Бакст Л.А.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие по курсу «МАТЕМАТИКА» ЛЕКЦИОННЫЙ КУРС

Для студентов экономических специальностей

Москва
2012

Теория вероятностей и математическая статистика. Учебное пособие по курсу «Математика». Лекционный курс. – М.: ИНО 2012.
Для студентов экономических специальностей.
Цикл общих математических и естественно-научных дисциплин.

ВВЕДЕНИЕ

Основной целью курса теории вероятностей и математической статистики для студентов экономических специальностей является изучение данного раздела математики и его приложений для задач экономического содержания, формирование и закрепление навыков по самостоятельному мышлению и умению применять изучаемые теоретические положения к анализу и решению конкретных практических проблем.

Пособие содержит необходимый теоретический материал (в соответствии с программой курса и требованиями Государственного Образовательного стандарта), дополняемый по ходу изложения, необходимыми примерами и контрольными вопросами по темам.

Особое внимание уделяется формированию у учащихся необходимых расчетных навыков, пониманию методов и алгоритмов решения задач, умению обосновывать и объяснять полученные результаты.

Наряду с теоретическим материалом, в пособии рассмотрены методы решения большого числа типовых примеров, позволяющих разобраться во всех аспектах выполнения контрольных заданий и необходимых математических расчетов. В конце каждой изучаемой темы помещены вопросы, позволяющие учащимся самостоятельно проверить свои знания.

Сложность курса требует от учащихся серьезного подхода, самостоятельной работы, обязательного выполнения практических заданий.

Контроль усвоения материала курса осуществляется путем *компьютерного тестирования*, включающего *рубежные и итоговые тесты*, содержащие вопросы, аналогичные приведенным в настоящем пособии.

ПРОГРАММА КУРСА

Тема 1. Основные понятия и определения

Место случайности в природе и в практической деятельности людей. Интуитивное понятие вероятности события.
Классификация событий. Классическое определение вероятности события, его недостатки. Статистическое определение вероятности. Геометрические вероятности. Теорема Я. Бернулли о сходимости по вероятности.

Тема 2. Элементы комбинаторики

Основные понятия комбинаторики. Правила произведения и суммы. Виды комбинаций элементов конечных множеств: размещения, перестановки, сочетания и их свойства.

Тема 3. Исчисление вероятностей событий

Основные понятия и соотношения алгебры событий. Теорема сложения для несовместных событий. Понятие зависимости событий и условная вероятность. Теорема умножения. Теорема сложения в общем виде. Формула полной вероятности. Формула Бейеса.

Тема 4. Дискретные случайные величины

Понятие случайной величины. Дискретная случайная величина и закон ее распределения. Функция распределения дискретной случайной величины. График функции распределения дискретной случайной величины. Схема повторения испытаний с бинарным исходом и биномиальное распределение. Формула Я. Бернулли. Наивероятнейшее число наступления события. Распределение Пуассона. Простейший поток событий. Геометрическое распределение.

Тема 5. Числовые характеристики дискретных случайных величин

О введении числовых характеристик случайных величин. Характеристики положения (математическое ожидание, медиана, мода) и характеристики рассеяния (дисперсия, среднеквадратическое отклонение, вероятное отклонение). Числовые характеристики случайных величин, имеющих биномиальное и геометрическое распределение. Числовые характеристики суммы и произведения дискретных случайных величин. Понятие о начальных и центральных моментах случайных величин. Функции дискретного случайного аргумента и их характеристики.

Тема 6. Непрерывные случайные величины и их вероятностные характеристики

Непрерывные случайные величины. Функция и плотность распределения. Кривая распределения. Распределение случайной величины по закону постоянной плотности (равномерное распределение). Числовые характеристики непрерывной случайной величины. Функции непрерывного случайного аргумента и их характеристики.

Тема 7. Некоторые типовые распределения случайных величин

Потоки событий. Распределение Пуассона и определяемый и простейший поток. Основные параметры и характеристики пуассоновского распределения. Распределение Пуассона как предельный случай биномиального. Общность пуассоновского распределения. Экспоненциальное распределение и его характеристики. Локальная и интегральная теоремы Лапласа. Нормальное распределение и центральная предельная теорема. Нормальная кривая, ее уравнение. Функция нормального распределения и интеграл вероятности. Вероятность попадания нормально распределенной случайной величины на заданный интервал. Правило трех сигм.

Тема 8. Выборочный метод. Общие вопросы

Выборочный метод, генеральная и выборочная совокупности, повторная и бесповторная выборки, репрезентативная выборка, способы отбора, эмпирическая функция распределения. Сплошное и выборочное наблюдения. Основные задачи теории выборки. Понятие выборочной оценки неизвестного параметра генерального распределения. Требования, предъявляемые к статистической оценке.

Вариационный ряд как результат первичной обработки результатов опыта (наблюдений). Дискретный и интегральный ряды. Средняя арифметическая и дисперсия, стандартное отклонение, мода, медиана, размах вариационного ряда. Графическое представление статистических данных.

Тема 9. Оценка доли признака и генеральной средней

Несмещенность и состоятельность выборочной доли как оценки генеральной доли. Формула для расчета доверительной вероятности.

Средняя квадратическая ошибка собственно случайной выборки при оценке доли признака при повторном и бесповторном отборе членов. Выборочная средняя как оценка генеральной средней. Несмещенность и состоятельность этой оценки. Формула для расчета доверительной вероятности. Средняя квадратическая ошибка собственно случайной выборки при оценке средней при повторном и бесповторном отборе членов.

Тема 10. Проверка статистических гипотез

Понятие статистической гипотезы. Основные этапы проверки гипотезы. Проверка гипотез о числовых значениях параметров нормального распределения. Проверка гипотез о равенстве математических ожиданий двух нормальных распределений с известными дисперсиями, а также с неизвестными, но равными дисперсиями. Проверка гипотезы о равенстве дисперсий двух нормальных распределений. Проверка гипотезы о числовом значении вероятности события. Проверка гипотезы о равенстве вероятностей. Проверка гипотезы о модели закона распределения. Критерий согласия Пирсона.

Тема 11. Элементы теории корреляции

Функциональная и статистическая зависимости. Корреляционные таблицы. Групповые средние. Понятие корреляционной зависимости. Основные задачи теории корреляции: выбор связи, оценка тесноты и существенности связи. Виды корреляционной связи: парная и множественная, линейная и нелинейная связи. Линейная корреляция. Уравнения прямых регрессии для парной корреляции. Определение параметров прямых регрессии методом наименьших квадратов. Коэффициент корреляции и его свойства. Оценка достоверности (значимости) выборочного коэффициента корреляции. Критерий Стъюдента. Понятие о множественной корреляционной зависимости.

О сколько нам открытий чудных
Готовят просвещенья дух
И Опыт, сын ошибок трудных,
И Гений, парадоксов друг,
И Случай, бог изобретатель…
А.С.Пушкин

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Тема 1. Основные понятия и определения

Основные вопросы темы

1. Введение.
2. Классификация событий.
3. Классическое, статистическое и геометрическое определение вероятности события.
4. Статистическое определение вероятности. Теорема Я.Бернулли о сходимости по вероятности.

ВВЕДЕНИЕ

Теория вероятностей – это раздел математики, изучающий закономерности случайных явлений, наблюдаемых при многократном повторении опыта, где под случайным явлением понимаются явления с неопределенным исходом.

Теория вероятностей является фундаментом связанного с ней раздела математики, называемого *математической статистикой*.

Математическая статистика - раздел математики, изучающий математические методы сбора, систематизации, обработки и интерпретации результатов наблюдений с целью выявления статистических закономерностей.

1.1 Классификация событий.

Дадим ряд важных определений.

Под *опытом* (*экспериментом, испытанием*) понимается воспроизведение определенного (неизменяемого) комплекса условий.

Результат опыта называется *событием* (*явлением, исходом*). События обозначаются заглавными буквами латинского алфавита: *A, B, C, …*

Под *случайным событием* (*случайным явлением*), понимается такое событие, которое при неоднократном воспроизведении данного опыта иногда происходит, а иногда - нет, причём исход опыта заранее предсказать нельзя.

Говорят, что *событие A влечет за собой событие B*, если появление события A в данном опыте всегда сопровождается появлением события B. Обозначение: $A \subset B$.

Два события A и B называются *равносильными* и обозначаются как $A=B$, если одновременно выполняются два условия: $A \subset B$ и $B \subset A$.

Событие называется *невозможным* (в данном опыте), если при любом повторении опыта оно произойти не может. Такое событие обозначается символом $\varnothing$.

Событие называется *достоверным* (в данном опыте), если при любом повторении опыта оно обязательно должно произойти. Такое событие обозначается символом Ω.

Несколько событий называются *равновозможными*, если природа их такова, что в рамках данного опыта ни одно из них не является объективно более возможным.

События называются *единственно возможными*, если в результате данного опыта обязательно должно произойти хотя бы одно из них.

События называются *несовместными* (*несовместимыми*), если природа их такова, что в рамках данного опыта наступление одного из них исключает наступление другого. В противном случае они называются *совместными* (*совместимыми*).

Несколько событий образуют *полную группу*, если в рамках данного испытания они являются единственно возможными и несовместными. Иными словами: несколько событий образуют *полную группу,* если в результате данного испытания обязательно должно произойти одно и только одно из этих событий.

Два события называются *противоположными*, если в результате опыта могут произойти только эти события, причем появление одно из них исключает появление другого. Иными словами: два события называются *противоположными*, если они образуют полную группу событий. Событие, противоположное событию A обозначается $\overline{A}$.

Возможные исходы опыта называются *элементарными событиями*, если они являются взаимно исключающими, и в результате опыта одно из них обязательно происходит.

1.2 Классическое определение вероятности.

Классическое определение вероятности базируется на понятии *классической схемы испытаний*, для которой число элементарных исходов *конечно* и все они *равновозможны.*

Рассматриваемое *классическое определение* использует, также, понятие случая*, благоприятствующего* данному событию:

Случай называется *благоприятствующим* событию А, если его появление влечет за собой событие *A*.

Классическое определение вероятности формулируется следующим образом.

ОПРЕДЕЛЕНИЕ (классическое определение вероятности). *Вероятность события A* равна отношению числа случаев, благоприятствующих ему, к общему числу случаев, т.е:

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

где: P(A) - вероятность события *A;*

m - число случаев, благоприятствующих событию *A;*

n - общее число случаев.

ЗАМЕЧАНИЕ. Современная математика рассматривает приведенное выше выражение не как определение вероятности, а как метод ее вычисления.

Свойства вероятности события.

Рассмотренное определение позволяет сформулировать следующие важнейшие свойства вероятности:

1. Вероятность достоверного события равна единице, т.е:

$$P(\Omega)=1 \;;$$

2. Вероятность недостоверного события равна нулю, т.е:

$$P(\varnothing)=0 \;;$$

3. Вероятность любого события заключена между нулем и единицей, т.е:

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

1.3. Статистическое определение вероятности.

Приведенное выше классическое определение вероятностей не является универсальным. Оно предполагает, в частности, *равновозможность* элементарных исходов, что не всегда выполняется на практике. Можно указать и ряд иных причин, ограничивающих область его возможного применения.

Однако, существует и другой подход к определению вероятностей, получивший название *статистического определения* вероятности.

ОПРЕДЕЛЕНИЕ (статистическое определение вероятности). Статистической вероятностью события А называется относительная частота (частость) появления этого события при проведении n испытаний (опытов), т.е:

$$\tilde{P}(A) = w(A) = \frac{m}{n} \quad ,$$

где: $\tilde{P}(A)$ - статистическая вероятность события A;

$w(A)$ - относительная частота (частость) события A;

m - число испытаний в которых появилось событие A;

n - общее число испытаний.

ЗАМЕЧАНИЕ 1. В отличии от классического определения, где речь идет о некотором теоретическом определении вероятности, понятие статистической вероятности, связано с практической реализацией серии опытов, т.е. статистическая вероятность является экспериментальной (опытной) характеристикой.

ЗАМЕЧАНИЕ 2. Статистическое определение применимо не ко всем событиям, а только к тем, которые удовлетворяют следующим условиям:

1. *Неограниченная* возможность проведения произвольного числа опытов (напомним, что под опытом (экспериментом, испытанием) понимается воспроизведение определенного и неизменяемого комплекса условий);

2. Число испытаний, в которых в которых наблюдалось событие А должно быть *достаточно велико*;

3. Частость появления события А не должна существенно меняться при проведении различных серий опытов (в этом случае говорят о *статистической устойчивости* события А).

ЗАМЕЧАНИЕ 3. Можно показать, что перечисленные ранее свойства вероятности, понимаемой в классической интерпретации, *не меняются* при переходе к статистической вероятности. Это утверждение основано на следующей теореме Бернулли.

ТЕОРЕМА БЕРНУЛЛИ (сформулирована Я.Бернулли, 1713г.).

Пусть вероятность появления события *A* при проведении повторных испытаний постоянна и равна *p(A)*. При проведении n независимых испытаний событие *A* наблюдается n_A раз. Тогда, для любого положительного числа ε (ε > 0) имеет место равенство:

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1,$$

иначе говоря относительная частота события *A* сходится по вероятности к вероятности *p(A)*.

1.4. Геометрическое определение вероятности.

Для ряда задач оказывается удобным использовать *геометрическое определение вероятности*, которое связывает понятие вероятности с определенной геометрической областью, благоприятствующей реализации рассматриваемого события.

ОПРЕДЕЛЕНИЕ (геометрическое определение вероятности). Геометрической вероятностью события *A* называется отношение меры области, благоприятствующей событию *A* к мере всей области, т.е:

$$P(A) = \frac{mes\ S_g}{mes\ S_G} ,$$

где: $mes\ S_g$ - мера области g, благоприятствующей событию A ;

$mes\ S_G$ - мера вей области G.

Рисунок ниже иллюстрирует геометрическое определение вероятности.

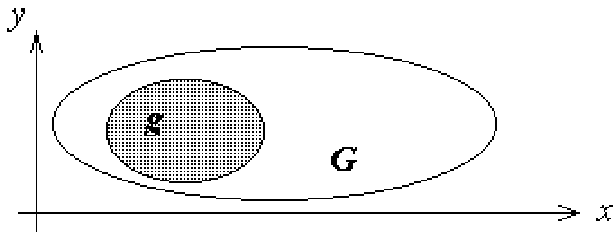


Рис. 1. Графическая иллюстрация к геометрическому определению вероятности.

Контрольные вопросы по Теме 1.

1. Является ли событие $2{\cdot}2=5$
   а) достоверным;
   б) невозможным ?
2. Сколько элементов содержит множество элементарных событий, описывающее результат бросания кубика?
3. Пусть событие $A=\{1, 2, 3\}$ и событие $B=\{1, 2, 3, 4, 5\}$. Можно ли сказать, что:
   а) событие A влечет событие B;
   б) событие B влечет событие A ?
4. Как называется числовая характеристика, описывающая объективную возможность наступления события?
5. Что такое относительная частота (частость) события?
6. Дайте классическое определение вероятности.
7. Дайте определение равносильных событий.
8. Равносильны ли события $\{1, 2, 3, 4, 5\}$ и $\{5, 4, 3, 2, 1\}$?
9. Запишите формулу классического определения вероятности и поясните смысл всех ее составляющих.
10. В каких случаях вероятность события отрицательна?
11. Может ли вероятность события быть равной 1.25?
12. Пусть событие $A=\{1, 2, 3\}$ и событие $B=\{1, 2, 3, 4, 5\}$. Можно ли сказать, что:
   а) событие A влечет событие B;
   б) событие B влечет событие A ?

Тема 2. ЭЛЕМЕНТЫ КОМБИНАТОРИКИ.

Основные вопросы темы

1. Основные определения.
2. Правила произведения и суммы.
3. Виды комбинаций элементов конечных множеств: перестановки, размещения и сочетания и их свойства.

2.1. Элементы комбинаторики. Основные определения.

Для вычисления вероятности, используя ее классическое определение, часто применяется раздел математики, называемый *комбинаторикой*. Рассмотрим его более подробно.

Комбинаторика[1] – это раздел математики, посвященный решению задач выбора и расположения элементов некоторого, обычно конечного, множества в соответствии с заданными правилами. Само слово комбинаторика происходит от латинского слова “*combinatio*” (соединение).

Основными и типичными операциями и задачами комбинаторики являются следующие:

1) образование упорядоченных множеств, путем установления определенного порядка следования элементов множества друг за другом, - *составление перестановок*;

2) образование подмножеств, состоящее в выделении из данного множества некоторой части его элементов, - *составление сочетаний*;

3) образование упорядоченных подмножеств - *составление размещений*.

Результатом выполнения этих операций является образование отдельных комбинаций (соединений) элементов исходного множества,

---

[1] Элементарные сведения комбинаторного характера были известны очень давно, но они носили разрозненный характер. Работы по их систематизация были сделаны в XVII веке Блезом Паскалем и Пьером Ферма. В том же XVII веке попытку рассматривать комбинаторику как единую теоретическую дисциплину предпринял Готфрид Лейбниц, который отводил комбинаторике роль универсального математического аппарата логических рассуждений. Существенный вклад в развитие комбинаторики внесли Рене Декарт и Леонард Эйлер.

Источник: http://combinatorika.by.ru/index.htm называемых соответственно: *перестановками*, *сочетаниями* и *размещениями*.

2.1.1. Перестановки.

ОПРЕДЕЛЕНИЕ. *Перестановками* называются различные комбинации, состоящие из одного и того же конечного множества элементов $\{a_1, a_2, a_3, \ldots, a_n\}$ и отличающиеся порядком расположения самих элементов. Если среди элементов $a_1, a_2, a_3, \ldots, a_n$ нет одинаковых, то рассматриваемые комбинации называют *перестановки без повторений*. В противном случае – *перестановки с повторениями*.

Перестановки без повторений.

В соответствии с определением, возьмем n различных элементов $a_1, a_2, a_3, \ldots, a_n$ и будем образовывать различные сочетания путем перестановки отдельных элементов. Общее число полученных таким образом перестановок, обычно обозначаемое как P_n (от французского слова *permutation* – перестановка), равно:

$$P_n = n!$$

Перестановки с повторениями.

Если среди элементов рассматриваемого множества есть одинаковые, в частности, если среди n элементов есть m_1 элементов одного вида, m_2 элементов другого вида и т.д., то можно показать, что число различных перестановок в этом случае $\tilde{P}_n(m_1, m_2, \ldots, m_k)$ будет равно:

$$\tilde{P}_n(m_1, m_2, \ldots, m_k) = \frac{(m_1 + m_2 + \ldots + m_k)!}{m_1!\, m_2! \ldots m_k!} = \frac{n!}{m_1!\, m_2! \ldots m_k!},$$

где: $m_1 + m_2 + \ldots + m_k = n$.

Например, перестановка книг на полке книжного шкафа. Если все книги различны, то любая перестановка меняет порядок в котором они находятся. Однако если на полке есть одинаковые книги, то перемена таких книг местами не приведет к изменению существовавшего ранее порядка.

2.2.2. Сочетания.

ОПРЕДЕЛЕНИЕ: Рассматривается конечное множество, состоящее из n элементов. *Сочетаниями из n элементов по* m называют соединения, содержащие m элементов исходного множества ($m < n$) и отличающиеся между собой хотя бы одним элементом. Если каждое, полученное таким образом соединение, не содержит одинаковых элементов, то рассматриваемые комбинации носят название *сочетаниями без повторений*. В противном случае – *сочетаниями с повторениями*.

Сочетания без повторений.

В соответствии с определением, будем выбирать отдельные элементы исходного множества и составлять из них различные подмножества, содержащие ровно m элементов, руководствуясь следующими правилами:

а) подмножество не содержит одинаковых элементов; и

б) все подмножества различаются хотя бы одним элементом.

Общее число полученных таким образом сочетаний, обычно обозначаемое как C_n^m (от французского слова *combinasion* - сочетание), равно:

$$C_n^m = \frac{n!}{m!(n-m)!}$$

Сочетания с повторениями.

В соответствии с определением, будем выбирать отдельные элементы исходного множества и составлять из них различные подмножества, содержащие ровно m элементов, руководствуясь следующими правилами:

а) подмножество может содержать одинаковые элементы; и

б) все подмножества различаются хотя бы одним элементом.

Можно показать, что общее число полученных таким образом сочетаний - $\tilde{C}_n^m$ равно:

$$\tilde{C}_n^m = C_{n+m-1}^m$$

Свойства сочетаний:

$$C_n^0 = 1 \ ;$$

$$C_n^m = C_n^{n-m} \ ;$$

$$C_n^m + C_n^{m+1} = C_{n+1}^{m+1} \ .$$

2.2.3. Размещения.

ОПРЕДЕЛЕНИЕ: Рассматривается конечное множество, состоящее из n элементов. *Размещениями из n элементов по m* называют различные комбинации, составленные из m элементов исходного множества и отличающиеся либо составом элементов, либо порядком их расположения. Если каждое полученное таким образом подмножество не содержит одинаковых элементов, то рассматриваемые комбинации называют *перестановки без повторений*. В противном случае – *перестановки с повторениями*.

Размещения без повторений.

Пусть рассматриваемое множество состоит из n различных элементов $a_1, a_2, a_3, \dots, a_n$. В соответствии с определением, будем составлять из них отдельные подмножества, содержащие ровно m элементов, руководствуясь следующими правилами:

а) все подмножества различаются либо составом элементов, либо порядком их расположения; и

б) подмножество не содержит одинаковых элементов.

Общее число полученных таким образом размещений обычно обозначается как A_n^m (от французского слова *arrangement*, что в переводе означает "размещение", "приведение в порядок"). Можно показать, что в этом случае:

$$A_n^m = \frac{n!}{(n-m)!} = (n-m+1)\cdot(n-m+2)\cdot\ldots\cdot n \ .$$

Размещения с повторениями.

Пусть исходное множество состоит из n различных элементов $a_1, a_2, a_3, \dots, a_n$. В соответствии с определением, будем составлять из них отдельные подмножества, содержащие ровно m элементов, руководствуясь следующими правилами:

а) все подмножества различаются либо составом элементов, либо порядком их расположения; и

б) подмножество может содержать любое число одинаковые элементов.

Можно показать, что общее число полученных таким образом размещений - $\tilde{A}_n^m$, равно:

$$\tilde{A}_n^m = n^m \ .$$

Два важных правила комбинаторики.

ПРАВИЛО СУММЫ. Если элемент a_1 может быть выбран n_1 способами, элемент a_2 - другими n_2 способами, элемент a_3 - отличными от первых двух n_3 способами, и т.д., элемент a_k - n_k способами, отличными от первых (k-1) способов, то выбор одного из элементов: a_1, или a_2, …, или a_k может быть осуществлен ($n_1 + n_2 + ... + n_k$) способами.

ПРАВИЛО ПРОИЗВЕДЕНИЯ. Если элемент a_1 может быть выбран n_1 способами, после каждого такого выбора элемент a_2 может быть выбран другими n_2 способами, …, и т.д., после каждого (k-1)-го выбора элемент a_k может быть выбран n_k способами, то выбор элементов a_1, a_2, …, a_k в указанном порядке может быть осуществлен $n_1 \cdot n_2 \cdot ... \cdot n_k$ способами.

