

КРАСНОЯРСКИЙ ГОСУДАРСТВЕННЫЙ
АГРАРНЫЙ УНИВЕРСИТЕТ
АЧИНСКИЙ ФИЛИАЛ

**«ТЕОРИЯ ВЕРОЯТНОСТЕЙ И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»**

Учебное пособие



Ачинск 2016

Министерство сельского хозяйства Российской Федерации
Красноярский Государственный Аграрный Университет
Ачинский филиал

Фёдорова И. А.

**«ТЕОРИЯ ВЕРОЯТНОСТЕЙ И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»**

Учебное пособие

Ачинск 2016

Рецензенты:

В.И. Иванов , к. ф-м. н., доцент, зав. кафедрой высшей и прикладной математики Красноярского государственного аграрного университета

Н.Н. Косырев , к. ф-м. н. научный сотрудник Института физики им. Л. В. Киренского

Составители:

Федорова И.А.

Федорова И.А.

Теория вероятностей и математическая статистика: учебное пособие/И.А. Федорова ; Краснояр. гос. аграр. ун-т. Ачинский ф-л.- Ачинск, 2016– 237 с.

Методические указания содержат теоретические сведения, практические задания, лабораторный практикум и методические указания к выполнению контрольной работы по теории вероятностей и математической статистике и предназначены для студентов всех форм обучения, обучающихся по направлению 38.03.01 «Экономика»

© Фёдорова И. А., 2016

© ФГБОУ ВО «Красноярский государственный аграрный университет» Ачинский филиал, 2016

Оглавление

Введение	5
Теоретическая часть	6
Практические задания.....	135
Лабораторные работы.....	167
Варианты контрольных заданий.....	217
Вопросы для подготовки к экзамену.....	225
Приложения.....	229
Заключение	235
Библиографический список.....	236

Введение

Слова «случай», «случайность», «случайно» едва ли не самые употребительные в любом языке. Случайность противопоставляется ясной и четкой информации, строгому логическому развитию событий. Однако так уж велика пропасть между случайным и неслучайным? Ведь случайность, когда она проявляется в поведении не одного объекта, а многих сотен и даже тысяч объектов, обнаруживает черты закономерности. Философы говорят: «путь, которым необходимость идет к цели, вымощен бесконечным множеством случайностей».

Мир – это бесконечное многообразие явлений. Непосредственное общение с миром приводит к мысли, что все явления разделяются на два вида: необходимые и случайные. Необходимые кажутся нам явлениями неизбежно происходящими, а случайные – явлениями, могущими как произойти так и не произойти в одно и то же время. Существование и изучение необходимых явлений представляется естественным, закономерным. А случайные явления в обыденном представлении кажутся нам крайне редкими, не имеющими закономерностей; они как бы нарушают естественный ход событий. Однако случайные явления происходят всюду и постоянно. В результате взаимодействия многих случайностей появляется ряд явлений, в закономерности которых мы не сомневаемся. Случайность и закономерность неотделимы друг от друга.

ТВ – математическая наука изучающая закономерность в массовых однородных случаях, явлениях и процессах.

Элементарные события – это простейшие не разложимые результаты опыта. Вся совокупность элементарных событий – пространство элементарных событий. Под опытом в ТВ понимается выполнение некоторого комплекса условий в результате которого происходят или не происходят некоторые события – факты.

Теоретические сведения

Раздел 1. Теория вероятностей.

Тема 1.1. Случайные события.

Виды случайных явлений.

Событие — это качественный и (или) количественный результат испытания (исход), осуществляемого при определённой совокупности условий. Для обозначения событий используются большие буквы латинского алфавита: A, B, C и т.д.

Различают следующие типы событий: случайные события, совместные или несовместные события, достоверные или невозможные события, зависимые или независимые события, равновозможные события, элементарные (простые, неразложимые) события, событие или совокупность событий (исходов), благоприятствующих какому-либо другому событию.

Случайное событие – это результат испытания (или величина), который нельзя заранее спрогнозировать, т.е. нельзя сказать, произойдёт это событие или не произойдёт, или, если событие произойдёт, то неизвестно, какое значение примет результат этого события.

Случайные события – первичные, неопределяемые (в строгом смысле) понятия в теории вероятностей, аналогичные понятиям точки и прямой – в геометрии.

Событие, которое обязательно наступит (никогда не произойдёт) в данном опыте, называется достоверным (невозможным). Достоверное событие обозначают символом Ω , а невозможное – $\emptyset$.

Два случайных события называются несовместными, если наступление одного из них исключает наступление другого в одном и том же испытании. (Таким образом, несовместные события не могут наступать одновременно). В противном случае, т.е. если наступление одного события не исключает наступление другого события в одном и том же испытании, то эти события называются совместными.

События называются равновозможными, если ни одно из них не является более возможным по сравнению с другими событиями.

