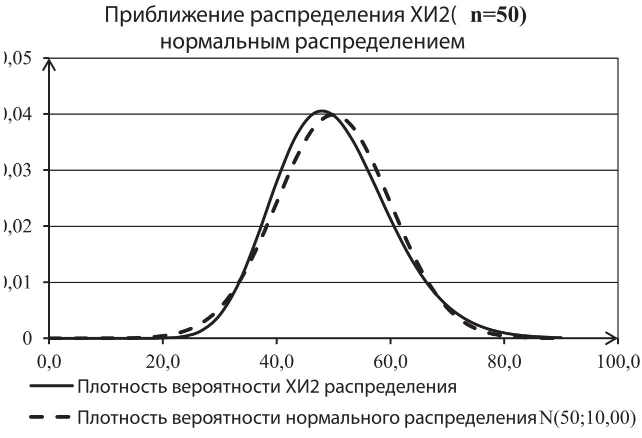


Рис. 10.2.1

10.3. Распределение Стьюдента

Распределение Стьюдента (распределение предложенное в 1908 г. английским статистом В. Госсетом, публиковавшим научные труды под псевдонимом Student) характеризует распределение случайной величины:

$$t = \frac{u_0}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_k^2)/k}}\ , \qquad (10.3.1)$$

где u_0, u_1, ..., u_0 взаимно независимые нормально распределенные случайные величины с нулевым средним и конечной дисперсией.

Аргумент i не зависит от дисперсии слагаемых. Функция плотности распределения Стьюдента:

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)}\left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}\ , \qquad (10.3.2)$$

Величина k характеризует количество степеней свободы. Плотность распределения – унимодальная и симметричная функция, похожая на нормальное распределение (рис. 10.3.1).

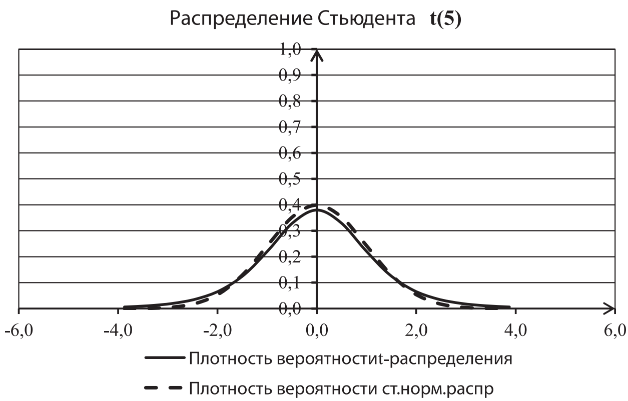


Рис. 10.3.1

Область изменения t от $-\infty$ до $+\infty$. Математическое ожидание и дисперсия равны (k>2) 0 и $k/(k-2)$ соответственно. По сравнению с нормальным распределение Стьюдента более пологое, оно имеет меньшее дисперсию.

Это отличие заметно при небольших значениях k, что следует учитывать при проверке статистических гипотез (критические значения аргумента распределения Стьюдента превышают аналогичные показатели нормального распределения).

Таблицы распределения содержат значения:

- для односторонней критической области:

$$\int_{t(k,\alpha)}^{\infty} f(t)\,dt = \alpha;$$

- для двухсторонней критической области:

$$\int_{-t(k,\alpha)}^{t(k,\alpha)} f(t)\,dt = \alpha.$$

Распределение Стьюдента применяется для описания ошибок выборки при $k > 30$. При $k > 100$ данное распределение практически соответствует нормальному, для $30 < k < 100$ различия между распределением Стьюдента и нормальным распределением составляют несколько процентов. Поэтому относительно оценки ошибок малыми считаются выборки объемом не более 30 единиц, большими – объёмом более 100 единиц.

При аппроксимации распределения Стьюдента нормальным распределением для односторонней критической области вероятность:

$$P\{t > t(k,\alpha)\} = u_{1-\alpha}\left(0, \frac{k}{k-2}\right),$$

где:

$$u_{1-\alpha}\left(0, \frac{k}{k-2}\right).$$

квантиль нормального распределения. Аналогичные соотношения можно составить и для двухсторонней критической области.

10.4. Распределение Фишера-Снедекора

Распределению Фишера (F распределению Фишера-Снедекора) подчиняется случайная величина:

$$x = \frac{\frac{y_1}{k_1}}{\frac{y_2}{k_2}}, \qquad (10.4.1)$$

равная отношению двух случайных величин y_1 и y_2, имеющих хи-квадрат распределение с k_1 и k_2 степенями свободы. Область изменения аргумента x от 0 до $+\infty$. Плотность распределения:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{k_1+k_2}{2}\right) k_1^{\frac{k_1}{2}} k_2^{\frac{k_2}{2}}}{\Gamma\left(\dfrac{k_1}{2}\right)\Gamma\left(\dfrac{k_2}{2}\right)} x^{\frac{k_1}{2}-1}(k_2+k_1x)^{-\frac{k_1+k_2}{2}}, & \text{при } x > 0, \\ 0, & \text{при } x \le 0. \end{cases} \qquad (10.4.2)$$

В этом выражении k_1 обозначает число степеней свободы величины y_1 с большей дисперсией, k_2 – число степеней свободы величины y_2 с меньшей дисперсией.

Плотность распределения – унимодальная, несимметричная (рис. 10.4.1).

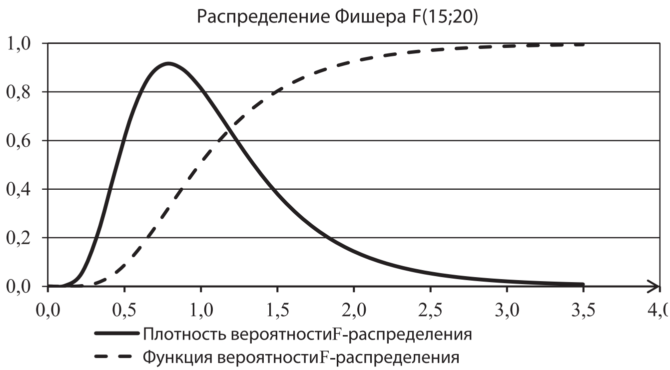


Рис. 10.4.1

Основная задача этой главы познакомить читателя с теоретическими законами распределения, которые в дальнейшем будут использоваться для оценки параметров и проверки гипотез.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что изучает математическая статистика?

2. Каковы задачи математической статистики?
3. Какие законы используются при проверке гипотез?
4. Какой закон является при этом эталоном?
5. В связи с чем возникли *Хи* – квадрат распределение, *T* – распределение Стьюдента, *F* – распределение Фишера-Снедекора?
6. Что определяет число степеней свободы в данных законах?
7. Как определяются значения вышеперечисленных распределений?
8. От чего не зависит аргумент функции распределения Стьюдента?
9. Какова связь между нормальным законом распределения и *Хи* - квадрат распределением?
10. Какова связь между нормальным законом распределения и распределением Стьюдента?

ЗАДАНИЯ

Предварительные пояснения к заданиям: необходимо пользоваться соответствующими таблицами приложения, где k – число степеней свободы, α – уровень значимости, $f(x^2)$ – значение распределения, *Хи* – квадрат (функции Пирсона), k_1 – число степеней свободы для случайной величины с большей дисперсией, k_2 – число степеней свободы для случайной величины с меньшей дисперсией, $f(t)$ – значение функции Стьюдента, $F(x)$ –значение функции Фишера-Снедекора.

1.1. Дано: $k=14$, $\alpha=0{,}05$. Найти $f(x^2)$.

1.2. Дано: $k=48$, $\alpha=0{,}05$. Найти $f(x^2)$.

1.3. Дано: $k=20$, $f(x^2)=12{,}443$. Найти α.

1.4. Дано: $k=41$, $f(x^2)=64{,}95$. Найти α.

1.5. Дано: $k=10$, $\alpha=0{,}05$. Найти $f(t)$.

1.6. Дано: $k=27$, $\alpha=0{,}4$. Найти $f(t)$.

1.7. Дано: $k=27$, $f(t)=0{,}711$. Найти α.

1.8. Дано: $k=7$, $f(t)=0{,}7111$. Найти α.

1.9. $k_1=6$, $k_2=12$, $\alpha=0{,}05$. Найти $F(x)$.

1.10. $k_1=11$, $k_2=18$, $\alpha=0{,}01$. Найти $F(x)$.

1.11. $k_1=2$, $k_2=9$, $F(x)=4{,}26$. Найти α.

1.12. $k_1=7$, $k_2=15$, $F(x)=4{,}14$. Найти α.

ГЛАВА 11
СТАТИСТИЧЕСКИЕ РЯДЫ И ИХ ХАРАКТЕРИСТИКИ

11.1. Понятие статистического ряда. Виды статистических рядов

Пусть нами проводится обследование элементов некоторого множества X. Если есть возможность и целесообразность или необходимость обследования без исключения всех элементов этого множества, то такое обследование называется сплошным.

По некоторым причинам, может быть нецелесообразно такое обследование (чаще всего по экономическим причинам), или отсутствует возможность обследования всех элементов множества X. В этом случае обследуется часть элементов этого множества – множество X_1. Такое обследование называется выборочным. Множество X называется **генеральной** совокупностью, а множество X_1, являющееся подмножеством множества X, называется **выборочной** совокупностью. Задача математической статистики состоит в том, чтобы на основании знания некоторых свойств элементов подмножества X_1, сделать утверждения о свойствах множества X. Эти утверждения могут иметь качественный или количественный характер.

Например, если проводится обследование множества валов на предмет соответствия их диаметров установленным стандартам, то этот признак имеет количественный характер. Если валы исследуются на предмет возможности их использования на определенном оборудовании, то это этот признак является качественным.

Объёмом выборки называется число объектов (наблюдений) в совокупности, генеральной или выборочной, и обычно обозначают через N и n соответственно.

Под **случайной выборкой** объема n понимается выбор n объектов из генеральной совокупности объема N, причем выбор отдельных элементов производится независимо один от другого.

Выборки разделяют на **повторные** и **бесповторные**. В первом случае отобранный объект возвращается в генеральную совокупность перед следующим отбором, во втором случае – нет.

Выборка, отражающая все пропорции и особенности генеральной совокупности, называется **репрезентативной** или представительной.

Числовые значения выборки, полученные в результате эксперимента (опыта, испытания), называются реализацией выборки и обозначаются строчными буквами с индексом: $x_1, x_2, ..., x_n$.

Естественно, в основе всех статистических исследований первоначальным является определение целей исследования и способов сбора и группировки статистических данных в соответствии с поставленными целями.

Другим весьма важным вопросом математической статистики является вопрос распределения исследуемой случайной величины. Чаще всего полагается, что случайная величина подчинена нормальному закону распределения. В этом случае задача математической статистики заключается в решении вопроса отыскания параметров распределения и их оценки.

Статистические данные (выборку) обычно представляют в виде таблицы, состоящей из двух строк или столбцов – номер наблюдений, наблюденные значения параметров (табл. 11.1.1).

Таблица 11.1.1

i - номер наблюдения	1	2	...	m	...	n
x_i - результат наблюдения	x_1	x_2	...	x_m	...	x_n

Пусть изучается некоторый признак заданной совокупности. Этот признак в совокупности может меняться от перехода от одного члена к другому. Изменение признака называют его вариацией, а значение признака произвольно взятого члена – его вариантой.

Если производится группировка вариант по отдельным значениям признака, и результат группировки представляется рядом, расположенный в порядке возрастания, то получим вариационный ряд **дискретной группировки**.

Если же производится группировка вариант по интервалам изменения признака, то получим вариационный ряд **интервальной группировки**.

Таким образом, ряд значений признака (вариантов), расположенных в порядке возрастания или убывания с соответствующими им весами, называется вариационным рядом (рядом распределения).

Следовательно, вариационные ряды бывают дискретными и интервальными.

В качестве весов выступают частоты или относительные частоты (частости), то есть числа n_i, показывающие, сколько раз встречаются варианты x_i в ряде наблюдений (измерений) или их отношение к объему выборки.

Таким образом, относительная частота (частость) определяется из формулы:

$$p_i^* = \frac{n_i}{N}, \tag{11.1.1}$$

где $N = \sum_{i=1}^{n} n_i$.

Перечень вариантов и соответствующие им частоты или относительные частоты (частости) называется статистическим рядом и записывается в виде таблицы – первая строка содержит варианты, а вторая их частоты n_i или относительные частоты (частости) p_i^*.

Пример 1. На юбилейном вечере разыграна беспроигрышная лотерея. В результате по рублю выиграли 20 человек, по два рубля 10 человек, по три рубля 30 человек, по четыре рубля 15 человек и по пять рублей 25 человек.

Решение. Запишем исходные данные примера в виде статистического ряда.

Всего разыграно N=100 выигрышей. По одному рублю выиграли 20 человек значит x_1=1, n_1=20, следовательно: x_2=2, n_2=10, x_3=3, n_3=30, x_4=4, n_4=15, ..., x_5=5, n_5=25.

Посчитаем относительные частоты для: x_1, $x_2 x_1 x_1 x_1$:

$$p_1^* = \frac{n_1}{N} = \frac{20}{100} = 0{,}2,\ p_2^* = \frac{10}{100} = 0{,}1,\ p_3^* = \frac{30}{100} = 0{,}3,$$

$$p_4^* = \frac{15}{100} = 0{,}15,\ p_5^* = \frac{25}{100} = 0{,}25$$

Статистические ряды имеют вид:

x_i	1	2	3	4	5
n_i	20	10	30	15	25

$$\left(\sum_{i=1}^{5} n_i = 100\right)$$

x_i	1	2	3	4	5
p_i	0,2	0,1	0,3	0,15	0,25

$$\left(\sum_{i=1}^{5} p_i^* = 1\right)$$

Статистический ряд выборки отражает закон изучаемого явления (наблюдения), при этом характер закона распределения неизвестен, но характер сходимости частоты и вероятности впервые изучен Я. Бернулли[12] и сформулирован в соответствующей теореме (см п.9.5).

[12] Якоб Бернулли (1654-1705)- швейцарский математик, профессор Базельского университета (с 1687). Фундаментальные достижения в теории вероятностей – разработка вероятностной модели независимых повторных испытаний (испытания Бернулли). Дал первое

Согласно теореме Я. Бернулли: частоты p_i^* при неограниченном увеличении числа опытов сходятся по вероятности к вероятностям p_i:

$$p_i^* = \xrightarrow[n \to \infty]{P} p_i\,, \tag{11.1.2}$$

Таким образом, согласно теореме Бернулли, можно утверждать, при больших значениях n статистическое распределение близка к истинному распределению.

Если объем выборки велик ($n > 50$), и при этом мы имеем дело с одномерной непрерывной случайной величиной или с одномерной дискретной случайной величиной, число возможных значений которой достаточно велико ($k > 10$), то для упрощения дальнейшей статистической обработки результатов наблюдений удобно перейти к так называемым «сгруппированным выборочным данным».

Алгоритм группирования выборочных данных:

1. Отмечаются наименьшее x_{min} и наибольшее x_{max} значения выборки;
2. Для определения числа равных интервалов k, на который следует разбить область изменения признака R= $x_{max} - x_{min}$, пользуются формулой Стерджеса:

   $$k \approx \log_2 n + 1 \approx 1 + 3{,}322 \lg n\,, \tag{11.1.3}$$

   где n — объем статистической совокупности. При этом следует учитывать, что число интервалов должно находиться в пределах от (8-10) до (20-25); $n \geq 50$;
3. Определяем величину интервала r:

   $$r = \frac{R}{k}\,, \tag{11.1.4}$$

   В ряде случаев интервалы берут разной длины (особенно, при неравномерном рассеивании экспериментальных данных).

   Замечание. Иногда при выпадении крайних значений экспериментальных величин, за начало первого интервала рекомендуется брать величину:

   $$x_0 = x_{min} - \frac{r}{2}; \tag{11.1.5}$$

4. Отмечаются крайние точки (границы) интервалов x_0, x_1, ..., x_{k-1}, x_k в порядке их возрастания, а также середины интервалов :

доказательство закона больших чисел. Основной труд по теории вероятностей “Искусство предположений” (опубликован в 1713 г).

$$\overset{o}{x}_k = \frac{x_{k+1} + x_i}{2}; \qquad (11.1.6)$$

5. Подсчитываются числа выборочных данных $n_1, n_2, \ldots, n_k$, попавших в каждый интервал (проверка: $n_1+n_2+\ldots n_k=n$). Выборочные данные, попавшие на границы интервалов, обычно относят к правому (левому) интервалу, а иногда равномерно распределяют между соседними интервалами.

Результаты группирования представляют в виде таблиц. Таблица $\{[x_j, x_{j+1}], n_j\}$ называется сгруппированным статистическим рядом (табл. 11.1.2).

Таблица 11.1.2

Интервалы R_j	$[x_0, x_1)$	$[x_1, x_2)$	...	$[x_{k-1}, x_k)$
Числа попаданий n_j	n_1	n_2	...	n_k

Статистический ряд распределения получается при замене n_j в таблице (11.1.3) на относительные частоты p_i^*:

Таблица 11.1.3

Интервалы R_j	$[x_0, x_1)$	$[x_1, x_2)$	...	$[x_0, x_1)$
Частости p_i^*	p_1^*	p_2^*	...	p_k^*

Ряд распределения (табл. 11.1.3) называется интервальным статистическим рядом распределения, при этом:

$$\sum_{i=1}^{k} p_i^* = p_1^* + p_2^* + \cdots + p_k^* = 1.$$

Пример 2. Путем опроса получены 25 значений веса студентов третьего курса МФПУ «Синергия». Результаты опроса таковы: 54, 72, 60, 83, 80, 65, 41, 70, 44, 55, 74, 63, 83, 74, 70, 59, 55, 60, 57, 60, 62, 83, 70, 75, 68. Необходимо построить интервальный статистический ряд.

Решение. Определяем минимальное и максимальные значения: $x_{min}=41$, $x_{max}=83$, следовательно $R= x_{max}-x_{min}= 83-41= 42$, $k\approx 1+3{,}322 \lg 25\approx 6{,}9$.

Берем $k=6$, тогда величина интервала $r=R/k=42/6=7$. Исходные данные разбиваем на 6 интервалов: [41, 48), [48, 55), [55, 62), [62, 69), [69, 76), [76, 83].

Посчитав число студентов (n_j), принадлежащих к конкретному промежутку, получим сгруппированный статистический и интервальный статистический ряды:

Вес R_j	[41, 48)	[48, 55)	[55, 62)	[62, 69)	[69, 76)	[76, 83)
n_j	2	1	7	4	7	4

Статистический ряд распределения получается при замене в таблице 11.1.2 на относительные частоты:

Вес R_j	[41, 48)	[48, 55)	[55, 62)	[62, 69)	[69, 76)	[76, 83)
n_j	0,08	0,04	0,28	0,16	0,28	0,16

11.2. Эмпирическая функция распределения

Одним из форм представления вариационных рядов является построение эмпирической функции распределения.

Пусть n_x – число случаев, в которых наблюдалось значение признака X, меньшее x, тогда при объеме выборки, равном n, относительная частота (частость) события $X<x$ равна n_x/n.

Функция, позволяющая определить для каждого x относительную частоту (частость) события $X<x$:

$$F^*(x) = p^*(X < x) = \frac{n_x}{n}, \tag{11.2.1}$$

называется эмпирической функцией распределения или выборочной функцией распределения.

Главное различие между теоретической функцией распределения $F(x)$ и эмпирической функцией распределения $F^*(x)$ состоит в том, что $F(x)$ определяет вероятность события $X<x$, а $F^*(x)$ – относительную частоту этого события.

Из теоретических результатов теории вероятностей (законы больших чисел) следует, что при больших n вероятность отличия этих функций друг от друга близка к единице:

$$F^*(x) = p^*(X < x) = \frac{n_x}{n}, \tag{11.2.2}$$

Из определения (11.2.1) имеем следующие свойства эмпирической функции:

1) $0 \le F^*(x) \le 1$;

2) $F^*(x)$ – неубывающая функция;

3) $F^*(-\infty)=0$, $F^*(\infty)=1$.

Функция $F^*(x)$ является «ступенчатой», имеются разрывы в точках, которым соответствуют наблюдаемые значения вариантов. Величина скачка равна относительной частоте варианта.

Аналитически $F^*(x)$ задается следующим выражением:

$$F^*(x) = p^*(X < x) = \begin{cases} 0, & \text{при } x \le x_1 \\ p_1^*, & \text{при } x_1 < x \le x_2 \\ p_1^* + p_2^*, & \text{при} x_2 < x \le x_3 \\ p_1^* + p_2^* + p_3^*, & \text{при } x_3 < x \le x_4 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ p_1^* + p_2^* + \ldots + p_{n-1}^*, & \text{при } x_{n-1} < x \le x_n \\ 1, & \text{при } \ x > x_n. \end{cases} \qquad (11.2.3)$$

Пример 3. Построить функцию $F^*(x)$, используя условие и результаты примера 1:

x_i	1	2	3	4	5
p_i^*	0,2	0,1	0,3	0,15	0,25

$$\left(\sum_{i=1}^{5} p_i^* = 1\right)$$

Решение. Вычислим значения эмпирической функции распределения $F^*(x)$ по формуле (11.1.3) и построим её график.

$$F^*(x) = p^*(\text{x} \le 1) = 0;$$

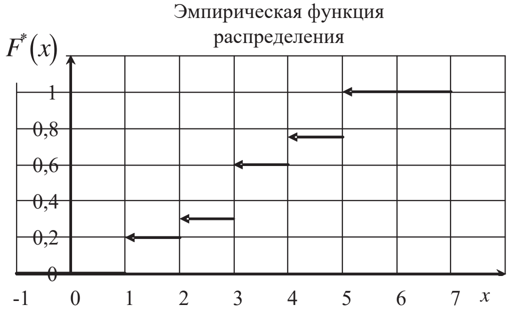
$$F^*(x) = p^*(1 < \text{x} \le 2) = 0{,}2;$$

$$F^*(x) = p^*(2 < \text{x} \le 3) = 0{,}2 + 0{,}1 = 0{,}3;$$

$$F^*(x) = p^*(3 < \text{x} \le 4) = 0{,}2 + 0{,}1 + 0{,}3 = 0{,}6;$$

$$F^*(x) = p^*(4 < \text{x} \le 5) = 0{,}2 + 0{,}1 + 0{,}3 + 0{,}15 = 0{,}75;$$

$$F^*(x) = p^*(\text{x} > 5) = 0{,}2 + 0{,}1 + 0{,}3 + 0{,}15 + 0{,}25 = 1.$$



11.3. Графическое представление статистического распределения

Для наглядного представления статистического распределения пользуются графическим изображением вариационных рядов — **полигоном частот, (относительных частот), гистограммой, кумулятой**.

Для построения полигона статистического ряда на оси абсцисс прямоугольной системы координат откладывают отдельные значения признака, в случае дискретного распределения, или интервалы значений признака, в случае интервального распределения признака.

В случае дискретного распределения, в точках значения признака восстанавливают перпендикуляры, длины которых пропорциональны соответствующим частотам (относительным частотам), концы соединяют отрезками прямых, концы крайних перпендикуляров соединяют с точками оси, частоты в которых равны нулю.

В случае интервального распределения, перпендикуляры восстанавливают в серединах интервалов. Полученная замкнутая фигура называется полигоном.

Для графического изображения интервального статистического ряда также пользуются гистограммой: на оси абсцисс откладывают интервалы значений признака, и на каждом из них, как на основании, строят прямоугольник с высотой, пропорциональной частоте интервала. Площадь гистограммы (площадь всех прямоугольников) равна единице.

Кумулята (кумулятивная функция) – это приближенная эмпирическая функция распределения, которая строится на базе интервального статистического ряда (таб. 11.1.3). Составляется таблица накопленных частот, а в качестве опорных точек берутся границы интервалов $[x_0, x_1)$, $[x_1, x_2),\ldots, [x_{k-1}, x_k)$.

При построении кумуляты по данным дискретного ряда по оси абсцисс откладываются значения признака (варианты), а по оси ординат — накопленные частоты или относительные частоты.

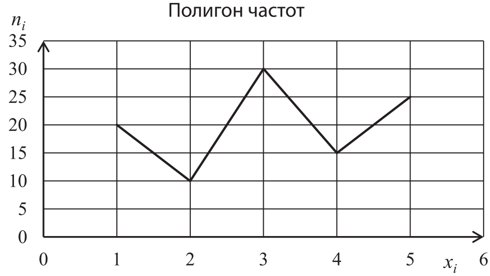
Продемонстрируем построение полигона частот, полигона относительных частот, гистограммы, интервального полигона и кумуляты на основе статистических таблиц, полученных в примерах 1 и 2.

Пример 4. Статистические ряды имеют вид:

x_i	1	2	3	4	5
n_i	20	10	30	15	25

$$\left(\sum_{i=1}^{5} n_i = 100\right)$$

Полигон частот

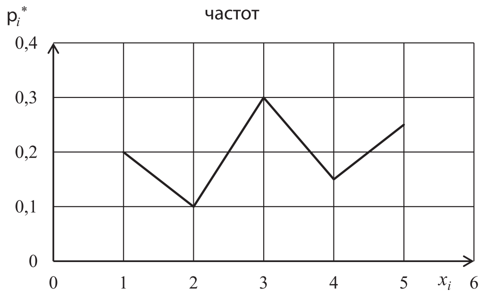


Пример 5.

x_i	1	2	3	4	5
p_i^*	0,2	0,1	0,3	0,15	0,25

$$\left(\sum_{i=1}^{5} p_i^* = 1\right)$$

Полигон относительных частот



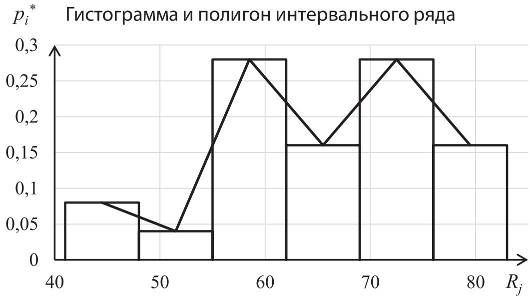
Пример 6. Построим гистограмму на основе данных таблицы, а для построения полигона относительных (серединных) частот найдем середины интервалов по формуле (11.1.6):

$$\overset{o}{x}_k = \frac{x_{k+1} + x_k}{2},$$

$$\overset{o}{x}_1 = \frac{41 + 48}{2} = 44{,}5,\ \overset{o}{x}_2 = \frac{48 + 55}{2} = 51{,}5,\ \overset{o}{x}_3 = \frac{55 + 62}{2} = 58{,}5,$$

$$\overset{o}{x}_4 = \frac{62 + 69}{2} = 65{,}5,\ \overset{o}{x}_2 = \frac{69 + 76}{2} = 72{,}5,\ \overset{o}{x}_3 = \frac{76 + 83}{2} = 79{,}5$$

Вес R_j	[41, 48)	[48, 55)	[55, 62)	[62, 69)	[69, 76)	[76, 83)
p_i^*	0,08	0,04	0,28	0,16	0,28	0,16



Пример 7. Для построения кумуляты вычислим накопленные относительные частоты:

$$p^*_{\text{нак}}(x) = p^*(\mathrm{x} \le 41) = 0;$$

$$p^*_{\text{нак}}(x) = p^*(41 < \mathrm{x} \le 48) = 0{,}08;$$

$$p^*_{\text{нак}}(x) = p^*(48 < \mathrm{x} \le 55) = 0{,}08 + 0{,}04 = 0{,}12;$$

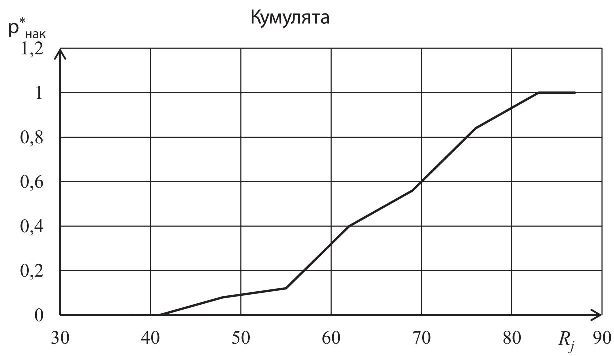
$$p^*_{\text{нак}}(x) = p^*(55 < \mathrm{x} \le 62) = 0{,}08 + 0{,}04 + 0{,}28 = 0{,}4;$$

$$p^*_{\text{нак}}(x) = p^*(62 < \mathrm{x} \le 69) = 0{,}08 + 0{,}04 + 0{,}28 + 0{,}16 = 0{,}56;$$

$$p^*_{\text{нак}}(x) = p^*(69 < \mathrm{x} \le 76) = 0{,}08 + 0{,}04 + 0{,}28 + 0{,}16 + 0{,}28 = 0{,}84;$$

$$p^*_{\text{нак}}(x) = p^*(76 < \mathrm{x} \le 83) = 0{,}08 + 0{,}04 + 0{,}28 + 0{,}16 + 0{,}28 + 0{,}16 = 1$$

$$p^*_{\text{нак}}(x) = p^*(\mathrm{x} > 83) = 1.$$



11.4. Числовые характеристики статистического распределения

Для статистического распределения можно определить ряд числовых характеристик, аналогичные тем, которые в теории вероятностей вводились для случайных величин.