Контрольные вопросы по Теме 2.

1. В чем отличие между сочетанием из пяти элементов по три и размещением из пяти по три?
2. Различаются ли понятия перестановки и пяти элементов и
3. размещением из пяти элементов по пять?
4. Равны ли числа C_7^3 и C_7^4?
5. Равны ли числа: а) C_7^1 и A_7^1; б) C_7^3 и A_7^3 ?
6. Какое число больше $10!$ или 10^{10} ?
7. Сформулируйте правило суммы.
8. Сформулируйте правило произведения.

Тема 3. ИСЧИСЛЕНИЕ СОБЫТИЙ.

Основные вопросы темы

1. Действия над событиями.
2. Теоремы сложения и умножения вероятностей.
3. Зависимость событий. Условная вероятность.
4. Формула полной вероятности. Формула Байеса

3.1. Действия над событиями.

Ниже определяются операции (действия) над событиями (*сумма*, *произведение* и *разность* событий).

ОПРЕДЕЛЕНИЕ. *Суммой* двух событий A и B называется новое событие, обозначаемое как $A+B$, и состоящее в наступлении хотя бы одного из этих событий (либо A, либо B, либо их обоих вместе).

Данное определение легко обобщается на случай суммы $n>2$ событий $A, B, C, \ldots$ и определяет новое событие, состоящее в наступлении хотя бы одного из них.

ОПРЕДЕЛЕНИЕ. *Произведением* двух событий A и B называется новое событие, обозначаемое как $A \cdot B$, и состоящее в их одновременном наступлении.

Данное определение также легко обобщается на случай произведения $n>2$ событий $A, B, C, \ldots$ и определяет новое событие, состоящее в одновременном наступлении всех этих событий.

ОПРЕДЕЛЕНИЕ. *Разностью* двух событий A и B называется новое событие, обозначаемое как: A-B, и состоящее в одновременном наступлении события A и события, противоположному B (иными словами: наступление события A с одновременным отсутствием события B).

СЛЕДСТВИЕ. Опираясь на приведенные выше определения операций, можно показать справедливость следующих выражений:

1) $A+A=A$; 2) $AA=A$;

3) $A+\Omega=\Omega$; 4) $A\Omega=A$;

5) $A\varnothing=\varnothing$; 6) $A+\varnothing=A$,

где: Ω - достоверное событие; и $\varnothing$ - невозможное событие.

Перечислим, также, свойства операций сложения и умножения событий:

1	$A+B = B+A$	коммутативность сложения
2	$A+(B+C) = (A+B)+C$	ассоциативность сложения
3	$AB = BA$	коммутативность умножения
4	$A(B+C) = AB+AC$	дистрибутивность умножения

3.1.1 Диаграммы Венна.

Удобно представить геометрическую интерпретацию действий над событиями, используя *диаграммы Венна*.

Рассмотрим некоторую прямоугольную область и реализуем опыт, состоящий в выборе произвольной точки внутри этого прямоугольника.

Рассмотрим, также, две области внутри данного прямоугольника. Если выбранная точка принадлежит первой области, то будем говорить, что реализовано событие A, в противном случае – событие обратное ему, т.е.: $\overline{A}$. Аналогичным образом (рассматривая вторую область) определим события B и $\overline{B}$.

Используя данную интерпретацию, можно графически представить сумму, произведение и разность событий A и B (что представлено на рисунке ниже).

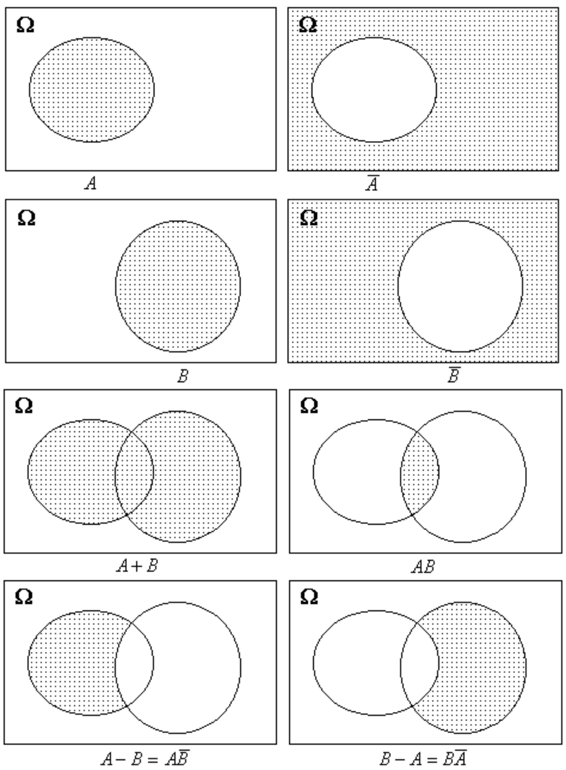


Рис. 2. Диаграммы Венна для суммы, произведения и разности событий A и B.

3.2. Теорема сложения вероятностей.

Введенные выше операции позволяют анализировать сложные события, являющиеся композицией (суммой, произведением и разностью) более простых событий. Какова вероятность таких сложных событий? Рассмотренные ниже теоретические положения дают ответ на этот вопрос.

ТЕОРЕМА. Вероятность суммы конечного числа несовместных событий равна сумме вероятностей этих событий, т.е.:

$$P(A+B+C+\ldots+L) = P(A)+P(B)+P(C)+\ldots+P(L).$$

СЛЕДСТВИЕ. Вероятность суммы событий, образующих полную группу, равна единице, т.е.:

$$P(A+B+C+\ldots+L) = 1.$$

В частности, вероятность суммы события A и противоположного ему события $\overline{A}$, также, равна единице (поскольку эти два события образуют полную группу):

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

ЗАМЕЧАНИЕ. Подчеркнем, что в рассмотренной выше теореме речь шла о *несовместных* событиях. Для случая совместных событий данная теорема неприменима.

Приведенное замечание показывает необходимость более детального рассмотрения совместных событий.

3.3. Условная вероятность событий.

В ряде случаев события могут быть некоторым образом связаны между собой. В частности, появление (реализация) одного из них может влиять на объективную возможность реализации другого.

Например, пусть в пачке лотерейных билетов находится один “счастливый” и два человека по очереди достают наугад по одному билету. Ясно, что шансы второго участника достать “счастливый” билет зависят от того, что вытянул первый: если он вытянул единственный “счастливый” билет, то шансов у второго участника нет никаких. Если первому участнику не повезло, то у второго остается шанс выиграть.

Для описания взаимосвязи событий, при которой реализация одного из них влияет на объективную возможность реализации другого, вводится понятие *условной вероятности*.

ОПРЕДЕЛЕНИЕ. *Условной вероятностью* $P_B(A)$ события *A* относительно события *B*, называется вероятность события *A*, вычисленную в предположении, что событие *B* уже наступило.

Для обозначения условная вероятности $P_B(A)$ обозначается также и другая символика: $P(A|B)$ или $P(A/B)$.

Можно показать, что для вычисления условной вероятности применима следующая формула:

$$P_B(A) = \frac{P(AB)}{P(B)}.$$

Естественно определить событие *A* *независимым* от события *B*, если его вероятность *P(A)* не зависит (не меняется) от того, произошло или нет событие *B*:

$$P(A) = P_B(A).$$

В противном случае, Событие *A* называется *зависимым* от события *B*.

Справедлива следующая теорема.

ТЕОРЕМА. Если событие *A* не зависит от события *B*, то и событие *B* не зависит от события *A*.

Суть данной теоремы состоит в том, что зависимость событий всегда носит взаимосвязанный характер. Опираясь на это утверждение, можно сформулировать следующее определение независимости событий.

ОПРЕДЕЛЕНИЕ. События *A* и *B* называются *независимыми* если реализация (наступление) одного из них не меняет вероятность наступления другого.

3.4 Общие формулы вероятности произведения и суммы событий

3.4.1 Правило умножения вероятностей.

Для независимых событий справедлива следующая теорема, которую также называют теоремой (правилом) умножения вероятностей.

ТЕОРЕМА. Вероятность произведения двух событий *A* и *B* равна произведению вероятностей этих событий:

$$P(AB)=P(A)\cdot P(B).$$

Обобщенная формула для вероятности произведения конечного числа <u>независимых</u> событий $A, B, C, \dots, L$ записывается аналогично:

$$P(ABC\dots L)=P(A)\cdot P(B)\cdot P(C)\dots\cdot P(L).$$

ЗАМЕЧАНИЕ. Часто утверждение данной теоремы применяется как определение независимости событий.

3.4.2 Вероятность суммы совместных событий.

Ранее сформулировано правило нахождения вероятности суммы *несовместных* событий. Правило нахождения вероятности суммы двух совместных событий определяется следующей теоремой.

ТЕОРЕМА. Вероятность суммы *совместных* событий A и B равна сумме вероятностей этих событий без вероятности их произведения:

$$P(A+B) = P(A) + P(B) - P(AB).$$

Обобщение для случая n ($n>2$) совместных событий определяется иной теоремой.

ТЕОРЕМА. Вероятность суммы *совместных* событий $A, B, C, \dots, L$ равна разности между единицей и вероятностью произведения противоположных событий:

$$P(ABC\dots L)=1-P(\overline{A}\overline{B}\overline{C}\dots\overline{L}).$$

3.5 Формула полной вероятности. Формула Байеса.

Формула полной вероятности.

Рассмотрим некоторое событие A, возможность наступления которого характеризуется *гипотезами* (событиями) $H_1, H_2, \dots, H_n$, образующих полную группу[2]. Как найти вероятность такого события? Ответ на этот вопрос дает следующая теорема.

ТЕОРЕМА. Пусть событие A может наступить только в результате реализации одной из гипотез (событий) $H_1, H_2, \dots, H_n$, образующих полную группу. Тогда вероятность наступления этого события (A) равна сумме попар-

[2] Напомним, что события образуют полную группу, если они попарно несовместны, а их сумма является достоверным событием, т.е: $H_i \cdot H_j = \varnothing, \quad (i \neq j; \quad i, j = 1,2,\dots,n)$; и $H_1 + H_2 + \dots + H_n = \Omega$.

ных произведений вероятности реализации каждой из этих гипотез, на соответствующую условную вероятность наступления события A :

$$P(A) = P(H_1)\cdot P_{H1}(A) + P(H_2)\cdot P_{H2}(A) + ... + P(H_n)\cdot P_{Hn}(A),$$

т.е:

$$P(A) = \sum_{i=1}^{n} P(H_i)\cdot P_{Hi}(A).$$

Данная формула получила название *формулы полной вероятности*.

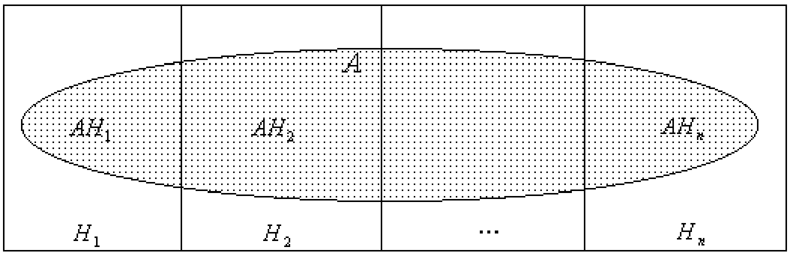


Рис. 3. Графическая иллюстрация к формуле полной вероятности.

Формула Байеса.

Как и ранее, рассмотрим некоторое событие A, возможность реализации которого характеризуется *гипотезами* (событиями) $H_1, H_2, ..., H_n$, образующих полную группу.

Пусть известно, что событие A наступило. Как оценить по этой информации какая из гипотез была реализована? Иными словами, как вычислить условные вероятности реализации каждой из гипотез при условии, что известно о наступлении события A ($P_A(H_1),\ \ P_A(H_2),\ \ ...,\ \ P_A(H_n)$)?

Ответ на этот вопрос дает следующая формула:

$$P_A(H_i) = \frac{P(H_i)\cdot P_{Hi}(A)}{\sum_{i=1}^{n} P(H_i)\cdot P_{Hi}(A)} = \frac{P(H_i)\cdot P_{Hi}(A)}{P(A)},$$

получившая название формулы Байеса.

ЗАМЕЧАНИЕ. Значение формулы Байеса состоит в том, что она связывает *априорные* вероятности (вероятности реализации гипотез и условные вероятности события A для каждой из них, полученные *до реализации самого события A*) с *апостериорными* вероятностями (вероятность реализации гипотез, вычисляемых при условии что *событие A уже наступило*.

Контрольные вопросы по теме 3.

1. Может ли элементарное событие, входящее в сумму событий A+B:

   а) входить в A, но не входить в B;

   б) входить в B, но не входить в A;

   с) одновременно входить и в A и в B;

   с) не входить ни в A ни в B ?

2. Может ли элементарное событие, входящее в сумму событий $A \cdot B$:

   а) входить в A, но не входить в B;

   б) входить в B, но не входить в A;

   с) одновременно входить и в A и в B;

   с) не входить ни в A ни в B ?

3. Пусть события A и B – несовместны. Каким событием будет их произведение и какова вероятность этого события?

4. Что такое событие, обозначаемое как $\overline{A}$ (по отношению к событию A), совместны ли эти события ?

5. Каким событием является событие $A+\overline{A}$ и какова вероятность этого события?

6. Каким событием является событие $A \cdot \overline{A}$ и какова вероятность этого события?

7. Даны два события: событие A=$\{1, 2, 3\}$ и событие B=$\{1, 4, 5\}$. Чему равна разность A-B этих событий?

8. В каких случаях вероятность суммы событий равна сумме их вероятностей?

9. Что больше $P(A+B)$ или $P(A)+P(B)$?

10. Что означает запись $P_B(A)$? Какое событие в данной записи предполагается достоверным?

11. Что больше $P(A \cdot B)$ или $P(A) \cdot P(B)$?

12. В каких случаях выполняется равенство:

    $P(A \cdot B) = P(A) \cdot P(B)$?

13. Что такое полная группа событий?

14. Образуют ли события A и $\overline{A}$ полную группу?

15. Пусть события A_1, A_2, ..., A_n . Чему равно:

    а) произведение этих событий $A_1 \cdot A_2 \cdot ... \cdot A_n$ и какова его вероятность;

    б) сумма этих событий $A_1 + A_2 + ... + A_n$ и вероятность их суммы?

16. Какие вероятности называют априорными и апостериорными в формуле Байеса и почему?

ТЕМА 4. Повторные испытания.

Основные вопросы темы

1. Повторные испытания. Формула Бернулли.
2. Асимптотические приближения Формулы Бернулли: формула Пуассона и локальная формула Муавра Лапласа.
3. Интегральная формула Муавра-Лапласа

4.1. Повторные испытания. Формула Бернулли[3].

Многие процессы носят повторяющийся характер. Так, например, автоматический станок, многократно реализует конкретную операцию в общей технологической цепочке конвейерного производства. Многократные испытания проходят различные изделия, и т.д. Кроме того, многие задачи и процессы могут быть описаны как многократное повторение определенных опытов (экспериментов) в которых исследуются отдельные явления, часто носящие случайный характер.

ОПРЕДЕЛЕНИЕ. Пусть многократно реализуются повторные испытания при неизменных условиях их проведения. В ходе испытания фиксируется появление некоторого случайного события A, вероятность появления которого – $P(A)$ не зависит от результатов предыдущих испытаний и остается неизменной ($P(A)=const$) при повторении опыта. Такие испытания называются *независимыми*, а схема проведения испытаний носит название *схемы Бернулли*.

Формула Бернулли.

ТЕОРЕМА. Вероятность $P_{m,n}$ того, что событие A наступит ровно m раз в n независимых испытаниях, удовлетворяющих условиям схемы Бернулли равна:

$$P_{m,n} = C_n^m p^m q^{n-m},$$

где: $p = P(A)$ – вероятность наступления события A; и $q=1-p$.

[3] Якоб Бернулли (1654-1705), Швейцария, профессор математики Базельского университета, один из первых исследователей в области теории вероятностей. Его главный труд «Искусство предположения» (Ars conjectandi), посвященный этой науке, опубликован посмертно в 1713г. В нем рассматриваются перестановки сочетания, другие элементы теории вероятностей, а главным результатом является теорема о биномиальных распределениях, названная затем его именем - «теорема Бернулли».

Пример 1. Вероятность выигрыша по лотерейному билету равна 0,05. Какова вероятность того, что среди купленных 10 билетов окажутся 2 выигрышных.

Решение. Данная задача описывается схемой Бернулли: проводится 10 испытаний, в ходе которых проверяется наличие выигрыша. Применяя формулу Бернулли, находим вероятность появления выигрыша в двух из десяти испытаний:

$$P_{2,10} = C_{10}^{2} \cdot 0{,}05^{2} \cdot (1-0{,}05)^{8} \approx 0{,}075.$$

Пример 2. При проведении маркетинговых исследований выявлено, что 20% опрошенных предпочитают использовать продукцию данной фирмы. Найти вероятности возможного числа пользователей продукцией фирмы в произвольно выбранной группе из пяти человек. Результат представить в графической форме.

Решение. Согласно условию задачи вероятность того, что человек использует продукцию данной фирмы равна 0,2 (т.е: $p = 0{,}2$ и $q = 1 - p = 0{,}8$). Искомые вероятности находим по формуле Бернулли $P_{m,n} = C_n^m p^m q^{n-m}$ (где: n=5; и m=1,...,5) и помещая их в таблицу.

№ п.п.	Число пользователей продукцией фирмы в группе из пяти человек	Вероятность рассматриваемого события ($P_{m,n} = C_n^m p^m q^{n-m}$)
1	Пользователей нет	$P_{0,5} = C_5^0 \cdot 0{,}2^0 \cdot 0{,}8^5 = 0{,}32768$
2	Один пользователь	$P_{1,5} = C_5^1 \cdot 0{,}2^1 \cdot 0{,}8^4 = 0{,}4096$
3	Два пользователя	$P_{2,5} = C_5^2 \cdot 0{,}2^2 \cdot 0{,}8^3 = 0{,}2048$
3	Три пользователя	$P_{3,5} = C_5^3 \cdot 0{,}2^3 \cdot 0{,}8^2 = 0{,}0512$
4	Четыре пользователя	$P_{4,5} = C_5^4 \cdot 0{,}2^4 \cdot 0{,}8^1 = 0{,}0064$
5	Все используют продукцию данной фирмы	$P_{5,5} = C_5^5 \cdot 0{,}2^5 \cdot 0{,}8^0 = 0{,}00032$

Нанесем расчетные значения на график соединяя полученные точки. Полученная ломанная носит название *многоугольника* (*полигона*) *распределения вероятностей*.

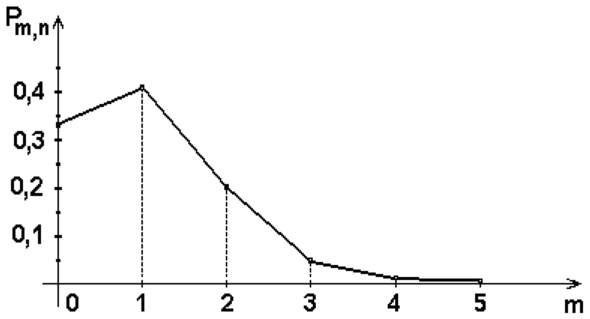


Рис.4. Графическое представление результатов решения – многоугольник распределения вероятностей.

Анализируя график можно выделить точку m_0 в которой достигается максимальное значение вероятности $P_{m,n}$ (в данном случае, $m_0=1$). Такие точки получили особое название.

ОПРЕДЕЛЕНИЕ. Величина m_0 при котором значение вероятности $P_{m,n}$ достигает своего максимального значения называется *наивероятнейшим*.

Можно показать, что справедливо следующее соотношение:

$$np - q \le m_0 \le np + p\ .$$

Отметим, что если $np - q$ и $np + p$ целые числа, то существует два значения m_0 и m_0^*, удовлетворяющие данному соотношению.

Пример 3. Результаты исследования показали, что вероятность - p того, что покупателя удовлетворит качество предлагаемого товара равно 0,8. Каково наивероятнейшее число покупателей в группе из девяти человек, которых полностью удовлетворяет качество предлагаемого ассортимента.

Решение. Найдем наивероятнейшее значение, используя соотношение $np - q \le m_0 \le np + p$ (где $n=9$; $p=0,8$; $q=0,2$):

$$9 \cdot 0{,}8 - 0{,}2 \le m_0 \le 9 \cdot 0{,}8 + 0{,}8\ ,$$

т.е:

$$7 \le m_0 \le 8\,.$$

Таким образом, существует два значения $m_0 = 7$ и $m_0^* = 8$, удовлетворяющие данному соотношению.

4.2. Асимптотические приближения формулы Бернулли.

Формула Бернулли дает точное значение вероятности того, что событие A наступит ровно m раз в n независимых испытаниях, определяемых схемой Бернулли. Однако, практическое применение этой формулы часто оказывается затруднительным, если числа m и n достаточно велики, а вероятность p – мала. Поэтому, были разработаны некоторые асимптотические приближения формулы Бернулли, позволяющие рассчитать значение вероятности $P_{m,n}$ в указанных затруднительных случаях с достаточно высокой степенью сложности.

4.2.1. Формула Пуассона.

Формула Пуассона[4] является асимптотическим приближением формулы Бернулли и определяется следующей теоремой.

ТЕОРЕМА. Если вероятность p наступления некоторого события A в каждом испытании, определяемом схемой Бернулли, стремится к нулю ($p \to 0$) при неограниченном увеличении числа испытаний n ($n \to \infty$), а произведение np стремится к некоторой константе λ ($np \to \lambda$), то вероятность $P_{m,n}$ того, что событие A наступит ровно m раз в n испытаниях удовлетворяет предельному равенству:

$$\lim_{n\to\infty} P_{m,n} = P_m(\lambda) = \frac{\lambda^m e^{-\lambda}}{m!}.$$

С практической точки зрения, условия и выводы данной теоремы означают, что при выполнении следующих трех условий:

а) вероятность p наступления некоторого события A в каждом испытании, определяемом схемой Бернулли, достаточно мала;

б) число испытаний n – велико; и

в) произведение np не превышает десяти ($np = \lambda \le 10$),

то с достаточно высокой степенью точности формула Бернулли может быть аппроксимирована следующей формулой (Пуассона):

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

где: $\lambda = np$.

---

[4] Пуассон Симеон Дени (Poisson Simeon 1781-1840) - французский математик, физик, механик. Член Парижской Академии наук (с 1812). Существенное значение имеют работы Пуассона, посвященные теории вероятностей, определенным интегралам, уравнениям в конечных разностях, дифференциальным уравнениями с частными производными, , вариационному исчислению, рядам.

4.2.2. Локальная формула Муавра-Лапласа.

Другое, применяемое на практике, асимптотическое приближение формулы Бернулли определяется теоремой Муавра-Лапласа.