События называются независимыми (зависимыми), если числовая характеристика возможности наступления одного события не зависит (зависит) от числовых характеристик наступления других событий.

Величина – это общее название различных количественных характеристик предметов и явлений. Длина, площадь, температура, давление и т. д. – примеры разных величин.

Величина, которая принимает различные числовые значения под влиянием случайных обстоятельств, называется случайной величиной. Примеры случайных величин: число больных на приеме у врача; точные размеры внутренних органов людей и т. д.

Различают дискретные и непрерывные случайные величины.

Случайная величина называется дискретной, если она принимает только определенные отделенные друг от друга значения, которые можно установить и перечислить.

Примерами дискретной случайной величиной являются:

- число студентов в аудитории – может быть только целым положительным числом: 0,1,2,3,4..... 20.....;

- цифра, которая появляется на верхней грани при бросании игральной кости – может принимать лишь целые значения от 1 до 6;

- относительная частота попадания в цель при 10 выстрелах – ее значения: 0; 0,1; 0,2; 0,3 ...1

- число событий, происходящих за одинаковые промежутки времени: частота пульса, число вызовов скорой помощи за час, количество операций в месяц с летальным исходом и т. д.

Случайная величина называется непрерывной, если она может принимать любые значения внутри определенного интервала, который иногда имеет резко выраженные границы, а иногда – нет*. К непрерывным случайным

величинам относятся, например, масса тела и рост взрослых людей, масса тела и объем мозга, количественное содержание ферментов у здоровых людей, размеры форменных элементов крови, рН крови и т. п.

Понятие случайной величины играет определяющую роль в современной теории вероятностей, разработавшей специальные приемы перехода от случайных событий к случайным величинам.

Если случайная величина зависит от времени, то можно говорить о случайном процессе.

Случайный процесс (СП) - это некоторый процесс или явление, поведение которого в течение времени и результат заранее предсказывать невозможно. Примеры случайных процессов: динамика изменения курса валют или акций, выручка или прибыль организации с течением времени, объемы продаж товара и т.д. Если случайный процесс может изменить своё состояние только в строго определённый момент времени, то он называется процессом с дискретным временем. Если же смена состояния возможна в произвольный момент времени, то это СП с непрерывным временем.

Если в любой момент времени СП представляет собой дискретную случайную величину (ее значение можно перечислить и выделить два соседних значения), то это процесс с дискретным состоянием.

Если же в любой момент времени состояние может меняться непрерывно, плавно и нельзя выделить два соседних состояния, то это СП с непрерывным состоянием.

Таким образом, возможно 4 вида СП:

1) СП с непрерывным временем и непрерывным состоянием (пример: температура воздуха в некоторый момент времени, изменяется плавно в любой момент времени).

2) СП с непрерывным временем и дискретным состоянием (пример: число посетителей в магазине, изменяется кратно одному в любой момент времени).

3) СП с дискретным временем и непрерывным состоянием (пример: динамика курса курс валюты, изменяется плавно в момент валютных торгов).

4) СП с дискретным временем и дискретным состоянием (пример: число пассажиров в транспорте изменяется кратно одному и только в определенные моменты времени, на остановках).

Аксиоматический подход.

Для построения математической теории важно не только установить естественно-научное значение основных понятий, но и формально перечислить их основные свойства. Так, в геометрии простейшие, лежащие в основе построения теории свойства точек и прямых, перечисляются в аксиомах, например: через две различные точки проходит одна и только одна прямая. Точно так же, логическое построение теории вероятностей основано на фиксации первичных, не подлежащих определению понятий данной теории. Их основные свойства формулируются в виде ряда аксиом. После этого все предложения теории (иначе говоря, все теоремы) выводятся из аксиом строго логическим путем, без обращения к посторонним понятиям опыта, наглядности, устойчивости частот и т.д.

Современная аксиоматика теории вероятностей принадлежит советскому математику А.Н. Колмогорову.

Приведём сначала некоторые предварительные соображения.

Заметим, что среди событий по отношению к данному опыту можно выделить такие, которые являются простейшими, элементарными.

Элементарные события характеризуются тем, что при каждом осуществлении опыта наступает одно и только одно из них и, что любое событие A , связанное с данным опытом, должно распадаться на элементарные, т.е. представляться в виде суммы некоторого множества элементарных событий.

Пример1. В опыте с бросанием игральной кости события $\omega_1, \omega_2, \dots, \omega_6$, где ω_i означает выпадение i очков, являются элементарными. Действительно, любое событие, представляющее интерес в связи с бросанием игральной кости, формулируется как некоторое условие на число очков (например: число

очков чётное, число очков превосходит 3 и т.д.). Отсюда ясно, что каждое событие, связанное с данным опытом, распадается на элементарные события ω_i .

Пример2. Пусть опыт заключается в «бросании» точки в области Ω на плоскости. Элементарное событие – это попадание точки в определенную точку области Ω . Ясно, что суммами таких событий исчерпываются все мыслимые исходы опыта. В данном примере, в отличие от предыдущего, множество элементарных событий бесконечно. Можно считать его совпадающим с множеством всех точек области Ω .