Пусть дан статистический ряд:

x_i	x_1	x_2	...	x_{k-1}	x_k	$\sum$
n_i	n_1	n_2	...	n_{k-1}	n_k	$\sum_{i=1}^{k} n_i = n$
p_i^*	p_1^*	p_2^*	...	p^*_{k-1}	p^*_k	$\sum_{i=1}^{k} p_i^* = 1$

Выборочной средней $\overline{x}_B$ называют среднее арифметическое значений признака выборочной совокупности:

а) Если все значения $x_1, x_2, ..., x_n$ признака выборки объёма различны, тогда выборочное среднее определяется по формуле:

$$\overline{x}_{\text{в}} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i\,. \quad (11.4.1)$$

б) Если статистические данные сгруппированы, то формула (1.1.1) примет следующий вид:

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i \frac{n_i}{n} = \sum_{i=1}^{k} x_i\ p_i^*\ . \quad (11.4.2)$$

в) В случае интервального ряда в качестве x_i берутся середины частичных интервалов:

$$\overset{o}{x}_k = \frac{x_{k+1} + x_k}{2},$$

и соответствующие им частоты (частости).

Выборочная средняя имеет те же единицы измерения, что и варианты x_i.

Выборочное среднее является статистическим аналогом математического ожидания случайной величины.

Выборочной дисперсией $D_{\text{В}}$ называется среднее арифметическое вариант от выборочной средней:

а) если все значения в выборке различны:

$$D_{\text{в}} = \frac{1}{n}((x_1 - \overline{x}_{\text{в}})^2 + (x_2 - \overline{x}_{\text{в}})^2 + \cdots + (x_n - \overline{x}_{\text{в}})^2) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_{\text{в}})^2\ . \quad (11.4.3)$$

б) если статистические данные сгруппированы, то формула (11.4.3) примет вид:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^2\ n_i = \sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^2 \frac{n_i}{n} = \sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^2\ p_i^* . \quad (11.4.4)$$

в) в случае интервального ряда в качестве x_i берутся середины частичных интервалов:

$$\overset{o}{x}_k = \frac{x_{k+1} + x_k}{2},$$

и соответствующие им частоты (частости).

Получим формулу, которую удобно использовать на практике:

$$D_{\text{в}} = \sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^2\ p_i^* = \sum_{i=1}^{k}\left(x_i^2 - 2x_i\overline{x}_{\text{в}} + \overline{x}_{\text{в}}^2\right) p_i^* =$$

$$\sum_{i=1}^{k} x_i^2\ p_i^* - 2\overline{x}_{\text{в}}\sum_{i=1}^{k} x_i\ p_i^* + \overline{x}_{\text{в}}^2\sum_{i=1}^{k}\ p_i^* = \overline{x^2} - 2\overline{x}_{\text{в}}\overline{x}_{\text{в}} + \overline{x}_{\text{в}}^2 \cdot 1 = \overline{x^2} - \overline{x}_{\text{в}}^2$$

Следовательно:

$$D_{\text{в}} = \overline{x^2} - \overline{x}_{\text{в}}^2. \quad (11.4.5)$$

Выборочная дисперсия является статистическим аналогом дисперсии теоретического распределения.

При решении практических задач используется и так называемая исправленная выборочная дисперсия, которая определяется из формулы:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^2\ n_i . \quad (11.4.6)$$

Тогда через выборочную дисперсию формула (11.4.6) имеет вид:

$$S^2 = \frac{n}{n-1} D_{\text{в}} . \quad (11.4.7)$$

Выборочное среднее квадратическое отклонение определяется как арифметический корень из выборочной дисперсии:

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}} . \quad (11.4.8)$$

Выборочное среднее квадратическое отклонение показывает надежность выборочной средней: чем меньше выборочное квадратическое отклонение, тем лучше выборочная средняя отражает собой всю представляемую совокупность.

Как было отмечено в пункте 6.3, для характеристики меры колеблемости изучаемого признака по относительно выборочной средней служит коэффициент вариации:

$$k_{var}^{\text{в}} = \frac{\sigma_{\text{в}}}{\overline{x}_{\text{в}}} \text{ или } k_{var}^{\text{в}} = \frac{\sigma_{\text{в}}}{\overline{x}_{\text{в}}} \cdot 100\% \ . \qquad (11.4.9)$$

В качестве описательных характеристик вариационного ряда или составленного на его основе статистического распределения используются размах выборки (вариации), мода, медиана и др.

Размахом вариации называется число (см. выше):

$$R = x_{max} - x_{min}. \qquad (11.4.10)$$

где x_{max} – наибольшей вариант ряда, x_{min} – наименьший вариант ряда.

Модой M_0^* вариационного ряда называется вариант, имеющий наибольшую частоту.

Медианой M_e^* называется значение признака, приходящееся на середину ранжированного ряда.

Если вариационный ряд имеет нечетное число членов n=2k+1, то $M_e^*=x_{k+1}$, если число членов ряда четно $n=2k$, то:

$$M_e^* = \frac{x_k + x_{k+1}}{2}. \qquad (11.4.11)$$

При рассмотрении вышеуказанных числовых характеристик статистического ряда предполагалось, что вариационный ряд составлен по данным выборки, поэтому они называются выборочными, если вариационный ряд составлен по данным генеральной совокупности, то характеристики называются генеральными.

Пример 8. По результатам примера 1 найти характеристики выборки.

x_i	1	2	3	4	5
n_i	20	10	30	15	25

$$\left(\sum_{i=1}^{5} n_i = 100\right)$$

Решение. По формулам (11.4.1), (11.4.4), (11.4.8), 11.4.7, 11.4.9 найдем требуемые числовые характеристики:

$$\overline{x}_{\text{в}} = \frac{1}{100}(1 \cdot 20 + 2 \cdot 10 + 3 \cdot 30 + 4 \cdot 15 + 5 \cdot 25) = 3{,}15$$

$$D_{\text{в}} = \frac{1}{100}((1 - 3{,}15)^2 \cdot 20 + (2 - 3{,}15)^2 \cdot 10 + \cdots + (5 - 3{,}15)^2 \cdot 25) = 2{,}03$$

$$\sigma_{\text{в}} = \sqrt{2{,}03} = 1{,}42$$

$$S^2 = \frac{100}{99} 2{,}03 = 2{,}05$$

$$S = \sqrt{2{,}05} = 1{,}43$$

$$R = 5 - 1 = 4$$

$$M_0^* = 3$$

$$M_e^* = 3$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Каковы основные вопросы математической статистики?
2. Что такое выборочный метод?
3. Что представляют генеральная и выборочная совокупность?
4. Что называется относительным показателем выборки?
5. Назовите основные виды выборок?
6. Что такое репрезентативная выборка?
7. Каковы основные показатели генеральной совокупности?
8. Каковы основные показатели выборочной совокупности?
9. Что такое эмпирическая функция?
10. Что такое полигон, гистограмма, кумулята?

ЗАДАНИЯ

11.1. В результате испытаний получены следующие статистические значения случайной величины X: – 0,9, 0,1, 3,1, – 0,9, – 2,9, – 8,9, 5,1, – 3,9, 1,39, 1,1, 5,1, 1,1, 6,1, 0,1, 6,1, – 2,9, 1,1, 3,1, 1,1, – 6,9, – 2,9, 0,1, – 0,9, – 0,1, – 3,9, 0,1, 0,1, – 2,9, 1,1, 0,1, 3,1, 0,1.

Требуется:

а) составить статистический ряд;

б) найти статистическую функцию распределения $F^*(x)$;

в) изобразить полигон относительных частот.

11.2. В результате испытаний получены следующие статистические значения случайной величины X: 1,40, 1,38, 1,35, 1,33, 1,37, 1,34, 1,38, 1,37, 1,39, 1,36, 1,35, 1,34, 1,37, 1,38, 1,36, 1,35, 1,31, 1,37, 1,39, 1,39, 1,41, 1,37, 1,35, 1,36, 1,34, 1,36, 1,35, 1,34, 1,40, 1,36.

Требуется:

а) составить статистический ряд;

б) найти статистическую функцию распределения $F^*(x)$;

в) изобразить полигон относительных частот.

11.3. В организации работает 20 сотрудников. Ниже приведены данные опроса этих сотрудников о размере денежных премий (тыс. руб.), полученных ими в течение года: 2, 2, 0,1, 0, 5, 0, 1, 0, 0, 0, 4, 2, 0, 0, 2, 0, 3, 0, 1.

1) Составить статистическое распределение выборки;

2) Найти моду вариационного ряда;

3) Построить полигон частот статистического распределения.

11.4. Получена выборка измерения массы тела 25 студентов-юношей: 55,3; 57,6; 54,8; 61,7; 58,4; 62,3; 47,1; 64,2; 65,3; 67,1; 68,7; 72,2; 66,8; 70,2; 67,7; 72,6; 70,8; 69,1; 68,6; 74,6; 77,2; 78,9; 81,1; 89,7; 84,8.

1) Составить интервальное статистическое распределение выборки;

2) Построить гистограмму частот статистического распределения.

11.5. В результате испытаний получены следующие статистические значения случайной величины X: 3,45. 3,50, 3,46, 3,48, 3,40, 3,46, 3,40, 3,44, 3,47, 3,41, 3,47, 3,49, 3,47, 3,44, 3,50, 3,44, 3,40, 3,43, 3,48, 3,43, 3,47, 3,47, 3,51, 3,47, 3,41, 3,40, 3,53, 3,50, 3,39, 3,47, 3,38, 3,47, 3,45, 3,46, 3,52, 3,46, 3,40, 3,41, 3,42, 3,43, 3,45, 3,44, 3,44, 3,50, 3,48, 3,50, 3,47, 3,50, 3,52, 3,48.

Построить:

а) интервальный статистический ряд;

б) статистический ряд, рассматривая в качестве значений середины интервалов;

в) статистическую функцию распределения $F^*(x)$;

г) гистограмму относительных частот.

11.6. При измерении диаметров ста подшипниковых шариков, выбранных из большой партии шариков для определения стандартности, получены следующие результаты: 8,31, 8,42, 8,37, 8,40, 8,30, 8,30, 8,42, 8,32, 8,29, 8,33, 8,36, 8,34, 8,37, 8,32, 8,36, 8,38, 8,33, 8,36, 8,40, 8,36, 8,32, 8,36, 8,36, 8,30, 8,30, 8,33, 8,35, 8,37, 8,37, 8,30, 8,40, 8,34, 8,33, 8,37, 8,34, 8,38, 8,29, 8,34, 8,31, 8,36, 8,37, 8,30, 8,41, 8,34, 8,37, 8,35, 8,40, 8,34, 8,36, 8,37, 8,37, 8,41, 8,35, 8,38, 8,33, 8,36, 8,36, 8,36, 8,37, 8,36, 8,40, 8,37, 8,34, 8,37, 8,32, 8,35, 8,36, 8,37, 8,41, 8,36, 8,36, 8,36, 8,40, 8,34, 8,40, 8,34, 8,33, 8,35, 8,37, 8,34, 8,36, 8,37, 8,37, 8,35, 8,36, 8,34, 8,42, 8,36, 8,33, 8,34, 8,35, 8,36, 8,32, 8,38, 8,32, 8,36, 8,37.

Построить:

а) интервальный статистический ряд;

б) статистический ряд, рассматривая в качестве значений середины интервалов;

в) статистическую функцию распределения $F^*(x)$;

г) гистограмму относительных частот.

11.7. По данной выборке признака X построить полигон (гистограмму) частот.

x_i^*	21	26	31	36	41	46	51
n_i	7	11	12	60	5	3	2

11.8. При изучении физико-механических свойств обувных кож испытано 10 образцов и получены значения предела прочности на разрыв X(H, мм) , представленные в таблице:

№ п/п	1	2	3	4	5	6	7	8	9	10
X(H, мм)	16,2	20,1	21,4	18,9	16,5	17,3	18,2	19,5	20,4	21,0

Требуется выполнить первичную статистическую обработку данных:

1. Определить выборочное среднее;
2. Найти «исправленное» стандартное отклонение;
3. Найти коэффициент вариации изучаемого признака.

Пояснить смысл полученных результатов.

11.9. При оценке свойств картофеля было обследовано 10 проб, и полученные значения содержания крахмала X% представлены в таблице:

№ п/п	1	2	3	4	5	6	7	8	9	10
X%	15,2	12,8	13,5	14,9	15,6	16,0	13,7	14,1	13,2	15,0

Требуется выполнить первичную статистическую обработку данных:

1. Определить выборочное среднее;
2. Найти «исправленное» стандартное отклонение;
3. Найти коэффициент вариации изучаемого признака.

Пояснить смысл полученных результатов.

11.10. С целью определения рациональной структуры размерного ассортимента детской одежды проведено выборочное обследование определенных групп детского населения, и получено распределение количества по величине обхвата груди X (см), представлены табл.9

Обхват груди X (см)	54-58	58-62	62-66	66-70	70-74	74-78	Σ
Количество детей n_i	21	40	59	62	26	14	225

Требуется:

1. Построить гистограмму относительных частот для приведенных в таблице значений признака;
2. Найти выборочное среднее, среднее квадратическое отклонение и коэффициент вариации признака X.

11.11. С целью определения рациональной структуры размерного ассортимента детской обуви проведено выборочное обследование определенных групп детского населения, и получено распределение количества детей по величине длины стопы X (см), представлены в таблице.

Длина стопы X (см)	170-175	175-180	180-185	185-190	190-195	195-200	Σ
Количество детей n_i	5	18	59	61	32	21	196

Требуется:

1. Построить гистограмму относительных частот для приведенных в таблице значений признака;
2. Найти выборочное среднее, стандартное отклонение и коэффициент вариации признака X.

ГЛАВА 12
ОСНОВНЫЕ ПОДХОДЫ К СТАТИСТИЧЕСКОМУ ОЦЕНИВАНИЮ

12.1. Понятие об оценке

Собранная статистическая информация, естественно, нуждается в оценке. Оценку можно производить точечно по результатам выборочного обследования. Задача оценивания параметров распределений по результатам наблюдений может решаться и в интересах сглаживания эмпирических (выборочных) кривых распределения подходящей теоретической кривой.

Любой закон распределения случайных величин X однозначно определяется рядом параметров – начальных и центральных моментов. Например, нормальный закон определяется двумя параметрами, показательный – одним, закон Пуассона – одним.

Обозначим $\Theta = (\Theta_1, \Theta_2, ..., \Theta_k)$ – вектор неизвестных параметров. Тогда для нормального закона – $\Theta_1 = m_x$, $\Theta_2 = \sigma_x$ для показательного – $\Theta_1 = \lambda = 1/m_x$, пуассоновского – $\Theta_1 = \lambda = m_x = D_x$.

Задача оценивания параметров распределений по результатам наблюдений может решаться и в интересах сглаживания эмпирических (выборочных) кривых распределения подходящей теоретической кривой.

Оценка неизвестных параметров закона распределения генеральной совокупности производится по выборке $(x_1, x_2, ..., x_n)$. Для определения неизвестных параметров $\Theta_1, \Theta_2, ..., \Theta_k$ по выборке используют функции результатов наблюдений, так называемые статистики или выборочные функции $X_B = X_B(x_1, x_2, ..., x_n)$.

Таким образом, **статистикой** будем называть любую функцию X_B результатов наблюдений $x_1, x_2, ..., x_n$ исследуемой случайной величины (параметра) X.

Оценками Θ^* неизвестных параметров будем называть наилучшие в некотором смысле статистики (выборочные функции) X_B. Оценка должна быть несмещенной, эффективной, состоятельной статистикой.

Оценка Θ^*называется **несмещенной**, если её математическое ожидание будет равно оцениваемому параметру:

$$M(\Theta^*) = \Theta \qquad (12.1.1)$$

Оценка Θ^*называется **смещенной**, если:

$$M(\Theta^*) \neq \Theta \qquad (12.1.2)$$

Оценка Θ^*называется э**ффективной**, если она среди всех прочих оценок того же самого параметра Θ обладает наименьшей мерой случайного разброса относительно истинного значения оцениваемого параметра, то есть обладает минимальной дисперсией.

Оценка Θ^* называется **состоятельной**, если по мере роста числа наблюдений n, то есть при $n \to \infty$ она сходится по вероятности к оцениваемому параметру Θ:

$$\lim_{n\to\infty} P(|\Theta^* - \Theta| < \varepsilon) = 1, \qquad (12.1.3)$$

где $\varepsilon > 0$ – сколь угодно малая величина.

12.2. Точечные статистические оценки параметров распределения

Пусть собранный и обработанный статистический материал представлен в виде статистического ряда.

Точечной статистической оценкой параметра Θ распределения случайной величины называется приближенное значение Θ^* этого параметра, вычисленное по статистическим данным.

Любая точечная статистическая оценка некоторого параметра, вычисляемая на основе статистического ряда, должна удовлетворять требованиям несмещенности, эффективности и состоятельности.

Применительно к точечной оценке, основная проблема заключается в выборе возможно лучшей оценки, отвечающей требованиям несмещенности, эффективности и состоятельности.

Рассмотрим вопрос, какими выборочными характеристиками по требованиям несмещенности, эффективности и состоятельности оцениваются математическое ожидание и дисперсия.

Статистической оценкой математического ожидания m_x называется среднее арифметическое статистических значений изучаемой случайной величины:

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{\kappa} x_i n_i\,, \qquad (12.2.1)$$

где $n + n_2 + \ldots + n_k = n$.

Эта оценка математического ожидания m_x обладает всеми свойствами оценок: состоятельности, несмещенности, эффективности.

Смещенной оценкой дисперсии D_x называется выборочная дисперсия:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x}_{\text{в}})^2 \ , \qquad (12.2.2)$$

Выборочная дисперсия является смещенной, так как:

$$M(D_{\text{в}}) = \frac{n-1}{n} D_x .$$

Несмещенной оценкой дисперсии D_x называется исправленная выборочная дисперсия:

$$S^2 = \frac{n}{n-1} D_{\text{в}} = \frac{1}{n-1} \cdot \sum_{i=1}^{k} n_i (x_i - \overline{x}_{\text{в}})^2 \qquad (12.2.3)$$

При расчете S^2 можно пользоваться более удобной формулой:

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} n_i \cdot x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} n_i \cdot x_i\right)^2\right) \qquad (12.2.4)$$

Выборочная дисперсия $D_в$ и исправленная выборочная дисперсия S^2 обладают свойством состоятельности.

Оценка S^2 не обладает свойством эффективности, но обладает свойством несмещенности, поэтому все чаще чем $D_в$ используют в качестве приближенного значения дисперсии D_x.

Оценкой среднего квадратического отклонения $_x$ называется квадратный корень из $D_в$ или S^2:

$$\sigma_в = \sqrt{D_в} \text{ или } \quad S = \sqrt{S^2} \ . \qquad (12.2.5)$$

Оценкой вероятности события A в n независимых испытаниях является относительная частота события A:

$$p^* = \frac{n_i}{n} \ . \qquad (12.2.6)$$

где n_i – появления события A i раз в n испытаниях.

Эта оценка вероятности события A в n независимых испытаниях обладает свойствами несмещенности, состоятельности и эффективности.

Если выборка состоит из вариант x_1, $x_2, \ldots, x_n$ громоздкого вида, то для упрощения расчета выборочных точечных оценок параметров следует воспользоваться свойствами выборочного среднего и выборочной дисперсии и применить метод упрощенных вычислений.

Свойства выборочного среднего и выборочной дисперсии:

1. Если варианты увеличить (уменьшить) на одно и то же число C, то выборочное среднее также увеличиться (уменьшиться) на это число, а выборочная дисперсия останется неизменной.

2. Если варианты увеличить (уменьшить) в h раз, то выборочное среднее увеличиться (уменьшиться) в h раз, а выборочная дисперсия увеличится (уменьшиться) в h^2 раз.

На основании свойств: вводим новые варианты:

$$u_i = \frac{x_i - C}{h}, \qquad (12.2.7)$$

где h – шаг между равноотстоящими вариантами, C – так называемый «ложный» нуль. В качестве числа C в основном выбирают варианту x_i которая расположена в середине статистического ряда или в крайнем случае имеет наибольшую частоту.

Для них произвести расчет точечных оценок параметров:

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{k} u_i n_i, \qquad (12.2.8)$$

$$D_в(u) = \frac{1}{n}\sum_{i=1}^{k} n_i(u_i - \overline{u})^2, \qquad (12.2.9)$$

$$S_u^2 = \frac{n}{n-1}\cdot \qquad (12.2.10)$$

Затем вычислить искомые точечные оценки:

$$\overline{x}_в = \overline{u}\cdot h + c\ , \qquad (12.2.11)$$

$$D_в = h^2 \cdot D_в(u), \qquad (12.2.12)$$

$$S_u^2 = \frac{n}{n-1}\cdot D_в(u)\ . \qquad (12.2.13)$$

12.3. Интервальные оценки параметров распределения

При выборке небольшого объема точечная оценка Θ^* может существенно отличаться от истинного значения параметра и привести к грубым ошибкам, поэтому в случае малой выборки часто используют интервальные оценки.

При этом по вычисленной точечной оценке Θ^* параметра Θ при заданной вероятности γ, называемой доверительной вероятностью, также по некоторому числу ε, зависящему от и Θ^*, строят интервал для истинного параметра Θ:

$$\Theta^* - \varepsilon < \Theta < \Theta^* + \varepsilon,$$

чтобы выполнялось равенство:

$$P(\Theta^* - \varepsilon < \Theta < \Theta^* + \varepsilon) = \gamma \qquad (12.3.1)$$

Число ε называется точностью сценки Θ^* границы интервала $\Theta^* - \varepsilon$ и $\Theta^* + \varepsilon$ – называются доверительными границами, интервал $(\Theta^* - \varepsilon; \Theta^* + \varepsilon)$ – доверительным интервалом, вероятность γ – доверительной вероятностью или надежностью интервальной оценки.

Интервальной оценкой математического ожидания m_x нормального распределения при известной дисперсии σ^2 называется интервал $(\overline{x}_B - \varepsilon;$ $\overline{x}_B + \varepsilon)$, где:

$$\varepsilon = z_\gamma \cdot \frac{\sigma}{\sqrt{n}} \qquad (12.3.2)$$

удовлетворяющий равенству:

$$P(\overline{x}_B - \varepsilon < m_x < \overline{x}_B + \varepsilon) = \gamma \qquad (12.3.3)$$

где γ – заданная доверительная вероятность, m_x – истинное математическое ожидание, $\overline{x}_B$ – точечная оценка мат ематического ожидания, n – объем выборки; число z_γ находится из уравнения:

$$\Phi_0(z_\gamma) = \frac{\gamma}{2} \qquad (12.3.4)$$

с помощью табл. 3 функции Лапласа , $\Phi_0(z)$ см. приложение 3.

Следовательно, интервальная оценка математического ожидания находится по формуле:

$$\overline{x}_в - z_\gamma \cdot \frac{\sigma}{\sqrt{n}} < m_x < \overline{x}_в + z_\gamma \cdot \frac{\sigma}{\sqrt{n}} \qquad (12.3.5)$$

Интервальной оценкой математического ожидания m_x нормального распределения при неизвестной дисперсии называется интервал

$$(\overline{x}_в - \varepsilon; \overline{x}_в + \varepsilon),$$

$$\varepsilon = t_\gamma \cdot \frac{S}{\sqrt{n}}, \qquad (12.3.6)$$

удовлетворяющий равенству:

$$P(\bar{x}_\text{в} - \varepsilon < m_x < \bar{x}_\text{в} + \varepsilon) = \gamma, \qquad (12.3.7)$$

где γ – заданная доверительная вероятность, m_x – истинное математическое ожидание, $\bar{x}_B$ – точечная оценка математического ожидания, S^2 – точечная оценка дисперсии, n – объем выборки; число t_γ вычисляется из уравнения :

$$\int_0^{t_\gamma} S(t;k)dt = \frac{\gamma}{2}, \qquad (12.3.8)$$

с помощью табл. 5 распределения Стьюдента (см. приложение 5).

Следовательно, интервальная оценка математического ожидания с доверительной вероятностью γ вычисляется по формуле:

$$\bar{x}_\text{в} - t_\gamma \cdot \frac{S}{\sqrt{n}} < m_x < \bar{x}_\text{в} + t_\gamma \cdot \frac{S}{\sqrt{n}}. \qquad (12.3.9)$$

Интервальной оценкой среднего квадратического отклонения σ нормального распределения называется интервал:

$$(S-\varepsilon;\, S+\varepsilon),\ \varepsilon = q_\gamma \cdot S, \qquad (12.3.10)$$

удовлетворяющий равенству:

$$P(S-\varepsilon<\sigma<S+\varepsilon)=\gamma$$

где γ – заданная доверительная вероятность, S^2 – исправленная выборочная дисперсия, n – объем выборки, число q_γ определяется из табл. 6 (см, приложение 6).

Следовательно, интервальная оценка среднего квадратического отклонения находится по формулам:

$$S(1 - q_\gamma) < \sigma < S(1 + q_\gamma),\ \text{если}\ q_\gamma < 1, \qquad (12.3.11)$$

$$0 < \sigma < S(1 + q_\gamma),\ \text{если}\ q_\gamma > 1. \qquad (12.3.11)$$

12.4. Метод моментов для точечной оценки параметров распределения

Оценка одного параметра

Пусть задан вид плотности распределения $f(x, \Theta)$, определяемый одним неизвестным параметром Θ.

Требуется найти точечную оценку параметра Θ.

$$M(\Theta^*) = \Theta$$

Для оценки одного параметра достаточно иметь одно уравнение относительно этого параметра. Следуя методу моментов, приравняем, например, начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка:$\alpha_1=M_1$.

Учитывая, что: $\alpha_1=m_x$ $M_1=\bar{x}_B$, получим:

$$m_x=\bar{x}_B. \tag{12.4.1}$$

Математическое ожидание m_x, как видно из соотношения: 677_

$$m_x = \int_{-\infty}^{\infty} (x, \Theta)dx = \varphi(\Theta), \tag{12.4.2}$$

есть функция от Θ, поэтому (12.4.1) можно рассматривать как уравнение с одним неизвестным Θ. Решив это уравнение относительно параметра Θ, тем самым найдем его точечную оценку Θ^*, которая является функцией от выборочной средней, следовательно, и от вариант выборки:

$$\Theta^*=\phi(x_1, x_2, ..., x_n).$$

Для экспоненциального распределения:

$$m_x=1/\lambda.$$

Тогда для оценки Θ^*:

$$1/\lambda^*= \bar{x}_B \Rightarrow \lambda^*=1/ \bar{x}_B.$$

Оценка двух параметров.

Пусть задан вид плотности распределения $f(x, \Theta_1 , \Theta_2)$, определяемый неизвестными параметрами Θ_1 и Θ_2. Для отыскания двух параметров необходимы два уравнения относительно этих параметров. Следуя методу моментов, приравняем, например, начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка и центральный теоретический момент второго порядка центральному эмпирическому моменту второго порядка:

$$\alpha_1= M_1 , \mu_2=m_2.$$

Учитывая, что $\alpha_1= m_x$, $M_1=\bar{x}_B$, $m_2= D_B$, $\mu_2=D(X)$, получим: $m_x=\bar{x}_B$, $D(X)= D_B$.

Математическое ожидание и дисперсия есть функция от Θ_1 и Θ_2, поэтому (12.4.2) можно рассматривать как систему двух уравнений с двумя неизвестными Θ_1 и Θ_2. Решив эту систему относительно неизвестных параметров, тем самым получим их точечные оценки Θ_1^* и Θ_2^*. Эти оценки являются функциями от вариант выборки:

$$\Theta_1^*=\phi(x_1, x_2, ..., x_n),\ \Theta_2^*=\phi(x_1, x_2, ..., x_n).$$

12.5. Метод наибольшего правдоподобия для точечной оценки параметров распределения

Метод максимального или наибольшего правдоподобия предложен Р. Фишером. С помощью этого метода производится точечная оценка неизвестных параметров априорно известного закона распределения случайной величины.

Рассмотрим сначала суть метода при оценке параметров дискретного распределения случайной величины.

Обозначим вероятность того, что в результате испытания случайная величина X примет значение x_i,(i=1, 2, ..., n), через $p_i(x_i,\Theta)$.

Функцией правдоподобия случайной дискретной величины X называют функцию аргумента Θ :

$$L(x_1, x_2, ..., x_n, \Theta) = p(x_1,\Theta)\, p(x_2,\Theta) \ldots p(x_n,\Theta), \qquad (12.5.1)$$

где $x_1, x_2, ..., x_n$ — фиксированные числа, полученные при измерении случайной величины X.