ТЕОРЕМА (локальная теорема Муавра-Лапласа). Если вероятность p наступления события A в каждом испытании, определяемом схемой Бернулли, постоянна отлична от 0 и 1 ($p \neq 0$ и $p \neq 1$), а число испытаний n достаточно велико, то вероятность $P_{m,n}$ того, что событие A наступит ровно m раз в n испытаниях приближенно равна:

$$P_{m,n} \approx \frac{f(x)}{\sqrt{npq}},$$

где: $f(x)$ - функция Гаусса, определяемая выражением

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

а x вычисляется по формуле:

$$x = \frac{m - np}{\sqrt{npq}}.$$

Точность данного приближения увеличивается с ростом n и на практике оно может быть использовано при значении n, удовлетворяющих условию:

$$npq > 20.$$

ЗАМЕЧАНИЕ. Значения функции Гаусса, фигурирующей в приведенной формуле, табулированы и могут быть найдены в справочной литературе по математике, учебно-методической литературе по теории вероятностей, а также в Приложении 1 данного пособия. Использование этих таблиц значительно упрощает вычисления.

При проведении соответствующих расчетов полезно иметь в виду следующие свойства функции Гаусса:

1. Для положительных значений x, функция Гаусса $f(x)$ является монотонно убывающей, стремящейся к нулю функцией, т. е:

$$\lim_{x \to \infty} f(x) = 0$$

(как видно из таблицы Приложения 1 при x>4 значение $f(x) \approx 0$).

2. Для области отрицательных значений аргумента, свойства функции Гаусса $f(x)$ определяются ее четностью:

$$f(-x) = f(x).$$

4.3. Интегральная теорема Муавра-Лапласа.

Формула Бернулли и рассмотренные выше ее асимптотические приближения позволяют вычислить вероятность $P_{m,n}$ того, что событие A наступит <u>ровно</u> m раз в n независимых испытаниях, удовлетворяющих условиями схемы Бернулли. Однако часто на практике необходимо найти вероятность того, что число наступлений события A находится в некотором диапазоне. Например, требуется найти вероятность того, что событие A произойдет от десяти пятнадцати раз. Ответ на этот вопрос помогает найти следующая теорема.

ТЕОРЕМА (интегральная теорема Муавра-Лапласа). Если вероятность p наступления некоторого события A в каждом испытании, определяемом схемой Бернулли, постоянна и отлична от 0 и 1 ($p \neq 0$ и $p \neq 1$), то вероятность того, что число m наступлений события A в n испытаниях заключено в пределах интервала [a, b] - $P_n(a \le m \le b)$, при достаточно большом n, приближенно равна:

$$P_n(a \le m \le b) \approx \frac{1}{2}\left[\Phi(x_2) - \Phi(x_1)\right],$$

где: $x_1 = \dfrac{a - np}{\sqrt{npq}}$, $x_2 = \dfrac{b - np}{\sqrt{npq}}$; и

$\Phi(x)$- функция, именуемая *интеграл вероятности Лапласа*, определяется выражением:

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

Приведенная выше формула вычисления вероятности $P_n(a \le m \le b)$ носит название *интегральной формулы Муавра-Лапласа*. Ее точность возрастает с ростом n, а ее применение на практике обычно выражается условием, аналогичным для локальной формулы:

$$npq > 20.$$

ЗАМЕЧАНИЕ. Существует несколько разновидностей представления функции $\Phi(x)$, например: $\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-\frac{t^2}{2}} dt$ и $\Phi^*(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

Значения этих функций табулированы (занесены в таблицы) и могут быть найдены в справочной литературе по математике, учебно-методической литературе по теории вероятностей, а также в Приложении 2 данного пособия. Использование этих таблиц значительно упрощает вычисления.

При проведении соответствующих расчетов полезно иметь в виду следующие свойства функции $\Phi(x)$:

1. Функция $\Phi(x)$ является монотонно возрастающей, стремящейся к единице функцией, т. е:

$$\lim_{x\to\infty}\Phi(x)=1$$

(как видно из таблицы Приложения 2 при x>4 значение $\Phi(x)\approx 1$).

2. Для области отрицательных значений аргумента, свойства функции $\Phi(x)$ определяются ее нечетностью:

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

СЛЕДСТВИЕ (интегральной теоремы Муавра-Лапласа). Если проведение испытаний удовлетворяет условиям интегральной теоремы Муавра-Лапласа, то:

1. Частость $w=\dfrac{m}{n}$ появления события A может быть оценена по следующей формуле:

$$P(|w-p|\le\varepsilon)\approx\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right),$$

где: ε - произвольное, сколь угодно малое, положительное число.

2. Число m появлений события A может быть оценено по следующей формуле:

$$P(|m-np|\le\varepsilon)\approx\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$

Контрольные вопросы по теме 4.

1. Можно ли считать схемой Бернулли многократное бросание кубика?
2. Что означает число $P_{m,n}$ в формуле Бернулли?
3. Может ли в схеме Берулли при n=30 и p=1/31 наивероятнейшее число успехов быть больше трех?
4. Как находится параметр λ для формулы Пуассона?
5. Как находится параметр x для формулы локальной теоремы Муавра-Лапласа?
6. С помощью какой формулы дается оценка вероятности в интегральной формуле Муавра-Лапласа?
7. По какой формуле оценивается отклонение частости события от его вероятности в схеме Бернулли?

ТЕМА 5. Дискретные случайные величины. Основные определения и характеристики.

Основные вопросы темы

1. Понятие случайной величины. Основные определения.
2. Дискретные случайные величины. Способы определения.
3. Математическое ожидание и дисперсия.

5.1. Понятие случайной величины. Основные определения.

ОПРЕДЕЛЕНИЕ. Случайной величиной X называется однозначная числовая функция $f(\omega)$, определенная на пространстве элементарных событий Ω, которая каждому элементарному событию ω ($\omega \in \Omega$) ставит в соответствие число $f(\omega)$:

$$X=f(\omega).$$

Для обозначения случайных величин обычно используются прописные буквы латинского алфавита X, Y, Z, … ; соответствующие им строчные буквы x, y, z, … обозначают конкретные значения, которые они принимают.

Случайные величины подразделяются на *непрерывные* и *дискретные*.

ОПРЕДЕЛЕНИЕ. Случайная величина, принимающая конечное или бесконечное, но счетное множество значений называется *дискретной*, в противном случае она называется *непрерывной*.

ОПРЕДЕЛЕНИЕ. Соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями называется *законом распределения*.

5.2. Дискретные случайные величины.

Существует несколько способов задания закона распределения дискретной случайной величины:

- *табличный*;
- *графический*;
- *аналитический.*

Рассмотрим эти способы.

Табличный способ задания закона распределения. В данном случае, связь между значениями случайной величины (обычно перечисляемых в порядке возрастания) и соответствующими им вероятностями определяется посредством *таблицы распределения*:

X	x_1	x_2	$\ldots$	x_n	$\ldots$
P	p_1	p_2	$\ldots$	p_n	$\ldots$

Здесь первая строка содержит <u>все</u> возможные значения случайной величины, а вторая – их вероятности. Такую таблицу называют *рядом распределения*.

Ряд распределения может быть записан и в матричной форме:

$$X = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}.$$

ЗАМЕЧАНИЕ. Согласно определению случайной величины и ряда ее распределения, таблица распределения содержит *все возможные* и *не равные друг другу* значения случайной величины. Иными словами, события $\{X = x_1\}$, $\{X = x_2\}$, …, $\{X = x_n\}$ образуют полную группу. Следовательно, сумма их вероятностей равна единице:

$$\sum_i p_i = 1 \ .$$

Графический способ задания закона распределения. В данном случае, связь между значениями случайной величины и соответствующими им вероятностями определяется посредством соответствующего графика. При этом по оси абсцисс откладываются возможные значения случайной величины, а по оси ординат – вероятности этих значений. Ломанную линию, соединяющую последовательные точки графика (x_1, p_1), (x_2, p_2), … называют *многоугольником* (или *полигоном*) *распределения*.

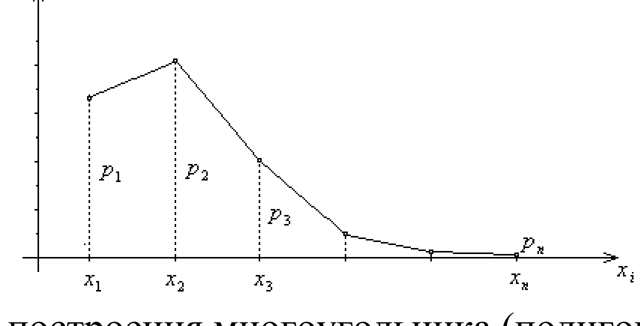


Рис. 6. Пример построения многоугольника (полигона) распределения вероятностей.

5.3. Математические операции над дискретными случайными величинами.

Сначала введем понятие *независимости случайных величин.*

ОПРЕДЕЛЕНИЕ. Две случайные величины X и Y называются *независимыми*, если события $(X = x_i)$ и $(Y = y_j)$ независимы для всех значений x_i и y_j. Это означает, в частности что:
$P[(X = x_i)\cdot(Y = y_j)] = P(X = x_i)\cdot P(Y = y_j)$.
Иными словами: случайные величины называются *независимыми*, если закон распределения одной из них остается неизменным и не зависит от значения другой случайной величины.

Пусть законы распределения двух независимых дискретных случайных величин X и Y заданы табличным способом:

X	x_1	x_2	…	x_n	…
P	p_1	p_2	…	p_n	…

Y	y_1	y_2	…	y_n	…
P	p_1	p_2	…	p_n	…

Произведение случайной величины *X* на постоянную величину *k*.

ОПРЕДЕЛЕНИЕ. Произведением kX случайной величины X на постоянную величину k называются случайная величина, которая принимает значения kx_i с теми же вероятностями p_i.

Таким образом, закон распределения случайной величины kX может быть записан как:

kX	kx_1	kx_2	…	kx_n	…
P	p_1	p_2	…	p_n	…

Возведение в степень случайной величины *X*.

ОПРЕДЕЛЕНИЕ. Возведением случайной величины X в степень m называется случайная величина, обозначаемая как X^m, которая принимает значения x_i^m с теми же вероятностями p_i.

Таким образом, закон распределения случайной величины X^m может быть записан как:

X^m	x_1^m	x_2^m	$\ldots$	x_n^m	$\ldots$
P	p_1	p_2	$\ldots$	p_n	$\ldots$

Сумма (разность, произведение) случайных величин *X* и *Y*.

ОПРЕДЕЛЕНИЕ. Суммой (разностью, произведением) случайных величин X и Y называется случайная величина, обозначаемая как $X+Y$ (X-Y, $X\cdot Y$) которая принимает все возможные значения вида $x_i + y_j$ ($x_i - y_j$, $x_i \cdot y_j$), где i=1,2,...,n и j=1,2,…,m с вероятностями p_{ij} того, что случайная величина X примет значение x_i, а случайная величина Y примет значение y_j:

$p_{ij} = P\left[(X = x_i)\cdot(Y = y_j)\right]$.

В случае независимости случайные величины X и Y:

$$p_{ij} = P\left[(X = x_i)\cdot(Y = y_j)\right] = P(X = x_i)\cdot P(Y = y_j).$$

ЗАМЕЧАНИЕ. При построении таблицы закона распределения суммы (разности или произведения) случайных X и Y может оказаться, что случайная величина X^m ($X+Y$, X-Y или $X\cdot Y$) принимает одно и тоже значение при различных значениях x_i и y_j. В этом случае данное значение заносятся в одну ячейку таблицы, а в соответствующую ей ячейку вероятности заносится суммарное значение соответствующих вероятностей.

5.4. Математическое ожидание дискретной случайной величины.

Пусть задан закон распределения случайной величины X.

X	x_1	x_2	x_3	$\ldots$	x_n
P	p_1	p_2	p_3	$\ldots$	p_n

ОПРЕДЕЛЕНИЕ. Математическое ожидание MX (или $M(X)$) дискретной случайной величины X определяется формулой:

$$MX = \sum_{i=1}^{n} x_i p_i$$

Свойства математического ожидания.

1. Если случайная величина x принимает одно и то же значение, то есть $x \equiv C$, то её математическое ожидание равно C.

2. Если $MX = a$, и k – константа, то $M(kX) = kMX$ (математическое ожидание случайной величины, умноженной на число, равно математическому ожиданию случайной величины, умноженному на это число). Иными словами: Постоянный множитель можно выносить за знак математического ожидания.

3. Если $MX = a$, и k - константа, то $M(k + X) = k + MX$ (математическое ожидание суммы случайной величины и числа равно сумме этого числа и математического ожидания случайной величины).

4. Математическое ожидание алгебраической суммы двух (или более) независимых случайных величин X и Y, определённых на одном и том же пространстве элементарных событий, равно алгебраической сумме их математических ожиданий:

$$M(X+Y)=MX+ MY$$

5. Математическое ожидание произведения двух (или более) <u>независимых</u> случайных величин X и Y, определённых на одном и том же пространстве элементарных событий, равно произведению их математических ожиданий:

$$M(XY) = MX \cdot MY$$

5.5. Дисперсия случайной величины.

ОПРЕДЕЛЕНИЕ. Дисперсия DX случайной величины X определяется формулой:

$$DX = M(X - MX)^2,$$

т.е: дисперсия случайной величины - это математическое ожидание квадрата отклонения случайной величины от её математического ожидания.

Используя определение дисперсии, для дискретной случайной величины формулу вычисления дисперсии можно представить в таком виде:

$$DX = \sum_{i=1}^{n} (x_i - MX)^2 p_i$$

Можно указать ещё одну формулу для вычисления дисперсии:

$$DX = M(X^2) - (MX)^2$$

Таким образом, дисперсия случайной величины равна разности математического ожидания квадрата случайной величины и квадрата её математического ожидания.

ЗАМЕЧАНИЕ: Дисперсия характеризует степень рассеяния значений случайной величины относительно её математического ожидания. Если все значения случайной величины тесно сконцентрированы около её математического ожидания и большие отклонения от математического ожидания маловероятны, то такая случайная величина имеет малую дисперсию. Если значения случайной величины рассеяны и велика вероятность больших отклонений от математического ожидания, то такая случайная величина имеет большую дисперсию.

Свойства дисперсии.

1. Дисперсия постоянной величины равна нулю:

$$D(C) = 0,$$

где $C = const$;

2. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(kX) = k^2 D(X).$$

3. Дисперсия алгебраической суммы конечного числа независимых случайных величин равна сумме их дисперсий:

$$D(X \pm Y) = DX + DY.$$

ТЕМА 6. Непрерывные случайные величины.

Основные вопросы темы

4. Непрерывные случайные величины.
5. Математическое ожидание, дисперсия и другие характеристики непрерывной случайной величины.
6. Функция распределения непрерывной случайной величины.

6.1. Непрерывные случайные величины.

Ранее было введено понятие непрерывной случайной величины X как случайной величины, принимающей бесконечное несчетное множество значений на некотором интервале (конечном или бесконечном) числовой оси.

Далее мы будем предполагать, что функция распределения непрерывной случайной величины обладает следующим свойством: *она непрерывна в любой точке и дифференцируема всюду, кроме, может быть, отдельных точек*.

ТЕОРЕМА. Можно показать, что вероятность любого отдельно взятого значения x^* непрерывной случайной величины X равна нулю:

$$p(X = x^*) = 0.$$

ОПРЕДЕЛЕНИЕ. *Плотностью вероятности* - $\varphi(x)$ (*плотностью распределения* или просто *плотностью*) случайной величины X называется производная ее функции распределения: $\varphi(x) = F'(x)$. График функции $\varphi(x)$ называют *кривой распределения*.

ЗАМЕЧАНИЕ. Плотность вероятности иногда также называют *дифференциальным законом распределения* или *дифференциальной функцией*.

6.1.1 Свойства плотности вероятности непрерывной случайной величины.

Свойство 1. Плотность вероятности $\varphi(x)$ – неотрицательная функция: $\varphi(x) \geq 0$.

Свойство 2. Связь между функцией распределения и плотностью распределения выражается как:

$$F(x) = \int_{-\infty}^{x} \varphi(x)dx$$

Свойство 3. Вероятность того, что непрерывная случайная величина X примет значение x, принадлежащее интервалу $[x_1, x_2]$ ($x \subset [x_1, x_2]$) равна приращению интегральной функции распределения на этом интервале, и, в силу свойства 2, может быть рассчитана как значение определенного интеграла от плотности распределения:

$$p(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} \varphi(x)dx .$$

Свойство 4. Интеграл плотности вероятности по всей области возможных значений случайной величины X равен 1. В частности, если случайная величина X определена на всей числовой оси, то:

$$\int_{-\infty}^{+\infty} \varphi(x)dx = 1 .$$

6.2. Математическое ожидание и дисперсия непрерывной случайной величины.

Введенные ранее понятия математического ожидания – $M(X)$ и дисперсии $D(X)$ для случая дискретной случайной величины могут быть распространены на непрерывный случай и рассчитаны по следующим формулам:

$M(X) = a = \int_{-\infty}^{+\infty} x\varphi(x)dx$ (если интеграл абсолютно сходится) и

$D(X) = \int_{-\infty}^{+\infty} (x-a)^2 \varphi(x)dx$ (если интеграл сходится).

6.2.1 Мода. Медиана. Квантили. Моменты случайной величины. Асимметрия и эксцесс.

ОПРЕДЕЛЕНИЕ. *Модой* - *Mo*(*X*) случайной величины *X* называется ее наиболее вероятное значение, т.е значение для которого вероятность - p_i (для дискретного случая) или плотность вероятности - $\varphi(x)$ (для случая непрерывной случайной величины) достигают своего максимального значения. Если функция плотности вероятности имеет несколько локальных максимумов, то распределение называют *полимодальным*.

ОПРЕДЕЛЕНИЕ. *Медианой* - *Me*(*X*) непрерывной случайной величины *X* называется такое ее значение для которого выполняется следующее условие:

$$p(X < Me(X)) = P(X > Me(X)) = \frac{1}{2} \ .$$

ОПРЕДЕЛЕНИЕ. *Квантилем уровня q* (*q-квантилем*) называется такое значение случайной величины, при котором функция ее распределения принимает значение равное *q*, т.е.:

$$F(x_q) = p(X < x_q) = q \, .$$

В частности, *медиана* является квантилем уровня 0,5 (т.е. $Me(X) = x_{0,5}$). Выделяют также квантили уровней 0,75 ($x_{0,75}$) и 0,25 ($x_{0,25}$), присвоив им названия *верхнего* и *нижнего квартилей*.

ОПРЕДЕЛЕНИЕ. *Начальным моментом k-ого порядка* случайной величины *X* называется математическое ожидание ее *k*-ой степени:

$$m_k = M\left(X^k\right).$$

Центральным моментом k-ого порядка случайной величины *X* называется математическое ожидание *k*-ой степени отклонения случайной величины *X* от ее математического ожидания:

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

В частности, математическое ожидание случайной величины есть ее начальный момент первого порядка, а дисперсия - ее центральный момент второго порядка.

Назначение (характер использования) начальных и центральных моментов различных порядков отображено в нижеследующей таблице.

Момент	Назначение (характер использования)
$m_1 = M(X)$	Начальный момент первого порядка (математическое ожидание – $M(X)$) характеризует среднее значение случайной величины, а также положение графика функции распределения относительно оси абсцисс.
$\mu_2 = D(X)$	Второй центральный момент (дисперсия – $D(X)$) характеризует степень рассеяния случайной величины относительно ее значения ее математического ожидания.
μ_3	Третий центральный момент характеризует характер асимметрии (скошенность) графика функции распределения относительно ее математического ожидания. Обычно используется безразмерная производная от этой характеристики, называемая *коэффициентом асимметрии*: $A = \frac{\mu_3}{\sigma^3}$. Если A=0, то распределение симметрично относительно математического ожидания.
μ_4	Четвертый центральный момент характеризует островершинность (плосковершинность) графика функции распределения. Используется производная от этой характеристики, называемая *эксцессом (коэффициентом эксцесса)*: $E = \frac{\mu_4}{\sigma^4} - 3$, которая позволяет сравнить данное распределение с нормальным распределением. Если $E > 0$, то исследуемое распределение является более островершинным, в противном случае ($E < 0$) – менее островершинным (т.е. плосковершинным).

6.3 Функция распределения случайной величины.

Выше, для описания *дискретной* случайной величины, использовано понятие закона распределения, устанавливающего соотношение между значениями случайной величины и соответствующими им вероятностями.

Это соотношение могло быть задано в *табличной* (*матричной*) форме, и в этом случае речь шла *о ряде распределения*.

При *графической форме* представления закона распределения используется *полигон* (*многоугольник*) *распределения вероятностей*.

Аналитическая формула, позволяющая осуществить непосредственный расчет вероятности значений дискретной случайной величины, также может быть использована для описания закона распределения.

Во всех этих случаях шла речь о установлении соответствия между *конкретными значениями* *дискретной* *случайной величины и соответствующими им вероятностями*.

Однако попытка использования этого подхода для *непрерывной* случайной величины сталкивается со значительными сложностями. Это связано, с тем, что непрерывная случайная величина имеет *бесконечное несчетное* множество возможных значений. С одной стороны, невозможно перечислить все эти значения, чтобы воспользоваться табличной или графической формой представления. С другой стороны, поскольку число возможных значений бесконечно велико, то вероятность того, что случайная величина примет некоторое конкретное значение (из всего бесконечного множества возможных) стремится к нулю. И этот факт, ограничивает применение аналитической формулы, устанавливавшей соотношение между конкретным значением случайной величины и его вероятностью.

Поэтому, наряду с описанным выше, разработан другой метод описания закона распределения случайной величины, применимый как для дискретного, так и непрерывного случая. Для этого подхода определяющим является введение понятия *функции распределения случайной величины* - $F(X)$, которая (в отличии от ранее рассмотренного) устанавливает связь конкретного значения x *не с вероятностью* $p(X=x)$, а с вероятностью того, что случайная величина X примет значение меньшее, чем x: $p(X<x)$.

ОПРЕДЕЛЕНИЕ. *Функцией распределения случайной величины* X называется функция $F(x)$, выражающая для каждого x вероятность того, что случайная величина X примет значение меньшее x:

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

Функцию $F(x)$ называют также *интегральной функцией распределения* (или *интегральным законом распределения*).

ЗАМЕЧАНИЕ. Функция распределения является разновидностью закона распределения для всех типов случайных величин.

Геометрическая интерпретация. Геометрически, функция распределения интерпретируется как вероятность того, что случайная величина X попадет левее заданной точки x числовой оси.

Графическое представление. Функция распределения дискретной случайной величины есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным значениям случайной величины и равны вероятностям этих значений. Сумма всех скачков функции $F(x)$ равна 1.

6.3.1 Свойства функции распределения.

Свойство 1. Функция распределения случайной величины $F(x)$ – неотрицательна, и ее значения заключены между нулем и единицей:

$$0 \le F(x) \le 1 .$$

Свойство 2. Функция распределения $F(x)$ есть неубывающая функция:

$$F(x_2) \ge F(x_1) \quad \textit{при} \quad x_2 > x_1 .$$

Свойство 3. Если непрерывная случайная величина X определена на всей числовой оси, то предельные значения функции распределения $F(x)$ при стремлении x к бесконечности (как '+' , так и '-') определяются как:

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

Свойство 4. Вероятность попадания случайной величины X в полуоткрытый интервал $[x_1, x_2)$ равна приращению ее функции распределения на этом интервале:

$$p(x_1 \le X \le x_2) = F(x_2) - F(x_1) .$$

Контрольные вопросы по темам 5, 6.