Возвращаясь к общей ситуации, обозначим через Ω множество всех элементарных событий для данного опыта. Каждому событию A , связанному с данным опытом, можно сопоставить **подмножество** множества Ω ; это подмножество состоит из тех элементарных событий, на которые распадается A . В аксиоматике Колмогорова событие A отождествляется с соответствующим подмножеством в Ω . В такой интерпретации, например: событие «на игральной кости выпало нечётное число очков» есть подмножество $\{\omega_1, \omega_3, \omega_5\}$ в множестве $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$, а событие «точка попала в подобласть D области Ω на Рис.1.» есть подмножество множества Ω , состоящее из всех точек D указанной подобласти.

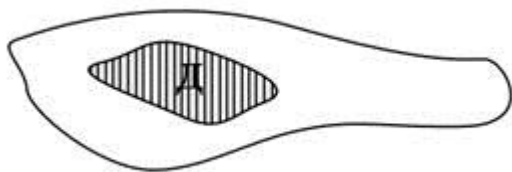


Рис.1

Такой подход к понятию события удобен тем, что, благодаря ему, понятия суммы событий, произведения событий и противоположного события приобретают естественно теоретико-множественный смысл, а именно: сумма событий A и B превращается в объединение соответствующих подмножеств,

произведение событий A и B – в пересечение этих подмножеств, а противоположное событие $\bar{A}$ – в дополнении к подмножеству A в Ω .

Таким образом, математическая формализация модели случайного опыта разбивается на две группы аксиом.

Алгебра событий.

Алгебра событий – алгебра подмножеств пространства элементарных событий, элементами которого служат элементарные события.

Как и положено алгебре множеств *алгебра событий* содержит невозможное событие (пустое множество) и замкнута относительно теоретико-множественных операций, производимых в конечном числе. Достаточно потребовать, чтобы *алгебра событий* была замкнута относительно двух операций, например, пересечения и дополнения, из чего сразу следует её замкнутость относительно любых других теоретико-множественных операций.

1. Суммой $A+B$ двух событий A и B называют событие, состоящее в том, что произошло хотя бы одно из событий A и B . Суммой нескольких событий, соответственно, называется событие, заключающееся в том, что произошло хотя бы одно из этих событий.

2. Произведением AB событий A и B называется событие, состоящее в том, что произошло и событие A , и событие B . Аналогично произведением нескольких событий называется событие, заключающееся в том, что произошли все эти события.

3. Разностью $A \setminus B$ событий A и B называется событие, состоящее в том, что A произошло, а B – нет.

4. События A и B называются совместными, если они могут произойти оба в результате одного опыта. В противном случае (то есть если они не могут произойти одновременно) события называются несовместными.

5. Говорят, что события $A_1, A_2, \dots, A_n$ образуют полную группу, если в результате опыта обязательно произойдет хотя бы одно из событий этой группы.

6. События называются равновозможными, если нет оснований считать, что одно из них является более возможным, чем другое.

Пространство элементарных событий.

Пространство элементарных событий — множество Ω всех различных исходов случайного эксперимента.

Элемент этого множества $\omega \in \Omega$ называется **элементарным событием** или **исходом**. Пространство элементарных событий называется **дискретным**, если число его элементов конечно или счётно. Любое пространство элементарных событий не являющееся дискретным, называется **недискретным**, и при этом, если наблюдаемыми результатами (нельзя произносить случайными событиями) являются точки того или иного числового арифметического или координатного пространства, то пространство называется **непрерывным**. Пространство элементарных событий Ω вместе с алгеброй событий $\mathcal{F}$ и вероятностью P образует тройку $(\Omega, \mathcal{F}, P)$, которая называется вероятностным пространством.

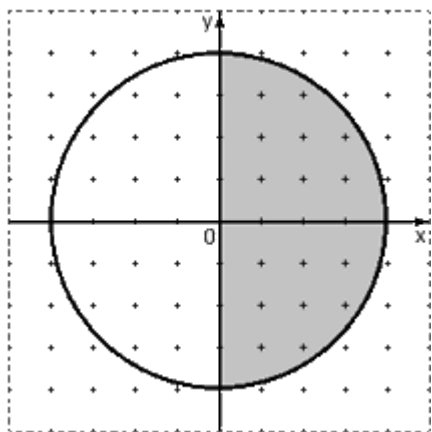
Непосредственный подсчет вероятностей.

Геометрическая вероятность.

Геометрическая вероятность события A , являющегося подмножеством множества Ω точек на прямой или плоскости — это отношение площади фигуры A к площади всего множества Ω :

$$P(A) = \frac{S(A)}{S(\Omega)}$$

Задача. Мишень имеет форму окружности радиуса 4. Какова вероятность попадания в ее правую половину, если попадание в любую точку мишени равновероятно? При этом промахи мимо мишени исключены.