В качестве точечной оценки параметра Θ принимают такое его значение $\Theta^*=\Theta^*(x_1, x_2, ..., x_n)$, при котором функция правдоподобия достигает максимума. Оценку Θ^* называют оценкой максимального правдоподобия.

Для упрощения расчетов в рассмотрение вводится логарифм функции правдоподобия $\ln L$, которую называют логарифмической функцией правдоподобия. Функции L и $\ln L$ достигают максимума при одном и том же значении своего аргумента, поэтому вместо отыскания максимума функции L ищут максимум функции $\ln L$. Записывая необходимое условие экстремума функции правдоподобия в случае скалярного параметра, получаем уравнения правдоподобия:

$$\frac{\partial L(\{x_n\}, \theta)}{\partial \theta} = 0, \qquad (12.5.2)$$

Или

$$\frac{\partial \ln L(\{x_n\}, \theta)}{\partial \theta} = 0, \qquad (12.5.3)$$

где $\{x_n\}$ — заданная выборка случайных величин.

Уравнение правдоподобия (12.5.3) с логарифмической функцией, как правило, более простое относительно функции правдоподобия (12.5.2).

Если распределение случайной величины X зависит от вектора параметров $\{\Theta\} = \Theta^*(x_1, x_2, ..., x_n)$, то уравнение (12.5.3) заменяется системой уравнений:

$$\frac{\partial \ln L(\{x_n\}, \theta)}{\partial \theta_k} = 0. \quad (k = 1,2, \dots, r) \tag{12.5.4}$$

Именно уравнения (12.5.3) и (12.5.4) принято называть уравнениями правдоподобия. Во многих случаях решение системы (12.5.4), являющейся, как правило, нелинейной, приходится искать численными методами.

Рассмотрим применение метода максимального правдоподобия для оценки параметров непрерывного распределения случайных величин генеральной совокупности $X°$.

Пусть Θ – непрерывная случайная величина, которая в результате испытаний n приняла значения $x_1, x_2, ..., x_n$. Предполагается, что вид плотности распределения $f(x)$ задан, но неизвестен параметр Θ, которым определяется эта функция.

Функцией правдоподобия непрерывной случайной величины X называют функцию аргумента Θ:

$$L=(x_1, x_2, ..., x_n°, \Theta)=L(x_1, x_2, ..., x_n, \Theta)=f(x_1, \Theta)\, f(x_2, \Theta), \dots, f(x_n, \Theta), \tag{12.5.5}$$

где $x_1, x_2, ..., x_n$ - фиксированные числа.

Оценку максимального правдоподобия неизвестного параметра Θ распределения непрерывной случайной величины ищут так же, как в случае дискретной величины.

Если плотность распределения непрерывной Θ_1 и Θ_2 случайной величины X определяется двумя неизвестными параметрами , то функция правдоподобия является функцией двух независимых аргументов Θ_1 и Θ_2:

$L=(x_1, x_2, ..., x_n°, \Theta_1, \Theta_2)=L(x_1, x_2, ..., x_n, \Theta)=f(x_1, \Theta_1, \Theta_2)\, f(x_2, \Theta_1, \Theta_2), \dots, f(x_n, \Theta_1, \Theta_2)$, 7.6

Как для дискретных распределений, так и для непрерывных, точку максимума логарифмической функции распределения $\ln L$ аргумента Θ можно искать через необходимое условие экстремума:

1) найти производную

$$\frac{d \ln L}{d\theta}, \ \frac{d^2 \ln L}{d\theta^2} \ ;$$

2) приравнять производную нулю и найти критическую точку — корень полученного уравнения (его называют уравнением правдоподобия);
3) найти вторую производную

$$\frac{d^2 \ln L}{d\theta^2};$$

4) если вторая производная при $\Theta = \Theta^*$ отрицательна, то Θ^* — точка максимума.

Найденную точку максимума Θ^* принимают в качестве оценки максимального правдоподобия параметра Θ.

Метод максимального правдоподобия имеет ряд достоинств: его оценки, вообще говоря, состоятельны (но они могут быть смещенными), распределены асимптотически нормально (при больших значениях *n* приближенно нормально) и имеют наименьшую дисперсию по сравнению с другими асимптотически нормальными оценками; если для оцениваемого параметра Θ существует эффективная оценка Θ^*, то уравнение правдоподобия имеет единственное решение Θ^*; этот метод наиболее полно использует данные выборки об оцениваемом параметре, поэтому он особенно полезен в случае малых выборок. Недостаток метода состоит в том, что он часто требует сложных вычислений.

Найдем методом наибольшего правдоподобия оценки параметров *a* и нормального распределения:

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

если в результате *n* испытаний величина *X* приняла значения $x_1, x_2, \ldots, x_n$.

Решение. Составим функцию правдоподобия, учитывая, что $\Theta_1 = \alpha$, $\Theta_2 = \sigma$

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x_1-a)^2/(2\sigma^2)} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_2-a)^2/(2\sigma^2)} \cdot \ldots$$

$$\ldots \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_n-a)^2/(2\sigma^2)}$$

Отсюда: $L = \dfrac{1}{\sigma^n(\sqrt{2\pi})^n} e^{-(\sum (x_i-a)^2/(2\sigma^2))}$.

Найдем логарифмическую функцию правдоподобия:

$$ln\,L = -n\,ln\,\sigma + ln\frac{1}{(\sqrt{2\boldsymbol{\pi}})^n} - \frac{\sum(x_i - a)^2}{2\sigma^2}.$$

Найдем частные производные по a и по σ:

$$\frac{\partial\,ln\,L}{\partial a} = \frac{\sum x_i - na}{\sigma^2},$$

$$\frac{\partial\,ln\,L}{\partial\sigma} = \frac{\sum(x_i - a)^2 - n\sigma^2}{4\sigma^4}.$$

Приравняв частные производные нулю и решив полученную систему двух уравнений относительно a и σ^2, получим:

$$\alpha=\Sigma x_i/n=\overline{x}_B;\ \sigma^2=(\Sigma(x_i-\overline{x}_B)^2/n)=D_B.$$

Итак, искомые оценки наибольшего правдоподобия: $\alpha^*=\overline{x}_B$; $\sigma^*=\sqrt{D_B}$. Заметим, что первая оценка несмещенная, а вторая смещенная.

12.6. Примеры точечных и интервальных оценок числовых характеристик

Пример 1. Найти доверительный интервал с надежностью 0,95 для оценки математического ожидания m_x нормально распределенной случайной величины *X,* если известно среднее квадратическое отклонение σ=2 , оценка математического ожидания $\overline{x}_B$=10, объем выборки n=36.

Решение.

Доверительный интервал для истинного математического ожидания m_x с доверительной вероятностью γ=0,95 при известной дисперсии σ находится по формуле (12.3.5):

$$\overline{x}_\text{в} - z_\gamma \cdot \frac{\sigma}{\sqrt{n}} < m_x < \overline{x}_\text{в} + z_\gamma \cdot \frac{\sigma}{\sqrt{n}}$$

где m_x - истинное математическое ожидание; $\overline{x}_B$ – оценка m_x по выборке; n – объем выборки; $z\gamma$ находится по доверительной вероятности γ=0,95 из равенства (12.3.40) при γ/2=0,475 :

$$\Phi_0(z_\gamma)=0{,}475.$$

Из табл. 3 приложения 3 находим z_γ=1,96. Следовательно, найден доверительный интервал для m_x:

$$10 - 1{,}96\frac{2}{\sqrt{36}} < m_x < 10 + 1{,}96\frac{2}{\sqrt{36}},$$

то есть $9{,}347 < m_x < 10{,}653$.

Ответ: (9,347; 10,653).

Пример 2. По данным испытаний построен статистический ряд:

x_i	3	2	1	0	1	2	3
n_i	1	3	4	5	3	2	2

Найти доверительный интервал для математического ожидания m_x с надежностью 0,95.

Решение. Воспользуемся формулой (12.3.9) для доверительного интервала математического ожидания при неизвестной дисперсии:

$$\overline{x}_{\text{в}} - t_\gamma \cdot \frac{S}{\sqrt{n}} < m_x < \overline{x}_{\text{в}} + t_\gamma \cdot \frac{S}{\sqrt{n}},$$

где n - объем выборки; $\overline{x}_B$ – оценка m_x; S– оценка среднего квадратического отклонения; t_γ находится по доверительной вероятности $\gamma=0{,}95$ из табл. 5 (Стьюдент) приложения 5.

По числам $\gamma=0{,}95$ и $k=n-1=20-1=19$ находим: $t_\gamma=2{,}093$.

Находим $\overline{x}_B$ и S^2:

$$\overline{x}_{\text{в}} = \frac{1}{20}\sum_{i=1}^{7} x_i \cdot n_i = \frac{1}{20}(-3-6-4+0+3+4+6) = \frac{1}{20}\cdot 0 = 0;$$

$$S^2 = \frac{1}{19}\left(9\cdot 1 + 4\cdot 3 + 1\cdot 4 + 0 + 1\cdot 3 + 4\cdot 2 + 9\cdot 2 - \frac{1}{20}\cdot 0\right) =$$

$$= \frac{1}{19}(9+12+4+3+8+18) = \frac{1}{19}\cdot 54 \approx 2{,}84$$

Тогда, $s \approx \sqrt{2{,}84} = 1{,}685$. Следовательно, искомый доверительный интервал математического ожидания задается формулой:

$$0 - 2{,}093\frac{1{,}685}{\sqrt{20}} < m_x < 0 + 2{,}093\frac{1{,}685}{\sqrt{20}} \Leftrightarrow -0{,}789 < m_x < 0{,}789$$

Ответ: (-0,789;0,789).

Пример 3. По данным десяти независимых измерений найдена оценка среднего квадратического отклонения $S = 0{,}88$. Найти доверительный интервал точности измерительного прибора с надежностью 95%.

Решение. Необходимо найти доверительный интервал для истинного среднего квадратического отклонения σ_x, так как точность прибора характеризуется средним квадратическим отклонением случайных ошибок измерений.

Доверительный интервал для среднего квадратического отклонения находим по формулам (12.3.11) и (12.3.12):

$$S(1 - q_\gamma) < \sigma < S(1 + q_\gamma), \text{ если } q_\gamma < 1,$$

$$0 < \sigma < S(1 + q_\gamma), \text{ если } q_\gamma > 1$$

где S=0,88 – оценка среднего квадратического отклонения, q_γ – число, определяемое из табл. 6 приложения 6 по заданной доверительной вероятности γ=0,95 и заданному объему выборки n=10.

Находим: q_γ=0,6<1.

Следовательно:

$$0{,}88(1 - 0{,}6) < \sigma_x < 0{,}88(1 + 0{,}6) \Leftrightarrow 0{,}352 < \sigma_x < 1{,}408.$$

Ответ: (0,352,1,408).

Пример 4. При изучении производительности труда в млн. руб. на одного работника за определенный период было обследовано n=10 однотипных магазинов. Полученные значения приведены в Таблице 1.

Таблица 1.

n=10	1	2	3	4	5	6	7	8	9	10
X_i млн. руб.	4,2	4,8	4,7	5,0	4,9	4,5	4,5	3,9	4,3	4,8

Требуется выполнить первичную статистическую обработку данных:

1. Определить выборочное среднее $\overline{x}_B$.
2. Найти «исправленное» стандартное отклонение S.
3. Найти коэффициент вариации $k_{var}^{\text{в}}$ изучаемого признака.
4. Полагая, что признак X имеет нормальное распределение, требуется найти доверительный интервал для средней выработки на одного работника с надежностью 0,95.
5. Пояснить смысл полученных результатов.

Обозначим через X изучаемый признак. Обработку результатов наблюдений будем производить с фиксацией всех промежуточных вычислений, занося их в расчетную таблицу 2.

Решение.

Чтобы найти выборочное среднее $\overline{x}_B$, «исправленное» стандартное заполним таблицу 2.

n	x_i	$x_i - \overline{x}_B$	$(x_i - \overline{x}_B)^2$
1	4,2	-0,32	0,1024
2	4,8	0,28	0,0784
3	4,7	0,18	0,0324
4	5,0	0,48	0,2304
5	4,9	0,38	0,1444
6	4,5	-0,02	0,0004
7	3,9	-0,62	0,3844
8	4,1	-0,42	0,1764
9	4,3	-0,22	0,0484
10	4,8	0,22	0,0784
n=10	$\sum_{i=1}^{10} x_i = 45{,}2$		$\sum_{i=1}^{10} (x_i - \overline{x}_\text{в})^2 = 1{,}276$

1) Найдем выборочное среднее по формуле:

$$\overline{x}_\text{в} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{45{,}2}{10} = 4{,}52$$

2) Найдем «исправленное» стандартное отклонение S. Так как мы имеем малую выборку $n = 10 < 30$, то «исправленное» стандартное отклонение S ищем по формуле:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} (x_i - \overline{x}_\text{в})^2} = \sqrt{\frac{1}{9} \cdot 1{,}276} \approx 0{,}38$$

3) Коэффициент вариации признака находим по формуле:

$$k_{var}^{\text{в}} = \frac{S}{\overline{x}_\text{в}} = \frac{0{,}38}{4{,}52} \cdot 100\% \approx 8{,}41\%$$

4) Найдем доверительный интервал для средней выработки на одного работника с надежностью 0,95, полагая, что признак X имеет нормальное распределение.

Так как задана малая выборка, то предельную ошибку выборки ищем по формуле:

$$\bar{x}_{\text{в}} - t_\gamma \cdot \frac{S}{\sqrt{n}} < m_x < \bar{x}_{\text{в}} + t_\gamma \cdot \frac{S}{\sqrt{n}}$$

По условию задачи $k=n-1=10-1=9$, $\gamma=0{,}95$. Значение t_γ - коэффициента доверия при значениях $k=9$, $\gamma=0{,}95$. Находим по таблице 5 приложения 5: $t_\gamma(9,\ 0{,}95)=2{,}262$.

Следовательно:

$$4{,}52 - 2{,}262 \cdot \frac{0{,}38}{\sqrt{10}} < m_x < 4{,}52 + 2{,}262 \cdot \frac{0{,}38}{\sqrt{10}}.$$

Отсюда находим, что доверительный интервал определятся неравенствами:

$$4{,}52 - 0{,}27 < m_x < 4{,}52 + 0{,}27,$$

или:

$$4{,}25 < m_x < 4{,}79.$$

Перейдем к пояснению полученных результатов.

1. Нами найдено, что выборочное среднее равно $\bar{x}_{\text{в}}=4{,}52$. Этот результат показывает, что средняя производительность труда на одного работника в выборочной совокупности, состоящей из 10 магазинов, составляет 4,5 млн. руб.

2. Значение «исправленного» стандартного отклонения $S \approx 0{,}38$ показывает, что среднее теоретическое отклонение производительности труда на одного работника от среднего выборочного отличается на 0,38 млн. руб.

3. Значение коэффициента вариации $k_{var}^{\text{в}} \approx 8{,}41\%$ показывает, что среднее относительное отклонение производительности труда на одного работника от среднего выборочного составляет 8,41%.

4. Нами найден с вероятностью 0,95 доверительный интервал для выборочной средней производительности на одного работника, в предположении, что исследуемый нами признак имеет нормальное распределение:

$$4{,}25 < m_x < 4{,}79.$$

Полученный результат позволяет утверждать, что с вероятностью 0,95 средняя производительность на одного работника данной выборки находится в пределах от 4,25 млн. руб. до 4.79 млн. руб., если исследуемый нами признак имеет нормальное распределение.

Задача 2. При изучении производительности труда в X млн. руб. на одного работника за определенный период было обследовано $n = 74$ однотипных магазинов. Полученное распределение их числа по величине признака X приведены в таблице 3.

Таблица 3

Rj(*X*)*X* млн. руб.	Количество магазинов *ni*
3,35 – 3,65	5
3,65 – 3,95	11
3,95 – 4,25	20
4,25 – 4,55	18
4,55 – 4,85	12
4,85 – 5,15	8
$\sum$	74

Требуется:

1. Найти выборочное среднее $\overline{x}_B$, стандартное отклонение $\sigma_в$ и коэффициент вариации $k_{var}^{в}$ признака X.

2. Полагая, что признак X имеет нормальное распределение, требуется найти:

а) доверительный интервал для ожидаемого среднего значения m_x производительности труда на одного работника приведенной выборки на уровне надежности γ=0,9544;

б) вероятность P того, что величина признака X у произвольно взятого работника магазина окажется в пределах от 3,70 млн. руб. до 4,30 млн. руб.

Решение. Для решения поставленной задачи найдем предварительно значения переменных, входящих в формулы (глава 11): 717

$$\overset{o}{x}_k = \frac{x_{k+1} + x_k}{2};$$

$$\overline{x}_в = \sum_{i=1}^{k} x_i \ p_i^*;$$

$$p_i^* = \frac{n_i}{n};$$

$$\sigma_{\text{в}} = \sqrt{\sum_{i=1}^{k} (x_i - \overline{x}_{\text{в}})^2 \, p_i^*} \,,$$

по которым определяются искомые величины.

Например:

$$\overset{o}{x}_k = \frac{3{,}35 + 3{,}65}{2} = 3{,}5, \qquad \overset{o}{x}_k = \frac{3{,}65 + 3{,}95}{2} = 3{,}8, \ldots,$$

$$p_1^* = \frac{5}{74} = 0{,}07, \quad p_1^* = \frac{11}{74} = 0{,}15, \ldots \text{и. т. д.}$$

Результаты поместим в таблицу 4.

Таблица 4.

R_j млн. руб.	n_i	$_i$	p_i^*	$_i p_i^*$	$_i - \overline{x}_B$	$(\ _i - \overline{x}_B)^2$	$(\ _i - \overline{x}_B) p_i^{*2}$
1	2	3	4	5	6	7	8
3,35 –3,65	5	3,50	0,07	0,245	–0,780	0,6084	0,042588
3,65 –3,95	11	3,80	0,15	0,570	–0,480	0,2304	0,034560
3,95 –4,25	20	4,10	0,27	1,107	–0,180	0,0324	0,008748
4,25 –4,55	18	4,40	0,24	1,056	0,120	0,0144	0,003456
4,55 –4,85	12	4,70	0,16	0,752	0,420	0,1764	0,028224
4,85 –5,15	8	5,00	0,11	0,550	0,720	0,5184	0,057024
Σ	74	–	1,00	4,280	–	–	0,174600

Выборочное среднее $\overline{x}_B$ находим по формуле:

$$\overline{x}_{\text{в}} = \sum_{i=1}^{6} x_i \; p_i^* = 4{,}28 \text{ млн. руб.}$$

Стандартное отклонение $\sigma_{\text{в}}$ находим по формуле:

$$\sigma_{\text{в}} = \sqrt{\sum_{i=1}^{6} (x_i - \overline{x}_{\text{в}})^2 \, p_i^*} = 0{,}418 \text{ млн. руб.}$$

Так как выборка состоит из числа магазинов, значительно превышающее число 30, то коэффициент вариации $k^{\text{в}}_{var}$ признака X вычисляем по формуле:

$$k^{\text{в}}_{var} = \frac{\sigma_{\text{в}}}{\overline{x}_{\text{в}}} \cdot 100\% = \frac{0{,}418}{4{,}280} \cdot 100\% = 9{,}766\%.$$

Полагая, что признак X имеет нормальное распределение, найдем доверительный интервал для ожидаемого среднего значения m_{x} производительности труда на одного работника приведенной выборки на уровне надежности $\gamma = 0{,}9544$.

Известно, что доверительный интервал для ожидаемого среднего значения m_x производительности труда на одного работника определяется неравенствами (12.3.5):

$$\overline{x}_{\text{в}} - z_{\gamma} \cdot \frac{\sigma_{\text{в}}}{\sqrt{n}} < m_x < \overline{x}_{\text{в}} + z_{\gamma} \cdot \frac{\sigma_{\text{в}}}{\sqrt{n}} \ .$$

Так как задана большая выборка, то параметр Z_{γ} является решением уравнения:

$$\Phi_0(z_{\gamma})(z) = \frac{\gamma}{2} \ .$$

По условию задачи γ=0,9544. По таблице 3 приложения 3 функции Лапласа находим, что:

$$\Phi_0(z_{\gamma})(z) = \frac{\gamma}{2} = \frac{0{,}9544}{2} = 0{,}4772 \, .$$

Следовательно, параметр z_{γ}=2.

Подставляя полученные значения в формулу, находим:

$$4{,}28 - 2 \cdot \frac{0{,}418}{\sqrt{74}} < m_x < 4{,}28 + 2 \cdot \frac{0{,}418}{\sqrt{74}}$$

$$4{,}28 - 0{,}097 < m_x < 4{,}28 + 0{,}097$$

Следовательно, доверительный интервал определятся неравенствами:

$$4{,}183 \text{ млн. руб.} < m_x < 4{,}377 \text{ млн. руб.}$$

Вероятность P того, что величина признака X у произвольно взятого работника магазина окажется в пределах от 3,70 млн. руб. до 4,30 млн. руб. найдем по формуле:

$$P(x_1 \leq X \leq x_2) = \Phi_0\left(\frac{x_2 - m_x}{\sigma}\right) - \Phi_0\left(\frac{x_1 - m_x}{\sigma}\right)$$

где $\Phi_0(z)$ — функция Лапласа.

По условию задачи x_2=4,30, x_2=3,70 . Значения параметров m_x и σ заменим соответственно на $\bar{x}_В$ и $\sigma_В$.

Такие оценки для большой выборки являются несмещенными.

Следовательно, искомую вероятность ищем по формуле:

$$P(3{,}70 \leq X \leq 4{,}30) = \Phi\left(\frac{4{,}30 - 4{,}28}{0{,}418}\right) - \Phi\left(\frac{3{,}70 - 4{,}28}{0{,}418}\right) =$$

$$= \Phi(0{,}05) - \Phi(-1{,}39) = \Phi(0{,}05) + \Phi(+1{,}39) = 0{,}016 + 0{,}4177$$

Поясним смысл полученных результатов.

Нами найдено, что выборочное среднее равно $\bar{x}_B = 4{,}28$. Этот результат показывает, что средняя производительность труда на одного работника в выборочной совокупности, состоящей из 74 магазинов, составляет 4,28 млн. руб.

1. Значение стандартного отклонения $\sigma_B = 0{,}418$ млн руб. показывает, что среднее теоретическое отклонение производительности труда на одного работника от среднего выборочного отличается на 0,418 млн руб.

2. Значение коэффициента вариации $k_{var}^{в}$ =9,77% показывает, что среднее относительное отклонение производительности труда на одного работника от среднего выборочного составляет 9,77%.

3. Нами найден с вероятностью 0,9544 доверительный интервал для выборочной средней производительности на одного работника, в предположении, что исследуемый нами признак имеет нормальное распределение:

4,183 млн.руб.$<m_x<$4,377 млн.руб.

Полученный результат позволяет утверждать, что с вероятностью 0,9544 средняя производительность на одного работника данной выборки находится в пределах от 4,183 млн руб. до 4,377 млн.руб., если исследуемый нами признак имеет нормальное распределение.

4. Вероятность P того, что производительность труда на одного работника, произвольно взятого магазина, окажется в пределах от 3,70 млн. руб. до 4,30 млн. руб. равна 0,4337.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Каковы основные подходы к статистическому оцениванию?
2. Что вы понимаете под понятием статистическая оценка?
3. Точечные оценки и их основные виды?
4. Что понимается под интегральной оценкой?
5. Что такое интервальная оценка?
6. Что такое доверительная вероятность g и доверительный интервал при оценке неизвестного математического ожидания нормально распределенной случайной величины?
7. Что такое математическое ожидание дискретной случайной величины, что оно характеризует, и по какой формуле оно определяется?
8. По каким формулам следует находить предельную ошибку выборки?
9. Что такое метод моментов?
10. Что такое метод наибольшего правдоподобия?

ЗАДАНИЯ

12.1. При изучении физико-механических свойств обувных кож испытано 10 образцов и получены значения предела прочности на разрыв *X*(Н, мм), представленные в таблице:

№ п/п	1	2	3	4	5	6	7	8	9	10
X(*Н*, мм)	16,2	20,1	21,4	18,9	16,5	17,3	18,2	19,5	20,4	21,0

Полагая, что признак *X* имеет нормальное распределение, требуется найти доверительный интервал для средней выработки на одного работника с надежностью 0,99.

12.2. При оценке свойств картофеля было обследовано 10 проб, и полученные значения содержания крахмала *X*% представлены в таблице:

№	1	2	3	4	5	6	7	8	9	10
X%	15,2	12,8	13,5	14,9	15,6	16,0	13,7	14,1	13,2	15,0

Полагая, что признак *X* имеет нормальное распределение, требуется найти доверительный интервал для средней выработки на одного работника с надежностью 0, 95.

12.3. С целью определения рациональной структуры размерного ассортимента детской одежды проведено выборочное обследование определенных групп детского населения, и получено распределение количества по величине обхвата груди *X* (см), представлены табл.9

Обхват груди X (см)	54-58	58-62	62-66	66-70	70-74	74-78	$\sum$
Количество детей n_i	21	43	59	62	26	14	225

Полагая, что признак X имеет нормальное распределение, требуется найти:

а) доверительный интервал для ожидаемого среднего значения обхвата груди у детей приведенной выборки на уровне надежности 0,95;

б) вероятность того, что величина признака X у произвольно взятого ребенка окажется в пределах от 66 см. до 70 см.

12.4. Распределение признака X в выборке дается следующим вариационным рядом:

3,0-3,6	3,6-4,2	4,2-4,8	4,8-5,4	5,4-6,0	6,0-6,6	6,6-7,2
2	8	35	43	22	15	5

При уровне значимости 0,01 проверить гипотезу о нормальности распределения X в генеральной совокупности.

12.5. С целью определения рациональной структуры размерного ассортимента детской обуви проведено выборочное обследование определенных групп детского населения, и получено распределение количества детей по величине длины стопы X (см), представлены в таблице.

Длина стопы X (см)	170-175	175-180	180-185	185-190	190-195	$\sum$
Количество детей n_i	5	18	59	61	32	175

Полагая, что признак X имеет нормальное распределение, требуется найти:

а) доверительный интервал для ожидаемого среднего значения α длины стопы X у детей приведенной выборки на уровне надежности 0,95;

б) вероятность того, что величина признака X у произвольно взятого ребенка окажется в пределах от 195 см. до 200 см.

12.6 С целью изучения размеров дневной выручки в сфере мелкого частного бизнеса была произведена 10%-ая случайная бесповторная выборка из 1 000 торговых киосков города. В результате были получены данные о средней дневной выручке, которая составила 500 у.е. В каких пределах с доверительной вероятностью

0,95 может находиться средняя дневная выручка всех торговых точек изучаемой совокупности, если среднее квадратическое отклонение составило 150 у. е.?

12.7. В результате выборочного обследования стажа работы из 1000 сотрудников предприятия получены данные, сведенные в таблицу:

Стаж работы (лет)	2	5	8	9	12	14
Число работников	7	9	2	6	4	7

Определить:

а) средний стаж работы и среднее квадратическое отклонение;

б) доверительный интервал, в котором с надежностью 0,99 заключен средний стаж работы сотрудников всего предприятия при повторном и бесповторном отборе;

в) доверительный интервал, в котором с надежностью 0,95 заключена доля сотрудников предприятия, имеющих стаж работы 9 лет и более.

12.8. Фирма, торгующая автомобилями в небольшом городе, собирает информацию о состоянии местного автомобильного рынка в текущем году. С этой целью из 8 745 лиц в возрасте 18 лет и старше, проживающих в этом городе, отобрано 500 человек. Среди них оказалось 30 человек, планирующих приобрести новый автомобиль в текущем году. Оцените долю лиц в генеральной совокупности в возрасте 18 лет и старше, планирующих приобрести новый автомобиль в текущем году, если α = 0,5.

12.9. Найти доверительный интервал для оценки с надежностью γ = 0,99 неизвестной генеральной средней в случае нормального распределения генеральной совокупности, если генеральное среднее квадратическое отклонение σ = 3, выборочная средняя $\bar{x}_В$ = 5 и объем выборки n = 39.

12.10. Для оценки числа безработных среди рабочих одного из районов города в порядке случайной повторной выборки отобраны 400 человек рабочих специальностей. 25 из них оказались безработными. Используя 95%-й доверительный интервал, оцените истинные размеры безработицы среди рабочих этого района.

12.11. Случайная величина X имеет нормальное распределение с неизвестным математическим ожиданием m_x и известной дисперсией $\sigma^2=100$. По выборке объема n = 150 вычислено выборочное среднее $\bar{x}_В$ = 119. Определить доверительный интервал для неизвестного параметра распределения m_x, отвечающий заданной доверительной вероятности γ = 0,95.