1. Что служит аргументом для функции, являющейся случайной величиной?
2. Вероятность какого события является значение функции распределения случайной величины?
3. Каким является множество значений дискретной случайной величины?
4. Какое условие определяет независимость случайных величин?
5. Перечислите известные Вам арифметические операции над случайными величинами.
6. Какая характеристика имеет смысл среднего значения случайной величины?
7. Какая характеристика оценивает степень рассеивания случайной величины?
8. Что определяет величина скачка функции распределения дискретной случайной величины?
9. Чему равны наименьшее и наибольшее значения функции распределения?
10. Как связаны функция распределения и плотность распределения непрерывной случайной величины?

ТЕМА 7. Основные законы распределения случайной величины.

7.1 Основные законы распределения дискретной случайной величины.

Основные вопросы темы

1. Биномиальный закон распределения.
2. Закон распределения Пуассона.
3. Геометрическое и гипергеометрическое распределение

7.1.1. Биномиальный закон распределения.

ОПРЕДЕЛЕНИЕ. Дискретная случайная величина X имеет *биномиальный закон распределения* с параметрами n и $p \in [0,1]$, что символически обозначается как $X \sim \mathrm{Bi}(n, p)$, если вероятность того, что она примет значение 0, 1, 2, ..., m, …, n соответственно равна:

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

где: q=1-p.

Отметим, что выражение $C_n^m p^m q^{n-m}$ представляет собой формулу Бернулли[5]. Следовательно биномиальный закон распределения описывает распределение числа наступления некоторого события - m при проведении n испытаний по схеме Бернулли.

Построим ряд распределения биномиального закона:

x_i	0	1	2	…	m	…	n
p_i	q^n	$C_n^1 p^1 q^{n-1}$	$C_n^2 p^2 q^{n-2}$	…	$C_n^m p^m q^{n-m}$	…	p^n

Проверим корректность записи закона распределения, для чего рассмотрим сумму всех вероятностей, представленных второй стро-

[5] Напомним, что формула Бернулли определяет вероятность того, что при проведении n испытаний (по схеме Бернулли) некоторое событие (вероятность появления которого равна p) произойдет ровно m раз.

кой таблицы. Заметим, что эта строка содержит элементы разложения бинома Ньютона[6]:

$$(q+p)^n = q^n + C_n^1 p^1 q^{n-1} + C_n^2 p^2 q^{n-2} + \ldots + C_n^m p^m q^{n-m} + \ldots + q^n = \sum_{i=1}^{n} C_n^i p^i q^{n-i}$$

Поскольку $(q+p)=1$, то:

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} C_n^i p^i q^{n-i} = (q+p)^n = 1.$$

Следовательно, сумма всех элементов второй строки равна единице, что завершает проверку корректности записи закона распределения.

Значения величины математического ожидания и дисперсии случайной величины, имеющей биномиальный закон распределения, определяются следующей теоремой.

ТЕОРЕМА. Математическое ожидание – $M(X)$ и дисперсия $D(X)$ дискретной случайной величины X, распределенной по биномиальному закону, могут быть найдены из следующих выражений:

$$M(X) = np,$$

$$D(X) = npq.$$

ЗАМЕЧАНИЕ. Биномиальный закон распределения находит свое применение в статистическом контроле качества продукции, описании систем массового обслуживания и других приложениях.

7.1.2. Закон распределения Пуассона.

ОПРЕДЕЛЕНИЕ. Дискретная случайная величина X имеет *закон распределения Пуассона* с параметром $\lambda > 0$, что символически обозначается как $X \sim \Pi(\lambda)$, если вероятность того, что она примет значение 0, 1, 2 ,..., m соответственно равна:

$$P(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}.$$

Построим ряд распределения закона Пуассона:

x_i	0	1	2	…	m	…

[6] Исаак Ньютон (1642-1726г). Величайший английский ученый: философ, физик и математик, автор известнейших физических законов, математических трудов и др. В знак признания его выдающихся заслуг избран президентом английского королевского общества (1703г), а в 1705г король Англии возвел его в рыцарское достоинство.

p_i	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\dfrac{\lambda^2 e^{-\lambda}}{2!}$	$\dots$	$\dfrac{\lambda^m e^{-\lambda}}{m!}$	$\dots$

Проверим корректность записи закона распределения, для чего рассмотрим сумму всех вероятностей, представленных второй строкой таблицы - $\sum_{i=1}^{n} p_i$:

$$\sum_{m=1}^{\infty} p_i = e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2 e^{-\lambda}}{2!} + \ldots + \frac{\lambda^m e^{-\lambda}}{m!} + \ldots = \sum_{m=1}^{\infty} \frac{\lambda^m e^{-\lambda}}{m!} = e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{m!}$$

Заметим, что выражение $\sum_{m=1}^{\infty} \frac{\lambda^m}{m!}$ представляет собой разложение в ряд функции e^x при $x = \lambda$, т.е:

$$\sum_{m=1}^{\infty} \frac{\lambda^m}{m!} = e^{\lambda}$$

Поэтому:

$$\sum_{m=1}^{\infty} p_i = e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} \cdot e^{\lambda} = 1 .$$

Следовательно, сумма всех элементов второй строки равна единице, что и завершает проверку корректности записи закона распределения.

Значения величины математического ожидания и дисперсии случайной величины, имеющей закон распределения Пуассона, определяются следующей теоремой.

ТЕОРЕМА. Числовые значения математического ожидания – $M(X)$ и дисперсия $D(X)$ дискретной случайной величины X, распределенной закону Пуассона, могут быть найдены из следующих выражений:

$$M(X) = \lambda,$$
$$D(X) = \lambda.$$

ЗАМЕЧАНИЕ 1. Хотя значения математического ожидания – $M(X)$ и дисперсии $D(X)$ дискретной случайной величины X, распределенной закону Пуассона, численно совпадают, однако физическая размерность этих величин – различна.

ЗАМЕЧАНИЕ 2. Закон распределения Пуассона находит широкое применение, в приложениях, связанных с описанием и моделированием систем массового обслуживания, выполнения приближенных вычислений, и др.

7.1.3. Геометрическое и гипергеометрическое распределения.

Геометрическое распределение.

ОПРЕДЕЛЕНИЕ. Дискретная случайная величина X имеет *геометрическое распределение*, если вероятность того, что она примет значение 1, 2, 3, ... (счетное множество значений) соответственно равна:

$$P(X=m)=pq^{m-1},$$

где: $p \in (0,1)$; и $q=1-p$.

Построим ряд распределения геометрического закона:

x_i	1	2	3	...	m	...
p_i	p	pq	pq^2	...	pq^{m-1}	...

Проверим корректность записи закона распределения, для чего рассмотрим сумму всех вероятностей, представленных второй строкой таблицы - $\sum_{i=1}^{n} p_i$:

$$\sum_{i=1}^{\infty} p_i = p + pq + pq^2 + \ldots + pq^{i-1} + \ldots = p\sum_{i=1}^{\infty} q^{i-1}$$

Заметим, что выражение $\sum_{i=1}^{\infty} q^{i-1}$ представляет собой сумму членов геометрической прогрессии (откуда и происходит название геометрического распределения) с первым членом, равным единице и знаменателем q. Используя известную формулу суммы бесконечного числа членов геометрической прогрессии, получим:

$$\sum_{i=1}^{\infty} q^{i-1} = \frac{1}{1-q} = \frac{1}{p}.$$

Откуда:

$$\sum_{i=1}^{\infty} p_i = p\sum_{i=1}^{\infty} q^{i-1} = p\cdot\frac{1}{p} = 1.$$

Следовательно, сумма всех элементов второй строки равна единице, что и завершает проверку корректности записи закона распределения.

Значения величины математического ожидания и дисперсии случайной величины, имеющей геометрический закон распределения, определяются следующей теоремой.

ТЕОРЕМА. Математическое ожидание – $M(X)$ и дисперсия $D(X)$ дискретной случайной величины X, распределенной по геометрическому закону, могут быть найдены из следующих выражений:

$$M(X) = \frac{1}{p}, \ \ и \ \ D(X) = \frac{q}{p^2},$$

где: $q=1-p$.

Ниже, на рисунке представлен пример геометрического распределения для значения p=0,2.

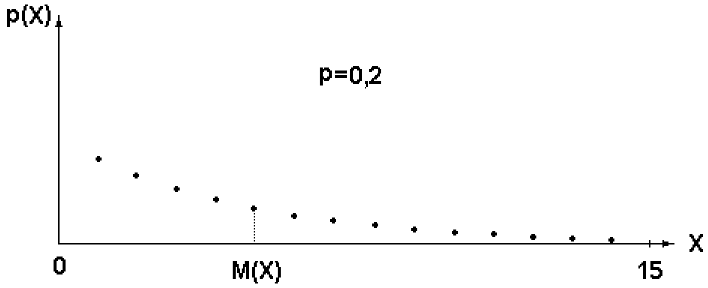


Рис. Пример геометрического распределения для значения p=0,2.

Геометрическое распределение тесно связано со схемой испытаний Бернулли, а следовательно с биномиальным распределением. Отличие состоит в том, что биномиальная случайная величина определяет вероятность m успехов в n испытаниях, а геометрическая - вероятность m испытаний до первого успеха (включая первый успех).

ПРИМЕР. Вероятность того, что покупатель ознакомился заранее с рекламой товара равна 0,6 (p=0,6). Осуществляется выборочный контроль качества рекламы путем опроса покупателей до первого, изучившего рекламу заранее. Составить ряд распределения количества опрошенных покупателей.

Решение. Согласно условию задачи p = 0,6. Откуда: q = 1-p = 0,4. Подставив данные значения, получим: $p(X=m) = p \cdot q^{m-1} = 0{,}6 \cdot 0{,}4^{m-1}$ и построим ряд распределения:

X	1	2	…	m	…
p_i	0,6	0,24	…	$0{,}6 \cdot 0{,}4^{m-1}$	…

ЗАМЕЧАНИЕ. Геометрическое распределение находит свое применение при решении задач, связанных с контролем качества продукции.

Гипергеометрическое распределение.

ОПРЕДЕЛЕНИЕ. Дискретная случайная величина X имеет *гипергеометрическое распределение* с параметрами n, M, е N если вероятность того, что она примет значение 1, 2, 3, ..., $min(n, M)$ соответственно равна:

$$P(X = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n},$$

где: n, M, е N - натуральные числа, $M \le N$; и $n \le N$.

Интерпретация гипергеометрического распределения. Пусть имеется множество N объектов, среди которых M обладают некоторым свойством. Из этого множества случайным образом извлекаются n объектов. Гипергеометрическое распределение описывает вероятность того, что среди извлеченных имеется некоторое количество m, обладающих отмеченным свойством - $p(X = m)$.

ПРИМЕР. В партии из N изделий имеется M ($M < N$) доброкачественных и N-M - с браком. Пусть, для контроля из всей партии отбирают случайным образом n изделий. Число доброкачественных изделий – m в такой контрольной партии - случайная величина, которую обозначим X. Можно показать, что распределение этой случайной величины описывается гипергеометрическим распределением:

$$p(X = m) = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}, \text{ где: } m = 0, 1, 2, \dots, \min(n, M)$$

Значения величины математического ожидания и дисперсии случайной величины, имеющей гипергеометрический закон распределения, определяются следующей теоремой.

ТЕОРЕМА. Математическое ожидание – $M(X)$ и дисперсия $D(X)$ дискретной случайной величины X, распределенной по гипергеометрическому закону, могут быть найдены из следующих выражений:

$$M(X) = n\frac{M}{N},$$

$$D(X)=n\frac{M}{N-1}\left(1-\frac{M}{N}\right)\left(1-\frac{n}{N}\right).$$

ЗАМЕЧАНИЕ. Гипергеометрическое распределение определяется тремя параметрами: N, M, n. Можно показать, что при $N>>n$ (практически при $N>10n$) данное распределение приближается к биномиальному с параметрами n и $p=\frac{M}{N}$.

ПРИМЕР. Согласно правилам спортивной лотереи "Спортлото 6 из 45", денежные призы получают участники, угадавшие от трёх до шести угаданных ими чисел (из отобранных ими из 45). Определить закон распределения, математическое ожидание и дисперсию числа угаданных чисел (случайной величины X) среди случайно отобранных шести. Какова вероятность получения денежного приза?

Решение. Случайная величина X, описывающая число угаданных чисел среди шести выбранных имеет гипергеометрическое распределение с параметрами $n=6$, $N=45$, $M=6$. Ряд распределения X, рассчитанный по формуле:

$$p(X=m)=\frac{C_M^m\cdot C_{N-M}^{n-m}}{C_N^n}=\frac{C_6^m\cdot C_{39}^{6-m}}{C_{45}^6}\ ,\ \text{где: } m=0,1,2,\dots,6$$

имеет вид:

X	0	1	2	3	4	5	6
p_i	0,40056	0,42413	0,15147	0,02244	0,00137	0,00003	0,0000001

Вероятность получения денежного приза:

$$P(3\le X\le 6)=P(X=3)+P(X=4)+P(X=5)+P(X=6).$$

Подставляя полученные значения, получим:

$$P(3\le X\le 6)=0{,}02244+0{,}00137+0{,}00003+0{,}0000001\approx 0{,}024.$$

Вычислим, также, значения математического ожидания и дисперсии:

$$M(X)=n\frac{M}{N}=6\cdot\frac{6}{45}=0{,}8,$$

$$D(X)=n\frac{M}{N-1}\left(1-\frac{M}{N}\right)\left(1-\frac{N}{M}\right)=6\cdot\left(1-\frac{39}{45}\right)\left(1-\frac{6}{45}\right)=0{,}6145$$

Таким образом, среднее число угаданных чисел достаточно мало (меньше единицы), а вероятность выигрыша составляет всего лишь 0,024 (около 2%).

ЗАМЕЧАНИЕ. Гипергеометрическое распределение используется в практике статистического контроля качества промышленной продукции, в задачах, связанных с организацией выборочных обследований и некоторых других областях.

7.2 Основные законы распределения непрерывной случайной величины.

Основные вопросы темы

1. Равномерный закон распределения.
2. Показательный (экспонециальный) закон распределения.
3. Нормальный закон распределения.
4. Правило трех сигм.

7.2.1. Равномерный закон распределения.

ОПРЕДЕЛЕНИЕ. Распределение вероятностей случайной величины X на интервале [a, b] называется *равномерным*, если плотность вероятности $\varphi(x)$ постоянна на этом интервале и равна нулю вне него (т.е.: $\varphi(x) = C = const$, если $x\in$[a, b]), и $\varphi(x) = 0$, если $x\notin$[a, b]).

Ранее было показано, что плотность вероятности обладает следующим свойством: $\int\limits_{-\infty}^{+\infty}\varphi(x)dx = 1$. Подставляя, получим:

$$\int\limits_{-\infty}^{+\infty}\varphi(x)dx = \int\limits_{-\infty}^{a}\varphi(x)dx + \int\limits_{a}^{b}\varphi(x)dx + \int\limits_{b}^{\infty}\varphi(x)dx =$$

$$= \int\limits_{-\infty}^{a}0dx + \int\limits_{a}^{b}Cdx + \int\limits_{b}^{\infty}0dx = C(b-a) = 1 \quad\Rightarrow\quad C = \frac{1}{(b-a)}.$$

Откуда:

$$\varphi(x) = \begin{cases} 0 & , \quad x < a \\ \dfrac{1}{(b-a)} & , \quad a \le x \le b \\ 0 & , \quad x > b \end{cases}.$$

График этой функции изображен ниже.

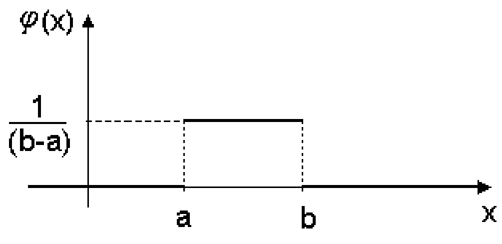


Рис. 7. Плотность вероятности $\varphi(x)$ для случая равномерного закона распределения

Функция распределения $F(X)$ может быть найдена путем интегрирования плотности вероятности: $F(x)=\int_{-\infty}^{x}\varphi(t)dt$. Таким образом:

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

График этой функции представлен на следующем рисунке.

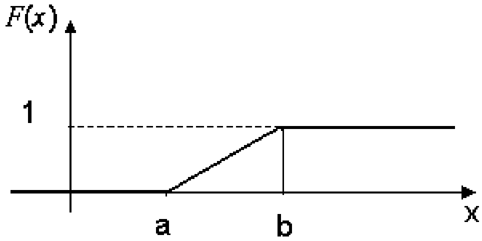


Рис. 8. Функция распределения $F(x)$ для случая равномерного закона распределения

Значения величины математического ожидания и дисперсии случайной величины, имеющей равномерный закон распределения, определяются следующей теоремой.

ТЕОРЕМА. Математическое ожидание непрерывной случайной величины с равномерным законом распределения вычисляется как:

$$M(X)=\frac{a+b}{2} \ ,$$

а ее дисперсия:

$$D(X)=\frac{(b-a)^2}{12} \ .$$

ЗАМЕЧАНИЕ. Равномерный закон распределения используется, в частности, при анализе ошибок округления, при проведении численных вычислений, а также, как исходная для моделирования иных законов распределения.

7.2.2. Показательный (экспонециальный) закон распределения

ОПРЕДЕЛЕНИЕ. *Показательный* (*экспонециальный*) *закон распределения* непрерывной случайной величины X задается плотностью вероятности вида:

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x} & при \quad x \geq 0 \\ 0 & при \quad x < 0 \end{cases}.$$

Можно показать, что функция распределения случайной величины в этом случае имеет вид:

$$F(x) = \begin{cases} 1 - \lambda e^{-\lambda x} & при \quad x \geq 0 \\ 0 & при \quad x < 0 \end{cases};$$

ее математическое ожидание равно:

$$M(X) = \frac{1}{\lambda};$$

дисперсия находится как:

$$D(X) = \frac{1}{\lambda^2};$$

а среднее квадратическое отклонение:

$$\sigma_x = \frac{1}{\lambda}.$$

Показательный закон распределение находит свое применение в теории массового обслуживания, и др.

7.2.3. Нормальный закон распределения.

Нормальный закон распределения (*гауссово распределение*) находит широкое применение при анализе различных реальных ситуаций. Этот связано с тем, что этот закон описывает предельное распределение, к которому стремятся другие законы распределения, часто применяемые на практике.

ОПРЕДЕЛЕНИЕ. *Нормальный закон распределения* (*закон Гаусса*) непрерывной случайной величины X задается плотностью вероятности $\varphi_N(x)$ вида:

$$\varphi_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где: a и σ - параметры распределения.

ЗАМЕЧАНИЕ 1. Можно показать, что параметры a и σ, входящие в формулу плотности распределения случайной величины X, равны, соответственно, ее математическому ожиданию и среднему квадратическому отклонению:

$$a = M(X),$$

$$\sigma = \sigma_x,$$

а, следовательно, дисперсия $D(X)$ равна:

$$D(X) = \sigma_x^2 = \sigma^2.$$

В случае нормального закона, график функции распределения (или плотности вероятности) называют *нормальной* (или *гауссовой*) кривой.

ЗАМЕЧАНИЕ 2. Для случайной величины, обладающей нормальным законом распределения, обычно используют специальное обозначение - $N(a, \sigma)$, или (имея в виду дисперсию) $N(a, \sigma^2)$.

ОПРЕДЕЛЕНИЕ. Нормальный закон распределения случайной величины с параметрами $a = 0$ и $\sigma^2 = 1$ называется *стандартным* (или *нормированным*) имеет обозначение $N(0,1)$, а соответствующая нормальная кривая – *стандартной*, или *нормированной* (это название тесно связано с введенным ранее понятием *стандартной случайной величины*, имеющей нулевое математическое ожидание единичную дисперсию).

Таким образом, плотность вероятности стандартного закона распределения - $\varphi_N^*(x)$ может быть записана как:

$$\varphi_N^*(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Данная функция является четной (в силу этого, симметричной относительно оси абсцисс), и достигает своего максимума, равного $\frac{1}{\sqrt{2\pi}}$, в точке $x = 0$. Далее функция монотонно убывает до нуля при х $\to \pm\infty$ ($\lim\limits_{x\to\pm\infty} \varphi_N^*(x) = 0$). График данной функции представлен ниже.

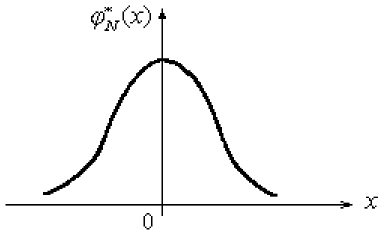


Рис. 9. Плотность вероятности $\varphi_N^*(x)$ для случая стандартного закона распределения.

На следующем рисунке представлен график плотности вероятности $N(a,\sigma^2)$ нормальный закон распределения, с параметрами $a \neq 0$ и $\sigma^2 \neq 1$ ($N(a,\sigma^2) \neq N(0,1)$).

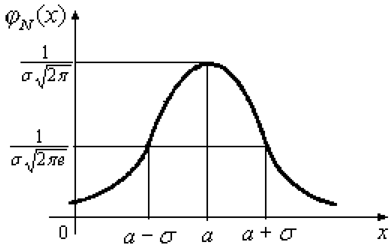


Рис. 10. Плотность вероятности $\varphi_N(x)$ для случая нормального закона распределения с параметрами $a \neq 0$ и $\sigma^2 \neq 1$ ($N(a,\sigma^2) \neq N(0,1)$).

Исследуем как меняется график плотность вероятности $\varphi_N(x)$ нормального закона распределения $N(a,\sigma^2)$ при изменении параметров *a и* σ. Рисунок ниже демонстрирует, что при изменении параметра *a* график функции сдвигается влево ($a < 0$) или вправо ($a > 0$).

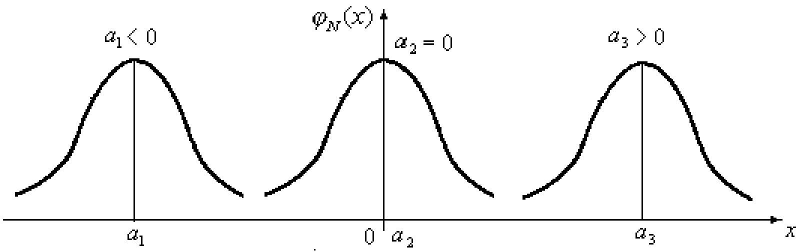


Рис. 11. Плотность вероятности $\varphi_N(x)$ нормального закона распределения $N(a,\sigma^2)$ для различных значений математического ожидания – *a*. График функции сдвигается влево ($a<0$) или вправо ($a>0$).

Следующий рисунок демонстрирует изменение графика функции $\varphi_N(x)$ нормального закона распределения $N(a,\sigma^2)$ при изменении параметра σ.

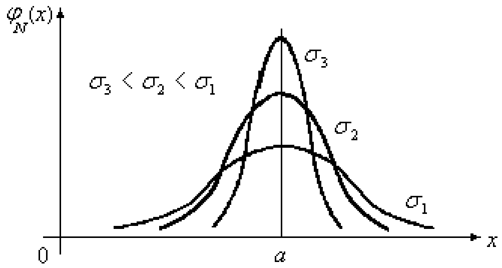


Рис. 11. Плотность вероятности $\varphi_N(x)$ нормального закона распределения $N(a,\sigma^2)$ для различных значений среднего квадратичного отклонения σ.