Взглянем на картинку: нас устроит любая точка из правого полукруга. Очевидно, площадь $S(A)$ этого полукруга составляет ровно половину площади всего круга, поэтому имеем:

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{0,5 \cdot S(\Omega)}{S(\Omega)} = 0,5$$

Как видите, ничего сложного в геометрической вероятности нет. Однако даже в Москве многие репетиторы по высшей математике стараются обойти эту тему стороной, поскольку считают ее необязательной. Результат — непонимание материала и, как следствие, проблемы на экзамене по теории вероятностей.

Чтобы наглядно представить себе, что такое геометрическая вероятность, возьмите лист бумаги и начертите произвольную фигуру. Треугольник, квадрат или окружность — что угодно. Затем возьмите острый, хорошо заточенный карандаш и ткните им в любую точку фигуры. Повторите этот нехитрый процесс несколько раз. Если исключить попадания за пределами фигуры, то получится вот что:

1. Вероятность попадания в фигуру равна $P(\Omega) = 1$. Это вполне логично, поскольку вся наша фигура — это и есть пространство элементарных событий Ω ;

2. Если некоторую точку (элементарное событие) отметить заранее, то вероятность попадания именно в нее равна нулю. Даже если специально «целиться», точного попадания не будет. Ошибка составит тысячные доли миллиметра, но не ноль;

3. Теперь возьмем две точки. Вероятность попадания в любую из них все равно ноль. Аналогично, если взять 3 точки. Или пять — без разницы.

Аксиоматическое определение вероятности.

Аксиоматическое определение вероятности. Пусть задано пространство элементарных событий E и каждому событию $A \subset E$ поставлено в соответствие единственное число $P(A)$ такое, что:

- 1) $P(A) \geq 0$,
- 2) для каждой пары несовместных событий $A, B \subset E$ имеет место равенство:
 $P(A \cup B) = P(A) + P(B)$,
- 3) $P(E) = 1$.

Тогда говорят, что на событиях в множестве E задана вероятность, а число $P(A)$ называется вероятностью события A .

Основные теоремы теории вероятностей.

Теорема сложения вероятностей

Случайные события $(A, B, C, D \dots)$ называются несовместными, если появление одного из них исключает появление других событий в одном и том же испытании.

Пример 1. Подброшена монета. При ее падении появление «герба» исключает появление «решки» (надписи, определяющей цену монеты). События «выпал герб» и «выпала решка» несовместные.

Пример 2. Получение студентом на одном экзамене оценки «2», или «3», или «4», или «5» — события несовместные, так как одна из этих оценок исключает другую на том же экзамене.

Для несовместных случайных событий выполняется теорема сложения вероятностей: вероятность появления одного, но все равно какого, из нескольких несовместных событий $A_1, A_2, A_3 \dots A_k$ равна сумме их вероятностей:

$$P(A_1 \text{ или } A_2 \dots \text{ или } A_k) = P(A_1) + P(A_2) + \dots + P(A_k). \quad (4)$$

Пример 3. В урне находится 50 шаров: 20 белых, 20 черных и 10 красных. Найдите вероятность появления белого (событие А) или красного шара (событие В), когда шар наугад достают из урны.

Решение: $P(A \text{ или } B) = P(A) + P(B)$;

$$P(A) = 20/50 = 0,4;$$

$$P(B) = 10/50 = 0,2;$$

$$P(A \text{ или } B) = P(\text{б. ш. или к. ш.}) = 0,4 + 0,2 = 0,6 = 60\%.$$

Условные вероятности. Теорема умножения вероятностей

В ряде случаев приходится рассматривать вероятности случайных событий, если известно, что некоторое случайное событие В уже произошло. Дадим следующее

Определение. Пусть событие В имеет положительную вероятность. Условной вероятностью события А при условии, что произошло событие В, называется число

Пример1. Последовательно брошены 3 монеты. Найти вероятность выпадения герба на первой монете при условии, что выпала хотя бы одна решка.

Решение. Множество элементарных исходов имеет вид поэтому $n = 8$. Событие $A = \{ГГГ, ГГР, ГРГ, ГРР\}$ — выпадение герба на первой т.е $k = 4$. Отсюда

Событие $B = \{ГГР, ГРГ, РГГ, ГРР, РГР, РРГ, РРР\}$ - выпадение хотя бы одной решки; здесь $k = 7$. Следовательно, $P(B) = 7/8$. Событие $= \{ГГР, ГРГ, ГРР\}$, т.е. $k = 3$. Отсюда . Используя определение условной вероятности, имеем

Непосредственно из определения условной вероятности сл теорема умножения вероятностей.

Теорема 1 (умножения вероятностей). Если $P(A) > 0$, $P(B) > 0$, то имеют место равенства

Доказательство теоремы следует из формул

Случайные события A и B называются независимыми, если Последнее равенство часто называют теоремой умножения вероятностей для независимых событий.