12.12. В результате выборочного обследования стажа работы из 1500 сотрудников предприятия получены данные, сведенные в таблицу:

Стаж работы (лет)	5	6	7	9	12	15
Число работников	9	7	5	10	8	12

Определить:
а) средний стаж работы и среднее квадратическое отклонение;
б) доверительный интервал, в котором с надежностью 0,99 заключен средний стаж работы сотрудников всего предприятия при повторном и бесповторном отборе;
в) доверительный интервал, в котором с надежностью 0,95 заключена доля сотрудников предприятия, имеющих стаж работы 7 лет и более.

12.13. Выборочные обследования показали, что доля покупателей, предпочитающих новую модификацию товара А, составляет 60% от общего числа покупателей данного товара. Каким должен быть объем выборки, чтобы можно было получить оценку генеральной доли с точностью не менее 0,05 при доверительной вероятности 0,90?

12.14. Случайная величина X имеет нормальное распределение с неизвестными математическим ожиданием и дисперсией. По данной выборке найти доверительный интервал для математического ожидания, отвечающий доверительной вероятности $\gamma = 0{,}98$.

x_i	19	16	22	28	34	40	46
n_i	2	14	16	50	10	3	5

12.15. Случайная величина X имеет нормальное распределение с неизвестным математическим ожиданием m_x и известной дисперсией $\sigma^2 = 144$. По выборке объема $n = 150$ вычислено выборочное среднее $\overline{x}_В = 120$. Определить доверительный интервал для неизвестного параметра распределения m_x, отвечающий заданной доверительной вероятности $\gamma = 0{,}9$.

12.16. Случайная величина X имеет нормальное распределение с неизвестными математическим ожиданием и дисперсией. По данной выборке найти доверительный интервал для математического ожидания, отвечающий доверительной вероятности $\gamma = 0{,}9$.

x_i	18	21	24	27	30	33	36
n_i	4	16	10	30	15	20	5

12.17. Случайная величина X имеет нормальное распределение с неизвестными математическим ожиданием и дисперсией. По данной выборке найти доверительный интервал для математического ожидания, отвечающий доверительной вероятности $\gamma = 0{,}8$.

x_i	65	70	75	80	85	90	95
n_i	3	7	10	40	20	12	8

ГЛАВА 13
ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

13.1. Статистическая гипотеза. Статистический критерий проверки нулевой гипотезы

На разных стадиях статистического исследования и моделирования возникает необходимость в формулировке и экспериментальной проверке некоторых предположительных утверждений (гипотез) относительно природы или величины неизвестных параметров анализируемой генеральной совокупности.

Статистической гипотезой называется предположение о виде распределения, о параметрах известных распределений.

Выдвинутая гипотеза называется **нулевой** (или основной) и обозначается H_0.

Гипотеза, которая противоречит нулевой, называется **конкурирующей гипотезой** (или альтернативной гипотезой) и обозначается H_1:

Если гипотеза содержит только одно предположение, то она называется **простой.**

Если гипотеза состоит из конечного или бесконечного числа предположений, то она называется **сложной.**

Основными типами гипотез, проверяемых в ходе статистического анализа и моделирования, являются:

1. Гипотезы о типе закона распределения исследуемой генеральной совокупности;

2. Гипотезы о числовых значениях параметров исследуемой генеральной совокупности;

3. Гипотезы об однородности двух или нескольких обрабатываемых выборок или некоторых характеристиках анализируемых совокупностей;

4. Гипотезы об общем виде модели, описывающей статистическую зависимость между анализируемыми параметрами выборочных данных.

Проверка статистических гипотез осуществляется с помощью того или иного статистического критерия – специально выбранной статистики (выборочной функции) $Z = Z(X_1, X_2, ..., X_n)$ – функции случайных (по множеству серий) результатов наблюдений $X_1, X_2, ..., X_n$, которые обычно предполагаются независимыми, одинаково распределенными случайными величинами. В качестве выборочной функции (критерия) Z обычно используются:

Z – критерий, распределенный по нормальному закону;

t – критерий, распределенный по закону Стьюдента;

χ^2 – критерий «хи-квадрат», имеющий «хи-квадрат» распределение, критерий согласия Пирсона;

F – критерий, распределенный по закону Фишера–Снедекора;

Λ – критерий, распределенный по закону Колмогорова.

13.2. Проверка гипотез. Ошибки первого и второго рода

Сопоставление выдвинутой гипотезы с выборочными данными называется проверкой гипотезы.

Правосторонней критической областью для проверки нулевой гипотезы H_0 с уровнем значимости α называется совокупность значений критерия проверки Z для которых выполняется равенство:

$$P(Z > Z_{\text{крит}}) = \alpha. \tag{13.2.1}$$

При этом число $Z_{\text{крит}}$ называется границей критической области.

Ясно, что критические области бывают правосторонняя, левосторонняя, двусторонняя, следовательно:

– правосторонняя критическая область определяется неравенством:

$$Z > Z_{\text{крит}}, \tag{13.2.2}$$

– левосторонняя критическая область определяется неравенством:

$$Z < -Z_{\text{крит}}, \tag{13.2.3}$$

– двусторонняя критическая область определяется неравенствами:

$$Z < Z_{1\text{крит}} \text{ и } Z > Z_{2\text{крит}}. \tag{13.2.4}$$

Схема проверки гипотезы:

1. Формирование нулевой гипотезы H_0, которая выдвигается на основе начального анализа выборочных данных, и конкурирующей гипотезы H_1.

2. Выбор некоторой вероятности α, называемой уровнем значимости нулевой гипотезы H_0.

3. Подбор по выборочным данным случайной величины Z, распределение которой называется критерием для проверки нулевой гипотезы.

4. Определение границы критической $Z_{крит}$ области для проверки нулевой гипотезы с уровнем значимости α.

5. Вычисление по данным выборки числа $Z_{набл}$.

Если оно попадает в критическую область нулевой гипотезы H_0, то есть $Z_{набл} > Z_{крит}$, то гипотеза H_0 отвергается и принимается гипотеза H_1, если наоборот $Z_{набл} < Z_{крит}$, то нет основания отвергать нулевую гипотезу H_0.

Если гипотеза принята, то это вовсе не означает, что выдвинутое нами предположение является наилучшей, единственно подходящей: просто эта гипотеза не противоречит имеющимся у нас в распоряжении выборочным данным, однако таким же свойством наряду с H_0 могут обладать и другие подобные гипотезы. Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость ее проверки. Поскольку проверку проводят статистическими методами, ее называют статистической.

При статистическом анализе результатов наблюдений в принципе возможен один из четырех исходов:

1. Верна гипотеза H_0, решение: верна H_0 (правильное решение);
2. Верна гипотеза H_0 , решение: верна H_1 (*ошибка первого рода*);
3. Верна гипотеза H_1, решение: верна H_0 (*ошибка второго рода*);
4. Верна гипотеза H_1, решение: верна H_1 (правильное решение).

В итоге статистической проверки гипотезы в двух случаях может быть принято неправильное решение, то есть могут быть допущены ошибки двух родов.

1. Ошибка I рода, заключающаяся в неоправданном отклонении правильной (истинной) нулевой гипотезы.

Вероятность α ошибки первого рода называется уровнем значимости статистического критерия.

2. Ошибка II рода, заключающаяся в неправильном принятии неверной нулевой гипотезы, что эквивалентно непринятию верной альтернативной гипотезы H_1.

Вероятность ошибки второго рода обозначается β. Вероятность $M = 1 - \beta$ несовершения ошибки второго рода называется мощностью статистического критерия.

Иначе говоря, мощность критерия есть вероятность отклонения нулевой гипотезы в случае, когда верна альтернативная гипотеза.

13.3. Критерий согласия Пирсона

Наиболее распространённым критерием проверки гипотезы о предполагаемом законе неизвестного распределения является критерий согласия Пирсона. Приведем описание алгоритма работы с критерием согласия Пирсона. Необходимым условием применения критерия Пирсона является наличие в каждом интервале не менее 5 наблюдений. Интервалы с частотами меньше 5 в интервальном вариационном ряде следует объединять, частоты при этом складываются. При вычислении числа степеней свободы считать количество интервалов уже после объединения.

Критерий Пирсона является наиболее состоятельным при большом числе наблюдений ($n \geq 50$) . Он почти всегда опровергает неверную гипотезу.

1. Выбрать закон распределения случайной величины.

2. По соответствующей формуле вычислить точечные или интервальные вероятности p_i.

3. Вычислить выравнивающие частоты $n_i^* = np_i$, где n – объем выборки.

4. Найти статистику:

$$\chi^2_{\text{набл}} = \sum_{i=1}^{m} \frac{(n_i - n_i^*)^2}{n_i^*}, \qquad (13.3.1)$$

Эта случайная величина при $n > \infty$ стремится к закону распределения χ^2 с k степенями свободы.

Число степеней свободы определяется по формуле $k = m - 1 - r$, где m – число интервалов, r – число параметров предполагаемого распределения.

Выражения для нахождения числа степеней свободы известных законов распределения представлены в табл. 13.3.1.

Таблица 13.3.1

Закон распределения	Число степеней свободы
Биномиальный закон	$k = m - 1$, если $P(A)$ известно
	$k = m - 2$, если $P(A)$ неизвестно
Закон распределения Пуассона	$k = m - 2$
Равномерный закон	$k = m - 3$
Показательный закон	$k = m - 2$
Нормальный закон	$k = m - 3$

5. Задавая уровень значимости α и учитывая k из таблицы 7 распределения χ^2 (приложение 7), находим критическое значение $\chi^2_{\text{крит}}$, при котором выполняется условие $P(\chi^2_{\text{набл}} > \chi^2_{\text{крит}}(\alpha, k)) = \alpha$.

6. Если $\chi^2_{набл} < \chi^2_{крит}$, то закон теоретического распределения не противоречит опытным данным, нет оснований отвергнуть гипотезу о выбранном законе распределения. В противном случае выдвинутая гипотеза отвергается.

Пример 1. На экзамене по теории вероятностей и математической статистике студенту задается только один вопрос по одной из четырех частей курса. Из 100 студентов 26 получили вопрос по первой части, 32 по второй, 17 по третьей и 25 по четвертой. Можно ли по этим результатам принять гипотезу, что с вероятностью 0,95 получить вопрос по любой из четырех частей одна и та же?

Решение: По условию задачи имеем:

$$m = 4, \quad \alpha = 0{,}05, \quad p_i = 1/4 = 0{,}25, \quad n = 100, \; n_1 = 26, \; n_2 = 32,$$

$$n_3 = 17, \quad n_4 = 25, \quad np_i = 25 \; (i = 1{,}2{,}3{,}4)$$

По формуле (13.3.1):

$$\chi^2_{набл} = \sum_{i=1}^{m} \frac{(n_i - n_i^*)^2}{n_i^*}$$

находим $\chi^2_{набл}$:

$$\chi^2_{набл} = \frac{(26-25)^2}{25} + \frac{(32-25)^2}{25} + \frac{(17-25)^2}{25} + \frac{(25-25)^2}{25} = 4{,}56.$$

Находим число степеней свободы по формуле:

$$k = m - 1 - r = 4 - 1 - 0 = 3.$$

В данной задаче $r = 0$, так как ни один из параметров предполагаемого распределения нами не находился по выборке.

Находим по табл. 7 приложения 7 границу критической области, являющейся корнем уравнения $P(\chi^2_{набл} > \chi^2_{крит}(\alpha, k)) = \alpha$ при значениях $\alpha = 0{,}05$, $k = 3$ – это значение равно $\chi^2_{крит} = 7{,}815$.

Сравниваем полученные значения – 4,56 и 7,82. Так как 4,56 < 7,82, то по критерию Пирсона наша гипотеза подтвердилась.

Пример 2. Путем опроса 50 матерей**,** имеющих четверых детей установлено количество мальчиков в семье: 0, 2, 1, 2, 1, 2, 2, 2, 3, 3, 3,2, 3, 0, 2, 3, 3, 3, 2, 4, 1, 3, 3, 1, 3, 2, 1, 0, 3, 2, 4, 3, 3, 2, 3,2,1, 2, 1, 1, 2, 2, 3, 2, 2, 1, 2, 0. 3, 1.

Требуется проверить на уровне значимости $\alpha = 0{,}01$ гипотезу о биномиальном распределении.

Решение: Всего у 50 матерей 200 детей. Случайной величиной X является число мальчиков в семьях из 4 детей. Построим вариационный ряд:

x_i	0	1	2	3	4	$\sum$
n_i	4	10	18	16	2	50

Рассчитаем среднюю частоту рождения мальчиков:

$$p_i^* = \frac{0 \cdot 4 + 1 \cdot 10 + 2 \cdot 18 + 3 \cdot 16 + 4 \cdot 2}{200} = \frac{102}{200} = 0{,}51\,.$$

По формуле биномиального закона $P_n(m) = C_n^m p^m q^{n-m}$ вычислим вероятности комбинаций рождения мальчика и девочки в семьях из четырех детей:

$$P_4(0) = (0{,}49)^4 = 0{,}0576,$$

$$P_4(1) = C_4^1 \cdot (0.51) \cdot (0{,}49)^3 = 4 \cdot (0.51) \cdot (0{,}49)^3 = 0{,}2401,$$

$$P_4(2) = 6 \cdot (0.51)^2 \cdot (0{,}49)^3 = 0{,}3747,$$

$$P_4(3) = 4 \cdot (0.51)^3 \cdot (0{,}49) = 0{,}2600,$$

$$P_4(4) = (0.51)^4 = 0{,}0676.$$

$$\sum_{i=1}^{5} P_n(m) = 0{,}0576 + 0{,}2401 + 0{,}3747 + 0{,}2600 + 0{,}0676 = 1\,.$$

Теоретические частоты равны $n_i^* = 50 \cdot P_n(m = i)$. Составим таблицу и рассчитаем $\chi^2_{\text{набл}}$.

x_i	n_i	n_i^*	$(n_i - n_i^*)^2$	$\frac{(n_i - n_i^*)^2}{n_i^*}$
0	4	3	1	$\frac{1}{15} = 0{,}667$
1	10	12		
2	18	19	1	$\frac{1}{19} = 0{,}053$
3	16	13	4	$\frac{4}{16} = 0{,}25$
4	12	3		
$\sum$	50	50		$\chi^2_{\text{набл}} = 0{,}37$

По табл.7 приложения 7 на уровне значимости $\alpha = 0{,}01$ и при числе степеней свободы $k = 3 - 1 - 1 = 1$ определяем $\chi^2_{крит} = 6{,}64$.

Так как $0{,}37 < 6{,}64$, то по критерию Пирсона нулевая гипотеза не отвергается.

Таким образом, число мальчиков в семье из 4 детей данной совокупности подчиняется биномиальному закону распределения.

Пример 3. По данным наблюдения в автомастерской, число несвоевременно обслуженных автомобилей за смену в 100 выборках, приводится в таблице:

Число автомобилей x_i	0	1	2	3	$\sum$
Число выборок n_i	85	11	3	1	100

На уровне значимости $\alpha = 0{,}05$ необходимо проверить гипотезу о законе Пуассона.

Решение: Определяем среднюю интенсивность числа автомобилей, несвоевременно обслуженных за одну выборку:

$$\bar{\lambda} = \frac{0 \cdot 85 + 1 \cdot 11 + 2 \cdot 3 + 3 \cdot 1}{100} = \frac{20}{100} = 0{,}2 .$$

По таблице приложения $e^{-\lambda}$ определяем $e^{-0,2} = 0{,}8117$.

По формуле закона Пуассона:

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda} .$$

вычисляем вероятности:

$$P_{100}(0) = \frac{0{,}2^0}{0!} e^{-0{,}2} = 0{,}8117,$$

$$P_{100}(1) = \frac{0{,}2^1}{1!} e^{-0{,}2} = 0{,}2 \cdot 0{,}8117 = 0{,}1637,$$

$$P_{100}(2) = \frac{0{,}2^2}{2!} e^{-0{,}2} = 0{,}02 \cdot 0{,}8117 = 0{,}0165,$$

$$P_{100}(3) = \frac{0{,}2^3}{3!} e^{-0{,}2} = \frac{0{,}008}{6} \cdot 0{,}8117 = 0{,}0011.$$

$$P_n(m) = 0{,}8117 + 0{,}1637 + 0{,}0165 + 0{,}0011 = 1.$$

Теоретические частоты равны $n_i^* = 100 \cdot P_n(m = i)$. Составим таблицу и рассчитаем $\chi^2_{\text{крит}}$.

<table>
<tr><th>x_i</th><th>n_i</th><th>n_i^*</th><th>$(n_i - n_i^*)^2$</th><th>$\dfrac{(n_i - n_i^*)^2}{n_i^*}$</th></tr>
<tr><td>0</td><td>85</td><td>82</td><td>9</td><td>$\frac{9}{82} = 0{,}108$</td></tr>
<tr><td>1</td><td>11</td><td>16</td><td>25</td><td>$\frac{25}{16} = 1{,}563$</td></tr>
<tr><td>2</td><td rowspan="2">3
1</td><td rowspan="2">2
0</td><td rowspan="2">4</td><td rowspan="2">$\frac{4}{2} = 2$</td></tr>
<tr><td>3</td></tr>
<tr><td>$\sum$</td><td>100</td><td>100</td><td></td><td>$\chi^2_{\text{набл}} = 3{,}671$</td></tr>
</table>

По табл. 7 приложения 7 на уровне значимости $\alpha = 0{,}05$ и при числе степеней свободы $k = 3 - 2 = 1$ определяем $\chi^2_{\text{крит}} = 3{,}84$.

Так как $0{,}3671 < 3{,}84$, то по критерию Пирсона нулевая гипотеза не отвергается.

Таким образом, число несвоевременно обслуженных автомобилей за смену из данной совокупности подчиняется закону распределения Пуассона.

Пример 4. В своей книге «Математические методы статистики» Г. Крамер приводит следующие данные распределения толщины (в мм) 12000 бобов:

Таблица 4.

Толщина	До 7.00	7.00–7.25	7.25–7.5	7.5–7.75	7.75–8.00	8.00–8.25
Количество бобов	32	103	239	624	1187	1650
Толщина	8.25–8.5	8.5–8.75	8.75–9.00	9.00–9.25	9.25–9.5	
Количество бобов	1883	1930	1638	1130	737	
Толщина	9.5–9.75	9.75–10.00	10.00–10.25	10.25–10.5	Свыше 10.5	
Количество бобов	427	221	110	57	32	

Проверить гипотезу о том, что толщина бобов подчиняется нормальному распределению на уровне значимости $\alpha = 0{,}02$.

Решение: Как видно левый и правый интервалы открытые, таким образом, различают интервалы открытые и закрытые.

Открытые – интервалы, у которых обозначена одна граница, верхняя или нижняя.

Закрытые – интервалы, у которых обозначены обе границы.

В нашем примере оба крайних интервала открытые, поэтому приведем правило закрытия открытых интервалов.

Правило закрытия открытых интервалов

Величина открытого интервала условно принимается равной величине последующего или предыдущего интервала. Если нижняя граница последующего интервала совпадает с верхней границей предыдущего, то нижняя граница считается **включительной**, а верхняя – **исключительной**.

Так как никаких сведений о параметрах нормального распределения по условию задачи у нас не имеется, то за математическое ожидание признака возьмем среднее выборочное $\overline{x}_B$ выборки X, а за генеральную дисперсию возьмем ее несмещенную оценку S^2.

Составим таблицу с промежуточными вычислениями:

$\overset{o}{x}_i$	p_i^*	$\overset{o}{x}_i\, p_i^*$	$(\overset{o}{x}_i - \overline{x}_B)^2$	$(\overset{o}{x}_i - \overline{x}_B)^2 p_i^*$
6,875	0,002667	0,018333	2,844846	0,007586
7,125	0,008583	0,061156	2,064012	0,017716
7,375	0,019917	0,146885	1,408179	0,028046
7,625	0,052	0,3965	0,877345	0,045622
7,875	0,098917	0,778969	0,471512	0,04664
8,125	0,1375	1,117188	0,190678	0,026218
8,375	0,156917	1,314177	0,034845	0,005468
8,625	0,160833	1,387188	0,004011	0,000645
8,875	0,1365	1,211438	0,098178	0,013401
9,125	0,094167	0,859271	0,317344	0,029883
9,375	0,061417	0,575781	0,661511	0,040628
9,625	0,035583	0,34249	1,130677	0,040233
9,875	0,018417	0,181865	1,724844	0,031766
10,125	0,009167	0,092813	2,44401	0,022403
10,375	0,00475	0,049281	3,288177	0,015619
10,625	0,002667	0,028333	4,257343	0,011353
$\sum$	1	8,561667	21,81751	0,383228

$$\overline{x}_{\text{в}} = 8{,}561667 \approx 8{,}56$$

$$S^2 = \frac{16-1}{16} \cdot 0{,}383228 = 0{,}359277$$

$$\sigma_{\text{в}} = \sqrt{0{,}359277} = 0{,}599397 \approx 0{,}60\,.$$

Находим величину np_i ($n = 1200$) для каждого из заданных интервалов по формуле:

$$p_i = P_i(x_i \le X \le x_{i+1}) = \Phi_0(z_{i+1}) - \Phi_0(z_i).$$

где

$$z_{i+1} = \frac{x_{i+1} - \overline{x}_{\text{в}}}{\sigma}, \qquad z_i = \frac{x_i - \overline{x}_{\text{в}}}{\sigma}\,.$$

Значения $\Phi_0(Z_{i+1})$ и $\Phi_0(Z_i)$ находим из табл. 3 (приложение 3)

Пример вычисления:

$$p_1 = P_1(6{,}75 \le X \le 7) = \Phi_0\left(\frac{7-8{,}56}{0{,}6}\right) - \Phi_0\left(\frac{6{,}75-8{,}56}{0{,}6}\right) =$$

$$= \Phi_0(-2{,}6) - \Phi_0(-3) = \Phi_0(3) - \Phi_0(2{,}6) = 0{,}499 - 0{,}4954 = 0{,}004\,.$$

Остальные значения вероятностей случайной величины найдем подобным образом и поместим в таблицу:

x_{i+1}	x_i	Z_{i+1}	$\Phi(Z_{i+1})$	Z_i	$\Phi(Z_i)$	$p_i = \Phi(Z_{i+1}) - \Phi(Z_i)$
7	6,75	–2,6	–0,4954	–3	–0,499	0,004
7,25	7	–2,2	–0,4857	–2,6	–0,495	0,0097
7,5	7,25	–1,8	–0,4617	–2,2	–0,486	0,0239
7,75	7,5	–1,4	–0,4122	–1,8	–0,462	0,0496
8	7,75	–0,9	–0,3256	–1,4	–0,412	0,0865
8,25	8	–0,5	–0,1985	–0,9	–0,326	0,1272
8,5	8,25	–0,1	–0,041	–0,5	–0,198	0,1575
8,75	8,5	0,31	0,1233	–0,1	–0,041	0,1643
9	8,75	0,73	0,2677	0,31	0,1233	0,1444
9,25	9	1,15	0,3746	0,73	0,2677	0,1069
9,5	9,25	1,57	0,4413	1,15	0,3746	0,0667
9,75	9,5	1,98	0,4763	1,57	0,4413	0,035
10	9,75	2,4	0,4918	1,98	0,4763	0,0155
10,25	10	2,82	0,4976	2,4	0,4918	0,0058

Продолжение таблицы

x_{i+1}	x_i	Z_{i+1}	$Ф(Z_{i+1})$	Z_i	$Ф(Z_i)$	$p_i = Ф(Z_{i+1}) – Ф(Z_i)$
10,5	10,25	3,23	0,4994	2,82	0,4976	0,0018
10,75	10,5	3,65	0,4999	3,23	0,4994	0,0012
$\sum p_i =$						1

n_i	n_i^*	$(n_i – n_i^*)^2$	$\frac{(n_i – n_i^*)^2}{n_i^*}$
32	48	256	5,333
103	117	196	1,675
239	287	2304	8,028
624	595	840,28	1,412
1187	1038	22131	21,32
1650	1526	15343	10,05
1883	1890	46,573	0,025
1930	1971	1719,2	0,872
1638	1733	8946	5,163
1130	1283	23327	18,19
737	800	3973	4,966
427	420	49	0,117
221	186	1222,3	6,57
110	70	1600	22,86
57	22	1225	55,68
32	14	324	23,14
$\sum = 1200$	$\sum = 1200$	770	$\chi^2_{набл} = 185{,}4$

Находим число степеней свободы по формуле k = 13 – 2 – 1 = 13.

Так как два параметра генерального распределения были найдены по выборке, то $r = 2$.

По табл. 7 приложения 7 на уровне значимости $\alpha = 0{,}02$ и при числе степеней свободы $k = 13$ определяем $\chi^2_{крит} = 25{,}50$.

Так как 185,4 > 25,50, то по критерию Пирсона нулевая гипотеза отвергается.

Таким образом, гипотеза о том, что толщина бобов подчиняется нормальному распределению на уровне значимости $\alpha = 0{,}02$ не подтверждается.

Пример 5.

По данным о производительности труда рассчитаны теоретические частоты в предположении нормального закона распределения. Результаты расчетов представлены следующей таблицей:

R_i	3,6 – 3,7	3,7 – 3,8	3,8 – 3,9	3,9 – 4,0	4,0 – 4,1	4,1 – 4,2	4,2 – 4,3
n_i	1	6	11	15	9	6	2
n_i^*	2	5	11	14	11	5	2

На уровне значимости $\alpha = 0{,}01$ проверить гипотезу о нормальном законе распределения.

Решение: Так как теоретические частоты даны вычисляем необходимые параметры для определения $\chi^2_{\text{набл}}$.

<table>
<tr><th>R_i</th><th>n_i</th><th>n_i^*</th><th>$(n_i - n_i^*)^2$</th><th>$\frac{(n_i - n_i^*)^2}{n_i^*}$</th></tr>
<tr><td>3,6 – 3,7</td><td rowspan="2">1
6</td><td rowspan="2">2
5</td><td rowspan="2">0</td><td rowspan="2">0</td></tr>
<tr><td>3,7 – 3,8</td></tr>
<tr><td>3,8 – 3,9</td><td>1</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>3,9 – 4,0</td><td>15</td><td>14</td><td>1</td><td>1/14 = 0,071</td></tr>
<tr><td>4,0 – 4,1</td><td>9</td><td>11</td><td>4</td><td>4/11 = 0,364</td></tr>
<tr><td>4,1 – 4,2</td><td rowspan="2">6
2</td><td rowspan="2">5
2</td><td rowspan="2">11</td><td rowspan="2">1/7 = 0,143</td></tr>
<tr><td>4,2 – 4,3</td></tr>
<tr><td>$\sum$</td><td>50</td><td>50</td><td></td><td>$\chi^2_{\text{набл}} = 0{,}578$</td></tr>
</table>

Находим число степеней свободы по формуле $k = 5 - 3 = 2$.

По табл. 7 приложения 7 на уровне значимости $\alpha = 0{,}01$ и при числе степеней свободы $k = 2$ определяем $\chi^2_{\text{крит}} = 9{,}21$.

Так как $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$ гипотеза H_0 не отвергается, то есть производительность труда для данной совокупности подчиняется нормальному закону распределения.

13.4. Критерий Колмогорова

Критерий согласия Колмогорова является наиболее простым критерием проверки гипотезы о виде закона распределения. Он связывает эмпирическую функцию распределения $\Phi^*(x)$ с теоретической функциией распределения $\Phi(x)$ некоторой случайной величиной (статистика Колмогорова):

$$D_n = \max_i |\Phi^*(x_i) - \Phi(x_i)|, \qquad (13.4.1)$$

где максимальная разность значений эмпирической и теоретической функций распределения, n – объем выборки, x_i – середины интервалов (интервальных рядов).

Согласно теории Колмогорова:

Если $x_1, x_2, \ldots, x_n$ – конкретная выборка из распределения с известной непрерывной функцией распределения $\Phi(x)$ и $\Phi^*(x)$ – эмпирическая функция распределения. Выдвигается простая гипотеза H_0: $\Phi^*(x) = \Phi(x)$; альтернативная: H_1: $\Phi^*(x) \neq \Phi(x)$, $x \in R$

Колмогоров доказал, что при $n \to \infty$ закон распределения случайной величины $\sqrt{n} \cdot Dn$ независимо от вида распределения случайной величины X стремится к закону распределения Колмогорова:

$$P\{\sqrt{n} \cdot D_n < \lambda\} \to F(\lambda),$$

α	0,2		0,1	0,05	0,02	0,01	0,001
λ_0	1,073	1,224		1.358	1,520	1,627	1,950

Найдем $D_{крит}$ такое, что $P(D_n > D_{крит}) = \alpha$.

Рассмотрим уравнение $F(\lambda) = 1 - \alpha$. С помощью функции Колмогорова найдем значение (корень λ_0) этого уравнения. Тогда по теореме Колмогорова:

$$P\{\sqrt{n} \cdot D_n < \lambda_0\} = 1 - \alpha, \ P\{\sqrt{n} \cdot D_n > \lambda_0\} = \alpha,$$

$$D_{крит} = \frac{\lambda_0}{\sqrt{n}}.$$

Если $Dn < D_{крит}$, то гипотезу H_0 нет оснований опровергнуть; в противном случае – ее опровергают.