ВЫВОД. Параметр *a* (математическое ожидание) характеризует положение центральной оси графика, а σ^2 (дисперсия) - форму нормальной кривой.

7.2.4. Правило трех сигм.

На практике часто используется следующее правило, получившее название "*Правило трех сигм*":

Если случайная величина X имеет нормальный закон распределения $N(a,\sigma^2)$ с параметрами a и σ, то практически достоверно, что ее значения заключены в интервале (a-3 σ, a+3 σ).

Данное утверждение может быть проиллюстрировано следующим образом:

Пусть имеется нормально распределённая случайная величина *x* с математическим ожиданием, равным *a* и дисперсией σ^2. Определим вероятность попадания ξ в интервал $(a - 3\sigma;\, a + 3\sigma)$, то есть вероятность того, что *x* принимает значения, отличающиеся от математического ожидания не более, чем на три среднеквадратических отклонения:

$$P(a - 3\sigma < x < a + 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3).$$

По таблице находим $\Phi(3)=0{,}49865$, откуда следует, что $2\Phi(3)$ практически равняется единице. Таким образом, можно сделать важный вывод: *Практически достоверно, что нормальная случайная величина принимает значения, отклоняющиеся от ее математического ожидания не более чем на 3σ.*

7.3 Некоторые употребительные распределения математической статистики.

Основные вопросы темы

1. Распределение χ^2 (хи квадрат) с k степенями свободы.
2. Распределение Стьюдента.
3. Распределением Фишера-Снедкора.
4. Распределение Вейбула.

7.3.1. Распределение χ^2 (хи-квадрат)

ОПРЕДЕЛЕНИЕ. *Распределением χ^2 (хи квадрат) с k степенями свободы* называется распределение суммы квадратов k независимых случайных величин, каждая из которых распределена по стандартному нормальному закону, т.е:

$$\chi^2 = \sum_{i=1}^{k} Z_i^2 ,$$

где: Z_i (i = 1, 2, …, n) - набор n независимых, нормально распределенных случайных величин, $Z_i \sim \boldsymbol{N}(0,1)$.

Плотность вероятности распределения χ^2 определяется выражением:

$$\varphi(x) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(\dfrac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} & при\ x \ge 0 \\ 0 & при\ x < 0 \end{cases},$$

где: $\Gamma(y) = \int\limits_0^{+\infty} e^{-t} t^{y-1} dt$ - гамма-функция Эйлера (можно показать, что для целых положительных значений аргумента гамма-функция Эйлера принимает более простой вид: $\Gamma(y) = (y-1)!$).

Графики функции $\varphi(x)$, называемые *кривыми Пирсона*[7], обладают положительной правосторонней асимметрией, и при $n \ge 2$ имеют один максимум в точке $x = n\text{-}2$ (см. рисунок ниже).

[7] Карл Пирсон (Pearson, 1857-1936) английский математик и биолог. С целью проверки теории Дарвина разработал статистический метод, получивший широкое распространение при исчислении коэффициента корреляции между различными переменными. В частности, в 1900 г. им предложен критерий '*хи-*

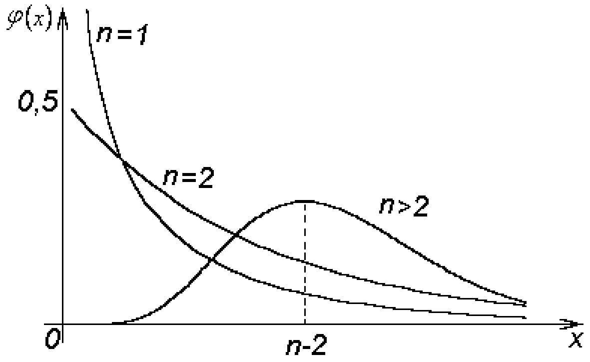


Рис. 12. Кривые Пирсона для различных значениях n.

ЗАМЕЧАНИЕ 1. Можно показать, что сумма любого числа - m независимых случайных величин $X_k, k=1,2,...,m$, имеющих распределение хи-квадрат с n_k степенями свободы, также имеет распределение хи-квадрат с n_Σ, степенями свободы, где:

$$n_\Sigma = \sum_{k=1}^{n} n_k \ .$$

ЗАМЕЧАНИЕ 2. Распределение хи-квадрат имеет многочисленные приложения в математической статистике.

7.3.2. Распределение Стьюдента.

ОПРЕДЕЛЕНИЕ. *Распределением Стьюдента*[8] (или t-распределением) называется распределение случайной величины

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}}$$

где: Z - случайная величина, распределенная по стандартному нормальному закону, т.е: $Z \sim \boldsymbol{N}(0,1)$;

χ^2 – независимая от Z случайная величина, имеющая распределение χ^2 с k степенями свободы.

квадрат'. В литературе часто упоминается осуществленный Пирсоном опыт по экспериментальной проверке вероятности выпадения герба при подбрасывании монеты. Из 24000 подбрасываний, герб выпал 12012 раз.

[8] Уильям Госсет (1876-1937) – английский статистик, писавший под псевдонимом “Student” (студент).

Плотность вероятности распределения Стьюдента определяется выражением:

$$\varphi(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{x^2}{n}\right)^{-\frac{k+1}{2}},$$

где: $\Gamma(y) = \int_0^{+\infty} e^{-t} t^{y-1} dt$ - гамма-функция Эйлера.

Графики функции $\varphi(x)$, называемые *кривыми Стьюдента* симметричны при всех n = 1,2,... относительно оси ординат (см рисунок ниже).

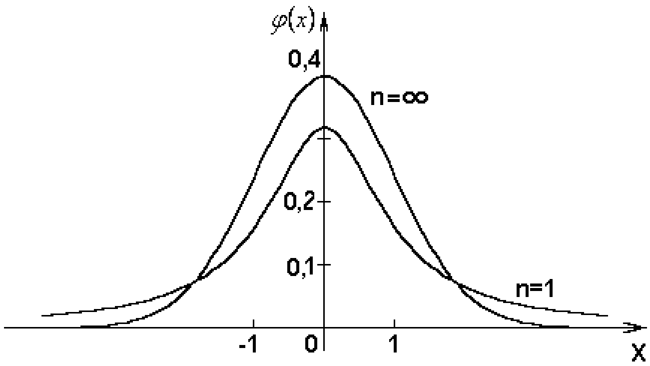


Рис. 13. Кривые Стьюдента.

ЗАМЕЧАНИЕ. При $n > 30$ распределение Стьюдента практически не отличается от **N**(0,1). Однако при $n \leq 30$ отличия существенны.

7.3.3. Распределение Фишера-Снедекора

ОПРЕДЕЛЕНИЕ. *Распределением Фишера-Снедкора (F-распределением)* называется распределение случайной величины:

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)},$$

где: $\chi^2(k_1)$ и $\chi^2(k_2)$ случайные величины, имеющие распределение χ^2 соответственно с k_1 и k_2 степенями свободы.

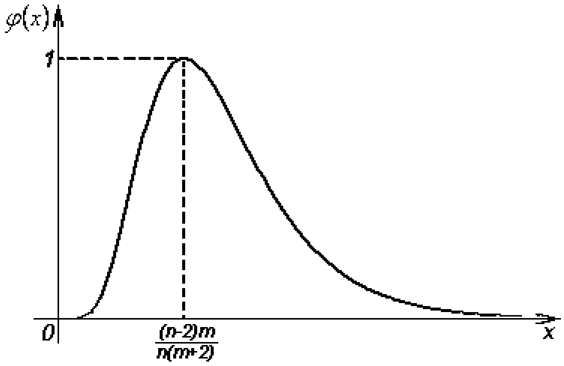
Плотность вероятности F-распределения имеет вид:

$$\varphi(x)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)k_1^{\frac{k_1}{2}}k_2^{\frac{k_2}{2}}}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}x^{\frac{k_1}{2}-1}\left(k_1x+k_2\right)^{-\frac{k_1+k_2}{2}},$$

где: $\Gamma(y)$ - гамма-функция Эйлера.

ЗАМЕЧАНИЕ. При $n\to\infty$ F-распределение приближается к нормальному.

Графики функции $\varphi(x)$ (см. рисунок ниже), называемые *кривыми Фишера*, асимметричны и достигают максимальных значений в окрестности точки $x=\frac{(n-2)m}{(m+2)n}$, близкой к единице.



Рису. 14. Распределение Фишера

7.3.4. Распределение Вейбула.

Распределение Вейбула[9] используется в различных приложениях, в частности, для: статистического контроля качества; описания распределения времени безотказной работы элементов и блоков различных устройств, машин и механизмов, 'усталости' и надежности конструкций, и др.

Данное распределение имеет две разновидности: двухпараметрическое и трехпараметрическое.

Плотность вероятности трехпараметрического распределения Вейбула описывается формулой:

[9] Доктор Е. Г. Валодди Вейбул (E. H. Waloddi Weibull, 1887-1979). Профессор Шведского Королевского Института Технологии. Автор ряда статистических исследований в области надежности различных устройств.

$$\varphi(x)=\frac{c}{b}\left(\frac{x-\theta}{b}\right)^{c-1}e^{-\left(\frac{x-\theta}{b}\right)^{c}},$$

а математическое ожидание – *M(X)* и дисперсия – *D(X)*, соответственно равны:

$$M(X)=\theta+b\Gamma\left(\frac{c+1}{c}\right),$$

$$D(X)=\left[\Gamma\left(\frac{c+1}{c}\right)-\Gamma^{2}\left(\frac{c+1}{c}\right)\right],$$

где $b, c, и\ \theta$ - параметры распределения.

Ниже приводится график функции распределения Вейбула.

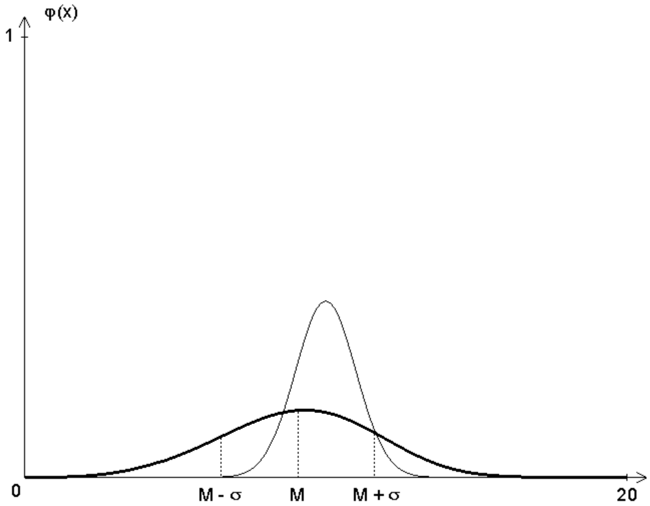


Рис. 15. График функции распаределения $\varphi(x)$ при $b=10, c=4, и\ \theta=0$.

Пример распределения Вейбула (приводится по интернет-источнику http://belyakov-as.narod.ru/Site/weibull.htm).

Известно, что "период полураспада" знаний о компьютерных технологиях составляет 2 года. Это значит, что каждые два года половина наших знаний безнадежно устаревает. Коэффициент изменения остаточных полезных знаний по прошествии *k* лет подчиняется распределению Вейбула.

Доказательство. Предположим, что *A* - первоначальное количество полезной информации. Тогда через 2 года количество полезной информации станет равным $A\times2^{-1}$, а по прошествии *k* лет примет вид $A\times2^{-\frac{k}{2}}$. Коэффициент изменения остаточных полезных знаний равен

$2^{-\frac{k}{2}} = \left(e^{\ln 2}\right)^{-\frac{k}{2}} = e^{-\frac{k}{2}\ln 2}$. Введем новую переменную $x = \frac{k}{2}\ln 2$. Коэффициент изменения остаточных полезных знаний примет следующий вид: e^{-x}. Итоговая формула совпадает с формулой плотности вероятности для распределения Вейбула при $b = 1, c = 1, \theta = 0$.

Контрольные вопросы по темам 7.1, 7.2, 7.3

1. Какие значения может принимать биномиально распределенная случайная величина?
2. Какими параметрами определяется биномиальное распределение?
3. Какие значения может принимать случайная величина, имеющая геометрическое распределение?
4. Какими параметрами определяется геометрическое распределение?
5. Какой смысл имеют вероятности p_m в гипергеометрическом распределении?
6. Какие значения может принимать пуассоновская случайная величина?
7. Какой смысл имеет параметр λ в пуассоновском распределении?
8. Чему равны математическое ожидание и дисперсия равномерной случайной величины?
9. Может ли случайная величина, распределенная по показательному закону, принимать отрицательные значения?
10. Как влияют параметры a и σ нормального распределения на форму графика его плотности распределения?

ЧАСТЬ II. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Тема 8.1. Предельные теоремы теории вероятностей.

Основные вопросы темы

1. Введение.
2. Предельных теоремы теории вероятностей и их значение.
3. Неравенство Чебышева. Неравенство Маркова.
4. Закон больших чисел. Центральная предельная теорема.

8.1.1. Введение

Как отмечалось ранее, *математическая статистика* - это раздел математики, изучающий математические методы сбора, систематизации, обработки и интерпретации результатов наблюдений с целью выявления *статистических закономерностей*: т.е: закономерностей достаточно общего характера, присущих данному классу наблюдений и изучаемых явлений. Таким образом, методы математической статистики используются для получения научно обоснованных выводов о *массовых явлениях и процессах* по данным наблюдений и экспериментов.

Поскольку результаты эксперимента заранее неизвестны, и по существу, часто носят непредсказуемый характер (в силу наличия неизвестных влияющих факторов, ошибок измерений, и.т.д.), то естественным фундаментом оценки такого рода измерений и экспериментов стала теория вероятностей. Возможность такого подхода обосновывается рядом специальных теорем, рассмотренных ниже. Полученные выводы относятся не к отдельным результатам эксперимента (испытаниям, из повторения которых складывается данное массовое явление), а представляют собой утверждения об общих статистических характеристиках, выражаемых посредством рассмотренных ранее понятий теории вероятностей, в частности: вероятностях, законах распределения, математических ожиданиях, дисперсиях и др. Поэтому, одной из важных задач математической статистики является оценка адекватности идеальных вероятностных моделей рассматриваемым *реальным* процессам.

8.1.2. Предельные теоремы теории вероятностей и их значение.

Важное значение имеет группа теорем, объединенная общим названием *предельных теорем теории вероятности*, которые являются теоретическим обоснованием (фундаментом) применения рассмотренных ранее методов теории вероятностей для статистической обработки результатов экспериментальных исследований. Она может быть условно разделена на две подгруппы, называемых, соответственно, *законом больших чисел* и *центральной предельной теоремой*.

Закон больших чисел устанавливает *статистическую устойчивость средних значений*. Это означает, что при большом числе испытаний среднее значение исследуемой величины перестает носить случайный характер и может быть предсказано с большой степенью определенности.

Центральная предельная теорема устанавливает условия при которых закон распределения суммы большого числа случайных величин неограниченно приближается к нормальному[10]. Согласно формулировке академика А.Н.Колмогорова[11], *совокупное действие большого числа случайных факторов приводит (при некоторых весьма общих условиях) к результату, почти независящему от случая*.

Неравенство Маркова[12]–Чебышева[13] (лемма Чебышева) и неравенство Чебышева используется для доказательства предельных терем, а также грубой оценки вероятностей событий случайной величины, распределение которой неизвестно.

---

[10] Первое доказательство центральной предельной теоремы при чрезвычайно общих условиях принадлежит ученику П.Л.Чебышева – А.М.Ляпунову (1857-1918). Для доказательства своей теоремы А.М.Ляпунов разработал специальный метод характеристических функций, широко используемый в теории вероятностей.

[11] А.Н.Колмогоров (1903-1987) – величайший математик XX столетия, внесший значительный вклад в теорию вероятностей и ряд других направлений математики. В истории российской науки его имя стоит рядом с именами М.В.Ломоносова, Д.И.Менделеева – ученых, всей своей жизнью прославивших Россию.

[12] А.А.Марков (1856-1922) – выдающийся математик, член-корреспондент АН СССР, плодотворно работавший в области теории вероятностей, алгебры, топологии, механики, математической логики и теории вычислимости, основоположник российской научной школы конструктивной математики.

[13] П.Л.Чебышев (1821–1894) – Выдающийся русский математик, член Императорской Академии Наук, основатель Петербуржской математической школы, внесший значительный вклад в развитие теории вероятностей, теории чисел, теории механизмов.

НЕРАВЕНСТВО МАРКОВА – ЧЕБЫШЕВА. Если неотрицательная случайная величина X имеет математическое ожидание, то для любого положительного числа A верно неравенство:

$$P(x > A) \le \frac{M(X)}{A} .$$

НЕРАВЕНСТВО ЧЕБЫШЕВА. Для любой случайной величины X, имеющей математическое ожидание $M(X)$ и дисперсию $D(X)$, справедливо неравенство:

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

ТЕОРЕМА ЧЕБЫШЕВА. Если случайные величины X_1, X_2, … , X_n независимы и одинаково распределены со средними значениями $M(X_i) = a$ и дисперсиями $D(X_i) = D(X)$, то справедливо следующее неравенство:

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \le \varepsilon\right\} \ge 1 - \frac{D(X)}{n\varepsilon^2} .$$

Закон больших чисел.

Основное утверждение закона больших чисел может быть получено из теоремы Чебышева при неограниченном увеличении числа n. Формулировка закона использует понятие *сходимости случайных величин по вероятности*.

ОПРЕДЕЛЕНИЕ. Случайные величины X_1, X_2, … , X_n, … *сходятся по вероятности* к величине A (случайной или неслучайной) если для любого $\varepsilon > 0$ вероятность события $\{|X_n - A| < \varepsilon\}$ при $n \to \infty$ стремится к единице:

$$\lim_{n\to\infty} P\{|X_n - A| < \varepsilon\} = 1 .$$

ЗАКОН БОЛЬШИХ ЧИСЕЛ (в форме Чебышева, 1886). Если случайные величины X_1, X_2, … , X_n независимы со средними значениями $M(X_i) = a$, то среднее арифметическое этих случайных величин сходится по вероятности к среднему арифметическому их математических ожиданий:

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \le \varepsilon\right\} = 1 .$$

Существует иная формулировка закона больших чисел, опирающаяся на теорему Бернулли.

ТЕОРЕМА БЕРНУЛЛИ. Пусть имеется n испытаний Бернулли с вероятностью успеха p, $q=1-p$, и m – число успехов. Тогда, для любого $\varepsilon > 0$ верно следующее неравенство:

$$P\left\{\left|\frac{m}{n}-p\right| \le \varepsilon\right\} \ge 1-\frac{pq}{n\varepsilon^2}.$$

ЗАКОН БОЛЬШИХ ЧИСЕЛ (в форме Бернулли). Закон больших чисел в форме Бернулли формулируется как предельное значение вероятности в теореме Бернулли при $n \to \infty$:

$$\lim_{n\to\infty} P\left\{\left|\frac{m}{n}-p\right| \le \varepsilon\right\} = 1.$$

Тема 8.2. Вариационные ряды и их характеристики.

Основные вопросы темы

1. Вариационный ряд. Основные определения.
2. Числовые характеристики вариационного ряда.

8.2.1 Понятие вариационного ряда. Основные определения.

Дадим основные определения, используемые при статистическом анализе массовых явлений.

Совокупность предметов или явлений, объединенных каким-либо общим признаком или свойством количественного или качественного характера, называется *объектом наблюдения*.

Всякий объект статистического наблюдения состоит из отдельных элементов — *единиц наблюдения.*

Результаты статистического наблюдения представляют собой числовую информацию - *данные.*

Статистические данные - это сведения о том, какие значения принял интересующий исследователя *признак* в статистической совокупности.

Признаки подразделяются на *количественные* и *качественные*.

Количественным называется признак, значения которого выражаются числами.

Качественным называется признак, характеризующийся некоторым свойством рассматриваемого объекта (элемента совокупности).

Значения признака, которые при переходе от одного элемента совокупности к другому изменяются (варьируют), называются *вариантами* и обычно обозначаются малыми латинскими буквами x, y, z.

Порядковый номер варианта (значения признака) называется *рангом*: x_1 — 1-й вариант (1-е значение признака), x_2 — 2-й вариант (2-е значение признака), x_i— i-й вариант (i-е значение признака).

Частота m_i показывает, сколько раз в статистической совокупности встречается значение признака, равное x_i .

Частость - w_i (или *относительная частота*) определяется как отношение частоты того или иного варианта к сумме всех частот ряда и показывает, какая часть единиц совокупности имеет данное значение признака (варианта):

$$w_i = \frac{m_i}{\sum_{i=1}^{k} m_i} = \frac{m_i}{n}.$$

Сумма всех частостей равна 1:

$$\sum_{i=1}^{k} w_i = \sum_{i=1}^{k} \frac{m_i}{n} = \frac{1}{n}\sum_{i=1}^{k} m_i = \frac{1}{n}\cdot n = 1.$$

Вариационным рядом (*рядом распределения*) называется ряд значений признака (вариантов), расположенных в порядке возрастания (или убывания) с соответствующими им весами. В качестве весов выступают *частоты* или *частости*.

ЗАМЕЧАНИЕ. Вариационный ряд является статистическим аналогом понятия распределения случайной величины *X* (распределения значений признака).

Вариационные ряды бывают дискретными и интервальными.

Вариационные ряды называются *дискретными*, если отдельные значения изучаемого признака отличаются друг от друга не менее чем на некоторую конечную величину (говорят, что дискретный вариационный ряд определяет *точечные значения признака*).

Общий вид дискретного вариационного ряда показан в табл. 8.1а и табл. 8.1б (где i = 1, 2, ..., k):

Значения признака - x_i	x_1	x_2	...	x_k
Частоты - m_i	m_1	m_2	…	m_k

Значения признака - x_i	x_1	x_2	...	x_k
Частости - w_i	w_1	w_2	…	w_k

Вариационные ряды называются *непрерывными* (*интервальными*), если отдельные значения изучаемого признака (варианты) могут отличаться друг от друга на сколь угодно малую величину. В этом случае возможные значения признака задаются в виде отдельных интервалов.

Общий вид интервального вариационного ряда показан в табл. 8.2а и табл. 8.2б (где i = 1, 2, ..., l):

Таблица 8.2а

Значения признака - x_i	$a_1 - a_2$	$a_2 - a_3$	...	$a_{i-1} - a_i$
Частоты - m_i	m_1	m_2	...	m_l

Таблица 8.2б

Значения признака - x_i	$a_1 - a_2$	$a_2 - a_3$	...	$a_{i-1} - a_i$
Частости - w_i	w_1	w_2	...	w_l

В интервальных вариационных рядах в каждом интервале выделяют *верхнюю* и *нижнюю границы*.

Разность между верхней и нижней границами интервала называется *величиной* (*длиной*) *интервала* (или *интервала интервальной разностью*). Таким образом, величина 1-го интервала Δ_1 определяется по формуле $\Delta_1 = a_2 - a_1$; 2-го - $\Delta_2 = a_3 - a_2$ и т.д. Если интервалы в вариационном ряде имеют одинаковую длину, их называют *равновеликими,* в противном случае — *неравновеликими*.