Следствие 1. Пусть $P(B) > 0$. События A и B независимы тогда и только тогда, когда $P(A/B) = P(A)$.

Доказательство.

Действительно, если A и B независимы, то

Докажем обратное утверждение: Пусть ; тогда и, следовательно, , т.е. события независимы.

Следствие доказано.

Следствие 1 часто используют в качестве определения независимости событий.

Следствие 2. Для произвольных событий $A_1, A_2, \dots, A_n$ таких, что имеем События $A_1, A_2, \dots, A_n$ называются попарно независимыми, если События $A_1, A_2, \dots, A_n$ называют независимыми в совокупности, если для любого k , и для любого набора индексов выполняется: где .

Следствие 3. Если события A и B независимы, то независимы и события A и B , и A , и B и .

Пример 3. Три стрелка, для которых вероятность попадания 0,7, 0,8 и 0,9, соответственно, стреляют в мишень.

Найти вероятность того, что в мишени окажется два попадания. Решение. Введем события $A_1 = \{\text{первый стрелок выстрелил и попал}\}$, $A_2 = \{\text{второй стрелок выстрелил и попал}\}$, $A_3 = \{\text{третий стрелок выстрелил и попал}\}$. Тогда $P(A_1) = 0,7$, $P(A_2) = 0,8$, $P(A_3) = 0,9$.

Искомое событие $B = (\text{после залпа в мишени два попадания})$ можно записать следующим образом: Применяя теорему умножения вероятностей для независимых событий, получим

$$P(B) = P(A_1) \times P(A_2) \times P(\bar{A}_3) + P(A_1) \times P(\bar{A}_2) \times P(A_3) + P(\bar{A}_1) \times P(A_2) \times P(A_3) =$$

$$P(A_1) \times P(A_2) \times (1 - P(A_3)) + P(A_1) \times (1 - P(A_2)) \times P(A_3) + (1 - P(A_1)) \times P(A_2) \times P(A_3) = 0,7 \times 0,8 \times (1 - 0,9) + 0,7 \times (1 - 0,8) \times 0,9 + (1 - 0,7) \times 0,8 \times 0,9 = 0,056 + 0,126 + 0,216 = 0,398$$

Пример 4. Партия товара содержит 12 изделий, среди которых 4 не стандартных. Наудачу вынимают 3 изделия одно за другим. Найти вероятность того, что все 3 изделия стандартны.

Решение. Введем события $A_1 = \{\text{первое взятое изделие стандартно}\}$, $A_2 = \{\text{второе изделие стандартно}\}$, $A_3 = \{\text{третье взятое изделие стандартно}\}$, $A = \{\text{все три взяты стандартны}\}$. Тогда очевидно, что Используя следствие 2, имеем

Рассчитаем вероятности этих событий. Очевидно, что , так как имеется 8 стандартных изделий среди 12. Если первое взятое изделие стандартно, вероятность вытащить второе стандартное изделие равна , так как после того, как событие A совершилось, осталось 7 стандартных изделий из 11 оставшихся в партии. Далее, после того, как совершились события A_1 и A_2 получим Общий результат будет следующим

Формула полной вероятности. Формула Байеса.

Случайные события образуют полную группу событий, если они попарно несовместны и в сумме дают достоверное событие

Пример 1. Эксперимент состоит в подбрасывании монеты один раз. Тогда и полную группу будут образовывать события $H_1 = \{\Gamma\}$, $H_2 = \{P\}$.

Теорема 1. Если $H_1, H_2, \dots, H_n$ - полная группа событий и $P(H_i) > 0$, $i = 1, 2, \dots, n$, то для любого случайного события A верна формула полной вероятности

Доказательство.

Заметим, что

Применим теорему сложения вероятностей для попарно несовместных событий и теорему умножения вероятностей: Теорема доказана. События H_1 ,

$H_2, \dots, H_n$, называют гипотезами а их вероятности - априорными (до опытными) вероятностями. Сумма вероятностей гипотез всегда равна единице.

Пример 2. В урну, содержащую 2 шара, опущен белый шар, после чего из нее наудачу извлечен один шар. Найти вероятность того, что извлеченный шар окажется белым, если равновозможны все возможные предположения о первоначальном составе шаров по цвету.