Пример 6. Опыт Бюффона (п. 1.3). Монету бросали 4040 раз (Бюффон). Получили $n_1 = 2048$ выпадений герба и $n_2 = 1992$ выпадений решётки. Проверить, используя:

а) критерий Колмогорова;

б) критерий Пирсона, согласуются ли эти данные с гипотезой H_0 о симметричности монеты ($\alpha = 0{,}05$).

Решение: Случайная величина X принимает два значения: $x_1 = -1$ (решётка); $x_2 = 1$ (герб). Гипотеза H_0:$\Phi^*(x) = \Phi(x) = 0{,}5$.

а) По таблице распределения Колмогорова находим корень уравнения $F(\lambda) = 1 - \alpha$, при $\alpha = 0{,}05$. Следует $\lambda x_0 = 1{,}358$. Тогда:

$$D_{\text{крит}} = \frac{\lambda_0}{\sqrt{n}} = \frac{1{,}358}{\sqrt{4040}} \approx 0{,}021\,.$$

Для нахождения по выборке Dn строим функции $\Phi(x)$ и $\Phi^*(x)$ и вычисляем величину:

$$D_n = \max_i |\Phi^*(x_i) - \Phi(x_i)|.$$

x_i	Решетка $x_1 = -1$	Герб $x_2 = 1$
p_i	0,5	0,5

$$\Phi(x) = \begin{cases} 0, & \text{при } x \le -1, \\ 0{,}5, & \text{при } x \in (-1, +1), \\ 1, & \text{при } x > 1. \end{cases}$$

x_i	Решетка $x_1 = -1$	Герб $x_2 = 1$
n_i	1992	2048
p_i	≈0,493	≈0,507

$$\Phi^*(x) = \begin{cases} 0, & \text{при } x \le -1 \\ 0{,}493, & \text{при } -1 < x \le 1 \\ 1. & \text{при } x > 1 \end{cases}$$

Максимальное отклонение $\Phi(x)$ от $\Phi^*(x)$ равно 0,007, то есть $D_n = 0{,}007$. Поскольку $D_n < D_{\text{крит}}$, то нет оснований отвергать гипотезу H_0; опытные данные согласуются с гипотезой H_0 о симметричности монеты.

б) Вычисляем статистику по проверочной формуле χ^2

$$\chi^2_{\text{набл}} = \sum_{i=1}^{6} \frac{n_i^2}{np_i} - n = \frac{1992^2}{\frac{1}{2}\cdot 4040} + \frac{2048^2}{\frac{1}{2}\cdot 4040} - 4040 = 0{,}776\,.$$

По табл. 7 приложения 7 χ^2-распределения находим критическую точку $\chi^2_{\text{крит}} = \chi^2_{0,05;1} = 3{,}8$; Так как $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то опытные данные согласуются с гипотезой о симметричности монеты.

13.5. Сравнение двух дисперсий нормальных генеральных совокупностей

На практике задача сравнения дисперсий возникает, если требуется сравнить точность приборов, инструментов, самих методов измерений и другое. Очевидно, предпочтительнее тот прибор, инструмент и метод, который обеспечивает наименьшее рассеяние результатов измерений, то есть наименьшую дисперсию.

Пусть генеральные совокупности X и Y распределены нормально. По независимым выборкам с объемами, соответственно равными n_1 и n_2, извлеченным из этих совокупностей, найдены исправленные выборочные дисперсии S_X^2 и S_Y^2. Требуется по исправленным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные дисперсии рассматриваемых совокупностей равны между собой:

$$H_0 : D_x = D_y, \qquad (13.5.1)$$

Исправленные дисперсии, полученные путем статистической обработки выборочных данных – оценки S_X^2 и S_Y^2 – являются несмещенными оценками генеральных дисперсий D_x и D_y (глава 12):

$$M(S_X^2) = D_x,\ M(S_Y^2) = D_y.$$

Таким образом, проверка гипотезы H_0 заключается в ответе на вопрос: значимо (существенно) или незначимо (несущественно) различаются исправленные дисперсии. При этом в случае, когда нулевая гипотеза H_0 справедлива (не противоречит опытным данным) – генеральные дисперсии одинаковы, то различие оценок S_X^2 и S_Y^2 незначимо и может быть объяснено случайными причинами, например нарушением репрезентативности выборки.

Если нулевая гипотеза H_0 отвергнута – генеральные дисперсии различаются, то различие оценок S_X^2 и S_Y^2 значимо и не может быть объяснено случайными причинами. Например, если различие оценок дисперсий результатов измерений, проведенных двумя различными приборами, оказалось значимым, то точности измерения этих приборов – различны.

В качестве критерия проверки нулевой гипотезы H_0 о равенстве генеральных дисперсий примем отношение большей исправленной дисперсии, например S_X^2 к меньшей, например S_Y^2 ($S_X^2 > S_Y^2$):

$$F_{\text{набл}} = \frac{S_X^2}{S_Y^2}\,. \qquad (13.5.2)$$

Величина $F_{\text{набл}}$, при условии справедливости нулевой гипотезы H_0, имеет распределение Фишера–Снедекора с двумя степенями свободы $k_1 = n_1 - 1$, $k_2 = n_2 - 1$. Распределение Фишера–Снедекора зависит только от чисел k_1 и k_2 степеней свободы и не зависит от других параметров (см. п.10.1).

Критическая область для проверки гипотезы строится в зависимости от вида конкурирующей гипотезы H_1, при этом различают два случая:

Первый случай. Нулевая гипотеза H_0:$D_x = D_y$, конкурирующая гипотеза H_1: $D_x > D_y$.

В этом случае строят правостороннюю критическую область, исходя из требования: вероятность попадания критерия $F_{\text{набл}}$ в критическую область при условии справедливости нулевой гипотезы H_0 должна быть равна принятому уровню значимости α: 791

$$P\left(F_{\text{набл}} \geq F_{\text{крит}}(\alpha, k_1, k_2)\right) = \alpha \qquad (13.5.3)$$

Критическая точка $F_{\text{крит}}(\alpha,k_1,k_2)$ определяется по табл.8 процентных точек распределения Фишера—Снедекора (см. приложение 8).

Правило проверки нулевой гипотезы H_0 формулируется следующим образом:

$$\begin{cases} \text{при } F_{\text{набл}} \geq F_{\text{крит}}(\alpha, k_1, k_2) \to H_1, \\ \text{при } F_{\text{набл}} < F_{\text{крит}}(\alpha, k_1, k_2) \to H_0. \end{cases} \qquad (13.5.4)$$

Приведем алгоритм работы для первого случая:

Пусть по двум независимым выборкам, объемы которых равны n_1 и n_2, соответственно, полученным из нормально распределенных генеральных совокупностей X и Y найдены исправленные выборочные дисперсии S_X^2 и S_Y^2. Требуется сравнить дисперсии генеральных совокупностей.

Алгоритм сравнения D_x и D_y:

1) Выдвигаем нулевую гипотезу: H_0:$D_x = D_y$. Тогда конкурирующей гипотезой будет H_1: $D_x > D_y$;

2) Задаем число α – уровень значимости нулевой гипотезы;

3) Находим из табл. 8 распределения Фишера–Снедекора (см. приложение 8) значение $F_{\text{крит}}$. по заданному α и числам степеней свободы $k_1 = n_1 - 1$, $k_2 = n_2 - 1$;

4) Находим критерий проверки нулевой гипотезы H_0:

$$F_{\text{набл}} = \frac{S_X^2}{S_Y^2}\,.$$

равное отношению большей исправленной дисперсии, например S_X^2 к меньшей, например S_Y^2 ($S_X^2 > S_Y^2$).

5) Сравниваем числа $F_{\text{крит}}$ и $F_{\text{набл}}$:

если $F_{\text{набл}} \geq F_{\text{крит}}$, то отвергаем гипотезу H_0 и принимаем гипотезу H_1,

если $F_{\text{набл}} < F_{\text{крит}}$, то нет оснований отвергать гипотезу H_0.

Пример 7. По двум независимым выборкам объемов $n_1 = 10$ и $n_2 = 15$, извлеченным из нормальных генеральных совокупностей X и Y найдены оценки дисперсий: $S_X^2 = 8{,}42$ и $S_Y^2 = 4{,}23$. При уровне значимости $\alpha = 0{,}05$ проверить гипотезу $H_0{:}D(X) = D(Y)$, при конкурирующей гипотезе $H_1{:}D(X) > D(Y)$;

Решение:

1) По данным выборки $S^2_{\text{наиб}} = S_X^2$, $S^2_{\text{наим}} = S_Y^2$, поэтому:

$$F_{\text{набл.}} = \frac{S^2_{\text{наиб}}}{S^2_{\text{наим}}} = \frac{8{,}42}{4{,}23} = 1{,}99\,.$$

2) По табл. 8 (см. приложение 8), учитывая значения

$\alpha = 0{,}05$, $k_1 = 10 - 1 = 9$, $k2 = 15 - 1 = 14$, находим число: $F_{\text{крит}} = 2{,}65$.

3) Сравниваем: так как $1{,}99 < 2{,}65$, то есть $F_{\text{набл}} < F_{\text{крит}}$, то нет основания отвергать гипотезу H_0.

Ответ: гипотеза $H_0{:}D(X) = D(Y)$ принимается.

Второй случай. Нулевая гипотеза $H_0{:}D(X) = D(Y)$, конкурирующая гипотеза $H_1{:}D(X) \neq D(Y)$.

В этом случае строят двустороннюю критическую область, исходя из требования, чтобы вероятность попадания критерия $F_{\text{набл}}$ в данную область при условии справедливости нулевой гипотезы H_0 была равна принятому уровню значимости α.

$$P\left(F_{\text{набл}} < F_{1\text{крит}}(\alpha, k_1, k_2)\right) = \frac{\alpha}{2}, \tag{13.5.5}$$

$$P\left(F_{\text{набл}} \geq F_{2\text{крит}}(\alpha, k_1, k_2)\right) = \frac{\alpha}{2}. \tag{13.5.6}$$

При определении критических точек достаточно определить правые критических точки для распределения Фишера–Снедекора.

Правило проверки нулевой гипотезы H_0 формулируется следующим образом:

$$P\begin{cases} \text{при } F_{\text{набл}} \geq F_{2\text{крит}}(\alpha, k_1, k_2) \to H_1, \\ \text{при } F_{\text{набл}} < F_{1\text{крит}}(\alpha, k_1, k_2) \to H_1, \\ \text{при } F_{1\text{крит}} \leq F_{\text{набл}} < F_{2\text{крит}} \to H_0. \end{cases} \tag{13.5.7}$$

Таким образом, для второго случая применительно к алгоритму для первого случая можно сделать следующие замечания.

Если для нулевой гипотезы $H_0{:}D(X) = D(Y)$ конкурирующая гипотеза $H_1{:}D(X)\neq D(Y)$, то строят двустороннюю критическую область. Для этого по табл. 8 (см. приложение 8) вычисляют правую границу $F_{2крит}$ критической области по уровню значимости $\alpha/2$ и числам степеней свободы $k_1 = n_1-1$, $k_2 = n_2-1$.

Тогда, если $F_{набл}\geq F_{2крит}$, то гипотеза H_0 отвергается и принимается гипотеза H_1, если $F_{набл} < F_{2крит}$, то нет основания отвергать гипотезу H_0.

Пример 8. По двум независимым выборкам объемов $n_1 = 10$ и $n_2 = 18$ и извлеченным из нормальных генеральных совокупностей X и Y, найдены оценки дисперсий: $S_X^2 = 1{,}23$ и $S_Y^2 = 0{,}41$. При уровне значимости $\alpha = 0{,}1$ проверить гипотезу $H_0{:}D(X) = D(Y)$ при конкурирующей гипотезе $H_1{:}D(X)\neq D(Y)$;

Решение:

1) По данным выборки $S^2_{наиб} = S_X^2$, $S^2_{наим} = S_Y^2$, поэтому

$$F_{набл.} = \frac{S^2_{наиб}}{S^2_{наим}} = \frac{1{,}23}{0{,}41} = 3.$$

2) По табл. 8 (см. приложение 8), учитывая значения:

здесь критическая область двухсторонняя, поэтому уровень значимости принимаем равным:

$$\boldsymbol{\alpha} = \frac{0{,}1}{2} = 0{,}05\cdot$$

Следовательно, $\alpha = 0{,}05$, $k_1 = 10 - 1 = 9$, $k_2 = 18 - 1 = 17$, находим число: $F_{крит} = 2{,}5$.

3) Сравниваем: так как $3 > 2{,}5$ то есть $F_{набл} > F_{крит}$, то нулевую гипотезу H_0 отвергаем.

Ответ: гипотеза $H_0{:}D(X) = D(Y)$ не принимается.

13.6. Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых неизвестны

Пусть даны две независимые выборки объемов n_1 и n_2 соответственно из нормально распределенных генеральных совокупностей X и Y. По выборкам найдены оценки математических ожиданий $\overline{x}_В$, $\overline{y}_В$ и исправленные выборочные дисперсии S_X^2 и S_Y^2, при этом генеральные дисперсии $D(X)$ и $D(Y)$ неизвестны, но предполагаются одинаковыми. Требуется сравнить $M(X)$ и $M(Y)$ генеральных совокупностей.

Алгоритм сравнения $M(X)$ и $M(Y)$:

1) Выдвигаем нулевую гипотезу: $H_0{:}D(X) = D(Y)$

В качестве конкурирующей гипотезы рассмотреть $H_1{:}D(X) \neq D(Y)$

2) Задаем число α – уровень значимости нулевой гипотезы;

3) Находим по табл. 5 распределения Стьюдента (см. приложение 5) значение $t_{\text{крит}}$ по заданному α и числу $k = n_1 + n_2 - 2$;

4) Находим число:

$$t_{\text{набл}} = \frac{\overline{x}_{\text{в}} - \overline{y}_{\text{в}}}{\sqrt{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2) - 2}{n_1 + n_2}}. \qquad (13.6.1)$$

5) Сравниваем числа $t_{\text{крит}}$ и $t_{\text{набл}}$:

если $|t_{\text{набл}}| > t_{\text{крит}}$, то отвергнуть гипотезу H_0,

если $|t_{\text{набл}}| < t_{\text{крит}}$, то нет основания отвергать гипотезу H_0.

13.7. Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых известны

Если необходимо проверить гипотезу $H_0{:}M(X) = M(Y)$ о равенстве средних двух нормальных генеральных совокупностей X и Y при условии известных дисперсий $D(X)$ и $D(Y)$ и при конкурирующей гипотезе $H_1{:}M(X) \neq M(Y)$, то в описанной выше схеме вместо $t_{\text{крит}}$ используют число $Z_{\text{крит}}$, определяемое с помощью табл.3 (см. приложение 3) по заданному α из равенства:

$$\Phi_0(z_{\text{крит}}) = \frac{1 - \alpha}{2}.$$

При конкурирующей гипотезе $H_1{:}M(X) > M(Y)$:

$$\Phi_0(z_{\text{крит}}) = \frac{12 - \alpha}{2}.$$

Вместо $t_{\text{набл}}$ по данным выборок вычисляют число:

$$z_{\text{набл}} = \frac{\overline{x}_{\text{в}} - \overline{y}_{\text{в}}}{\sqrt{\dfrac{D(X)}{n_1} + \dfrac{D(Y)}{n_2}}}. \qquad (13.7.1)$$

Если $|Z_{\text{набл}}| < Z_{\text{крит}}$, то нет основания отвергать гипотезу H_0.

Если $|Z_{набл}| > Z_{крит}$, то гипотезу H_0 отвергают.

Пример 9. По двум независимым выборкам объемов $n_1 = 10$ и $n_2 = 16$, извлеченным из нормальных генеральных совокупностей X и Y, найдены оценки математических ожиданий $\bar{x}_B = 2{,}5$ и $\bar{y}_B = 3{,}1$ и исправленные выборочные дисперсии $S_X^2 = 0{,}62$ и $S_Y^2 = 0{,}43$. Проверить нулевую гипотезу $H_0{:}D(X) = D(Y)$ при конкурирующей гипотезе: $H_1{:}D(X) \neq D(Y)$ и уровне значимости $\alpha = 0{,}01$.

Решение:

1) Так как $S_X^2 \neq S_Y^2$, то предварительно проверим гипотезу: $H_0{:}D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1{:}D(X) \neq D(Y)$ Для этого поступим по аналогии с решением примера 1:

а) По данным выборки вычисляем:

$$F_{\text{набл.}} = \frac{S^2_{\text{наиб.}}}{S^2_{\text{наим.}}} = \frac{0{,}62}{0{,}43} = 1{,}44;$$

б) По табл. 8 (см. приложение 8), учитывая значения:

$\alpha = 0{,}05$, $k_1 = 10 - 1 = 9$, $k_2 = 16 - 1 = 15$, находим число $F_{крит} = 2{,}59$;

в) Сравниваем: так как $1{,}44 < 2{,}59$, то есть $F_{набл} < F_{крит}$, то гипотеза о равенстве генеральных дисперсий принимается, следовательно различие между $S_X^2 = 0{,}62$ и $S_Y^2 = 0{,}43$ считаем незначительным.

2) Проверяем гипотезу $H_0{:}D(X) = D(Y)$ о равенстве средних при конкурирующей гипотезе : $H_1{:}D(X) \neq D(Y)$.

а) Находим по табл. 5 (см. приложение 5) значение $t_{крит}$ по заданному $\alpha = 0{,}01$ и числу $k = 10+16 - 2 = 24$ $t_{крит} = 2{,}8$.

б) Находим число $t_{набл}$:

$$t_{\text{набл.}} = \frac{\bar{x}_{\text{в}} - \bar{y}_{\text{в}}}{\sqrt{(n_1 - 1)\cdot S_X^2 + (n_2 - 1)\cdot S_Y^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} =$$

$$= \frac{2{,}5 - 3{,}1}{\sqrt{9 \cdot 0{,}62 + 15 \cdot 0{,}43}} \cdot \sqrt{\frac{10 \cdot 16 \cdot (10 + 16 - 2)}{10 + 16}} =$$

$$= -\frac{0{,}6}{\sqrt{12{,}03}} \cdot 12{,}153 = -\frac{0{,}6}{3{,}47} \cdot 12{,}153 = -2{,}101$$

в) Сравниваем числа $t_{крит}$ и $|t_{набл}|$, так как $2{,}101 < 2{,}8$, то $|t_{набл}| < t_{крит}$ и гипотеза $H_0{:}D(X) = D(Y)$о равенстве средних принимается.

Ответ: гипотеза $H_0{:}D(X) = D(Y)$ принимается.

Пример 10. По двум независимым выборкам объемов $n_1 = 18$ и $n_2 = 15$, извлеченным из нормальных генеральных совокупностей X и Y с дисперсиями $D(X) = 9$, $D(Y) = 12$, вычислены оценки математических ожиданий: $x_В = 12{,}7$, $y_В = 10{,}2$. При уровне значимости $\alpha = 0{,}05$ проверить гипотезу $H_0{:}D(X) = D(Y)$.

Решение:

1) Вычислим:

$$z_{\text{набл}} = \frac{\overline{x}_{\text{в}} - \overline{y}_{\text{в}}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} = \frac{12{,}7 - 10{,}2}{\sqrt{\frac{9}{18} + \frac{12}{15}}} = \frac{2{,}5}{\sqrt{0{,}5 + 0{,}8}} = \frac{2{,}5}{\sqrt{1{,}3}} = 2{,}19\,.$$

2) Находим $Z_{\text{крит}}$ из уравнения:

$$\Phi_0\left(z_{\text{крит}}\right) = \frac{1-\alpha}{2} = \frac{1-0{,}05}{2} = \frac{0{,}95}{2} = 0{,}475\,\cdot$$

то есть $\Phi_0^{-1}\left(z_{\text{крит}}\right) = 0{,}475$, используя табл. 3 (см. приложение 3).

Следовательно, $Z_{\text{крит}} = 1{,}96$.

3) Сравниваем: так как $2{,}19 > 1{,}96$, то есть $Z_{\text{набл}} > Z_{\text{крит}}$, то гипотезу H_0 отвергают. Значит, различие генеральных математических ожиданий значительное.

Ответ: гипотеза $H_0{:}D(X) = D(Y)$ отвергается.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что называют статистической гипотезой?
2. Какая гипотеза называется нулевой?
3. Какая гипотеза называется альтернативной?
4. Какая гипотеза называется сложной?
5. Какими методами проверяются гипотезы?
6. Что относят к ошибкам первого рода?
7. Что относят к ошибкам второго рода?
8. Для проверки какой гипотезы используется критерий согласия Пирсона?
9. Для проверки какой гипотезы используется критерий Колмогорова?
10. Что такое число степеней свободы?
11. Как сравнить дисперсии нормальных генеральных совокупностей?
12. Как сравнить средние нормальных генеральных совокупностей?

ЗАДАНИЯ

13.1. Распределение признака X в выборке дается следующим вариационным рядом:

3,0–3,6	3,6–4,2	4,2–4,8	4,8–5,4	5,4–6,0	6,0–6,6	6,6–7,2
2	8	35	43	22	15	5

При уровне значимости α = 0,01 проверить гипотезу о нормальности распределения X в генеральной совокупности.

13.2. По двум независимым выборкам объемов n_1 = 15 и n_2 = 10, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии S_X^2 = 1,9 и S_Y^2 = 3,2. При уровне значимости α = 0,1 проверить гипотезу H_0:D(X) = D(Y) при конкурирующей гипотезе H_1:D(X) $\neq$ D(Y).

13.3. По двум независимым выборкам объемов n_1 = 20 и n_2 = 30, извлеченным из нормальных генеральных совокупностей *X* и *Y* и с дисперсиями D(X) = 25 и D(Y) = 32, найдены выборочные средние $\bar{x}_В$ = 52 и $\bar{y}_В$ = 60. При уровне значимости α = 0,1 проверить нулевую гипотезу H_0:D(X) = D(Y) при конкурирующей гипотезе: H_1:D(X) $\neq$ D(Y).

13.4. В итоге проверки получено эмпирическое распределение. Случайная величина *X* принимает значения в интервалах с соответствующими частотами (см. таблицу). Требуется при уровне значимости α = 0,9 проверить по критерию χ^2 гипотезу о том, что случайная величина *X* распределена по нормальному закону.

R_i	[0–1)	[1–2)	[2–3)	[3–4)	[4–5)
n_i	11	26	33	24	19

13.5. В итоге проверки получено эмпирическое распределение. Случайная величина *X* принимает значения x_i с соответствующими частотами n_i (см. таблицу). Требуется при уровне значимости α = 0,9 проверить по критерию χ^2 гипотезу о том, что случайная величина X распределена по закону Пуассона.

x_i	0	1	2	3	4
n_i	123	54	35	22	3

13.6 В итоге проверки получено эмпирическое распределение. Случайная величина *X* принимает значения в интервалах с соответствующими частотами (см. таблицу). Требуется при уровне значимости α = 0,95 проверить по критерию χ^2 гипотезу о том, что случайная величина X распределена по нормальному закону.

R_i	[0–1)	[1–2)	[2–3)	[3–4)	[4–5)
n_i	20	45	50	19	15

13.7. В итоге проверки получено эмпирическое распределение. Случайная величина X принимает значения x_i с соответствующими частотами n_i (см. таблицу). Требуется при уровне значимости α = 0,1 проверить по критерию χ^2 гипотезу о том, что случайная величина X распределена по закону Пуассона.

x_i	0	1	2	3	4
n_i	150	98	46	15	5

13.8. По двум независимым выборкам объемов n_1 = 18 и n_2 = 28, извлеченным из нормальных генеральных совокупностей X и Y и с дисперсиями D(X) = 24 и D(Y) = 31, найдены выборочные средние $\bar{x}_B$ = 52 и $\bar{y}_B$ = 60. При уровне значимости α = 0,1 проверить нулевую гипотезу H_0: D(X) = D(Y), при конкурирующей гипотезе H_1:D(X) $\neq$ D(Y).

13.9. По двум независимым выборкам объемов n_1 = 19 и n_2 = 14 извлеченным из нормальных генеральных совокупностей X и Y найдены выборочные средние $\bar{x}_B$ = 11,1 и $\bar{y}_B$ = 15,6 и исправленные выборочные дисперсии S_X^2 = 0,57 и S_Y^2 = 0,82. При уровне значимости α = 0,05 проверить нулевую: H_0:D(X) = D(Y), при конкурирующей гипотезе H_1:D(X) $\neq$ D(Y).

13.10. По двум независимым выборкам объемов n_1 = 11 и n_2 = 19, извлеченным из нормальных генеральных совокупностей X и Y найдены исправленные выборочные дисперсии S_X^2 = 25,31 и S_Y^2 = 10,23. При уровне значимости α = 0,05 проверить гипотезу H_0:D(X) = D(Y), при конкурирующей гипотезе H_1:D(X) $\neq$ D(Y).

13.11. По двум независимым выборкам объемов n_1 = 10 и n_2 = 12, извлеченным из нормальных генеральных совокупностей X и Y найдены выборочные дисперсии D(X) = 12,2 и D(Y) = 18,2. При уровне значимости α = 0,02 проверить гипотезу H_0:D(X) = D(Y), при конкурирующей гипотезе H_1:D(X) $\neq$ D(Y).

13.12. По двум независимым выборкам объемов, извлеченным из нормальных генеральных совокупностей X и Y и с дисперсиями D(X) = 82 и D(Y) = 72 найдены выборочные средние оценки математического ожидания: $\bar{x}_B$ = 120 и $\bar{y}_B$ = 115, значимости α = 0,01 проверить нулевую гипотезу H_0:D(X) = D(Y) при конкурирующей гипотезе H_1: D(X) $\neq$ D(Y).

13.13. В итоге проверки получено эмпирическое распределение. Случайная величина *X* принимает значения в интервалах с соответствующими частотами (см. таблицу). Требуется при уровне значимости α = 0,3 проверить по критерию χ^2 гипотезу о том, что случайная величина *X* распределена по нормальному закону.

R_i	[0–1)	[1–2)	[2–3)	[3–4)	[4–5)
n_i	13	25	33	20	18

13.14. В итоге проверки получено эмпирическое распределение. Случайная величина *X* принимает значения x_i с соответствующими частотами n_i (см. таблицу). Требуется при уровне значимости α = 0,01 проверить по критерию χ^2 гипотезу о том, что случайная величина *X* распределена по закону Пуассона.

x_i	0	1	2	3	4
n_i	178	101	65	40	40

13.15. По двум независимым выборкам объемов n_1 = 14 и n_1 = 18, извлеченным из нормальных генеральных совокупностей *X* и *Y* найдены исправленные выборочные дисперсии S_X^2 = 1,64 и S_Y^2 = 3. При уровне значимости α = 0,01 проверить гипотезу H_0:D(X) = D(Y), при конкурирующей гипотезе H_1:D(X) $\neq$ D(Y).

ГЛАВА 14
КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

14.1. Основные понятия

Остановимся несколько подробнее на вопросе зависимости между случайными величинами. Для простоты понимания рассмотрим зависимость между двумя случайными величинами X и Y.

Целью задач корреляционного анализа является установление и оценка зависимости между случайными величинами X и Y.

Между X и Y зависимость может быть функциональной, статистической (стохастической), либо отсутствовать зависимость вообще.

Функциональной называется такая зависимость, если задается закон сопоставления между X и Y, то есть каждому значению независимой (факторной) случайной величины X ставится в соответствие **одно** возможное (результативное) значение Y. Данные в этом случае называются **несгруппированными**. Как известно, функциональная зависимость может быть аналитической, табличной и графической.

Функциональная зависимость при несгруппированных данных в общем случае записывается так:

$$y = f(x). \tag{14.1.1}$$

Статистической (или стохастической) называется такая зависимость между X и Y, когда каждому значению независимой случайной величины X ставятся в соответствие среднее вероятное значение Y. Данные в этом случае называются **сгруппированными.**

В частности, статистическая вероятность может задаваться функциональной зависимостью между независимыми значениями X и средними значениями Y или наоборот. Такая зависимость называется **корреляционной.** Записывается так:

$$\overline{y}_x = f(x). \tag{14.1.2}$$

или:

$$\overline{x}_y = \varphi(y). \tag{14.1.3}$$

В приведенных равенствах левые части называются условными средними:

y_x – среднее арифметическое наблюдающихся значений случайной величины Y;

x_y – среднее арифметическое наблюдающихся значений случайной величины X.

При этом, правые части этих равенств – выборочными регрессиями Y на X и X на Y соответственно.

Равенство (14.1.2) называется выборочным уравнением регрессии Y на X, а равенство (14.1.3) – выборочным уравнением регрессии X на Y.

Графики функций $y_x = f(x)$, $x_y = \varphi(y)$ называются выборочными линиями регрессии.