Интервал, включающий обе границы, то его называют *закрытым*, в противном случае он называется *открытым.*

Часто, при анализе вариационного ряда, оказывается целесообразным разбить его на несколько подгрупп (интервалов). Для определения оптимальной величины равновеликих интервалов используется формула Стэрджесса:

$$k = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \ln n},$$

где n — число единиц совокупности; $x_{\max}$ *и* $x_{\min}$ наибольшее и наименьшее значения вариантов ряда.

Для характеристики вариационного ряда наряду с понятиями частоты и частости используются понятия *накопленной частоты* - $m_i^{нак}$ и накопленной *частости* - $w_i^{нак}$, которые показывают, сколько единиц совокупности (какая их часть) меньше заданного значения (варианта) x.

При анализе интервального вариационного ряда накопленные частоты (частости) для каждого интервала находят путем суммирования частот (частостей) всех предшествующих интервалов, включая рассматриваемый:

$$m_i^{нак} = m_i + m_{i-1} + ... + m_1 = \sum_{j=1}^{i} m_j \ .$$

Для графического представления вариационных рядов обычно используется *полигон*, *гистограмма* и *кумулятивная кривая*.

Дискретный вариационный ряд графически обычно представляют с помощью *полигона распределения частот или частостей*. Соответствующий график, представляющий собой ломанную линию, соединяющую точки с координатами $(x_i\ , m_i)$ или $(x_i\ , w_i)$, представлен на рис. 8.1.

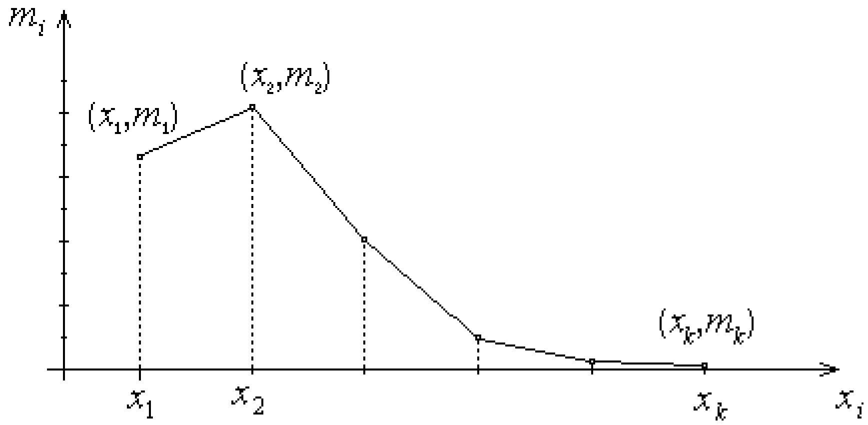


Рис. 16. Полигон распределения частот.

Графическое представление интервальных вариационных рядов реализуется с помощью *гистограммы,* т. е, столбчатой диаграммы (рис. 8.2).

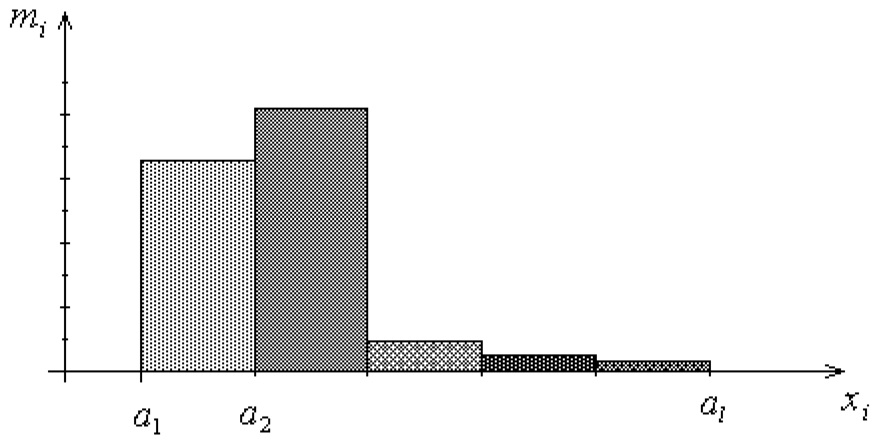


Рисунок 17. Гистограмма (столбчатая диаграмма).

При ее построении по оси абсцисс откладываются значения изучаемого признака (границы интервалов), а по оси координат - частоты или частости. Сам график представляет собой ступенчатую фигуру, состоящую из прямоугольников, с основанием, равным величине соответствующего интервала, и высотой, равной частоте (частости) интервала.

Графическое представление интервальных вариационных рядов реализуется не только с помощью гистограммы, но и используя полигон распределения. Соединяя середины верхних оснований прямоугольников отрезками прямых, получим полигон того же распределения.

В том случае, если интервалы одинаковой величины, по оси ординат можно откладывать частоты или частости. Если же интервалы имеют разную величину, по оси ординат необходимо откладывать значения абсолютной или относительной плотности распределения. *Абсолютная плотность* — отношение частости интервала к его величине:

$$f(a)_i = \frac{m_i}{k_i},$$

где: $f(a)_i$ - абсолютная плотность i-го интервала; m_i - его частота; и

k_i - величина интервала (интервальная разность).

ЗАМЕЧАНИЕ. Ранее указывалось, что вариационный ряд является статистическим аналогом распределения случайной величины *X*. В этом смысле полигон и гистограмма являются аналогами кривой распределения.

Как дискретные, так и интервальные вариационные ряды графически можно представить в виде *кумуляты* и *огивы*. При построении первой по данным дискретного ряда по оси абсцисс откладываются значения признака (варианты), а по оси ординат - накопленные частоты или частости. Нанося на график точки с координатами $\left(x_i\ , m_i^{нак}\right)$ или $\left(x_i\ , w_i^{нак}\right)$ и соединяя их отрезками прямой, ломаную линию, называемую *кумулятивной кривой* (*кумулятой*) .

Для интервального ряда процесс построения кумуляты аналогичен. Однако в этом случае на график наносятся точки с координатами $\left(a_i\ , m_i^{нак}\right)$ или $\left(a_i\ , w_i^{нак}\right)$, где a_i - соответствуют конечным значениям соответствующих интервалов. Часто добавляют еще одну точку, абсциссой которой является нижняя граница первого интервала, а ордината равна нулю. Соединяя точки отрезками или кривой, получаем *кумуляту* интервального ряда (рис. 8.3).

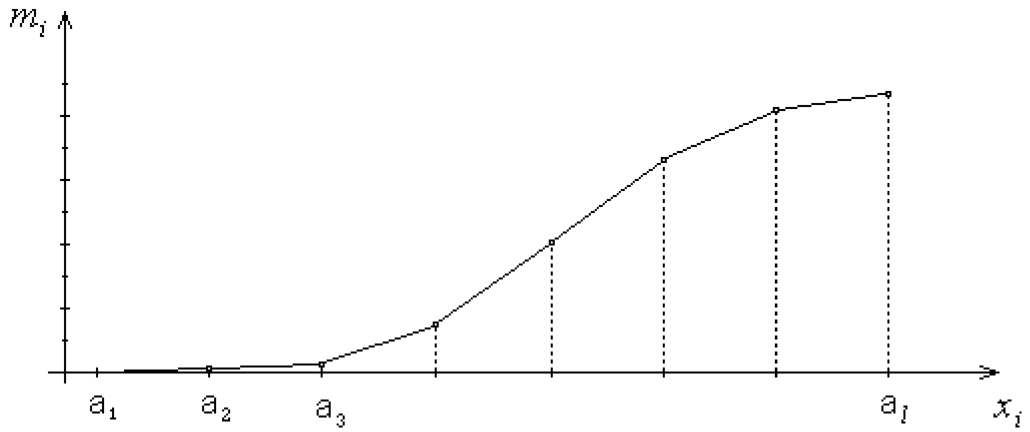


Рис. 18. Пример построения кумуляты.

8.2.2. Числовые характеристики вариационного ряда.

Одной из важнейших числовых характеристик вариационного ряда (ряда распределения) является *средняя арифметическая.* Для ее расчета могут быть использованы две формулы: *простая* и *взвешенная.*

Первую (простую среднюю арифметическую) обычно применяют, когда данные наблюдения не сведены в вариационный ряд (а также, если все частоты равны единице или одинаковы):

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

где: x_i - i-*е* значение признака;

n - объем вариационного ряда.

Если частоты отличны друг от друга, расчет производится по формуле *средней арифметической взвешенной*. При этом в качестве весов могут выступать частоты или частости i-*го* значения признака. В первом случае применяется формула:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i m_i}{\sum_{i=1}^{k} m_i}$$

где: x_i - i-*е* значение признака;

m_i - частота i-*го* значения признака;

k - число его возможных значений (вариантов).

Если в качестве весов используются частости значения признака, то соответствующая формула дается в виде:

$$\bar{x} = \sum_{i=1}^{k} x_i w_i$$

где: x_i - i-е значение признака;

w_i - частость i-го значения признака;

k - число его возможных значений (вариантов).

СВОЙСТВА СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ. Основные свойства средней арифметической вариационного ряда аналогичны рассмотренным ранее свойствам средней арифметической случайной величины:

1. Средняя арифметическая постоянной равна самой постоянной.
2. Если ко всем вариантам добавить постоянное число, то значение средней арифметической увеличиться (уменьшится) на это же число.
3. Если все варианты умножить на одно и то же число k, то значение средней арифметической измениться (увеличится или уменьшиться) в k раз.
4. Средняя арифметическая отклонений вариантов от средней арифметической равна нулю:

$$\overline{(x-\bar{x})}=0\,.$$

5. Средняя арифметическая алгебраической суммы нескольких признаков равна сумме средних арифметических этих признаков:

$$\overline{(x+y)}=\bar{x}+\bar{y}\,.$$

6. В случае если вариационный ряд содержит несколько групп, то общая средняя вариационного ряда $\bar{x}$ равна средней арифметической всех групповых средних $\bar{x}_j$, причем весовыми коэффициентами являются объемы групп n_j:

$$\bar{x}=\frac{\sum_{j=1}^{l}\bar{x}_j n_j}{n}\,.$$

Степень рассеяния (колеблемость) изучаемого признака относительно его среднего значения можно охарактеризовать с помощью *показателей вариации.* К наиболее употребительным относятся: *дисперсия*, *среднее квадратическое отклонение*, и *коэффициент вариации*.

Формула, определяющая дисперсию дается в виде:

$$s^2=\overline{(x-\bar{x})^2}\,.$$

Как и в случае среднего арифметического, для расчета *дисперсии* может быть использована простая или взвешенная формулы:

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

$$s^2=\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2 m_i}{\sum_{i=1}^{k}m_i}\,.$$

Среднее квадратическое отклонение рассчитывается по формуле:

$$s=\sqrt{s^2}\,.$$

Коэффициент вариации определяется формулой:

$$VX=\frac{s}{x}100\%\,.$$

СВОЙСТВА ДИСПЕРСИИ. Основные свойства дисперсии вариационного ряда аналогичны рассмотренным ранее свойствам дисперсии случайной величины:

1. Дисперсия постоянной равна нулю.
2. Если ко всем вариантам добавить постоянное число, то значение дисперсии не изменится.
3. Если все варианты умножить на одно и то же число k, то значение дисперсии измениться (увеличится или уменьшиться) в k^2 раз.
4. Дисперсия равна разности между средней арифметической квадратов вариантов и квадратом средней арифметической:
$$s^2 = \overline{x^2} - \bar{x}^2 .$$
5. (*правило сложения дисперсий*). Если вариационный ряд содержит несколько (l) групп наблюдений, то общая дисперсия равна сумме средней арифметической групповых дисперсий и межгрупповой дисперсии:
$$s^2 = \overline{s_i^2} + \delta^2 ,$$

где: s^2 - общая дисперсия вариационного ряда;

s_i^2 - дисперсия i-ой группы:

$$s_i^2 = \frac{\sum_{j=1}^{k} \left(x_j - \bar{x}_i\right)^2 m_j}{m_i};$$

$\overline{s_i^2}$ - средняя арифметическая групповых дисперсий:

$$\bar{s}_i^2 = \frac{\sum_{i=1}^{l} \bar{s}_i^2 m_i}{n};$$

δ^2 - межгрупповая дисперсия:

$$\delta^2 = \frac{\sum_{i=1}^{l} \left(\bar{x}_i - \bar{x}\right)^2 m_i}{n}.$$

ЗАМЕЧАНИЕ. Средняя арифметическая $\bar{x}$, дисперсия - s^2, среднее квадратическое отклонение - s вариационного ряда являются статистическими аналогами аналогичных характеристик случайной величины X (MX, σ^2, σ).

Кроме указанных выше характеристик, в статистическом анализе применяются и другие, в частности: *мода* и *медиана*.

Медианой - $\tilde{Me}$ вариационного ряда называется значение признака, приходящегося на его середину.

Модой - $\tilde{Mo}$ вариационного ряда называется вариант, которому соответствует наибольшая частота.

В таблице 8.3 приведены терминология, обозначения и формулы, используемые при описании характеристик случайной величины, с одной стороны, и вариационного ряда – с другой.

Табл.8.3

Случайная величина		Вариационный ряд	
Используемый Термин	Обозначения, Формулы	Используемый Термин	Обозначения, Формулы
Дискретная случайная Величина		Дискретный ряд	
Непрерывеая Случайная Величина		Интервальный ряд	
Значение случайной величины	x, x_i	Вариант	x_i
Вероятность	P, p, p_i	частость	w„ w_i
Полигон распределения вероятностей, кривая распределения		Полигон, гистограмма	
Математическое Ожидание	$M(X)=\sum_{i=1}^{k} x_i p_i$	Средняя арифметическая	$\bar{x}=\sum_{i=1}^{k} x_i w_i$
Дисперсия	$\sigma^2 = M[X - M(X)]^2$	Дисперсия	$s^2 = \overline{(x-\bar{x})^2}$
Среднее квадратическое отклонение	$\sigma = \sqrt{\sigma^2}$	Среднее квадратическое отклонение	$s = \sqrt{s^2}$
Мода	Mo	Мода	$\tilde{Mo}$
Медиана	Me	Медиана	$\tilde{Me}$

ВЫВОД. Как видно из приведенной таблицы, существует определенная аналогия понятий случайной величины и вариационного ряда, а также их характеристик. Поэтому*, вариационный ряд может рассматриваться как одна из реализаций распределения случайной величины (признака) X.*

Тема 9. ВЫБОРОЧНЫЙ МЕТОД

Основные вопросы темы

1. Выборочный метод. Основные определения.
2. Статистическое оценивание.
3. Ошибки выборки
4. Определение объема выборки.
5. Интервальное оценивание

9.1. Выборочный метод. Основные определения.

Прежде чем приступить к рассмотрению выборочного метода, сформулируем основные определения, используемые в данном разделе.

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ.

Часто, при проведении социологических, экономических, или иных исследований, возникает необходимость распространения выводов, сделанных на основе изучения свойств части совокупности (*случайной выборки*), на всю совокупность (*генеральную совокупность*). Подобные задачи являются определяющими при рассмотрении *выборочного метода*.

ОПРЕДЕЛЕНИЕ. *Генеральной совокупностью* называется совокупность объектов или наблюдений, все элементы которой подлежат изучению при статистическом анализе. Генеральная совокупность может быть конечной или бесконечной.

ЗАМЕЧАНИЕ. Понятие генеральной совокупности аналогично понятию случайной величины, изучаемого в разделе теории вероятностей.

Часть объектов генеральной совокупности, отобранная для целей ее изучения называется *выборкой*[14] (*выборочной совокупностью*).

Число объектов (наблюдений) генеральной (выборочной) совокупности называется ее *объемом*. Объем генеральной совокупности

[14] Наряду с рассматриваемой терминологией часто используются также термины «набор данных» и «элемент выборки».

обозначается N, а объем выборочной совокупности – n, он может быть конечным или бесконечным.

ЗАМЕЧАНИЕ. В математической статистике понятие генеральной совокупности трактуется как совокупность всех мыслимых наблюдений, которые могли бы быть произведены при определенных условиях. В данном статистическом контексте понятие генеральной совокупности является аналогом понятия случайной величины, изучаемой при выполнении некоторого комплекса условий.

Выборочная совокупность, позволяющая делать выводы о свойствах генеральной совокупности в целом, называется *представительной* (*репрезентативной*).

Репрезентативность выборки обеспечивается:

- необходимым объем данных; и
- случайным и равновероятным характером выборки отдельных элементов.

ЗАМЕЧАНИЕ. *Несоблюдение вышеозначенных условий часто является источником разнообразных ошибок*.

Например, при проведении социологических исследований очень важно обеспечить широкий охват самых разнообразных групп населения. Иначе, полученные выводы могут оказаться ошибочными.

Имеются две схемы образования выборки:

- *случайная повторная* (когда каждый случайно отобранный элемент после изучения возвращается в исходную генеральную совокупность и, вообще говоря, может быть случайным образом выбран вновь);
- *случайная бесповторная* (когда каждый случайно отобранный элемент после изучения не возвращается в исходную генеральную совокупность и, поэтому, не может быть выбран вновь).

9.2. Статистическое оценивание. Типы статистических оценок.

Для получения правильных выводов о свойствах генеральной совокупности по данным выборки, необходимо, чтобы рассматриваемая выборка представляла собой некую ‘уменьшенную копию’ генеральной совокупности, т. е. была *представительной (репрезентативной).*

Соблюдение данного требования проверяется путем использования соответствующих статистических оценок.

В статистике различают *точечные оценки* и *интервальные*.

9.2.1 Точечные оценки.

ОПРЕДЕЛЕНИЕ. *Точечной статистической оценкой* $\tilde{\theta}$ некоторого параметра θ генеральной совокупности (или теоретического распределения случайной величины X) называется приближенное значение этого параметра, рассчитанное по данным выборки.

Иными словами, *точечная статистическая оценка* $\tilde{\theta}$ некоторого параметра θ генеральной совокупности (теоретического распределения случайной величины X) есть некоторая функция выборки (результатов наблюдений за случайной величиной):

$$\tilde{\theta} = \tilde{\theta}(x_1, x_2, ..., x_n),$$

где: $x_1, x_2, ..., x_n$ - n отобранных элементов генеральной совокупности, (реализации случайной величины X в первом, втором, ..., n-ом опытах).

Функция выборки (функция результатов наблюдений) $\tilde{\theta} = \tilde{\theta}(x_1, x_2, ..., x_n)$ называется *статистикой*.

Используя понятие *статистики*, можно сформулировать рассматриваемое определение в следующем виде:

Точечной статистической оценкой $\tilde{\theta}$ некоторого параметра θ генеральной совокупности (или теоретического распределения случайной величины X) называется статистика, которая близка (в определенном смысле) истинному значению оцениваемого параметра.

ЗАМЕЧАНИЕ. Поскольку точечная статистическая оценка $\tilde{\theta}$ рассчитывается по данным выборки, а сама выборка осуществляется случайным образом, то можно сказать, что и сама точечная оценка является некоторой случайной величиной, характеризуемой определенным распределением. Учет этого обстоятельства делает необходимой проверку качества статистической оценки.

Существуют определенные критерии, по которым можно судить о качестве статистической оценки. К ним относятся:
- *несмещенность*;
- *состоятельность*;
- *эффективность*.

ОПРЕДЕЛЕНИЕ. Статистическая оценка $\tilde{\theta}$ некоторого параметра θ называется *несмещенной* если математическое ожидание оценки $\tilde{\theta}$ равно значению оцениваемого параметра θ, т.е.:

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

В противном случае (т.е., если: $M\tilde{\theta} \neq \theta$), она называется *смещенной*.

ЗАМЕЧАНИЕ. Статистическая оценка $\tilde{\theta}$ может зависеть от объема выборки - n; в этом случае она обозначается как $\tilde{\theta}_n$. Учитывая это, введем понятие *асимптотически несмещенной оценки*.

ОПРЕДЕЛЕНИЕ. Статистическая оценка $\tilde{\theta}_n$ некоторого параметра θ называется *асимптотически несмещенной* если с увеличением объема выборки, математическое ожидание оценки $\tilde{\theta}_n$ стремится к значению оцениваемого параметра θ, т.е.:

$$M\tilde{\theta}_n \to \theta .$$

Перейдем к понятию *состоятельности* оценки.

ОПРЕДЕЛЕНИЕ. Статистическая оценка $\tilde{\theta}_n$ некоторого параметра θ называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру, т.е. для любого $\varepsilon > 0$ выполняется следующее предельное равенство:

$$\lim_{n\to\infty} p\left(\left|\tilde{\theta}_n - \theta\right| < \varepsilon\right) = 1 .$$

ЗАМЕЧАНИЕ. Свойство состоятельности *обязательно* для любого критерия оценивания (*несостоятельные оценки не используются*).

Для проверки состоятельности оценки может быть использована следующая теорема.

ТЕОРЕМА. Если статистическая оценка $\tilde{\theta}_n$ является несмещенной, а ее дисперсия стремится к нулю при увеличении объема выборки ($D\tilde{\theta}_n \to 0$ при $n \to \infty$), то $\tilde{\theta}_n$ является состоятельной оценкой.

Эффективность оценки определяется следующим образом.

ОПРЕДЕЛЕНИЕ. Несмещенная статистическая оценка $\tilde{\theta}_n$ некоторого параметра θ называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных оценок параметра θ. Иными словами:
Несмещенная статистическая оценка $\tilde{\theta}_n$ некоторого параметра θ называется *эффективной*, если ее дисперсия минимальна.

ЗАМЕЧАНИЕ. На практике *не всегда* удается удовлетворить одновременно всем трем перечисленным свойствам (несмещенности, состоятельности и эффективности).

Точечные оценки математического ожидания и дисперсии.

ОПРЕДЕЛЕНИЕ. Средние арифметические распределения признака в генеральной и выборочной совокупностях называются соответственно *генеральной* и *выборочными средними*, а дисперсии этих распределений – *генеральной* и *выборочной дисперсиями.*

Отношение числа элементов генеральной и выборочной совокупностей, обладающих некоторым признаком *A*, к их объемам, называются соответственно *генеральной* и *выборочной долями* [15].

Следующая теорема характеризует возможность применения выборочного среднего для оценки математического ожидания генеральной совокупности (случайной величины *X*).

ТЕОРЕМА. Пусть $x_1, x_2, ..., x_n$ - выборка из генеральной совокупности (реализации случайной величины *X*). Тогда выборочное среднее

[15] Теоретическое обоснование возможности использования приведенных выборочных оценок для суждений о характеристиках и свойствах генеральной совокупности дают закон больших чисел и центральная предельная теорема Ляпунова.

$\overline{X}_{выб} = \frac{1}{n}\sum_{i=1}^{n} x_i$ - является несмещенной и состоятельной оценкой математического ожидания MX. Кроме того, если случайная величина X имеет нормальное распределение, то эта оценка является также эффективной.

Таким образом, как показывает данная теорема, для оценки математического ожидания применима обычная формула нахождения среднего, и эта оценка является *несмещенной*, *состоятельной*, а в случае нормального распределения, *эффективной*.