Решение. Обозначим $A = \{\text{извлечен белый шар}\}$. Тогда возможны следующие гипотезы (предположения) относительно первоначального состава урны: $H_1 = \{\text{белых шаров нет}\}$; $H_2 = \{\text{один белый шар}\}$; $H_3 = \{\text{два белых шара}\}$. Поскольку, по условию задачи, все три гипотезы равновозможны, и сумма вероятностей гипотез равна 1, то

Так как событие A может произойти с одной из заданных гипотез, то вероятность события A будем вычислять по формуле полной вероятности

Вероятности гипотез $P(H_i)$ нами определены. Остается вычислить $P(A/H_i)$, $i=1,2,3$. $P(A/H_1)$ - вероятность того, что будет извлечен белый шар при условии, что в урне первоначально не было белых шаров. Она равна 0, так как среди трех шаров, находящихся в урне после того, как туда опустили один белый шар, имеется только один белый шар. Аналогично, - вероятность того, что будет извлечен белый шар, при условии, что первоначально в урне находился один белый шар. Поэтому $P(A/H_2) = 1/2$. И, наконец, - вероятность того, что будет извлечен белый шар, при условии, что первоначально в урне находятся два белых шара, равна 1. В итоге имеем

Английский математик Т. Бейес (или Байес) доказал одну из основных теорем теории вероятностей.

Теорема 2. Если $\{H_i\}$ - полная группа событий $P(H_i) > 0$, $i = 1, 2, \dots, n$, то для любого случайного события A , такого, что $P(A) > 0$, верна формула Байеса

Доказательство.

Используя определение условной вероятности, теорему умножения и формулу полной вероятности, получим Теорема доказана.

Смысл формулы Байеса заключается в следующем. Предположим, что проведен опыт, в результате которого наступило событие A . Формула Байеса позволяет переоценить гипотезы. Условные вероятности $P(H_k / A)$ называют апостериорными (после опытными) вероятностями.

Пример3. Два автомата производят одинаковые детали, которые поступают на общий конвейер. Производительность первого автомата вдвое больше производительности второго. Первый автомат производит в среднем 60% деталей отличного качества, а второй — 84%. Наудачу взятая с конвейера деталь оказалась отличного качества. Найти вероятность того, что эта деталь произведена первым автоматом.

Решение. Пусть событие A = (деталь отличного качества). Можно сделать два предположения (две гипотезы): H_1 = (деталь произведена первым автоматом), H_2 = (деталь произведена вторым автоматом). Поскольку первый автомат производит вдвое больше деталей, чем второй, очевидно, что $P(H_1) = 2/3$, $P(H_2) = 1/3$. Условная вероятность того, что деталь будет отличного качества, если она произведена первым автоматом, равна $P(A/H_1) = 0,6$, а условная вероятность того, деталь отличного качества, если она произведена вторым автоматом, равна $P(A/H_2) = 0,84$. Тогда можно вычислить вероятность того, что наудачу взятая деталь - отличного качества. По формуле полной вероятности

Задача заключается в переоценке вероятности гипотезы H_1 , а именно: вычислить вероятность того, что взятая деталь отличного качества произведена первым автоматом. По формуле Байеса получаем

Пример. Батарея из трех орудий произвела залп, причем два снаряда попали в цель. Найти вероятность того, что первое орудие попало в цель, если вероятности попадания в цель первым, вторым и третьим орудиями, соответственно, равны $p_1 = 0,4$; $p_2 = 0,45$; $p_3 = 0,3$.

Решение. Обозначим через A событие: A = (два орудия попали в цель). Сделаем предположения (гипотезы): H_1 = (первое орудие попало в цель), H_2

= (первое орудие не попало цель). По условию $P(H1) = 0,4$; $P(H2) = 1 - P(H1) = 0,6$. Тогда по формуле полной вероятности получаем, что

Найдем условную вероятность $P(A/H1)$, т.е. условную вероятность того, что в цель попало два снаряда, причем один из них был послан первым орудием и, следовательно второй снаряд послан либо вторым, либо третьим орудием. Тогда $P(A/H1) = p_2(1 - p_3) + (1 - p_2) p_3 = 0,45 \cdot 0,7 + 0,55 \cdot 0,3 = 0,48$.

Теперь найдем условную вероятность $P(A/H2)$, т.е. вероятность того, что в цель попало два орудия, причем первое орудие дало промах. Это возможно тогда, когда второе и третье орудия пали в цель. Поэтому

$$P(A/H2) = p_2 \cdot p_3 = 0,45 \cdot 0,3 = 0,135,$$

и теперь можно вычислить

$$P(A): P(A) = 0,4 \cdot 0,48 + 0,6 \cdot 0,135 = 0,273.$$

Независимые испытания.

С понятием независимых событий связано и имеет широкое применение понятие независимых опытов или испытаний.

Опыты $\alpha_1, \alpha_2, \dots, \alpha_n$ называются независимыми, если любая комбинация их исходов является совокупностью независимых событий. Иначе, если в задаче проводится ряд многократно повторяющихся испытаний $\alpha_1, \alpha_2, \dots, \alpha_n$ при неизменном комплексе условий и в каждом испытании некоторое событие A может наступить с некоторой вероятностью $p = p(A)$ не зависящей от других испытаний, и не наступить с вероятностью $p(\bar{A})$, то указанные испытания называются независимыми. Данная схема независимых испытаний носит название схемы Бернулли.