14.2. Основные задачи корреляционного анализа

Если функции зависят от одного аргумента, то, проведя корреляционный анализ, можно установить взаимный рост или убывание функций друг от друга в зависимости от изменения аргумента.

О наличии или отсутствии корреляции между двумя случайными величинами качественно можно судить по виду поля корреляции, поместив экспериментальные точки на координатную плоскость, что будет продемонстрировано решением конкретного примера.

Основные задачи корреляционного анализа:

1) Найти $y_x = f(x)$ или $x_y = \varphi(y)$;

2) Оценить тесноту корреляционной связи.

Если функции $y_x = f(x)$ или $x_y = \varphi(y)$ являются линейными, то корреляционная связь называется линейной. В остальных случаях она называется нелинейной.

Вид корреляционной связи определяется положением точек (x_i, y_i). Если эти точки на плоскости лежат вдоль некоторой прямой, то, естественно, следует положить, что зависимость линейная; если эти точки определяют форму параболы, то зависимость квадратичная и т.д.

Теснота корреляционной зависимости определяется по величине отклонения (рассеивания) точек (x_i, y_i) вокруг графиков функций $y_x = f(x)$, $x_y = \varphi(y)$: чем меньше отклонения, тем теснее зависимость.

Для отыскания прямой линии регрессии необходимо получить n пар чисел (x_i, y_i):

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

В линейном случае зависимость $y_x = f(x)$ выглядит так:

$$\overline{y}_x = kx + \mathrm{a}. \qquad (14.2.1)$$

Коэффициент k называется выборочным коэффициентом регрессии и обозначается ρ_{xy}.

Следовательно, выборочное уравнение прямой регрессии ищется в виде:

$$\overline{y}_x = \rho_{xy} X + \text{a}. \qquad (14.2.2)$$

Если данные несгруппированы, то есть каждому значению X поставлено в соответствие единственное значение y, то в равенстве (14.2.2) нет необходимости использовать понятие условной средней. В рассматриваемом случае вместо условной средней y_x берется Y:

$$Y = \rho_{xy} X + \text{a}. \qquad (14.2.3)$$

Естественно, неизвестные параметры ρ_{xy} и α желательно подобрать так, чтобы экспериментальные точки $(x_1, y_1), (x_2, y_2), \ldots, (x_\text{n}, y_\text{n})$ лежали как можно ближе к прямой (14.2.3).

Доказывается, что наивероятнейшими значениями параметров ρ_{xy} и α являются те, которые обеспечивают минимум функции:

$$S = \sum_{i=1}^{n} \left(\rho_{xy} x_i + a - y_i\right), \qquad (14.2.4)$$

по параметрам ρ_{xy} и α. В этой связи уравнение (14.2.3) называют среднеквадратическим.

Функция (14.2.3) является функцией двух переменных ρ_{xy} и α. По необходимому условию экстремума функции многих переменных должны выполняться равенства:

$$\frac{\partial S}{\partial \rho_{xy}} = 0, \qquad \frac{\partial S}{\partial a} = 0\cdot \qquad (14.2.5)$$

С учетом равенств (14.2.5) отыскание значений параметров ρ_{xy} и α сводится к решению системы двух линейных уравнений с двумя неизвестными относительно ρ_{xy} и α:

$$\begin{cases} na + \left(\sum_{i=1}^{n} x_i\right)\rho_{xy} = \sum_{i=1}^{n} y_i\,, \\ \left(\sum_{i=1}^{n} x_i\right) a + \left(\sum_{i=1}^{n} x_i^2\right)\rho_{xy} = \sum_{i=1}^{n} (x_i y_i). \end{cases} \qquad (14.2.6)$$

Вычисление параметров ρ_{xy} и α можно упростить, если уравнение регрессии искать в виде прямой не с угловым коэффициентом, а прямой, проходящей чрез точку: (x, b)

$$Y = b + \rho_{xy}(x - \overline{x}) \cdot \qquad (14.2.7)$$

Искомые параметры регрессии в этом случае определяются по формулам:

$$b = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad (14.2.8)$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad (14.2.9)$$

$$\rho_{xy} = \frac{Q_{xy}}{Q_x}, \qquad (14.2.10)$$

где:

$$Q_x = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2, \qquad (14.2.11)$$

$$Q_y = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2, \qquad (14.2.12)$$

$$Q_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right). \qquad (14.2.13)$$

14.3. Выборочный коэффициент корреляции

Количественной оценкой тесноты связи между признаками Y и X служит **выборочный коэффициент корреляции** ρ_B, который вычисляется по формуле:

$$\rho_B = \frac{Q_{xy}}{\sqrt{Q_x Q_y}} \; . \qquad (14.3.1)$$

Выборочный коэффициент корреляции обладает следующими свойствами:

1) $\rho_B \leq 1$;
2) если за уравнение регрессии выбрана линейная функция, то при $\rho_B = 0$ между переменными Y и X отсутствует корреляционная зависимость;
3) при $\rho_B = \pm 1$ между Y и X имеется строгая линейная функциональная зависимость;
4) при возрастании $\rho_{\text{В}}$ от от нуля до единицы с увеличением признака X увеличивается в среднем значение признака Y (положительная корреляция);
5) при $\rho_B < 0$ имеется отрицательная корреляция, заключающаяся в том, что увеличение признака X приводит в среднем к уменьшению признака Y.

Для определения степени корреляционной зависимости пользуются таблицей Чеддока:

Таблица 14.3.1

ρ_B	0,1– 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Степень связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Остановимся подробнее на вопросе определения параметров выборочного уравнения прямой линии регрессии при сгруппированных данных.

Обычно, на практике, если опыт проводится при одних и тех же условиях, при большом числе испытаний одни и те же значения признаков X и Y могут встречаться по несколько раз.

Обозначим через n_{xi} частоту появления x_i, через n_{yj} – частоту появления y_i , а через $n_{x_i y_j}$ – частоту появления одной и той же точки (x_i , y_i).

Данные записывают в таблицу (табл. 14.3.2), которая называется корреляционной.

Таблица 14.3.2

X \ Y	x_1	x_2	x_3		n_y
y_1	n_{x_1,y_1}	n_{x_2,y_1}	n_{x_3,y_1}		n_{y_1}
y_2	n_{x_1,y_2}	n_{x_2,y_2}	n_{x_3,y_2}		n_{y_2}
y_3	n_{x_1,y_3}	n_{x_2,y_3}	n_{x_3,y_3}		n_{y_3}
n_x	n_{x_1}	n_{x_2}	n_{x_3}		n

Выборочное уравнение прямой линии регрессии Y на X имеет вид:

$$\overline{Y}_x - \overline{y} = \rho_B \frac{\sigma_y}{\sigma_x}(x - \overline{x}), \tag{14.3.2}$$

где σ_x — выборочная дисперсия признака X, σ_y — выборочная дисперсия признака Y.

Для определения ρ_B введем промежуточный параметр B

$$B = \sum_{i=1}^{n}\left(n_{x_{i,}y_j} \cdot x_i \cdot y_j\right) - n\overline{x}\,\overline{y}$$

$$\rho_B = \frac{B}{n\sigma_x\sigma_y}\ . \tag{14.3.3}$$

По аналогии записывается выборочное уравнение прямой регрессии X на Y:

$$\overline{X}_y - \overline{x} = \rho_B \frac{\sigma_x}{\sigma_y}(y - \overline{y}). \tag{14.3.4}$$

Отметим дополнительные свойства выборочного коэффициента корреляции:

1) Если выборка имеет достаточно большой объем и хорошо представляет генеральную совокупность, то есть репрезентативна, то тесноту линейной зависимости выборки можно распространить и на всю генеральную совокупность.

2) Если объем выборки $n \geq 50$ и генеральная совокупность является нормально распределенной, то коэффициент корреляции генеральной совокупности $r_Г$ имеет следующий доверительный интервал:

$$r_B - 3\frac{1 - r_B^2}{\sqrt{n}} \leq r_{\text{Г}} \leq r_B + 3\frac{1 + r_B^2}{\sqrt{n}}\ . \tag{14.3.5}$$

3) Выборочный коэффициент корреляции r_B равен среднему геометрическому выборочных коэффициентов регрессии ρ_{xy}, ρ_{yx}:

$$r_B = \pm\sqrt{\rho_{xy} \cdot \rho_{yx}}\ . \tag{14.3.6}$$

Так как знак выборочного коэффициента корреляции совпадает со знаком выборочного коэффициента регрессии, то знак перед радикалом совпадает со знаком коэффициента регрессии.

Пример 1. Экономист, изучая зависимость выработки Y (млн. руб.) на одного работника торговли от величины товарооборота X (млн. руб.) магазина за определенный период, получил следующие данные по 15 магазинам одинакового профиля (таблица 14.3.2):

Таблица 14.3.2

№	1	2	34	4	5	6	7	8	9	10	11	12	13	14	15
X	150	38	85	28	146	34	95	50	134	120	74	140	110	60	86
Y	7,2	5,8	7,5	4,4	8,4	4,5	7,0	5,0	6,4	8,0	6,0	7,8	6,2	5,8	6,0

По диаграмме рассеивания (рис. 14.3.1) можно проследить линейную зависимость между признаками *Y* и *X*. В этом предположении требуется найти выборочное уравнение регрессии и выборочный коэффициент линейной корреляции; оценить ожидаемое значение признака *Y* при значении показателя *X*, равным 90 млн. руб.

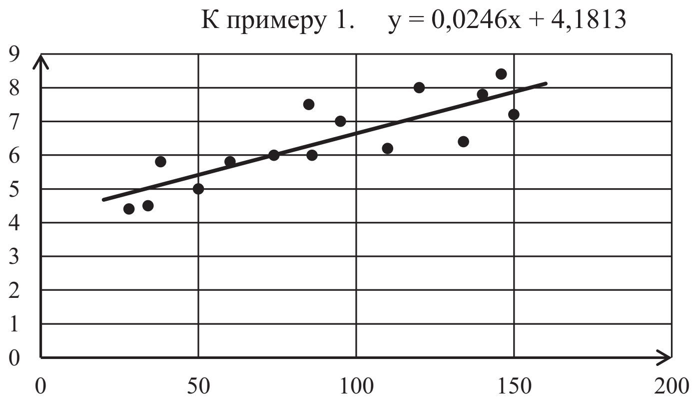


Рис. 14.3.1

Решение:

1. Составим таблицу данных и промежуточных вычислений (таб. 14.3.3).

Таблица 14.3.3

№	*X*	*Y*	X^2	Y^2	*XY*
1	150	7,2	22500	51,84	1080,00
2	38	5,8	1444	33,64	220,40

Продолжение таблицы

№	*X*	*Y*	*X*²	*Y*²	*XY*
3	85	7,5	7225	56,25	637,50
4	28	4,4	784	19,36	123,20
5	146	8,4	21316	70,56	1226,40
6	34	4,5	1156	20,25	153,00
7	95	7,0	9025	49,00	665,00
8	50	5,0	2500	25,00	250,00
9	134	6,4	17956	40,96	857,00
10	120	8,0	14400	64,00	960,00
11	74	6,0	5476	36,00	444,00
12	140	7,8	19600	60,84	1092,00
13	110	6,2	12100	38,44	682,00
14	60	5,8	3600	33,64	384,00
15	86	6,0	7396	36,00	515,00
∑	1350	96,0	146478	635,78	9255,10

2. Найдем значения исходных статистических характеристик по формулам (14.2.7) – (14.2.12):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1350}{15} = 90, b = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{96}{15} = 6{,}4,$$

$$Q_x = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 146478 - \frac{(1350)^2}{15} = 24978$$

$$Q_y = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 635{,}78 - \frac{(96)^2}{15} = 21{,}38$$

$$Q_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = 9255{,}1 - \frac{1350 \cdot 96}{15} = 615{,}1$$

1. Найдем значение выборочного коэффициента регрессии по формуле (14.2.10):

$$\rho_{xy} = \frac{Q_{xy}}{Q_x} = \frac{615{,}1}{24978} = 0{,}02463\cdot$$

2. Подставим найденные значения параметров $b = 6{,}4$, в уравнение прямой линии регрессии:

$$Y = 6{,}4 + 0{,}02463(x - 90) = 0{,}02463x + 4{,}1873.$$

3. Для оценки тесноты линейной связи, вычислим выборочный коэффициент линейной корреляции по формуле (14.3.1):

$$\rho_B = \frac{Q_{xy}}{\sqrt{Q_x Q_y}} = \frac{615{,}1}{\sqrt{24978 \cdot 21{,}38}} = 0{,}84.$$

Оценим ожидаемое значение признака Y при значении показателя $X = 90$ млн руб.:

$$Y(90) = 0{,}02463 \cdot 90 + 4{,}1873 = 6{,}4010 \text{ млн. руб.}$$

Выводы

1) Из уравнения регрессии следует, что при увеличении товарооборота на 1 млн. руб. выработка на одного работника в среднем увеличивается на 24,6 тыс. руб.

2) На основе значения выборочного коэффициента корреляции по таблице 1 Чеддока можно заключить, что связь между товарооборотом X магазина и выработкой Y на одного работника является положительной и высокой.

Пример 2. С целью выяснения линейной корреляционной зависимости между стоимостью готовой продукции (признак Y млн. руб.) предприятий и их основных фондов (признак X млн. руб.) собрана информация по 200-ам предприятий, сгруппированная в таблицу 14.3.4.

Таблица 14.3.4

X \ Y	10	20	30	40	50	60	n_y
15	5	7	–	–	–	–	12
25	–	20	23	–	–	–	43
35	–	–	30	47	2	–	79
45	–	–	10	11	20	6	47
55	–	–	–	9	7	3	19
n_x	5	27	63	67	29	9	$n = 200$

Требуется:

1) Найти значение коэффициента корреляции и оценить тесноту связи между рассматриваемыми признаками;

2) Составить уравнения прямых линий регрессии *Y* на *X* и *X* на *Y*.

Решение: как известно (см. формулы 14.3.1–14.3.3), выборочное уравнение прямой линии регрессии *Y* на *X* и *X* на *Y* таковы:

$$\overline{Y}_x - \overline{y} = \rho_B \frac{\sigma_y}{\sigma_x}(x - \overline{x}),$$

$$\overline{X}_y - \overline{x} = \rho_B \frac{\sigma_x}{\sigma_y}(y - \overline{y}),$$

где $\overline{Y}_x$, $\overline{X}_y$ – условные средние признаков *Y* и *X*;
$\overline{y}$, $\overline{x}$ – выборочные средние признаков *Y* и *X*;
σ_x – выборочная дисперсия признака *X*;
σ_y – выборочная дисперсия признака *Y*;
ρ_B – выборочный коэффициент корреляции.

1) По имеющимся данным (таблица 14.3.4) выборка большая, так как $n > 50$. В таких случаях для упрощения вычислений переходят к условным вариантам: вместо *X* вводят варианту *u*, а вместо *X* — разность *Y* – *u*, которые вычисляются по следующим формулам:

$$u_i = \frac{x_{i-}C_1}{h_1}, \quad v_j = \frac{y_{j-}C_2}{h_2},$$

где C_1 – варианта признака *X*, имеющая наибольшую частоту;
C_2 – варианта признака *Y*, имеющая наибольшую частоту;
h_1 – шаг варианты признака *X*;
h_2 – шаг варианты признака *Y*.

В этих обозначениях выборочный коэффициент корреляции вычисляется по формуле:

$$\rho_B = \frac{\sum\sum n_{uv}uv - \overline{u}\,\overline{v}}{n\sigma_u\sigma_v},$$

где $\overline{u}$ – выборочное среднее значение условной варианты *u*;
$\overline{v}$ – выборочное среднее значение условной варианты *v*;
σ_u – выборочное среднее квадратическое отклонение варианты *u*;
σ_v – выборочное среднее квадратическое отклонение варианты *v*,
которые вычисляются по формулам:

$$\overline{u} = \frac{\sum n_u\, u}{n}, \quad \overline{v} = \frac{\sum n_v\, v}{n},$$

$$\sigma_u = \sqrt{\overline{u^2} - (\overline{u})^2}, \quad \sigma_v = \sqrt{\overline{v^2} - (\overline{v})^2}.$$

Здесь $\overline{u^2}$ – выборочное среднее значение квадрата условной варианты u, $\overline{v^2}$ – выборочное среднее значение квадрата условной варианты v.

По заданной корреляционной таблице 14.3.4 наибольшая частота n_{xy} равна 47, которая соответствует значениям $X = 40$ и $Y = 35$. Эти значения и берутся за C_1 и C_2: $C_1 = 40$; $C_2 = 35$.

Шаг h_1 равен разности между двумя соседними значениями вариант признака X, т.е. $h_1 = 20 - 10$;

h_2 равен разности между двумя соседними значениями вариант признака Y, т.е. $h_2 = 25 - 15$.

Следовательно, $h_1 = h_2 = 10$.

В общем случае эти шаги не обязательно должны быть равными. Это сделано для простоты вычислений.

Составим корреляционную таблицу 17.

1) В первой строке вместо ложного нуля C_1 (варианты 40) пишут 0. Слева от нуля последовательно записывают –1; –2; –3, а справа от нуля значения 1; 2.

2) В первом столбце вместо ложного нуля C_2 (варианты 35) пишут 0. Над нулем последовательно записывают –1; –2; а под нулем значения 1; 2.

3) Все остальные данные переписывают из таблицы 14.3.4. В итоге получим корреляционную таблицу 14.3.5 в условных вариантах.

Таблица 14.3.5.

u \ v	–3	–2	–1	0	1	2	n_v
–2	5	7	–	–	–	–	12
–1	–	20	23	–	–	–	43
0	–	–	30	47	2	–	79
1	–	–	10	11	20	6	47
2	–	–	–	9	7	3	19
n_u	5	27	63	67	29	9	$n = 200$

4) С помощью таблицы 14.3.5 находим средние выборочные условных вариант:

$$\overline{u} = \frac{\sum n_u u}{n} = \frac{5 \cdot (-3) + 27 \cdot (-2) + 63 \cdot (-1) + 67 \cdot 0 + 29 \cdot 1 + 9 \cdot 2}{200} = -0{,}425$$

$$\overline{v} = \frac{\sum n_v v}{n} = \frac{12 \cdot (-2) + 43 \cdot (-1) + 79 \cdot 0 + 47 \cdot 1 + 19 \cdot 2}{200} = 0{,}090$$

5) Находим средние выборочные квадратов условных вариант:

$$\overline{u^2} = \frac{\sum n_u u^2}{n} = \frac{5 \cdot 9 + 27 \cdot 4 + 63 \cdot 1 + 67 \cdot 0 + 29 \cdot 1 + 9 \cdot 4}{200} = 1{,}405$$

$$\overline{v^2} = \frac{\sum n_v v^2}{n} = \frac{12 \cdot 4 + 43 \cdot 1 + 63 \cdot 1 + 79 \cdot 1 + 47 \cdot 1 + 19 \cdot 4}{200} = 1{,}0470,$$

6) Находим:

$$\sigma_u = \sqrt{\overline{u^2} - (\overline{u})^2} = \sqrt{1{,}405 - (0{,}425)^2} = 1{,}106,$$

$$\sigma_v = \sqrt{\overline{v^2} - (\overline{v})^2} = \sqrt{1{,}070 - (0{,}090)^2} = 1{,}209.$$

7) Находим среднеквадратические отклонения истинных вариант x и y.

$$\sigma_x = h_1 \sigma_u = 10 \cdot 1{,}106 = 11{,}06,$$

$$\sigma_y = h_2 \sigma_v = 10 \cdot 1{,}209 = 12{,}09.$$

8) Находим среднее выборочное истинных вариант:

$$\overline{x} = h_1 \overline{u} + C_1 = 10 \cdot (-0{,}425) + 40 = 35{,}75,$$

$$\overline{y} = h_2 \overline{v} + C_2 = 10 \cdot 0{,}090 + 35 = 35{,}9$$

9) Находим двойную сумму $U = \sum\sum n_{uv} uv$ с помощью расчетной таблицы 14.3.6.

Таблица 14.3.6

u \ v	–3	–2	–1	0	1	2	$U = \sum n_{uv} u$	vU
–2	–15 5 –10	–14 7 –14	–	–	–	–	–29	58
–1	–	–40 20 –20	–23 23 –23	–	–	–	–63	63
0	–	–	–30 30 0	0 47 0	2 2 0	–	–28	0
1	–	–	–10 10 10	0 11 11	20 20 20	12 6 6	22	22
2	–	–	–	0 9 18	7 7 14	6 3 6	13	26
$V = \sum n_{uv} v$	–10	–34	–13	29	34	12	Контроль	$\sum uV = 169$
uV	30	68	13	0	34	24	$\sum uV = 169$	

Пояснения к заполнению таблицы 14.3.6:

а) Количество строк и столбцов увеличиваем на единицу по сравнению с таблицей 17;

б) В каждой клетке записываем три числа: в центре клетки записана варианта n_{uv}; в правом верхнем углу ее произведение на соответствующую варианту u из второй строки. Например, в первой клетке в правом верхнем углу записываем произведение $5 \cdot (-3) = -15$; в левом нижнем углу записывается произведение частоты n_{uv} на соответствующую варианту v, стоящую в первом столбце: $5 \cdot (-2) = -10$;

в) Складываем все числа, помещенные в правых верхних углах клеток одной и той же строки, а их сумму записываем в клетку этой же строки, но в предпоследний столбец. U.

Например, для первой строки $U = -15 + (-14) = -29$;

г) Умножаем варианту v из первого столбца на соответствующее значение U из предпоследнего столбца и полученное произведение записываем в клетку этой же строки последнего столбца vU. Например, для первой строки варианта $v = -2$, значение параметра $U = -29$. Следовательно, $vU = (-2) \cdot (-29) = 58$. Это число и записываем для первой строки в последний столбец;

д) Складываем все числа последнего столбца vU. Получаем сумму $\sum vU = 169$, которая и равна искомой сумме $U = \sum n_{uv}uv$;

е) Для контроля аналогичные вычисления производят по столбцам. Правильность вычислений проверяется справедливостью равенства:

$$\sum_v vU = \sum_u uU;$$

10) Находим значение выборочного коэффициента корреляции:

$$\rho_B = \frac{\sum\sum n_{uv}uv - \overline{u}\,\overline{v}}{n\sigma_u\sigma_v} = \frac{169 - 200 \cdot (-0{,}425) \cdot 0{,}09}{200 \cdot 1{,}106 \cdot 1{,}209} = 0{,}603;$$

11) Подставляем все найденные значения в уравнения прямых линий регрессии:

$$\overline{Y}_x - \overline{y} = \rho_B \frac{\sigma_y}{\sigma_x}(x - \overline{x}),$$

$$\overline{X}_y - \overline{x} = \rho_B \frac{\sigma_x}{\sigma_y}(y - \overline{y}).$$

После выполнения простейших арифметических вычислений, окончательно находим:

$$\overline{Y}_x - 35{,}9 = 0{,}603\frac{12{,}09}{11{,}06}(x - 35{,}75) \Rightarrow \overline{Y}_x = 0{,}603\frac{12{,}09}{11{,}06}(x - 35{,}75) + 35{,}9 \Rightarrow$$

$$\overline{Y}_x = 0{,}66x + 12{,}34.$$

$$\overline{X}_y - 35{,}75 = 0{,}603\frac{11{,}06}{12{,}09}(y - 35{,}9) \Rightarrow \overline{X}_y = 0{,}603\frac{11{,}06}{12{,}09}(y - 35{,}9) + 35{,}75 \Rightarrow$$

$$\overline{X}_y = 0{,}55 + 15{,}95$$

Таким образом

$$\overline{Y}_x = 0{,}66x + 12{,}34,$$

$$\overline{X}_y = 0{,}55 + 15{,}95\ \cdot$$

14.4. Ранговая корреляция

Если n объектов какой-либо совокупности N пронумерованы в соответствии с возрастанием или убыванием какого-либо признака X, таким образом, что они выстроены в ряд (по рангу), то есть ранжированы по рассматриваемому признаку. Ранг x_i указывает место, которое занимает i-й объект среди остальных n объектов, расположенных в соответствии с признаком X ($i = 1{,}2, \ldots, n$). Например, при исследовании вопроса пополнения каталога библиотеки необходимой литературой приходится учитывать предпочтения читателей, какой жанр литературы их больше интересует (Э – Эпос, М – мелодрама, Д – детектив, С – сказки, П – поэма, Н – научная фантастика).

Следуя результатам опроса, мы можем составить некоторые наборы ранжированных рядов и можем попытаться установить степень линейной связи между ними. Предположим, имеется 6 указанных жанров, расположенных по порядку предпочтений от 1 до 6 в соответствии с двумя характеристиками Ж – жизненность жанра, и У – увлекательность (табл. 14.4.1).

Характеристики для ранжирования	Вид литературы					
	Э	М	Д	С	П	Н
Ж	4	5	4	1	3	2
У	2	4	6	3	4	4

Для определения наличия взаимосвязи между ранговыми оценками используется коэффициент ранговой корреляции Спирмена ρ, который основан на различии между рангами и рассчитывается по формуле:

$$\rho = 1 - \left(6 \sum_{i=1}^{n} D^2\right) \Big/ \left(n(n^2 - 1)\right), \qquad (14.4.1)$$

где n – число пар ранжированных наблюдений.

В нашем случае $n = 6$, тогда:

$$\sum_{i=1} D^2 = (4-2)^2 + (5-4)^2 + (4-6)^2 + (1-3)^2 + (3-4)^2 + (2-4)^2 =$$

$$= 4 + 1 + 4 + 4 + 1 + 4 = 18$$

По формуле (14.4.1) получим коэффициент ранговой корреляции Спирмена:

$$\rho = 1 - \frac{6 \cdot 18}{6(36-1)} = \frac{108}{210} = 0{,}514.$$

Таким образом, коэффициент Спирмена показывает, что между признаками есть достаточно сильная линейная связь. Значения коэффициента Спирмена определяются неравенством $-1 \leq \rho \leq 1$ и интерпретируется так же, как и коэффициент Пирсона с разницей в том, что он применяется для ранжированных данных.

Значимость коэффициента Спирмена проверяется на основе t критерия Стьюдента по формуле:

$$t_{\text{набл}} = \rho\sqrt{\frac{n-1}{1-\rho^2}}. \qquad (14.4.2)$$

Значение коэффициента считается существенным, если:

$$t_{\text{набл}} > t_{\text{крит}}(\alpha, k = n - 2).$$

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какую цель преследует теория корреляции?
2. Какие виды зависимостей могут быть между X и Y?
3. Каковы основные задачи корреляционного анализа?
4. Что показывает коэффициент корреляции?
5. В каких пределах меняется коэффициент корреляции?
6. Как определяется выборочный коэффициент корреляции?
7. Что понимается под уравнением регрессии?
8. Что понимается под диаграммой рассеивания?
9. Что показывает коэффициент ранговой корреляции?
10. По какой формуле определяется коэффициент ранговой корреляции?

ЗАДАНИЯ

14.1. На плоскости даны шесть точек, координаты которых занесены в таблицу. Пусть случайная величина X – абсцисса точек, а случайная величина Y– ордината. Найти коэффициент линейной корреляции. Написать уравнения линейной регрессии Y на X и X на Y.

x_i	1,1	2,2	2,9	3,1	3,5	3,8
y_i	2,6	3,5	4,1	4,2	4,4	4,7

14.2. Методами корреляционно – регрессионного анализа изучить влияние факторного признака на результативный признак исходные данные представлены в таблице: Зависимость потребления некоторого продукта Y (кг) от среднедушевого дохода X (тыс. руб.)

X	1	2	3	4	5	6
Y	2,6	5,7	9,3	11,9	15,1	17,9

1) определить уравнение регрессии;
2) вычислить выборочные показатели тесноты связи между признаками;
3) проверить гипотезу о значимости показателя тесноты связи в генеральной совокупности при α = 0,01 и n^* = 100;
4) оценить показатель тесноты связи в генеральной совокупности.

14.3. Получено распределение n магазинов по размерам торговой площади X (м^2) и по товарообороту Y (млн. руб.), представленное корреляционной таблицей.

X \ Y	4	9	14	19	24	20	n_y
10	2	3	–	–	–	–	5
20	–	7	3	–	–	–	10
30	–	–	2	50	2	–	54
40	–	–	1	10	6	–	17
50	–	–	–	4	7	3	14
n_x	2	10	6	64	15	3	n = 100

Полагая, что между признаками X и Y существует линейная корреляционная зависимость, требуется:

1) вычислить коэффициент корреляции и оценить тесноту связи между признаками X и Y;

2) составить уравнения прямых линий регрессии Y на X и X на Y.

14.4. Методами корреляционно-регрессионного анализа изучить влияние факторного признака на результативный признак, исходные данные представлены в таблице:

Зависимость удельного постоянного расхода Y (руб.) от объема выпускаемой продукции X (тыс. шт.)