Оценка дисперсии осуществляется более сложным образом. Это объясняется тем, что выборочная дисперсия является *смещенной оценкой* дисперсии исходной случайной величины. Однако этот недостаток может быть устранен путем использования, так называемой, *исправленной выборочной дисперсией* - S^2, получаемой путем умножения выборочной дисперсии - $D_{выб}$ на поправочный коэффициент $\frac{n}{n-1}$:

$$S^2 = \frac{n}{n-1} D_{выб},$$

или:

$$S^2 = \frac{n}{n-1} \sigma^2_{выб}.$$

Теоретическим обоснованием использования данной формулы этого служит следующая теорема.

ТЕОРЕМА. Пусть $x_1, x_2, ..., x_n$ - выборка из генеральной совокупности (реализации случайной величины *X*). Тогда исправленная выборочная дисперсия $S^2 = \frac{n}{n-1} D_{выб}$ является *несмещенной состоятельной оценкой* дисперсии DX.

ЗАМЕЧАНИЕ. При больших значениях *n*, величина поправочного коэффициента стремится к единице ($\frac{n}{n-1} \to 1$ при $n \to \infty$), и разница между S^2 и $D_{выб}$ становится пренебрежимо малой. На практике это означает, что использование поправочного коэффициента целесообразно при малых значениях *n* (*n*<*30*). Для значений *n*>*30* поправочный коэффициент можно не использовать, полагая $S^2 \approx D_{выб}$.

Генеральное среднее квадратическое отклонение $\sigma_{ген}$ также имеет две точечные оценки: $\sigma_{выб}$ — выборочное среднее квадратическое отклонение и S — исправленное выборочное среднее квадратическое отклонение. Первое - $\sigma_{выб}$ используется для оценивания $\sigma_{ген}$ при $n > 30$, а второе - S для оценивания $\sigma_{ген}$ при $n < 30$. При этом, используются следующие формулы:

$$\sigma_{выб} = \sqrt{\sigma^2_{выб}}\ ,$$

$$S_{выб} = \sqrt{S^2_{выб}}\ .$$

Указанные числовые характеристики генеральной совокупности, вычисленные по данным выборки называют *статистиками*[16].

Эмпирическая функция распределения.

Построение функции распределения требует использования значений вероятности события. Следующая теорема характеризует возможность применения относительной частоты (частости) для оценки значений вероятности появления события A.

ТЕОРЕМА. Относительная частота (частость) $w_A = \frac{n_A}{n}$ появления некоторого события A в n независимых испытаниях является *несмещенной* и *эффективной* оценкой вероятности появления этого события в каждом испытании ($p = P(A)$).

ОПРЕДЕЛЕНИЕ. Перечень вариантов x_i и соответствующих им частот n_i или частостей w_i называется *статистическим распределением выборки* или *статистическим рядом*.

Для оценки функции распределения случайной величины X (генеральной совокупности) по данным выборки используется понятие эмпирической (статистической функцией распределения.

[16] Для того чтобы любые статистики служили хорошими оценками параметров генеральной совокупности, они должны обладать рядом свойств: несмещенности, эффективности и состоятельности..

ОПРЕДЕЛЕНИЕ. *Эмпирической* (*статистической*) *функцией распределения* называется функция $F_n^*(x)$, определяющая для каждого значения x частость события $\{X < x\}$:

$$F_n^*(x) = w\{X < x\}.$$

Иными словами, *эмпирическая* (*статистическая*) *функция распределения* $F_n^*(x)$ для каждого значения x определяет накопленную частость $w_x^{нак} = \dfrac{n_x^{нак}}{n}$ (являясь статистическим аналогом функции распределения случайной величины X):

$$F_n^*(x) = w\{X < x\} = w_x^{нак}.$$

Следующая теорема характеризует возможность применения эмпирической функции распределения выборки $F_n^*(x)$ для оценки функции $F_n(x)$ распределения генеральной совокупности (случайной величины X).

ТЕОРЕМА. Эмпирическая функция распределения $F_n^*(x)$ является *несмещенной состоятельной* оценкой функции $F_n(x)$ распределения случайной величины X.

9.3. Ошибки выборки.

Любая выборка представляет собой лишь часть генеральной совокупности, и поэтому выборочные характеристики не будут точно совпадать с соответствующими генеральными.

ОПРЕДЕЛЕНИЕ. Среднее квадратическое отклонение выборочной средней $\sigma_{\bar{x}}$ и выборочной доли σ_w собственно случайной выборки называется *средней квадратической* (*стандартной*) *ошибкой выборки*.

Формулы расчета средних квадратических ошибок выборки для случаев повторного и бесповторного отбора приведены в Табл. 9.1.

Таблица 9.1

Формулы расчета средних квадратических ошибок выборки

Оцениваемый параметр	Собственно-случайный отбор	
	Повторная выборка	Бесповторная выборка
Генеральная средняя	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} \approx \sqrt{\frac{s^2}{n}}$	$\sigma'_{\bar{x}} \approx \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)}$
Генеральная доля	$\sigma_w = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{w(1-w)}{n}}$	$\sigma''_w \approx \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$

Здесь: σ^2 - выборочная дисперсия значений признака;
n - объем выборки;
N - объем генеральной совокупности;
n/N - доля обследованной совокупности;
(1- n/N) - поправка на конечность совокупности (в литературе *(1 - n/N)* иногда называется «поправкой на бесповторность отбора»).

9.4. Определение объема (численности) выборки.

Важной проблемой, решаемой при использовании выборочного метода, является определение необходимого объема выборки, т.к. от него зависит размер полученной средней ошибки и необходимые затраты на реализацию выборочного наблюдения (чем больше объем выборки, тем больше затраты на изучение элементов выборки, но тем меньше ошибка выборки).

Необходимый объем выборки для случаев повторного (*n*) и бесповторного (*n'*) отбора из генеральной совокупности объема *N* может быть определен по формулам приведенным в Табл.9.2

Таблица 9.2

Формулы расчета необходимого объема выборки.

Оцениваемый параметр	Собственно-случайный отбор	
	Повторная выборка	Бесповторная выборка
Генеральная Средняя	$n = \frac{t^2\sigma^2}{\Delta^2}$	$n' = \frac{t^2\sigma^2 N}{N\Delta^2 + t^2\sigma^2}$
Генеральная доля	$n = \frac{t^2 w(1-w)}{\Delta^2}$	$n' = \frac{t^2 N w(1-w)}{N\Delta^2 + t^2 w(1-w)}$

Объемы выборки для случаев повторного (n) и бесповторного (n') отбора связаны соотношением:

$$n' = n\frac{N}{n+N}.$$

ЗАМЕЧАНИЕ. Поскольку величина $\frac{N}{n+N} < 1$, то из приведенной формулы следует, что $n' < n$ (т.е. требуемый объем бесповторной выборки меньше объема повторной выборки). Поэтому, на практике, обычно используют бесповторную выборку.

9.5. Интервальное оценивание.

Наряду с точечными оценками (приближенными численными значениями исследуемых параметров), используются также *интервальные оценки*.

ОПРЕДЕЛЕНИЕ. *Интервальной оценкой* параметра θ называют числовой интервал (определяемый его начальной $\tilde{\theta}_n^{(1)}$ и конечной $\tilde{\theta}_n^{(2)}$ точками — концами интервала), который с заданной вероятностью - γ накрывает (охватывает) неизвестной значение исследуемого параметра θ генеральной совокупности. Интервал, содержащий оцениваемый параметр генеральной совокупности, называют *доверительным интервалом*, а вероятность γ - *доверительной вероятностью*, *уровнем доверия* или *надежностью оценки*.

ЗАМЕЧАНИЕ. Границы $\tilde{\theta}_n^{(1)}$ и $\tilde{\theta}_n^{(2)}$ и величина интервальной оценки вычисляются по данным выборки, и потому, являются случайными величинами, в отличии от самого исследуемого параметра θ значение которого является неслучайной величиной.

Величина $\Delta = \left|\tilde{\theta} - \theta\right|$ характеризует отклонение оценки $\tilde{\theta}$ от оцениваемого параметра θ характеризует *ошибку репрезентативности выборки*, а ее максимальное значение называют *предельной ошибкой выборки*. Ошибки репрезентативности характеризуется соответствующим уровнем вероятности, значение которой определяется следующей теоремой.

ТЕОРЕМА. Вероятность того, что отклонение выборочной средней (или доли) от генеральной средней (или доли) не превзойдет по абсолютной величине указанной ошибки репрезентативности Δ, равна:

$$P\left(\left|\bar{x}-\bar{x}_0\right|<\Delta\right)=\Phi(t)=\gamma \text{ и}$$
$$P\left(\left|w-p\right|<\Delta\right)=\Phi(t)=\gamma,$$

где: $t=\dfrac{\Delta}{\sigma_x}$ (или $t=\dfrac{\Delta}{\sigma_w}$ - для второй формулы);

$\Phi(t)$ - функция Лапласа (интеграл вероятностей Лапласа).

Контрольные вопросы по теме 9

1. Как называется совокупность всех объектов или наблюдений, подлежащих изучению в статистическом анализе?
2. Является ли генеральной совокупностью множество всех значений случайной величины?
3. Может ли генеральная совокупность состоять из двух объектов?
4. В чем различие между выборочной и генеральной совокупностями?
5. Как можно из частоты вариации получить ее частость?
6. Какие величины составляют вариационный ряд?
7. Как связаны значения эмпирической функции распределения с накопленными частотами?
8. Как определяются выборочные средние для дискретного и интервального вариационных рядов:
9. Как изменится выборочное среднее при умножении всех его вариаций на 10?
10. Может ли выборочная дисперсия быть отрицательной?

Тема 9. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Основные вопросы темы

1. Введение
2. Основные определения.
3. Проверка статистических гипотез.

10.1 Введение.

В процессе статистического анализа часто оказывается необходимым осуществить проверку некоторых предположений (гипотез) относительно полученных величин параметров или закона распределения изучаемой генеральной совокупности. Например, оценить как полученные данные согласуются с предположением (гипотезой) о преимуществе того или иного рода инвестиций, метода управления, эффективности технологического процесса, и др.

Для решения такого рода задач используется *метод проверки статистических гипотез*.

10.2 Основные определения.

ОПРЕДЕЛЕНИЕ. Статистической гипотезой называется любое предположение о виде неизвестного закона распределения и значений его параметров. Для обозначения гипотез используют символ H.

Выдвинутую (проверяемую) гипотезу обычно называют *нулевой* (*основной*). Ее принято обозначать H_0.

По отношению к высказанной (основной) гипотезе формулируется также *альтернативная (противоположная ей, конкурирующая),* гипотеза, являющаяся ее логическим отрицанием. Альтернативную (конкурирующую) гипотезу принято обозначать H_1. Отметим, что для основной гипотезы H_0 может существовать несколько альтернативных гипотез H_1, H_2, …, H_k . В дальнейшем, однако, ограничимся случаем одной альтернативной гипотезы H_1.

По своему содержанию статистические гипотезы можно подразделить на несколько основных типов:

1. Гипотезы о виде закона распределения исследуемой случайной величины;
2. Гипотезы о числовых значениях параметров случайной величины;
3. Гипотезы об общем виде математической модели, описывающей статистическую зависимость между признаками;
4. Гипотезы о принадлежности некоторого признака к определенному классу.

Гипотеза называется *простой*, если она, на основании данных выборки, полностью определяет теоретическое распределение случайной величины (генеральной совокупности). В противном случае, гипотеза называется сложной. Обычно основная гипотеза является простой, а альтернативная – сложной.

Привило (критерий), по которому может быть принята или отвергнута выдвинутая гипотеза, называется *статистическим критерием* или *статистическим тестом* и обозначается символом K.

Цель статистической проверки гипотез состоит в том, чтобы на основании выборочных данных и значения используемого статистического критерия принять (или отвергнуть) выдвинутую гипотезу H_0. Реализация данной цели (сущность метода *проверки статистических гипотез*) состоит в сопоставлении выдвинутой гипотезы H_0 (или альтернативной ей - H_1) относительно вида распределения и параметров генеральной совокупности с имеющимися данными выборки. Это сопоставление выражается в виде количественной оценки степени достоверности полученного вывода, оцениваемого с помощью используемого статистического критерия. Если степень достоверности мала – то гипотеза отвергается. В противном случае она может быть принята с заданной степенью уверенности (достоверности).

Поскольку проверка статистических гипотез осуществляется на основании данных выборки (т. е. ограниченного ряда наблюдений) принятое решение относительно нулевой гипотезы H_0 может содержать ошибки.

Виды ошибок.

Различают ошибки двух видов.

1. Отвергнута верная гипотеза. Такая ошибка называется *ошибкой первого рода*, а ее вероятность (обозначаемая символом α) – *уровнем значимости*:

$$\alpha = P\{H_1|H_0\}.$$

Величина γ=(1-α) называется *уровнем доверия* и определяет вероятность принять верную гипотезу:

$$\gamma = (1-\alpha) = P\{H_0|H_0\} = \gamma .$$

2. Принята неверная гипотеза. Такая ошибка называется *ошибкой второго рода*. Вероятность ошибки второго рода обозначается как β:

$$\beta = P\{H_0|H_1\}.$$

Величина (1-β) определяет вероятность принятия гипотезы H_1 если она верна. Это значение носит название *мощность критерия*:

$$1-\beta = P\{H_1|H_1\}.$$

Данные об ошибках первого и второго рода сведены в Табл. 10.1.

Таблица 10.1

Нулевая гипотеза H_0	Принятое решение относительно нулевой гипотезы H_0	
	Отклонена	Принята
Верна	Ошибка 1-го рода, ее вероятность $P(H_1/H_0) = \alpha$	Правильное решение, его вероятность $P(H_0/H_0) = 1 - \alpha$
Неверна	Правильное решение, его вероятность $P(H_1/H_1) = 1 - \beta$	Ошибка 2-го рода, ее вероятность $P(H_0/H_0) = \beta$

В качестве примера, приведем интерпретацию параметров α и β в различных сферах деятельности. Так, в статистическом контроле качества, параметр α отражает *риск поставщика*, а параметр β - *риск потребителя*. В юридической практике параметр α отражает вероятность осуждения невиновного, а параметр β - вероятность неправомочного оправдания преступника.

Естественно стремление снизить вероятности появления ошибок первого и второго рода. Однако, значения величин α и β связаны между собой. Добиться одновременного снижения вероятностей ошибок первого и второго рода удается лишь увеличением объем выборки. Поэтому, при решении задач проверки статистических гипотез можно выбрать по своему усмотрению величину вероятности только одной из них: α или β. Обычно задается значение вероятности ошибки 1-го рода α (уровня значимости). При этом часто пользуются ‘стандартными’ значениями уровня α: 0,1; 0,05; 0,025; 0,01; 0,005; 0,001.

10.3 Проверка статистических гипотез.

Статистический критерий – K определяется правилом или формулой, по которой оценивается мера близости (расхождения) данных выборки с высказанной гипотезой H_0. Как и всякая функция результатов наблюдения, он является случайной величиной (а следовательно, может принимать различные значения, описываемые некоторым законом распределения).

Выбор критерия для проверки статистических гипотез может быть осуществлен на основании различных принципов. В частности, удобно использовать критерий для которого известен (табулирован) закон распределения (плотность распределения - $f(k)$) в предположении справедливости нулевой гипотезы H_0.

Может возникнуть ситуация, когда существуют несколько критериев, характеризующихся одной и той же вероятностью α (отклонение верной гипотезы H_0). В этом случае следует принять тот, которому соответствует меньшая ошибка 2-го рода β, т.е. большая мощность, что поясняется следующей теоремой.

Теорема (лемма) Неймона-Пирсона. Среди всех критериев заданного уровня значимости α, проверяющего простую гипотезу H_0 против альтернативной H_1, *принцип отношения правдоподобия* является наиболее мощным.

Принцип отношения правдоподобия позволяет выбрать наиболее мощный среди всех возможных критериев. Суть его сводится к использованию такого критерия K (с известной функцией плотности $f(k)$ при условии справедливости гипотезы H_0), для которого при заданном уровне значимости α можно было бы найти критическую точку $K_{кр}$ распределения *f(k),* которая разделила бы область значений критерия на две области: *область допустимых значений*, для которой результаты выборочного наблюдения полагаются правдоподобными, и *критическую область*, в которой данные выборки отвергают реальность нулевой гипотезы H_0.

Критические точки, разделяющие совокупность возможных значений критерия на область допустимых значений (наиболее правдоподобных в отношении нулевой гипотезы H_0) и критическую область (область значений, менее правдоподобных в отношении нулевой гипотезы H_0) определяют для заданного уровня значимости α (по известным формулам или по таблицам распределения критерия K).

В области *допустимых значений* принимается нулевая гипотеза H_0, а в критической области она отклоняется в пользу конкурирующей H_1.

Различают *правостороннюю, левостороннюю* и *двустороннюю критические области.*

Если конкурирующая гипотеза - правосторонняя, например, H_1: $a>a_0$, то и критическая область - правосторонняя. При правосторонней конкурирующей гипотезе критическая точка принимает положительные значения.

Если конкурирующая гипотеза — левосторонняя, например, H_1: $a<a_0$, то и критическая область - *левосторонняя*. При левосторонней конкурирующей гипотезе критическая точка принимает отрицательные значения.

Если конкурирующая гипотеза - двусторонняя, например. H_1: $a \neq a_0$, то и критическая область - *двусторонняя*. При двусторонней конкурирующей гипотезе определяются две критические точки: левая и правая.

Выбрав тип статистического критерия, реализуется расчет его числового значения (при заданном уровне значимости α) и сопоставление с областью допустимых значений.

Укажем на рекомендуемые типы критериев, применяемых при решении различных статистических задач:

- проверка гипотезы о виде закона распределения случайной величины может быть осуществлена с помощью *критерия согласия Пирсона* χ^2*;*
- проверка гипотезы о равенстве неизвестных значений дисперсий двух генеральных совокупностей — с помощью *критерия Фишера F;*

- ряд гипотез о неизвестных значениях параметров генеральных совокупностей проверяется с помощью критерия - нормальной распределенной случайной величины и *критерия t-Стьюдента* и т. д.

При правосторонней конкурирующей гипотезе:

- при $K_{выч} \leq K_{кр.}$ нулевая гипотеза H_0 не отклоняется. В противном случае ($K_{выч} > K_{кр}$) нулевая гипотеза H_0 отклоняется в пользу конкурирующей H_1.

При левосторонней конкурирующей гипотезе:

- при $K_{вычл} \geq -K_{кр}$, то нулевую гипотезу H_0 нельзя отклонить. В противном случае ($K_{выч} < -K_{кр}$) нулевая гипотеза H_0 отклоняется в пользу конкурирующей H_1.

При двусторонней конкурирующей гипотезе:

- при $-K_{кр} \leq K_{выч} \leq K_{кр}$, нулевая гипотеза H_0 не отклоняется. В противном случае ($K_{выч} > K_{кр}$ или $K_{набл} < -K_{кр}$) нулевая гипотеза H_0 отклоняется в пользу конкурирующей H_1.

Алгоритм проверки статистических гипотез.

Приведем алгоритм проверки статстических гипотез.

1) сформулировать нулевую H_0 и альтернативную H_1 гипотезы;

2) выбрать уровень значимости α (по возможности, используя стандартные уровни);

3) в соответствии с видом выдвигаемой нулевой гипотезы H_0 выбрать статистический критерий для ее проверки, с известным (точным или приближенным) распределением;

4) по таблицам распределения случайной величины K, выбранной в качестве статистического критерия, найти критическое значение $K_{кр}$ (критическую точку или точки);

5) на основании данных выборки вычислить значение критерия $K_{выч}$;

6) по виду конкурирующей гипотезы H_1 определить тип критической области;

7) сопоставляя вычисленное значение критерия $K_{выч}$ с областью допустимых значений, принять решение относительно нулевой гипотезы H_0.

ЗАМЕЧАНИЕ. Если в результате проверки нулевую гипотезу H_0 нельзя отклонить, то это означает, что имеющиеся выборочные данные не позволяют с достаточной уверенностью отклонить нулевую гипотезу H_0: вероятность нулевой гипотезы H_0 больше α, а конкурирующей H_1 - меньше (1-α). Если в результате проверки нулевая гипотеза H_0 отклоняется в пользу конкурирующей H_1, то это означает, что имеющиеся выборочные данные не позволяют с достаточной уверенностью принять нулевую гипотезу H_0: вероятность нулевой гипотезы H_0 меньше α, а конкурирующей H_1 больше (1 - α).

Контрольные вопросы по теме 10

1. Какие виды ошибок рассматриваются при проверке статистических гипотез?
2. Как связаны ошибки первого рода и уровень доверия?
3. Как связаны ошибки второго рода и мощность критерия?
4. В какие области попадает статистика при принятии или непринятии основной гипотезы?
5. Какого вида бывают критические области?
6. Сформулируйте теорему (лемма) Неймона-Пирсона.
7. Сформулируйте алгоритм проверки статистических гипотез.

Тема 11. ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ

Основные вопросы темы

1. Виды и формы связи.
2. Задачи и этапы корреляционного анализа.
3. Расчет коэффициента корреляции.
4. Регрессионный анализ.

11.1. Виды и формы связи.

Одним из важных приложений методов математической статистики является установление связи (зависимости) между двумя или более наблюдаемыми величинами (явлениями и их признаками). При исследовании таких зависимостей различают два типа связи:

- *функциональную* (или *жестко детерминированную*); и
- *статистическую* (или *стохастически детерминированную*).

Функциональная связь - это вид причинной зависимости, при которой каждому допустимому значению факторного признака (аргумента) соответствует одно или несколько строго определенных значений результативного признака (значение функции). Функциональная связь определяется аналитическим, табличным, графическим или алгоритмическим образом.

Статистическая связь (называемая также *стохастическая* или *вероятностная*) - это вид причинной зависимости, при котором каждому допустимому значению факторного признака соответствует определенное (условное) распределение значений результативного признака. Иными словами, статистическая связь проявляется не в каждом отдельном случае, а в общем, в среднем, при большом числе наблюдений.

Для исследования статистической связи широко используются понятия *корреляции* и *регрессии*[17] , и связанные с ними методы корреляционного и регрессионного анализа данных, которые появились в середине XIX в. благодаря работам английских статистиков Ф.Гальтона и К.Пирсона.

Рассмотрим эти понятия более подробно.

---

[17] Термин *корреляция* произошел от английского слова “correlation” – соотношение, взаимосвязь. Термин *регрессия* (от лат. “regressio” – движение назад) введен Ф.Гальтоном, который изучая зависимость между ростом родителей и их детей, обнаружил явление “регрессии к среднему” – у детей, родившихся у очень высоких родителей, рост имел теденцию быть ближе к среднему значению.

11.2. Задачи и этапы корреляционного анализа.

Корреляционная связь — это зависимость среднего значения результативного признака - Y от изменения факторного признака - X (заметим, что каждому отдельному значению факторного признака может соответствовать множество различных значений результативного).

Наличие корреляционной связи может служить проявлением:

- односторонней причинно-следственной связи;
- взаимных прямых (или опосредствованных) причинно следственных связей;
- ложной корреляции.

Причинно-следственная связь факторов.