Схема названа в честь Якоба Бернулли – родоначальника семьи выдающихся швейцарских учёных. (Якоб Б., Иоганн Б., Николай Б., Даниил Б. и др.). Якоб Бернулли доказал так называемую теорему Бернулли – важный частный случай закона больших чисел (см. п. 3.11). Указанная теорема относится к рассматриваемой здесь последовательности независимых испытаний.

Примерами независимых испытаний являются: а) многократное (n раз) подбрасывание монеты; б) извлечение (n раз) одинаковых на ощупь шаров из урны с их последующим возвращением; в) любая совокупность независимых испытаний (опытов), в каждом из которых вероятность успешных исходов одинакова, например, серия выстрелов по мишени, выбор n деталей из их совокупности, изучение n анализов горной породы определённого свойства и т.д.

В схеме Бернулли наступление события A с вероятностью $p = p(A)$ условно называется успехом, а его ненаступление (противоположное событие $\bar{A}$) – неудачей. Вероятность неудачи в каждом опыте такого типа равна $q = 1 - p$.

На практике обычно возникают задачи со сложными событиями, в которых из n опытов, составляющих схему Бернулли, в m опытах ($m < n$) событие A наступает (т.е. завершается успехом), а в $(n - m)$ опытах это событие не наступает (завершается неудачей). Пусть $P_n(k)$ – обозначает вероятность того, что при производстве n опытов успех наступает в k опытах (успех реализуется k раз). Ставится следующая задача: *пусть в n испытаниях, соответствующих схеме Бернулли, k испытаний завершились успехом. Требуется найти вероятность $P_n(k)$ (читается: « P из n испытаний k успешных»).* Данная вероятность рассчитывается по формуле Бернулли, которой соответствует одноименная теорема.

Теорема Бернулли. Если вероятность p наступления события A в каждом из последовательности n испытаний $\alpha_1, \alpha_2, \dots, \alpha_n$ постоянна, то вероятность того, что событие A наступит k раз и не наступит $n - k$ раз, вычисляется по формуле Бернулли:

$$P_n(k) = C_n^k p^k q^{n-k}, \quad (2.1)$$

где $q = 1 - p$.

Доказательство. Действительно, пусть события A_i и $\bar{A}_i$ — появление и не появление соответственно события A в i -ом испытании α_i ($i = 1, 2, \dots, n$). Пусть также B_k обозначает событие, состоящее в том, что в n независимых

испытаниях событие A появилось k раз. При $n = 3$ и $k = 2$ событие B_2 выражается через элементарные события A_i ($i = 1, 2, 3$) по формуле :

$$B_2 = A_1 A_2 \bar{A}_3 + A_1 \bar{A}_2 A_3 + \bar{A}_1 A_2 A_3.$$

В общем виде последняя формула будет такой

(2.2)

т.е каждый член суммы (2.2) соответствует появлению события A k раз и $(n - k)$ раз неоявлений. Число всех комбинаций (слагаемых) в (2.2) равно числу способов выбора из n испытаний k испытаний, в которых событие A произошло, т.е числу сочетаний C_n^k . Вероятность каждой такой комбинации по теореме умножения вероятностей независимых событий равна $p^k \times q^{n-k}$, так как $p(A_i) = p$, $p(\bar{A}_i) = q$, $i = 1, 2, \dots, n$. Но комбинации в (2.2) являются несовместными событиями. Поэтому по теореме сложения вероятностей получим

$$P_n(k) = P(B_k) = p^k q^{n-k} + \dots + p^k q^{n-k} = C_n^k p^k q^{n-k}.$$

Таким образом, имеет место формула Бернулли

$$P_n(k) = C_n^k p^k q^{n-k}.$$

Что и требовалось доказать.

Тема 1.2. Случайные величины.

Понятие случайной величины.

Случайная величина является одним из основных понятий теории вероятностей. Роль случайной величины, как одного из основных понятий теории вероятностей, впервые была чётко осознана П. Л. Чебышевым, который обосновал общепринятую на сегодня точку зрения на это понятие (1867). Понимание случайной величины как частного случая общего понятия функции, пришло значительно позднее, в первой трети 20 века. Впервые полное формализованное представление основ теории вероятностей на базе теории меры было разработано А. Н. Колмогоровым(1933), после которого

стало ясным, что случайная величина представляет собой измеримую функцию, определенную на вероятностном пространстве. В учебной литературе эта точка зрения впервые последовательно проведена У. Феллером.

Случайные величины используются для математического представления таких сторон объектов, систем и событий, количественную характеристику которых до проведения опыта по их измерению, однозначно определить принципиально невозможно. Примером таких систем могут служить микроскопические объекты, состояние которых описывается квантовой механикой. Случайными величинами описываются события передачи наследственных признаков от родительских организмов к их потомкам. Следует также отметить, что существует ряд задач математического анализа и теории чисел для которых участвующие в их формулировках функции целесообразно рассматривать как случайные величины, определенные на подходящих вероятностных пространствах.