X	2	3	4	5	6	7
Y	4,5	5,7	7,3	8,6	9,5	11,4

1) определить уравнение регрессии;
2) вычислить выборочные показатели тесноты связи между признаками;
3) проверить гипотезу о значимости показателя тесноты связи в генеральной совокупности при $\alpha = 0{,}01$ и $n^* = 100$;
4) оценить показатель тесноты связи в генеральной совокупности.

14.5. Опрос случайно выбранных 10 студентов, проживающих в общежитии университета, позволяет выявить зависимость между средним баллом по результатам предыдущей сессии и числом часов в неделю, затраченных студентом на самостоятельную подготовку.

Средний балл	4,6	4,3	3,8	3,8	4,2	4,3	3,8	4,0	3,1	3,9
Число часов	25	22	9	15	15	30	20	30	10	17

Постройте график исходных данных и определите по нему характер зависимости. Рассчитайте выборочный коэффициент линейной корреляции, проверьте его значимость при $\alpha = 0{,}05$. Постройте уравнение регрессии и дайте пояснения полученных результатов.

14.6 Компанию по прокату автомобилей интересует зависимость между пробегом автомобилей (X) и стоимостью ежемесячного технического обслуживания (Y). Для выяснения характера этой связи было отобрано 15 автомобилей.

X	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Y	12	15	14	19	18	22	25	23	29	31	20	34	33	39	38

Требуется построить график исходных данных, и определите по нему характер зависимости. Рассчитать выборочный коэффициент линейной корреляции, проверить его значимость при $\alpha = 0{,}05$. Построить уравнение регрессии и дать пояснения полученным результатам.

14.7. На плоскости даны шесть точек, координаты которых занесены в таблицу. Пусть случайная величина X – абсцисса точек, а случайная величина Y – ордината. Найти коэффициент линейной корреляции. Написать уравнения линейной регрессии Y на X и X на Y.

x_i	–2,1	–2,1	–1,7	–1,3	0,1	0,9
y_i	5,7	5,1	5,4	5,1	4,1	3,6

14.8. Методами корреляционно-регрессионного анализа изучить влияние факторного признака на результативный признак, исходные данные представлены в таблице:

Зависимость объема продаж Y (тыс.руб.) от расходов на рекламу X (тыс. руб.).

X	1	2	3	4	5	6
Y	0,3	1,5	3,3	4,6	6,1	7,5

1) определить уравнение регрессии;
2) вычислить выборочные показатели тесноты связи между признаками;
3) проверить гипотезу о значимости показателя тесноты связи в генеральной совокупности при α = 0,01 и n^* = 100;
4) оценить показатель тесноты связи в генеральной совокупности.

14.9. Имеются следующие статистические данные об уровне механизации работ X(%) и производительности труда Y(т/ч) для 14 однотипных предприятий:

x_i	32	30	36	40	41	47	56	54	60	55	61	67	69	76
y_i	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Требуется:

а) оценить тесноту и направление связи между переменными с помощью ко эффициента корреляции; проверить значимость коэффициента корреляции на уровне α = 0,05;

б) найти уравнения прямых регрессии.

14.10. Найти коэффициент корреляции между производительностью труда Y (тыс. руб.) и энерговооруженностью труда X (кВт) (в расчете на одного работающего) для 14 предприятий региона по следующим данным, представленным в таблице

x_i	2,8	2,2	3,0	3,5	3,2	3,7	4,0	4,8	6,0	5,4	5,2	5,4	6,0	9,0
y_i	6,7	6.9	7.2	7.3	8,4	8,8	9,1	9,8	10.6	10,7	11,1	11,8	12,1	12,4

ЛИТЕРАТУРА

1. *Вентцель Е.С., Овчаров Л.А.* Теория вероятностей и ее инженерные приложения. – М.: Наука. Гл. ред. физ. – мат. лит., 1988.
2. *Вентцель Е.С., Овчаров Л.А.* Теория случайных процессов и её инженерные приложения. М.: Наука. Гл. ред. физ.-мат. лит., 1991.
3. *Гмурман В.Е.*, Теория вероятностей и математической статистика. Учеб. пособие для вузов. Изд. 7-е, стер – М.: Высш. шк., 2000.
4. *Гмурман В.Е.*, Руководство к решению задач по теории вероятностей и математической статистике. Учеб. пособие для студентов вузов Изд. 5-е, стер. – М.: Высш. шк., 2001.
5. *Гнеденко Б.В.* Курс теории вероятностей. 6-е изд., – М.: Наука. Гл. ред. физ.-мат. лит., 1988.
6. *Гулиян Б.Ш., Гулиян Г.Б., Хамидуллин Р.Я.* Теория вероятностей и математическая статистика в примерах и задачах. – М.: ИД Академия Жуковского, 2015.
7. *Гулиян Б., Хамидуллин Р.* Математика Базовый курс. – 4-е изд., перераб. и доп. – М.: МФПА, 2014.
8. *Крамер Г.* Математические методы статистики. – М.: Мир, 1975.
9. *Красс М.С., Чупрынов Б.П.* Математика для экономистов. — СПб.: Питер, 2005.
10. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. Учебник для вузов. – М.: ЮНИТИ-ДАНА, 2000.
11. *Кремер Н.Ш., Путко Б.А.* Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. – М.:ЮНИТИ-ДАНА, 2002.
12. *Монсик В.Б.* Основы теории вероятностей и математической статистики: учебное пособие. – М.: ИД Академия Жуковского, 2004.
13. *Монсик В.Б., Скрытников А.А.* Вероятность и статистика: учебное пособие. – М.: Бином. Лаборатория знаний. 2011.
14. *Мхитарян В.С., Астафьева Е.В., Миронкина Ю.Н., Трошин Л.И.* Теория вероятностей и математическая статистика. – 2-е изд., перераб. и доп. – М.: Московский финансово-промышленный университет «Синергия», 2013.
15. *Ниворожкина П. П., Морозова З. А., Герасимова П. А., Житников П. В.* Основы статистики с элементами теории вероятностей для экономистов. – Ростов н/Д: Феникс, 1999.

16. *Письменный Д.Т.* Конспект лекций по теории вероятностей и математической статистике. – М.: Айрис-пресс. 2002.
17. *Примаков Д. А., Хамидуллин Р. Я.* Геометрия и топология. – М.: Маркет ДС, 2009.
18. *Рудык Б.М., Ермаков В.И., Гринцевичюс Р.К., Бобрик Г.И., Матвеев В.И., Гладких И.М., Сагитов Р.В., Шершнев В.Г.* Общий курс высшей математики для экономистов: Учебник/Под ред. В.И. Ермакова – М.: ИНФРА-М, 2000.

ПРИЛОЖЕНИЯ

Приложение 1 (табл. 1)

Таблица значений нормированной нормальной плотности вероятности

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

x	φ(*x*)	*x*	φ(*x*)	*x*	φ(*x*)	*x*	φ(*x*)	*x*	φ(*x*)
0	0,398942	**0,21**	0,390242	**0,42**	0,365263	**0,63**	0,327133	0,84	0,280344
0,01	0,398922	**0,22**	0,389404	**0,43**	0,363714	**0,64**	0,325062	0,85	0,277985
0,02	0,398862	**0,23**	0,388529	**0,44**	0,362135	**0,65**	0,322972	0,86	0,275618
0,03	0,398763	**0,24**	0,387617	**0,45**	0,360527	**0,66**	0,320864	0,87	0,273244
0,04	0,398623	**0,25**	0,386668	**0,46**	0,35889	**0,67**	0,318737	0,88	0,270864
0,05	0,398444	**0,26**	0,385683	**0,47**	0,357225	**0,68**	0,316593	0,89	0,268477
0,06	0,398225	**0,27**	0,384663	**0,48**	0,355533	**0,69**	0,314432	0,9	0,266085
0,07	0,397966	**0,28**	0,383606	**0,49**	0,353812	**0,7**	0,312254	0,91	0,263688
0,08	0,397668	**0,29**	0,382515	**0,5**	0,352065	**0,71**	0,31006	0,92	0,261286
0,09	0,39733	**0,3**	0,381388	**0,51**	0,350292	**0,72**	0,307851	0,93	0,258881
0,1	0,396953	**0,31**	0,380226	**0,52**	0,348493	**0,73**	0,305627	0,94	0,256471
0,11	0,396536	**0,32**	0,379031	**0,53**	0,346668	**0,74**	0,303389	0,95	0,254059
0,12	0,39608	**0,33**	0,377801	**0,54**	0,344818	**0,75**	0,301137	0,96	0,251644
0,13	0,395585	**0,34**	0,376537	**0,55**	0,342944	**0,76**	0,298872	0,97	0,249228
0,14	0,395052	**0,35**	0,37524	**0,56**	0,341046	**0,77**	0,296595	0,98	0,246809
0,15	0,394479	**0,36**	0,373911	**0,57**	0,339124	**0,78**	0,294305	0,99	0,24439
0,16	0,393868	**0,37**	0,372548	**0,58**	0,33718	**0,79**	0,292004	1	0,241971
0,17	0,393219	**0,38**	0,371154	**0,59**	0,335213	**0,8**	0,289692	1,01	0,239551
0,18	0,392531	**0,39**	0,369728	**0,6**	0,333225	**0,81**	0,287369	1,02	0,237132
0,19	0,391806	**0,4**	0,36827	**0,61**	0,331215	**0,82**	0,285036	1,03	0,234714
0,2	0,391043	**0,41**	0,366782	**0,62**	0,329184	**0,83**	0,282694	1,04	0,232297
1,05	0,229882	**1,26**	0,180371	**1,47**	0,135418	**1,68**	0,097282	**1,89**	0,066871
1,06	0,22747	**1,27**	0,178104	**1,48**	0,133435	**1,69**	0,095657	**1,9**	0,065616
1,07	0,22506	**1,28**	0,175847	**1,49**	0,131468	**1,7**	0,094049	**1,91**	0,064378

Продолжение таблицы 1

x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$
1,08	0,222653	**1,29**	0,173602	**1,5**	0,129518	**1,71**	0,092459	**1,92**	0,063157
1,09	0,220251	**1,3**	0,171369	**1,51**	0,127583	**1,72**	0,090887	**1,93**	0,061952
1,1	0,217852	**1,31**	0,169147	**1,52**	0,125665	**1,73**	0,089333	**1,94**	0,060765
1,11	0,215458	**1,32**	0,166937	**1,53**	0,123763	**1,74**	0,087796	**1,95**	0,059595
1,12	0,213069	**1,33**	0,16474	**1,54**	0,121878	**1,75**	0,086277	**1,96**	0,058441
1,13	0,210686	**1,34**	0,162555	**1,55**	0,120009	**1,76**	0,084776	**1,97**	0,057304
1,14	0,208308	**1,35**	0,160383	**1,56**	0,118157	**1,77**	0,083293	**1,98**	0,056183
1,15	0,205936	**1,36**	0,158225	**1,57**	0,116323	**1,78**	0,081828	**1,99**	0,055079
1,16	0,203571	**1,37**	0,15608	**1,58**	0,114505	**1,79**	0,08038	**2**	0,053991
1,17	0,201214	**1,38**	0,153948	**1,59**	0,112704	**1,8**	0,07895	**2,01**	0,052919
1,18	0,198863	**1,39**	0,151831	**1,6**	0,110921	**1,81**	0,077538	**2,02**	0,051864
1,19	0,19652	**1,4**	0,149727	**1,61**	0,109155	**1,82**	0,076143	**2,03**	0,050824
1,2	0,194186	**1,41**	0,147639	**1,62**	0,107406	**1,83**	0,074766	**2,04**	0,0498
1,21	0,19186	**1,42**	0,145564	**1,63**	0,105675	**1,84**	0,073407	**2,05**	0,048792
1,22	0,189543	**1,43**	0,143505	**1,64**	0,103961	**1,85**	0,072065	**2,06**	0,0478
1,23	0,187235	**1,44**	0,14146	**1,65**	0,102265	**1,86**	0,07074	**2,07**	0,046823
1,24	0,184937	**1,45**	0,139431	**1,66**	0,100586	**1,87**	0,069433	**2,08**	0,045861
1,25	0,182649	**1,46**	0,137417	**1,67**	0,098925	**1,88**	0,068144	**2,09**	0,044915

Приложение 2 (табл. 2)

Таблица значений нормированной нормальной функции распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

x	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,5000	0,5039	0,5079	0,5119	0,5159	0,5199	0,5239	0,5279	0,5318	0,5358
0,1	0,5398	0,5438	0,5477	0,5517	0,5556	0,5596	0,5635	0,5674	0,5714	0,5753
0,2	0,5792	0,5831	0,5870	0,5909	0,5948	0,5987	0,6025	0,6064	0,6102	0,6140
0,3	0,6179	0,6217	0,6255	0,6293	0,6330	0,6368	0,6405	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6627	0,6664	0,6700	0,6736	0,6772	0,6808	0,6843	0,6879
0,5	0,6914	0,6949	0,6984	0,7019	0,7054	0,7088	0,7122	0,7156	0,7190	0,7224
0,6	0,7257	0,7290	0,7323	0,7356	0,7389	0,7421	0,7453	0,7485	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7733	0,7763	0,7793	0,7823	0,7852
0,8	0,7881	0,7910	0,7938	0,7967	0,7995	0,8023	0,8051	0,8078	0,8105	0,8132
0,9	0,8159	0,8185	0,8212	0,8238	0,8263	0,8289	0,8314	0,8339	0,8364	0,8389
1,0	0,8413	0,8437	0,8461	0,8485	0,8508	0,8531	0,8554	0,8576	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8707	0,8728	0,8749	0,8769	0,8790	0,8810	0,8829
1,2	0,8849	0,8868	0,8887	0,8906	0,8925	0,8943	0,8961	0,8979	0,8997	0,9014
1,3	0,9032	0,9049	0,9065	0,9082	0,9098	0,9114	0,9130	0,9146	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9250	0,9264	0,9278	0,9292	0,9305	0,9318
1,5	0,9331	0,9344	0,9357	0,9369	0,9382	0,9394	0,9406	0,9417	0,9429	0,9440
1,6	0,9452	0,9463	0,9473	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9544
1,7	0,9554	0,9563	0,9572	0,9581	0,9590	0,9599	0,9608	0,9616	0,9624	0,9632
1,8	0,9640	0,9648	0,9656	0,9663	0,9671	0,9678	0,9685	0,9692	0,9699	0,9706
1,9	0,9712	0,9719	0,9725	0,9732	0,9738	0,9744	0,9750	0,9755	0,9761	0,9767
2,0	0,9772	0,9777	0,9783	0,9788	0,9793	0,9798	0,9803	0,9807	0,9812	0,9816
2,1	0,9821	0,9825	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9853	0,9857
2,2	0,9861	0,9864	0,9867	0,9871	0,9874	0,9877	0,9880	0,9884	0,9887	0,9889
2,3	0,9892	0,9895	0,9898	0,9901	0,9903	0,9906	0,9908	0,9911	0,9313	0,9915
2,4	0,9918	0,9920	0,9922	0,9924	0,9926	0,9928	0,9930	0,9932	0,9934	0,9936
2,5	0,9937	0,9939	0,9941	0,9943	0,9944	0,9946	0,9947	0,9949	0,9950	0,9952
2,6	0,9953	0,9954	0,9956	0,9957	0,9958	0,9959	0,9960	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9972	0,9973
2,8	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9978	0,9979	0,9980	0,9980
x	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**

Продолжение таблицы 2

2,9	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,0	0,9986	0,9986	09887	0,9987	0,9988	0,9988	0,9988	0,9989	0,9989	0,9990
3,1	0,9990	0,9990	0,9991	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992
3,2	0,9993	0,9993	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995
3,3	0,9995	0,9995	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996
3,4	0,9996	0,9996	0,9996	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997
3,5	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,7	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Приложение 3 (табл. 3)

Таблица значений функции Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}} dz$$

x	Сотые доли									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,0000	0,0040	0,0000	0,0120	0,0160	0,0199	0,0239	0,0278	0,0319	0,0359
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	2703	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	3588	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	4032	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	4393	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	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4928	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

Продолжение таблицы 3

x	Сотые доли									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4982	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	4492	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	4947	4997	4997	4997	4998
3,5	4998	4998	4998	4998	4998	4998	4998	4998	4998	4998

Приложение 4 (табл. 4)

Таблица значений вероятностей *Pn*(*m*) (закон Пуассона)

$$P_n(m) = \frac{\lambda^m}{m!}e^{-\lambda}$$

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.0905	0.1637	0.2223	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659	0.3679
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1216	0.1438	0.1647	0.1839
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494	0.0613
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111	0.0153
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	0.0007	0.0012	0.0020	0.0031
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	0.0005
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

Продолжение таблицы 4.

λ	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
0	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001	0.0001
1	0.2707	0.1494	0.0733	0.0337	0.0149	0.0064	0.0027	0.0011	0.0005
2	0.2707	0.2240	0.1465	0.0842	0.0446	0.0223	0.0107	0.0050	0.0023
3	0.1805	0.2240	0.1954	0.1404	0.0892	0.0521	0.0286	0.0150	0.0076
4	0.0902	0.1681	0.1954	0.1755	0.1339	0.0912	0.0572	0.0337	0.0189
5	0.0361	0.1008	0.1563	0.1755	0.1606	0.1277	0.0916	0.0607	0.0378
6	0.0120	0.0504	0.1042	0.1462	0.1606	0.1490	0.1221	0.0911	0.0631
7	0.0034	0.0216	0.0595	0.1045	0.1377	0.1490	0.1396	0.1171	0.0901

Продолжение таблицы 4.

λ	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
8	0.0009	0.0081	0.0298	0.0653	0.1033	0.1304	0.1396	0.1318	0.1126
9	0.0002	0.0027	0.0132	0.0363	0.0689	0.1014	0.1241	0.1318	0.1251
10	0.0000	0.0008	0.0053	0.0181	0.0413	0.0710	0.0993	0.1186	0.1251
11	0.0000	0.0002	0.0019	0.0082	0.0225	0.0452	0.0722	0.0970	0.1137
12	0.0000	0.0001	0.0006	0.0034	0.0113	0.0264	0.0481	0.0728	0.0948
13	0.0000	0.0000	0.0002	0.0013	0.0052	0.0142	0.0296	0.0504	0.0729
14	0.0000	0.0000	0.0001	0.0005	0.0022	0.0071	0.0169	0.0324	0.0521

Продолжение таблицы 4

λ	**2.0**	**3.0**	**4.0**	**5.0**	**6.0**	**7.0**	**8.0**	**9.0**	**10.0**
15	0.0000	0.0000	0.0000	0.0002	0.0009	0.0033	0.0090	0.0194	0.0347
16	0.0000	0.0000	0.0000	0.0000	0.0003	0.0015	0.0045	0.0109	0.0217
17	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0021	0.0058	0.0128
18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009	0.0029	0.0071
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014	0.0037
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0006	0.0019
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009
22	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004
23	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
24	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
25	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Приложение 5 (табл.5)

Таблица значений $t_{\gamma,\kappa}$ – критерия Стьюдента

γ	Вероятность γ									
	0,4	**0,25**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,0025**	**0,001**	**0,0005**
1	0,3249	1,0000	3,0777	6,3138	12,706	31,820	63,656	127,321	318,308	636,619
2	0,2887	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	14,089	22,327	31,599
3	0,2767	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	10,214	12,924
4	0,2707	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	7,1732	8,6103
5	0,2672	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8934	6,8688
6	0,2648	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588
7	0,2632	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	4,7853	5,4079
8	0,2619	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0413
9	0,2610	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2968	4,7809
10	0,2602	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869
11	0,2596	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370
12	0,2590	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178
13	0,2586	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2208
14	0,2582	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7874	4,1405
15	0,2579	0,6912	1,3406	1,7530	2,1314	2,6025	2,9467	3,2860	3,7328	4,0728
16	0,2576	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150
17	0,2573	0,6892	1,3334	1,7396	2,1098	2,8982	2,8982	3,2224	3,6458	3,9651
18	0,2571	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9216
19	0,2569	0,6876	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,5794	3,8834
20	0,2567	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8495
21	0,2566	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193
22	0,2564	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921
23	0,2563	0,6851	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676
24	0,2562	0,6348	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,4668	3,7454
25	0,2561	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251
26	0,2560	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066
27	0,2559	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896
28	0,2558	0,6834	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,4081	3,6739
29	0,2557	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3962	3,6594
30	0,2556	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460
40	0,2550	0,6807	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510
50	0,2547	0,6794	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,2614	3,4960

Продолжение таблицы 5

γ	Вероятность γ									
	0,4	**0,25**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,0025**	**0,001**	**0,0005**
60	0,2545	0,6786	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,2317	3,4602
70	0,2543	0,6780	1,2938	1,6669	2,9944	2,3808	2,6479	2,8987	3,2108	3,4350
80	0,2542	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163
90	0,2541	0,6772	1,2910	1,6620	2,9867	2,3685	2,6316	2,8779	3,1833	3,4019
100	0,2540	0,6770	1,2901	1,6602	2,9840	2,3642	2,6259	2,8707	3,1737	3,3905
200	0,2537	0,6757	1,2858	1,6525	2,9719	2,3451	2,6006	2,8385	3,1315	3,3398
300	0,2536	0,6753	1,2844	1,6499	2,9679	2,3388	2,5923	2,8279	3,1176	3,3233
500	0,2535	0,6750	1,2832	1,6479	1,9647	2,3338	2,5857	2,8195	3,1066	3,3101

Приложение 6 (табл. 6)

Таблица значений $q = q(\gamma, n)$ для определения доверительного интервала σ

n	γ			n	γ		
	0,95	**0,99**	**0,999**		**0,95**	**0,99**	**0,999**
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,2	2,06	40	0,24	0,35	0,50
10	0,6	1,08	1,80	45	0,22	0,32	0,46
11	0,58	0,98	1,60	50	0.21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,270
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

Приложение 7 (табл. 7)

Таблица значений $\chi^2_{\alpha\kappa}$ критерия Пирсона

k	Уровень значимости α										
	99	**98**	**95**	**90**	**80**	**70**	**50**	**30**	**20**	**10**	**5**
1	0,0002	0,001	0,004	0,016	0,064	0,148	0,455	1,07	1,642	2,706	3,841
2	0,020	0,040	0,103	0,211	0,446	0,71	1,39	2,41	3,219	4,605	5,991
3	0,115	0,185	0,352	0,584	1,01	1,42	2,37	3,66	4,642	6,251	7,815
4	0,297	0,429	0,711	1,06	1,65	2,19	3,36	4,88	5,989	7,779	9,488
5	0,554	0,752	1,15	1,61	2,34	3,00	4,35	6,06	7,289	9,236	11,07
6	0,872	1,13	1,64	2,20	3,07	3,83	5,35	7,23	8,558	10,64	12,59
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,38	9,803	12,02	14,07
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51
9	2,09	2,53	3,33	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92
10	2,56	3,06	3,94	4,87	6,18	7,27	9,34	11,78	13,44	15,99	18,31
11	3,05	3,61	4,57	5,58	6,99	8,15	10,34	12,90	14,63	17,28	19,68

Продолжение табл. 7.

k	Уровень значимости α										
	99	**98**	**95**	**90**	**80**	**70**	**50**	**30**	**20**	**10**	**5**
12	3,57	4,18	5,23	6,30	7,81	9,03	11,34	14,01	15,81	18,55	21,03
13	4,11	4,77	5,89	7,04	8,63	9,93	12,34	15,12	16,98	19,81	22,36
14	4,66	5,37	6,57	7,79	9,47	10,82	13,34	16,22	18,15	21,06	23,68
15	5,23	5,98	7,26	8,55	10,31	11,72	14,34	17,32	19,31	22,31	25,00
16	5,81	6,61	7,96	9,31	11,15	12,62	15,34	18,42	20,47	23,54	26,30
17	6,41	7,26	8,67	10,09	12,00	13,53	16,34	19,51	21,61	24,77	27,59
18	7,01	7,91	9,39	10,86	12,86	14,44	17,34	20,60	22,76	25,99	28,87
19	7,63	8,57	10,12	11,65	13,72	15,35	18,34	21,69	23,90	27,20	30,14
20	8,26	9,24	10,85	12,44	14,58	16,27	19,34	22,77	25,04	28,41	31,41
21	8,90	9,91	11,59	13,24	15,44	17,18	20,34	23,86	26,17	29,62	32,67
22	9,54	10,60	12,34	14,04	16,31	18,10	21,34	24,94	27,30	30,81	33,92
23	10,20	11,29	13,09	14,85	17,19	19,02	22,34	26,02	28,43	32,01	35,17
24	10,86	11,99	13,85	15,66	18,06	19,94	23,34	27,10	29,55	33,20	36,42
25	11,52	12,70	14,61	16,47	18,94	20,87	24,34	28,17	30,68	34,38	37,65
26	12,20	13,41	15,38	17,29	19,82	21,79	25,34	29,25	31,79	35,56	38,89
27	12,88	14,13	16,15	18,11	20,70	22,72	26,34	30,32	32,91	36,74	40,11
28	13,56	14,85	16,93	18,94	21,59	23,65	27,34	31,39	34,03	37,92	41,34
29	14,26	15,57	17,71	19,77	22,48	24,58	28,34	32,46	35,14	39,09	42,56
30	14,95	16,31	18,49	20,60	23,36	25,51	29,34	33,53	36,25	40,26	43,77

Продолжение табл. 7.

k	Уровень значимости α							
	2	**1**	**0,5**	**0,2**	**0,1**	**0,05**	**0,02**	**0,01**
1	5,412	6,635	7,879	9,550	10,83	3,84	5,41	6,64
2	7,824	9,210	10,60	12,43	13,82	5,99	7,82	9,21
3	9,837	11,34	12,84	14,80	16,27	7,82	9,84	11,30
4	11,67	13,28	14,86	16,92	18,47	9,49	11,70	13,30
5	13,39	15,09	16,75	18,91	20,52	11,10	13,40	15,10
6	15,03	16,81	18,55	20,79	22,46	12,60	15,00	16,80
7	16,62	18,48	20,28	22,60	24,32	14,10	16,60	18,50
8	18,17	20,09	21,95	24,35	26,12	15,50	18,20	20,10
9	19,68	21,67	23,59	26,06	27,88	16,90	19,70	21,70
10	21,16	23,21	25,19	27,72	29,59	18,30	21,20	23,20