Пусть, например, значение параметра X характеризует балл оценки плодородия почв, а параметр Y - урожайность сельскохозяйственной культуры. В данном случае, корреляционная зависимость является проявлением причинно-следственной связи рассматриваемых факторов: зависимость урожайности от плодородия почвы.

Наличие взаимных причинно-следственных связей.

Корреляция возникает и в случае наличия взаимных причинно-следственных связей. Так, например, рост инвестиций в развивающуюся отрасль приводит к росту ее конкурентоспособности. С другой стороны, повышение конкурентоспособности является стимулирующим фактором для привлечения новых инвестиций.

Ложная корреляция.

Наличие корреляционной связи не всегда означает наличие причинно–следственных отношений между рассматриваемыми признаками. В частности, если значения двух независимых величин X и Y определяются воздействием некоторого общего фактора Z, то совместное рассмотрение множества значений, принимаемых величинами X и Y может указать на наличие корреляционной связи, хотя сами величины X и Y являются независимыми. Проявление корреляционной связи между независимыми величинами называют *ложной корреляцией*.

ПРИМЕР 1 (классический пример, приведенный известным статистиком начала XX в. А.А.Чупровым).

Если величина X характеризует число пожарных команд в городе, а величина Y - сумму убытков от пожаров, то совместный анализ параметров X и Y выявляет достаточно высокую корреляцию, которую нельзя интерпретировать в рамках причинно-следственных отношений (чем больше в городе пожарных команд, тем больше убытки от пожаров). На самом деле выявленная в данном случае корреляция является лишь следствием зависимости обоих величин *X* и *Y от третьего фактора* – *Z*, характеризующего размер города: чем больше город, тем чаще в нем возникают пожары (а следовательно, тем больше от них убытки), и тем больше нужно пожарных команд для их ликвидации и профилактического обслуживания городского хозяйства.

ПРИМЕР 2 (приводится по книге Л.И.Ниворожкиной[18]).

Сразу после 17 августа 1998 г. в России резко возросли цена зарубежной валюты и объем ее покупки частными лицами. В данном случае также нельзя рассматривать оба эти явления в рамках причинно-следственной связи (возрос объем покупки валюты т.к. возросла ее цена). Общая причина (третий фактор) - обострение финансового кризиса - привело и к росту курсовой стоимости валюты, и к стремлению населения сохранить свои накопления в твердой валюте.

Рассмотренные примеры позволяют сформулировать *основные задачи корреляционного анализа*:

1) изучение степени тесноты связи отдельных явлений;
2) выявление факторов, оказывающих наиболее существенное влияние на результативный признак;
3) выявление неизвестных причинных связей.

11.2.2 Этапы корреляционного анализа.

Исследование корреляционных зависимостей включает ряд этапов:

1) предварительный анализ свойств совокупности;
2) установление факта наличия связи, определение ее направления и формы;
3) измерение степени тесноты связи между признаками;
4) построение регрессионной модели (нахождение аналитического выражения связи);
5) оценку адекватности модели, ее экономическую интерпретацию и практическое использование.

---

[18] Л.И.Ниворожкина З.А.Морозова. Основы статистики с элементами теории вероятностей для экономистов. Руководство для решения задач. Ростов на Дону., изд. ФЕНИКС, 1999. – 320с.

На этапах предварительного анализа свойств совокупности и факта наличия связи (этапы 1 и 2 корреляционного анализа) полезными оказываются качественные методы анализа, включая *графический метод, метод приведения параллельных данных*, а также, *метод аналитических группировок*.

Графический метод используется для наглядного отображения формы связи между изучаемыми признаками. Для этого в прямоугольных осях координат строят график, по оси ординат которого откладывают значения результативного признака, а по оси абсцисс - значения факторного признака.

Совокупность точек результативного и факторного признаков называется *полем корреляции*.

По направлению выделяют связь *прямую* и *обратную*, по аналитическому выражению — *прямолинейную* и *нелинейную* (см. рисунок ниже).

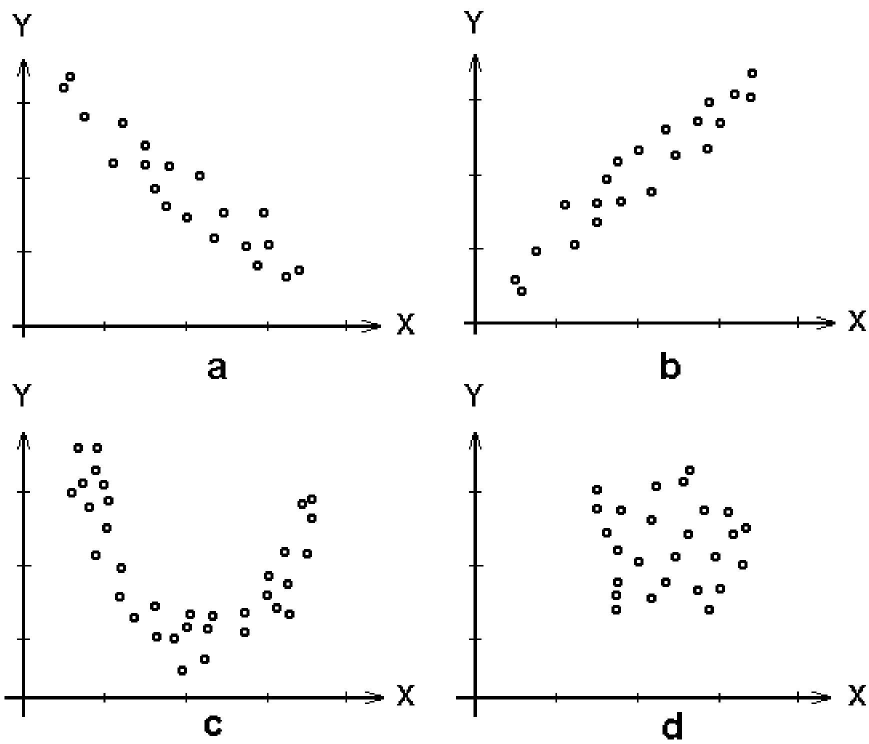


Рис. 19. Примеры корреляционных связей

a) прямолинейная обратная. b) прямолинейная прямая.
c) нелинейная. d) некоррелированные признаки.

Метод приведения параллельных данных основан на сопоставлении числовых данных 2-х или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие и направление связи (прямая или обратная). Сопоставление удобно проводить, занося данные в таблицу. Например, данных таблицы 11.1 позволяет сделать вывод о наличии обратной зависимости (связи) Y и X.

Таблица 11.1

X	1	2	3	4	5	6	7	8	9
Y	92	78	73	59	41	37	32	18	10

Метод аналитических группировок используется при необходимости анализа влияния качественных признаков на показатели вариации количественных признаков. В этом случае в таблице отдельно группируются средние значения результативных признаков для каждого качественного значения (факторного) признака. Изменение этого признака при переходе от одной группы к другой вызывает соответствующие изменения результативного признака, что и определяет наличие и направление связи.

11.3. Измерение степени тесноты связи между признаками. Расчет коэффициента корреляции.

Дальнейшим этапом корреляционного анализ является *измерение степени тесноты связи между признаками*.

Методы оценки тесноты связи обычно используют численные критерии *совместной вариации* признаков, учитывающие отклонения изучаемых величин от их средних значений. К числу таких критериев относятся коэффициент ковариации и *коэффициент линейной корреляции Пирсона* (часто называемый просто *коэффициентом корреляции*).

11.3.1 Коэффициент ковариации.

ОПРЕДЕЛЕНИЕ. Ковариацией (мерой *совместной вариации*) K_{xy} случайных величин X и Y называется математическое ожидание произведения отклонений этих величин от их математических ожиданий:

$$K_{xy} = M[(X - M(X))(Y - M(Y))],$$

записываемое также, как:

$$K_{xy} = M[(X - a_x)(Y - b_y)],$$

где: $a_x = M(X)$ и $b_y = M(Y)$ - математические ожидания случайных величин *X* и *Y*.

Используя формулу вычисления математического ожидания для случая дискретных случайных величин *X* и *Y*, получим:

$$K_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - a_x)(y_j - b_y)p_{ij},$$

где: p_{ij} - вероятность того, что двумерная случайная величина (*XY*) приняла значение ($x_i y_j$).

В случае непрерывных случайных величин *X* и *Y*, соответствующая формула имеет вид:

$$K_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_i - a_x)(y_j - b_y)\varphi(x,y)dxdy.$$

ЗАМЕЧАНИЕ. Ковариацию называют также *вторым смешанным центральным моментом случайных величин X и Y*. Для обозначения ковариации также используются следующие обозначения: *cov*(*X*,*Y*) и σ_{xy}.

Свойства ковариации.

1. Как следует из определения:
$$K_{xy} = K_{yx}.$$
2. Ковариация случайной величины с самой собой есть ее дисперсия:
$$K_{xx} = M[(X - M(X))(Y - M(X))] = M[X - M(X)]^2 = D(X).$$
3. Ковариация двух независимых случайных величин *X и Y* равна нулю: $K_{xy} = 0$.
4. Ковариация двух случайных величин равна математическому ожиданию их произведения минус произведение их математических ожиданий:
$K_{xy} = M(X \cdot Y) - M(X) \cdot M(Y)$.
5. Ковариация двух случайных величин по абсолютной величине не превосходит произведению их среднеквадратических отклонений, т.е.:
$$|K_{xy}| \le \sigma_x \sigma_y.$$

Являясь мерой *совместной вариации* случайных величин *X* и *Y* коэффициент ковариации K_{xy} позволяет судить о наличии связи между ними (в частности, в случае независимых случайных величин, коэффициент ковариации равен нулю). Однако, к сожалению, этот показатель нельзя использовать непосредственно для определения степени тесноты этой связи, т.к. он имеет смешанную размерность ($X \cdot Y$) и не нормирован (а следовательно, нельзя сравнивать коэффициенты ковариации разных пар переменных). Поэтому, наряду с данным показателем, используется и другой – *коэффициент корреляции*. Данный показатель, называемый также *линейный коэффициент корреляции Пирсона* характеризует тесноту и направление связи между двумя случайными величинами в случае наличия между ними линейной зависимости.

11.3.2 Коэффициент корреляции.

ОПРЕДЕЛЕНИЕ. Коэффициентом корреляции случайных величин X и Y называется отношение их ковариации к произведению средних квадратических отклонений этих величин:

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}.$$

Свойства коэффициента корреляции.

1. Коэффициент корреляции является безразмерной величиной и обладает следующим свойством:
$$r_{xy} = r_{yx}.$$
2. Значения коэффициента корреляции находятся в диапазоне $[-1, +1]$:
$$-1 \le r_{xy} \le 1 .$$
3. Коэффициент корреляции двух независимых случайных величин X *и* Y равен нулю: $r_{xy} = 0$.
4. Если коэффициент корреляции двух случайных величин равен единице (минус единице), то эти случайные величины связывает линейная функциональная зависимость.

ЗАМЕЧАНИЕ. Подчеркнем, что значение коэффициента корреляции отражает степень линейной зависимости между величинами. Поэтому, даже при наличии ярко выраженной зависимости другого вида (например, квадратичной) его значение может оказаться близким к нулю. Иными словами: равенство нулю коэффициента корреляции означает отсутствие линейной связи, но не означает зависимости нелинейной.

Последовательность вычисления эмпирического значения коэффициента корреляции.

1. Определяются средние значения $\bar{x}$ и $\bar{y}$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{и} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \ .$$

2. Вычисляются значения дисперсий $D(X)$, $D(Y)$ и средних квадратических отклонений σ_x, σ_y:

$$D(X) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2 \ ;$$

$$D(Y) = \left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) - \bar{y}^2 \ ;$$

$$\sigma_x = \sqrt{D(X)}\,;$$

$$\sigma_y = \sqrt{D(Y)}\,.$$

3. Значение ковариации определяется из выражения:

$$K_{xy} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\, y_i\right) - \bar{x}\cdot\bar{y}\,.$$

4. Вычисляется коэффициент корреляции:

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y} \ .$$

Линейный коэффициент корреляции широко применяется при исследовании социальных, экономических и др. процессов и явлений.

11.4. Регрессионный анализ.

Ранее были рассмотрены вопросы предварительного анализа данных на наличие связи, определения ее направления и формы, изучения степени тесноты связи.

Дальнейшие процедуры, связанные с уточнением формы связи, нахождением ее аналитического выражения являются предметом регрессионного анализа.

11.4.1. Задачи регрессионного анализа.

К *задачам регрессионного анализа* относятся:

1) установление формы зависимости;
2) определение функции регрессии;
3) оценка неизвестных значений зависимой переменной.

Регрессия - это односторонняя статистическая зависимость результативного признака *Y* от одного (или нескольких) факторных признаков *X*.

Функция регрессии графически представляет собой кривую (поверхность) которая дает наилучшее (в смысле некоторого критерия) приближение к исходным данным (для чего обычно используется *метод наименьших квадратов*).

Уравнение *парной регрессии* определяет среднее значение результативного признака (*Y*) в зависимости от значений факторного признака (*X*), в предположении, что остальные факторы, влияющие на *Y* и не связанные с *X*, остаются неизменными. Уравнение взаимосвязи может быть записано как (см. рисунок ниже):

$$Y = \varphi(X) + \varepsilon,$$

где: $\varphi(X)$ - функция парной регрессии; и

ε - *случайное отклонение* (*возмущение*) от функции регрессии.

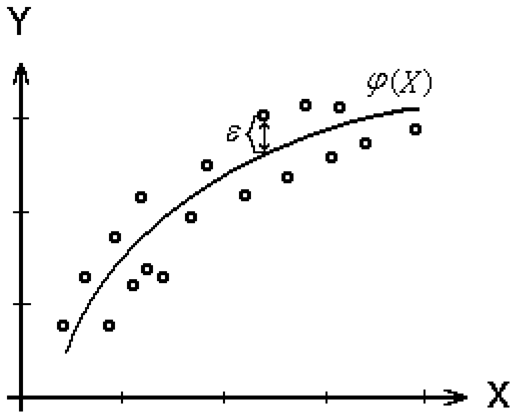


Рис. 20. Пример парной регрессии.
$\varphi(X)$ - функция парной регрессии; ε - случайное отклонение.

Будем полагать, что случайное отклонение (возмущение) ε удовлетворяет следующим условиям:

- математическое ожидание величины данного возмущения равно нулю:

$$M(\varepsilon) = 0;$$

- случайное возмущение ε имеет нормальный закон распределения:

$$N(0, \sigma_{\varepsilon}^{2}).$$

Тогда уравнение парной регрессии однозначной функции $\varphi(X)$ может быть записано как:

$$M(Y) = \varphi(X).$$

В зависимости от вида аналитического выражения $\varphi(X)$ различают *прямолинейную* и *криволинейную связи*.

Прямолинейная связь имеет место, когда с возрастанием (или убыванием) значений X наблюдается тенденция возрастания (убывания) значения Y, при этом изменение Y носит более или менее равномерный характер. В этом случае уравнение связи записывается так:

$$\bar{y}_x = b_0 + b_1 x.$$

Криволинейная форма связи может выражаться различными кривыми, среди которых часто используются:

- парабола второго порядка:

$$\bar{y}_x = b_0 + b_1 x + b_2 x^2;$$

- гипербола:

$$\bar{y}_x = b_0 + \frac{b_1}{x};$$

- показательная:

$$\bar{y}_x = b_0 b_1^x;$$

- логарифмическая (показательная регрессия в логарифмическом масштабе):

$$\ln \bar{y}_x = \ln b_0 + x \ln b_1.$$

Как было указано выше, функция регрессии должна обеспечить наилучшее приближение к исходным данным, для чего обычно используется метод наименьших квадратов (в котором предполагается, что сумма квадратов отклонений теоретических значений от эмпирических должна быть минимальной). Поэтому, после определения формы связи и выбора реализующего его уравнения, производят уточнение параметров этого уравнения (значений входящих в него коэффициентов), используя данный (или иной) метод.

11.4.2. Парная линейная регрессия

Определим параметры линейного уравнения регрессии вида:

$$\bar{y}_x = b_0 + b_1 x \bullet$$

Запишем условие минимизации суммы квадратов отклонений - S_T эмпирических данных от линии регрессии:

$$S_T = \sum (y - \bar{y}_x)^2 \to \min \;;$$

т.е:

$$S_T = \sum (y - b_0 - b_1 x)^2 \to \min .$$

Рассматривая S_T как функцию переменных b_0 и b_1 найдем их значения, обеспечивающих минимум S_T. Для этого определим и приравняем к нулю соответствующие частные производные:

$$\frac{\partial S_T}{\partial b_0} = 2b_0 n - 2\sum_{i=1}^{n} y_i + 2b_1 \sum_{i=1}^{n} x_i = 0;$$

$$\frac{\partial S_T}{\partial b_1} = 2b_1 x^2 - 2b_0 \sum_{i=1}^{n} x_i + 2\sum_{i=1}^{n} x_i y_i = 0 .$$

Отсюда получаем систему уравнений:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases},$$

где: n - число измерений;

$\sum_{i=1}^{n} x_i$ - сумма значений факторного признака;

$\sum_{i=1}^{n} y_i$ - сумма значений результативного признака.

Решая данную систему, находим значения параметров b_0 *и* b_1, обеспечивающих условие минимизации[19]:

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Можно привести и иные (в некоторых случаях, более удобные) формулы вычисления b_1. В частности, разделив числитель и знаменатель первой формулы на n^2, получим:

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\cdot\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2} = \frac{K_{xy}}{D(X)} ,$$

[19] Отметим, что в настоящее время существует множество компьютерных программ, реализующих методы регрессионного анализа. Поэтому, нет необходимости решать эти уравнения вручную.

т.е.:

$$b_1 = \frac{K_{xy}}{D(X)},$$

где: K_{xy} - эмпирическое значение ковариации; и

$D(X)$ - эмпирические значение дисперсии *X*.

Приведем также и другие применяемые формулы вычисления b_1:

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{(x)^2} - (\bar{x})^2},$$

$$b_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2},$$

$$b_1 = r_{yx} \frac{\sigma_y}{\sigma_x} .$$

ЗАМЕЧАНИЕ 1. Параметр b_1 называется *коэффициентом регрессии Y по X* и часто обозначается как b_{yx}. Он обладает размерностью отношения *y* к *x* и показывает, на сколько единиц изменится в среднем *Y* при изменении *X* на 1 единицу. Если $b_1 > 0$, то наблюдаем положительную связь. Если $b_1 < 0$, то связь — отрицательная.

Смысловая интерпретация параметра b_0 определяется смысловым содержанием изучаемых признаков.

ЗАМЕЧАНИЕ 2. Укажем на следующие, употребительные в статистике, формы записи уравнения линейной регрессии:

$$y_x - \bar{y} = b_{yx}(x - \bar{x}),$$

или:

$$y_x - \bar{y} = r_{yx} \frac{\sigma_y}{\sigma_x}(x - \bar{x}) .$$

Аналогично, уравнение регрессии *X* на *Y* имеет вид:

$$x_y - \bar{x} = r_{xy} \frac{\sigma_x}{\sigma_y}(y - \bar{y}).$$

Контрольные вопросы по теме 11

1. В чем заключается задача регрессионного анализа?
2. Из каких точек состоит поле корреляции?
3. Как строится график эмпирической линии регрессии?
4. Какой коэффициент стоит при независимой переменной в уравнении линейной регрессии?

ЛИТЕРАТУРА

1. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для бакалавров. (Гриф МО РФ) – М.: Юрайт, 2012
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. Учебное пособие. (Гриф МО РФ) – М.: Высшее образование, 2010
3. Ермаков В.И. Сборник задач по высшей математике для экономистов. Учебное пособие. (Гриф МО РФ) – М.: Инфра-М, 2009
4. Кремер Н.Ш.Теория вероятностей и математическая статистика. /под ред. *Н.Ш. Кремер, Б.А. Путко, И.М. Тришин* и др./ 3-е изд. перераб. и доп. (Гриф УМО МО РФ) – М.: Юнити-ДАНА, 2010
5. Красс М.С. Чупрынов Б.П. Математика для экономистов. (Гриф МО РФ), Питер, 2010
6. Макаров С.И. Математика для экономистов [Электронный ресурс] : электронный учебник. - 2-е изд., стереотип. – М.: Кнорус, 2009 (1CD-ROM)
7. Плис А.А., Сливина Н.А. MATHCAD математический практикум. (Гриф МО РФ). – М.: Финансы и статистика, 2003

Интернет – ресурсы

1. http://exponenta.ru/educat/class/class.asp (Internet-класс по высшей математике)
2. http://www.teorver-online.narod.ru/teorver73.html (лекции по теории вероятностей и математической статистике)
3. http://www.mathem.h1.ru/ (математика On-Line)

Приложение 1

Таблица значений функции $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ (кривая вероятностей)

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0.2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0.1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551

2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,1	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,2	0,0001	0,0001	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Приложение 2

Таблица значений функции Лапласа $\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{\frac{-t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670

2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997
4,0	0,499968									
4,5	0,49997									
5,0	0,4999997									

Приложение 3

Таблица значений функции Пуассона:

$$P(X = m) = P_{n,m} \approx \frac{\lambda^m}{m!} e^{-\lambda}$$

T \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0.9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1638	0,2222	0,2681	0,3033	0,3293	0,3476	0,3596	0,3696
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4	-	-	0,0002	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5	-	-	-	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6	-	-	-	-	-	-	0,0001	0,0002	0,0003

m \ λ	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0055
3	0,0313	0,1804	0,2240	0,1954	0,1404	0,0892	0.0521	0,0286	0,0150
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337
5	0,0081	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607

m \ λ	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911
7	0,0001	0,0034	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1318
8	-	0,0009	0,0081	0,0298	0,0655	0,1033	0,1304	0,1396	0,1318

9	-	0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,0318
10	-	-	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1180
11	-	-	0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970
12	-	-	0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728
13	-	-	-	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504
14	-	-	-	0,0001	0,0005	0,0022	0,0071	0,0169	0,0324
15	-	-	-	-	0,0002	0,0009	0,0033	0,0090	0,0194
16	-	-	-	-	-	0,0003	0,0014	0,0045	0,0109
17	-	-	-	-	-	0,0001	0,0006	0,0021	0,0058
18	-	-	-	-	-	-	0,0002	0,0009	0,0029
19	-	-	-	-	-	-	0,0001	0,0004	0,0014
20	-	-	-	-	-	-	—	0,0002	0,0006
21	-	-	-	-	-	-	-	0,0001	0,0003
22	-	-	-	-	-	-	-	-	0,0001

Приложение 4

Критические точки распределения χ^2

Число степеней свободы *k*	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63

20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Приложение 5

Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,00	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,70
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,28	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88

20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
Число степеней свободы *k*	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Число степеней свободы *k*	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,07	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
Число степеней свободы *k*	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Приложение 6

Критические точки распределения Фишера-Снедекора (K_1 — число степеней свободы большей дисперсии, K_2 — число степеней свободы меньшей дисперсии)

	Уровень значимости α = 0,01											
	K_1											
K_2	*1*	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	38,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,96	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4.71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,72	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,44	4,10	3,93	3,79	3,68	3,59	3,52	3,45

	Уровень значимости $\alpha = 0{,}05$											
	K_1											
K_2	*1*	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38