Таблица значений $F_{\alpha,k1,k2}$ критерия Фишера-Снедекора

(5%-ные значения – верхняя цифра, 1%-ные – нижняя цифра)

k_2 – степени свободы для меньшей дисперсии

k_2 – степени свободы для меньшей дисперсии

k_2*	k_1 – степени свободы для большей дисперсии											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	161 4052	200 4999	216 5403	225 5625	230 5764	234 5889	237 5928	239 5981	241 6022	242 6056	243 6082	244 6106
2	18,51 98,49	19,00 99,01	19,16 99,17	19,25 99,25	19,30 99,30	19,33 99,33	19,36 99,34	19,37 99,36	19,38 99,38	19,39 99,40	19,40 99,41	19,41 99,42
3	10,13 32,12	9,55 30,81	9,28 29,46	912 28,71	9,01 28,24	8,94 27,91	8,88 27,67	8,84 27,49	8,81 27,34	8,78 27,23	8,76 27,13	8,74 27,05
4	7,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,26 15,52	6,16 15,21	6,09 14,98	6,04 14,80	6,00 14,66	5,96 14,54	5,93 1445	5,91 1447
5	6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	5,05 10,97	4,95 10,67	4,88 10,45	4,82 10,27	4,78 10,15	4,74 10,05	4,70 9,96	4,68 9,89
6	5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,39 8,75	4,28 8,47	4,21 8,26	4,15 8,10	4,10 7,98	4,06 7,87	4,03 7,79	4,00 7,72
7	5,59 12,25	4,74 9,55	4,35 8,45	4,12 7,85	3,97 7,46	3,87 7,19	3,79 7,00	3,73 6,84	3,68 6,71	3,63 6,62	360 6,54	3,57 6,47
8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,69 6,63	3,58 6,37	3,50 6,19	3,44 6,03	3,39 5,91	3,31 5,82	3,31 5,74	3,28 5,67
9	5,12 10,56	4,26 8,02	3,86 6,99	3,63 6,42	3,48 6,06	3,37 5,80	3,29 5,62	3,23 5,47	3,18 5,35	3,13 5,26	3,10 5,18	3,07 5,11
10	4,96 10,04	4,10 7,56	3,71 6,55	3,48 5,99	3,33 5,64	3,22 5,39	3,14 5,21	3,07 5,06	3,02 4,95	2,97 4,85	2,94 4,78	2,91 4,71
11	4,84 9,85	3,98 7,20	3,59 6,22	3,36 5,67	3,20 5,32	3,09 5,07	3,01 4,88	2,95 4,74	290 4,63	2,86 4,54	2,82 4,46	2,79 4,40
12	4,75 9,33	3,88 6,93	3,49 5,95	3,26 5,41	3,11 5,06	3,00 4,82	2,92 4,65	2,85 4,50	2,80 4,39	2,76 4,30	2,72 4,22	2,69 4,16
13	4,67 9,07	3,80 6,70	3,41 5,74	3,18 5,20	3,02 4,86	2,92 4,62	2,84 4,44	2,77 4,30	2,72 4,19	2,67 4,10	2,63 4,02	2,60 3,96
14	4,60 8,86	3,74 6,51	3,34 5,56	3,11 5,03	2,96 4,69	2,85 4,46	2,77 4,28	2,70 4,14	2,65 4,03	2,60 3,94	2,56 3,86	2,53 3,80
15	4,54 8,68	3,68 6,36	3,29 5,42	3,06 4,89	2,90 4,56	2,79 4,32	2,70 4,14	2,64 4,00	2,59 3,89	2,55 3,80	2,51 3,73	2,48 3,67
16	4,49 8,53	3,63 6,23	3,24 5,29	3,01 4,77	2,85 4,44	2,74 4,20	2,66 4,03	2,59 3,89	2,54 3,78	2,49 3,69	2,45 3,61	2,42 3,55
17	445 8,40	3,59 6,11	3,20 5,18	2,96 4,67	2,81 4,34	2,70 4,10	2,62 3,93	2,55 3,79	2,50 3,68	2,45 3,59	2,41 3,52	2,38 3,45
18	4,41 8,28	3,55 6,01	3,16 5,09	2,93 4,58	2,77 4,25	2,66 4,01	2,58 3,85	2,51 3,71	2,46 3,60	2,41 3,51	2,37 3,44	2,34 3,37
19	4,38 8,18	3,52 5,93	3,13 5,01	2,90 4,50	2,74 4,17	2,63 3,94	2,55 3,77	2,48 3,63	2,43 352	2,38 3,43	2,34 3,36	2,31 3,30
20	4,35 8,10	3,49 5,85	3,10 4,94	2,87 4,43	2,71 4,10	2,60 3,87	2,52 3,71	2,45 3,56	2,40 3,45	2,35 3,37	2,31 3,30	2,28 3,23
21	4,32 8,02	3,47 5,78	3,07 4,87	2,84 4,37	2,68 4,04	2,57 3,81	2,49 3,65	2,42 3,51	2,37 3,40	2,32 3,31	2,28 3,24	2,25 3,17
22	4,30 7,94	3,44 5,72	3,05 4,82	2,82 4,31	2,66 3,99	2,55 3,76	2,47 3,59	2,40 3,45	2,35 3,35	2,30 3,26	2,26 3,18	2,23 3,12
23	4,28 7,88	3,42 5,66	3,03 4,76	2,80 4,26	2,64 3,94	2,53 3,71	2,45 3,54	2,38 3,41	2,32 3,30	2,28 3,21	2,24 3,14	2,20 3,07
24	4,26 7,82	3,40 5,61	3,01 4,72	2,78 4,22	2,62 3,90	2,51 3,67	2,43 3,50	2,36 3,36	2,30 3,25	2,26 3,17	2,22 3,09	2,18 3,03
25	4,24 7,77	3,38 5,57	2,99 4,68	2,76 4,18	2,60 3,86	2,49 3,63	2,41 3,46	2,34 3,32	2,28 3,21	2,24 3,13	2,20 3,05	2,16 2,99

Приложение 8 (табл. 8)

k_1 – степени свободы для большей дисперсии												k_2	k_2 – степени свободы для меньшей дисперсии
14	**16**	**20**	**24**	**30**	**40**	**50**	**75**	**100**	**200**	**500**	**1000**		
245 6142	246 6169	248 6208	249 6234	250 6258	251 6286	252 6302	253 6323	253 6334	254 6352	254 6361	254 6366	**1**	
19,42 99,43	19,43 99,44	19,44 99,45	9,45 99,46	19,46 99,47	19,47 99,48	19,47 99,48	19,48 9,49	19,49 99,49	19,49 99,49	1950 99,50	19,50 99,50	**2**	
8,71 26,39	8,69 26,83	8,66 26,69	8,64 26,60	8,62 26,50	8,60 26,41	8,58 26,35	8,57 26,27	8,56 26,23	8,54 26,18	8,54 26,14	8,53 26,12	**3**	
5,87 14,24	5,84 14,15	5,80 14,02	5,77 13,93	5,74 13,83	5,71 13,74	5,70 13,69	5,68 13,61	5,66 13,57	5,65 13,52	5,64 13,48	5,63 13,46	**4**	
4,64 9,77	4,60 9,68	4,56 9,55	4,53 9,47	4,50 9,38	4,46 9,29	4,44 9,24	4,42 9,17	4,40 9,13	4,38 9,07	4,37 9,04	4,36 9,02	**5**	
3,96 7,60	3,92 7,52	3,87 7,39	3,84 7,31	3,81 7,23	3,77 7,14	3,75 7,09	372 7,02	3,71 6,99	3,69 6,94	3,68 6,90	3,67 6,88	**6**	
3,52 6,35	3,49 6,27	3,44 6,15	3,41 6,07	3,38 5,98	3,34 590	3,32 5,85	3,29 5,78	3,28 5,75	3,25 5,70	3,24 5,67	3,23 5,65	**7**	
3,23 5,56	320 5,48	3,15 5,36	3,12 5,28	3,08 5,20	3,05 5,11	3,03 5,06	3,00 5,00	2,98 4,96	2,96 4,91	2,94 4,88	2,93 4,86	**8**	
3,02 5,00	2,98 492	2,93 4,80	2,90 4,73	2,86 464	2,82 4,56	2,80 451	2,77 4,45	2,76 4,41	2,73 4,36	2,72 4,33	2,71 4,31	**9**	
2,86 4,60	2,82 4,52	2,77 4,41	2,74 4,33	2,70 4,25	2,67 4,17	2,64 4,12	2,61 4,05	2,59 4,01	2,56 3,96	2,55 3,93	2,54 3,91	**10**	
2,74 4,29	2,70 4,21	2,65 4,10	2,61 4,02	2,57 3,94	2,53 3,86	2,50 3,80	2,47 3,74	2,45 3,70	2,42 3,66	2,41 362	2,40 3,60	**11**	
2,64 4,05	2,60 3,98	2,54 3,86	2,50 3,78	2,46 3,70	2,42 3,61	2,40 3,56	2,36 3,49	2,35 3,46	2,32 3,41	2,31 3,38	2,30 3,36	**12**	
2,55 3,85	2,51 3,78	2,46 3,67	2,42 3,59	2,38 3,51	2,34 3,42	2,32 3,37	2,28 3,30	2,26 3,27	2,24 3,21	2,22 3,18	2,21 3,16	**13**	
2,48 3,70	2,44 3,62	2,39 3,51	2,35 3,43	2,31 3,34	2,27 3,26	2,24 3,21	2,21 3,14	2,19 3,11	2,16 3,06	2,14 3,02	2,13 3,00	**14**	
2,43 3,56	2,39 3,48	2,33 3,36	2,29 3,29	2,25 3,20	2,21 3,12	2,18 3,07	2,15 3,00	2,12 2,97	2,10 2,92	2,08 2,89	2,07 2,87	**15**	
2,37 3,45	2,33 3,37	2,23 3,25	2,24 3,18	2,20 3,10	2,16 3,01	2,13 2,96	2,09 2,89	2,07 2,86	2,04 2,80	2,02 2,77	2,01 2,75	**16**	
2,33 3,35	2,29 3,27	2,23 3,16	2,19 3,08	2,15 3,00	2,11 2,92	2,08 2,86	2,04 2,79	2,02 2,76	1,99 2,70	1,97 2,67	1,96 2,65	**17**	
2,29 3,27	2,25 3,19	219 3,07	2,15 3,00	2,11 2,91	2,07 2,83	2,04 2,78	2,00 2,71	1,98 2,68	1,95 2,62	1,93 2,59	1,92 2,57	**18**	
2,26 3,19	2,21 3,12	2,15 3,00	2,11 2,92	2,07 2,84	2,02 2,76	2,00 2,70	1,96 2,63	1,94 2,60	1,91 2,54	1,90 2,51	1,88 2,49	**19**	
2,23 3,13	2,18 3,05	2,12 2,94	2,08 2,86	2,04 2,77	1,99 2,69	1,96 2,63	1,92 2,56	1,90 2,53	1,87 2,47	1,85 2,44	1,84 2,42	**20**	
2,20 3,07	2,15 2,99	2,09 2,88	2,05 2,80	2,00 2,72	1,96 2,63	1,93 258	1,89 2,51	1,87 2,47	1,84 2,42	1,82 2,38	1,81 2,36	**21**	
2,18 3,02	2,13 2,94	2,07 2,83	2,03 2,75	1,98 2,67	1,93 2,58	1,91 2,53	1,87 2,46	1,84 2,42	1,81 2,37	1,80 2,33	1,78 2,31	**22**	
2,14 2,97	2,10 2,89	2,04 2,78	2,00 2,70	1,96 2,62	1,91 2,53	1,88 2,48	1,84 2,41	1,82 2,37	1,79 2,32	1,77 2,28	1,76 2,26	**23**	
2,13 2,93	2,09 2,85	2,02 2,74	1,98 2,66	1,94 2,58	1,89 2,49	1,86 2,44	1,82 2,36	1,80 2,33	1,76 2,27	1,74 2,23	1,73 2,21	**24**	
2,11 2,89	2,06 2,81	2,00 2,70	1,96 2,62	1,92 2,54	1,87 2,45	1,84 2,40	1,80 2,32	1,77 2,29	1,74 2,23	1,72 2,19	1,71 2,17	**25**	

Продолжение табл. 8

k_2 – степени свободы для меньшей дисперсии

k_2	k_1 – степени свободы для большей дисперсии											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
26	4,22 7,72	3,37 5,53	2,98 4,64	2,74 4,14	2,59 3,82	2,47 3,59	2,39 3,42	2,32 3,29	2,27 3,17	2,22 3,09	2,18 3,02	2,15 2,96
27	4,21 7,68	3,35 5,49	2,96 4,60	2,73 4,11	2,57 3,79	2,46 3,56	2,37 3,39	2,30 3,26	2,25 3,14	2,20 3,06	2,16 2,98	2,13 2,93
28	4,20 7,64	3,34 5,45	2,95 4,57	2,71 4,07	2,56 3,76	2,44 3,53	2,36 3,36	2,29 3,23	2,24 3,11	2,19 3,03	2,15 2,95	2,12 2,90
29	4,18 7,60	3,33 5,42	2,93 4,54	2,70 4,04	2,54 3,73	2,43 3,50	2,35 3,33	2,28 3,20	2,22 3,08	2,18 3,00	2,14 2,92	2,10 2,87
30	4,17 7,56	3,32 5,39	2,92 4,51	2,69 4,02	2,53 3,70	2,42 3,47	234 3,30	2,27 3,17	2,21 3,06	2,16 2,98	2,12 2,90	2,09 2,84
32	4,15 7,50	3,30 5,34	2,90 4,46	2,67 3,97	2,51 3,66	2,40 342	2,32 3,25	2,25 3,12	2,19 3,01	2,14 2,94	2,10 2,86	2,07 2,80
34	4,13 7,44	3,28 5,29	2,88 4,42	2,65 3,93	2,49 3,61	2,38 3,38	2,30 3,21	2,23 3,08	2,17 2,97	2,12 2,89	2,08 2,82	2,05 276
36	4,11 7,39	3,26 5,25	286 4,38	2,63 3,89	2,48 3,58	2,36 3,35	2,28 3,18	2,21 3,04	2,15 2,94	2,10 2,86	2,06 3,78	2,03 2,72
38	4,10 7,35	3,25 5,21	2,85 4,34	2,65 3,86	2,46 3,54	2,35 3,32	2,26 3,15	219 3,02	2,14 2,91	2,09 2,82	2,05 2,75	2,02 2,69
40	4,08 7,31	3,23 5,18	2,84 4,31	2,61 3,83	2,45 3,51	2,34 3,29	2,25 3,12	2,18 2,99	2,12 2,88	2,07 2,80	2,04 2,73	2,00 2,66
42	4,07 7,27	3,22 5,15	2,83 4,29	2,59 3,80	2,44 3,49	2,32 3,26	2,24 3,10	2,17 2,96	2,11 2,86	2,06 2,77	2,02 2,70	1,90 2,64
44	4,06 7,24	3,21 5,12	2,82 4,26	2,58 3,78	2,43 3,46	2,31 3,24	2,23 3,07	2,16 2,94	2,10 2,84	2,05 2,75	2,01 2,68	1,98 2,62
46	4,05 7,21	3,20 5,10	2,81 4,24	2,57 3,76	2,42 3,44	2,30 3,22	2,22 3,05	2,14 2,92	2,09 2,82	2,04 273	200 2,66	1,97 2,60
48	4,04 7,19	3,19 5,08	2,80 4,22	256 3,74	2,41 3,42	2,30 3,20	2,21 3,04	2,14 2,90	2,08 2,80	2,03 2,71	1,99 2,64	1,96 2,58
50	4,03 7,17	3,18 5,06	2,79 4,20	2,56 3,72	2,40 3,41	2,29 3,18	2,20 3,02	2,13 2,88	2,07 2,78	2,02 2,70	1,98 2,62	1,95 2,56
55	4,02 7,12	3,17 5,01	2,78 4,16	2,54 3,68	2,38 3,37	2,27 3,15	2,18 2,98	2,11 2,85	2,05 2,75	2,00 2,66	1,97 2,59	1,93 2,53
60	4,00 7,08	3,15 4,98	2,76 4,13	2,52 3,65	2,37 3,34	2,25 3,12	2,17 2,95	2,10 2,82	2,04 2,72	1,99 2,63	1,95 2,56	1,92 2,50
65	3,99 7,04	3,14 4,95	2,75 4,10	2,51 3,62	2,36 3,31	2,24 3,09	2,15 2,93	2,08 2,79	2,02 2,70	1,98 2,61	1,94 2,54	1,90 2,47
70	3,98 7,01	3,13 4,92	2,74 4,08	2,50 3,60	2,35 3,29	2,23 3,07	2,14 2,91	2,07 2,77	2,01 2,67	1,97 2,59	1,93 2,51	1,89 2,45
80	3,96 6,96	3,11 4,88	2,72 404	2,48 3,56	2,33 3,25	2,21 3,04	2,12 2,87	2,05 2,74	1,99 2,64	1,95 2,55	1,91 2,48	1,89 2,41
100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,30 3,20	2,19 2,99	2,10 2,82	2,03 2,69	1,97 2,59	1,92 2,51	1,88 2,43	1,85 2,36
125	3,92 6,84	3,07 4,78	2,68 3,94	2,44 3,47	2,29 317	2,17 2,95	2,08 2,79	2,01 2,65	1,95 2,56	1,90 2,47	1,86 2,40	1,83 2,33
150	3,91 6,81	3,06 4,75	2,67 3,91	2,43 3,44	2,27 3,14	2,16 2,92	2,07 2,76	2,00 2,62	1,94 2,53	1,89 2,44	1,85 2,37	1,82 2,30
200	3,89 6,76	3,04 4,71	2,65 3,88	2,41 3,41	2,26 3,11	2,14 2,90	2,05 2,73	1,98 2,60	1,92 2,50	1,87 241	1,83 2,34	1,80 2,28
400	3,86 6,70	3,02 4,66	2,62 3,83	2,39 3,36	2,23 3,06	2,12 2,85	2,03 2,69	1,96 2,55	1,90 2,46	1,85 2,37	1,81 2,29	1,78 2,23
1000	3,85 6,66	3,00 4,62	2,61 3,80	2,38 3,34	2,22 3,04	2,10 2,82	2,02 2,66	1,95 2,53	1,89 2,43	1,84 2,34	1,80 2,26	1,76 2,20
∞	3,84 6,64	2,99 4,60	2,60 3,78	2,37 3,32	2,21 3,02	2,09 2,80	2,01 2,64	1,94 251	1,88 2,41	1,83 2,32	1,79 2,24	1,75 2,18

Продолжение табл. 8

k_1 – степени свободы для большей дисперсии												k_2 – степени свободы для меньшей дисперсии
14	**16**	**20**	**24**	**30**	**40**	**50**	**75**	**100**	**200**	**500**	**1000**	k_2
2,10 2,86	205 2,77	1,99 2,66	1,95 2,58	1,90 2,50	1,85 241	1,82 2,36	1,78 2,28	1,76 2,25	1,72 2,19	1,70 2,15	1,69 2,13	**26**
2,08 2,83	2,03 2,74	1,97 2,63	1,93 2,55	1,88 2,47	1,84 2,38	1,80 2,33	1,76 2,25	1,74 2,21	1,71 2,16	1,68 2,12	1,67 2,10	**27**
2,06 2,80	2,02 2,71	1,96 2,60	1,91 2,52	1,87 2,44	1,81 2,35	1,78 2,30	1,75 2,22	1,72 2,18	1,69 2,13	1,67 2,09	1,65 2,06	**28**
2,05 2,77	2,00 2,68	1,94 2,57	1,90 2,49	1,85 2,41	1,80 2,32	1,77 2,27	1,73 2,19	1,71 215	1,68 2,10	1,65 2,06	1,64 2,03	**29**
2,04 2,74	1,99 2,66	1,93 2,55	1,89 2,47	1,84 2,38	1,79 229	1,76 2,24	1,72 2,16	1,69 213	1,66 207	1,64 2,03	1,62 2,01	**30**
2,02 2,70	1,97 2,62	1,91 2,51	1,86 2,42	1,82 2,34	1,76 2,25	1,74 2,20	1,69 2,12	1,67 2,08	1,64 2,02	1,61 1,98	1,59 1,96	**32**
2,00 2,66	1,95 2,58	1,89 2,47	1,84 2,38	1,80 2,30	1,74 2,21	1,71 2,15	1,67 2,08	1,64 2,04	1,61 1,98	1,59 1,94	1,57 1,91	**34**
1,98 2,62	1,93 2,54	1,87 2,43	1,82 2,35	1,78 2,26	1,72 2,17	1,69 2,12	1,65 2,04	1,62 2,00	1,59 1,94	1,56 1,90	1,55 1,87	**36**
1,96 2,59	1,92 2,51	1,85 2,40	1,80 2,32	1,76 2,22	1,71 2,14	1,67 2,08	1,63 2,00	1,60 1,97	1,57 1,90	1,54 1,86	1,53 1,84	**38**
1,95 2,56	1,90 2,49	1,84 2,37	1,79 2,29	1,74 2,20	1,69 2,11	1,66 2,05	1,61 1,97	1,59 1,94	1,55 1,88	1,53 1,84	1,51 1,81	**40**
1,94 2,54	1,89 2,46	1,82 2,35	1,78 2,26	1,73 2,17	1,68 2,08	1,64 2,02	1,60 1,94	1,57 1,91	1,54 1,85	1,51 1,80	1,49 1,78	**42**
1,92 2,52	1,88 2,44	1,81 2,32	1,76 2,24	1,72 2,15	1,66 2,06	1,63 2,00	1,58 1,92	1,56 1,88	1,52 1,82	1,50 1,78	1,48 1,75	**44**
1,91 2,50	1,87 242	1,80 2,30	1,75 2,22	1,71 2,13	1,65 2,04	1,62 1,98	1,57 1,90	1,54 1,86	1,51 1,80	1,48 1,76	1,46 1,72	**46**
1,90 248	1,86 2,40	1,79 2,28	1,74 220	1,70 2,11	1,64 2,02	1,61 1,96	1,56 1,88	1,53 1,84	1,50 1,78	1,47 1,73	1,45 1,70	**48**
1,90 2,46	1,85 2,39	1,78 2,26	1,74 2,18	1,69 2,10	1,63 2,00	1,60 1,94	1,55 1,86	1,52 1,82	1,48 1,76	1,46 1,71	1,44 1,68	**50**
1,88 2,43	1,83 2,35	1,76 2,23	1,72 2,15	1,67 2,06	1,61 1,96	1,58 1,90	1,52 1,82	1,50 1,78	1,46 1,71	1,43 1,66	1,41 1,64	**55**
1,86 2,40	1,81 2,32	1,75 2,20	1,70 2,12	1,65 2,03	1,59 1,93	1,56 1,87	1,50 1,79	1,48 1,74	1,44 1,68	1,41 1,63	1,39 1,60	**60**
1,85 2,37	1,80 230	1,73 2,18	1,68 2,09	1,63 2,00	1,57 1,90	1,54 1,84	1,49 1,76	1,46 1,71	1,42 1,64	1,39 1,60	1,37 1,56	**65**
1,84 2,35	1,79 2,28	1,72 2,15	1,67 2,07	1,62 1,98	1,56 1,88	1,53 1,82	1,47 1,74	1,45 1,69	1,40 1,62	1,37 1,56	1,35 1,53	**70**
1,82 2,32	1,77 2,24	1,70 2,11	1,65 2,03	1,60 1,94	1,54 1,84	1,51 1,78	1,45 1,70	1,42 1,65	1,38 1,57	1,35 1,52	1,32 1,49	**80**
1,79 2,26	1,75 2,19	1,68 2,06	1,63 1,98	1,57 1,89	1,51 1,79	1,48’ 1,73	1,42 1,64	1,39 1,59	1,34 1,51	1,30 1,46	1,28 1,43	**100**
1,77 2,23	1,72 2,15	1,65 2,03	1,60 1,94	1,55 1,85	1,49 1,75	1,45 1,68	1,39 1,59	1,36 1,54	1,31 1,46	1,27 1,40	1,25 1,37	**125**
1,76 2,20	1,71 2,12	1,64 2,00	1,59 1,91	1,54 1,83	1,47 1,72	1,44 1,66	1,37 1,56	1,34 1,51	1,29 1,43	1,25 1,37	1,22 1,33	**150**
1,74 2,17	1,69 2,09	162 1,97	1,57 1,88	1,52 1,79	1,45 1,69	1,42 1,62	1,35 1,53	1,32 1,48	1,26 1,39	1,22 1,33	1,19 1,28	**200**
1,72 2,12	1,67 2,04	1,60 1,92	1,54 1,84	1,49 1,74	1,42 1,64	1,38 1,57	1,32 1,47	1,28 1,42	1,22 1,32	1,16 1,24	1,13 1,19	**400**
1,70 2,09	1,65 2,01	1-58 1,89	1,53 181	1,47 1,71	1,41 1,61	1,36 1,54	1,30 1,44	1,26 1,38	1,19 1,28	1,13 1,19	1,08 1,11	**1000**
1,69 2,07	1,64 1,99	1,57 1,87	1,52 1,79	1,46 1,69	1,40 1,59	1,35 1,52	1,28 1,41	1,24 1,36	1,17 1,25	1,11 1,15	1,00 109	∞

КРАТКОЕ СОДЕРЖАНИЕ

СОДЕРЖАНИЕ

Хамидуллин Равгат Явдатович

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Учебное пособие

Редактор *Издательский дом «Научная библиотека»*
Корректор *Издательский дом «Научная библиотека»*
Компьютерная верстка *И. Н. Глазунова*
Дизайн обложки *Издательский дом Университета «Синергия»*

Подписано в печать 30.08.2019. Формат 60x90 $^{1}/_{16}$.
Усл. п.л. 17,25. Тираж 1500 экз.

Университет «Синергия»
125190, Москва, Ленинградский просп., д. 80, корп. Г
Тел.: +7 (495) 800-10-01

ИСТОРИЯ

Матюхин
Андрей
Викторович

Доктор политических наук, заведующий кафедрой Философии и истории Университета «Синергия». Специалист в области истории общественной мысли, политических идеологий, концепций модернизации.

Давыдова
Юлия
Александровна

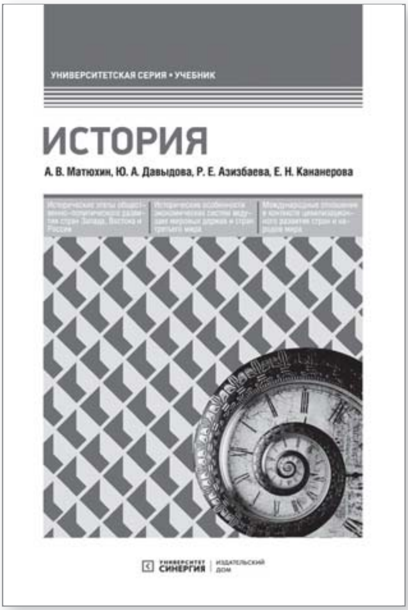
Кандидат исторических наук, доцент кафедры Политологии и социологии Российского экономического университета им. Плеханова. Специалист в области методики преподавания отечественной и мировой истории в высшей школе.

Азизбаева
Раиса
Евгеньевна

Кандидат исторических наук, доцент кафедры Философии и истории Университета «Синергия». Специалист в области истории русской культуры XVIII века, исследователь истории призрения сирот и незаконнорожденных в России.

Кананерова
Елена
Николаевна

Кандидат исторических наук, доцент кафедры Философии и истории Университета «Синергия». Специалист в области аграрной истории России.



Учебное пособие содержит необходимый материал для изучения курса «История» в высших учебных заведениях Российской Федерации. Целью данного учебника является формирование у обучающихся базовой системы знаний в области российской и мировой истории, выработка навыков исторического мышления, воспитание у студентов чувства гражданственности и патриотизма. Предназначен для студентов вузов всех направлений подготовки, аспирантов социально-гуманитарных направлений и многочисленных любителей истории.

УПРАВЛЕНИЕ ЧЕЛОВЕЧЕСКИМИ РЕСУРСАМИ ОРГАНИЗАЦИИ

Алавердов
Ашот
Робертович

Почетный работник высшего профессионального образования Российской Федерации, доктор экономических наук, профессор, заведующий кафедрой Управления человеческими ресурсами Московского финансово-промышленного университета «Синергия». Профессиональный стаж работы в области управления персоналом – 32 года, педагогический стаж – 23 года. Автор 3 монографий, 5 учебников и более 70 научных работ.

Предметная область предлагаемого учебника - теория и практика управления персоналом современной организации. Целями изучения являются формирование у обучающихся осознанного понимания роли человеческого капитала организации и необходимости управления им на системно-формализованной основе, а также приобретение обучающимися профессиональных компетенций, необходимых эффективному менеджеру в области управления человеческими ресурсами.
Учебник предназначен для обучающихся по программам бакалавриата, магистратуры и аспирантуры.

ОСНОВЫ ОЦЕНОЧНОЙ ДЕЯТЕЛЬНОСТИ

Кацман
Валерий
Евельевич

Доктор техн. наук, профессор кафедры Информационных систем в экономике, директор Учебно-методического центра подготовки профессиональных экспертов Университета «Синергия», академик Международной академии информатизации. Специализируется на разработке экономико-математических методов для решения задач большой размерности.

Косорукова
Ирина
Вячеславовна

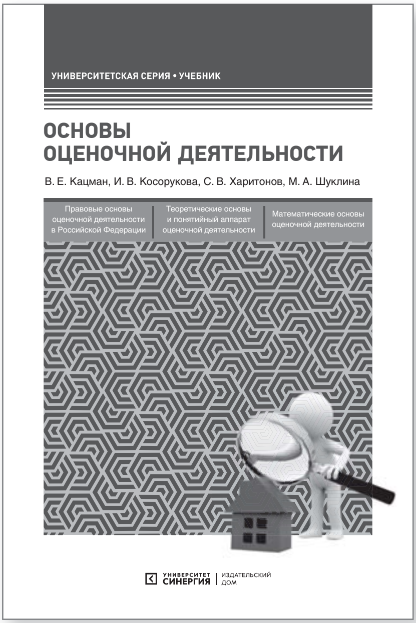
Доктор экон. наук, профессор, зав. кафедрой Оценочной деятельности и корпоративных финансов, директор Центра профессиональной переподготовки Университета «Синергия», Почетный работник сферы образования Российской Федерации.

Харитонов
Сергей
Владимирович

Кандидат экон. наук, доцент, Руководитель направления методического обеспечения и развития образовательных проектов АО «ИнфоВотч», доцент Базовой кафедры АО «ИнфоВотч» МИЭМ НИУ ВШЭ, доцент кафедры Управления информационными системами и программирования РЭУ им.Г.В. Плеханова.

Шуклина
Мария
Александровна

Доцент кафедры Оценочной деятельности и корпоративных финансов Университета «Синергия», зам. директора Центра профессиональной переподготовки, кандидат сельскохозяйственных наук. Специалист в области оценки стоимости ценных бумаг, корпоративного управления. Опыт работы в Росимуществе, Минэкономразвития России.



В учебнике рассматриваются вопросы правового регулирования оценочной деятельности, понятийный аппарат, теоретические основы оценочной деятельности и математические методы оценки. В приложениях даются основные правовые и нормативные материалы, табличные материалы.
Предназначен для широкой аудитории — от студентов до преподавателей образовательных учреждений и практикующих оценщиков, предпринимателей и экономистов.

УНИВЕРСИТЕТ СИНЕРГИЯ | ИЗДАТЕЛЬСКИЙ ДОМ