Случайной называется величина, которая в результате испытаний принимает то или иное (но при этом только одно) возможное значение, заранее неизвестное, меняющееся от испытания к испытанию и зависящее от случайных обстоятельств. В отличие от случайного события, являющегося качественной характеристикой случайного результата испытания, случайная величина характеризует результат испытания количественно. Примерами случайной величины могут служить размер обрабатываемой детали, погрешность результата измерения какого-либо параметра изделия или среды. Среди случайных величин, с которыми приходится встречаться на практике, можно выделить два основных типа: дискретные и непрерывные.

Дискретной называется случайная величина, принимающая конечное или бесконечное счетное множество значений. Например: частота попаданий при трех выстрелах; число бракованных изделий в партии из

np

штук; число вызовов, поступающих на телефонную станцию в течение суток; число отказов элементов прибора за определенный промежуток времени при испытании его на надежность; число выстрелов до первого попадания в цель и т. д.

Непрерывной называется случайная величина, которая может принимать любые значения из некоторого конечного или бесконечного интервала. Очевидно, что число возможных значений непрерывной случайной величины бесконечно. Например: ошибка при измерении дальности радиолокатора; время безотказной работы микросхемы; погрешность изготовления деталей; концентрация соли в морской воде и т. д.

Случайные величины обычно обозначают буквами

X, YX, Y

и т. д., а их возможные значения —

x, yx, y

и т. д. Для задания случайной величины недостаточно перечислить все ее возможные значения. Необходимо также знать, как часто могут появиться те или иные ее значения в результате испытаний при одних и тех же условиях, т. е. нужно задать вероятности их появления. Совокупность всех возможных значений случайной величины и соответствующих им вероятностей составляет распределение случайной величины.

Распределение Пуассона

Иногда на практике встречаются испытания Бернулли, в которых число испытаний n относительно $=\lambda$ велико, вероятность успеха p относительно мала, а их произведение $n \cdot p$ не мало и не велико. В таких случаях вместо биномиального распределения пользуются его приближением —

распределением Пуассона. При больших n : Обозначим через p_k вероятность $P(X = k)$. Следовательно, или: Далее по индукции Это и есть распределение Пуассона.

Пример 1. На курсе 100 студентов. Каждый может выиграть билет на концерт популярной музыкальной группы с вероятностью $1/20$. Какова вероятность, что 6 человек с курса попадут на концерт? Свяжем испытания Бернулли с каждым из студентов, $n = 100$, $p = 1/20$, $\lambda = 5$. ♦ Найдем математическое ожидание и дисперсию для случайной величины X , распределенной по закону Пуассона. Производящая функция

Пример 2. Математическое ожидание и дисперсия числа студентов, выигравших билет на концерт, совпадают с параметром распределения Пуассона: $MX = DX = 5$.

Пример 3. В начале XX столетия инженер Эрланг заметил, что число звонков, поступающих на телефонную станцию за единицу времени, имеет распределение Пуассона. Параметр этого распределения равен среднему количеству звонков, поступающих на телефонную станцию за эту единицу времени.

Равномерное распределение.

Непрерывное равномерное распределение— в теории вероятностей распределение, характеризующееся тем, что вероятность любого интервала зависит только от его длины.

Равномерное распределение полезно при описании переменных, у которых каждое значение равновероятно, иными словами, значения переменной равномерно распределены в некоторой области.

Определение

Говорят, что случайная величина имеет непрерывное равномерное распределение на отрезке $[a, b]$, где $a, b \in \mathbb{R}$, если её *плотность* имеет вид:

$$f(x|a,b) = \begin{cases} \frac{1}{b-a}, x \in [a,b] \\ 0, x \notin [a,b] \end{cases}$$

Пишут: $X \sim U(a,b)$ или $X \sim R[a,b]$

Иногда значения плотности в граничных точках $x = a$ и $x = b$ меняют на другие, например 0 или $\frac{1}{2(b-a)}$. Так как интеграл Лебега от плотности не зависит от поведения последней на множествах меры нуль, эти вариации не влияют на вычисления связанных с этим распределением вероятностей.

Если $L(\xi) = U(a,b)$, то $L\left(\frac{\xi - a}{b - a}\right) = U(0,1) = Be(1,1)$

Равномерное распределение $U(a,b)$ описывает процесс «выбора точки наудачу» в интервале $[a,b]$. Так, если $[a,b]$ – интервал между последовательными отправлениями автобуса от остановки, то время ожидания пассажира, не знающего расписания и пришедшего на остановку, есть случайная величина с распределением $U(0,1)$. Распределение $U(0,1)$ играет особую роль в методах моделирования с помощью компьютеров случайных величин с заранее заданными распределениями. Такие методы широко используют для приближенных вычислений интегралов, решений дифференциальных и интегральных уравнений и т.д.

Стандартное равномерное распределение

Если $a = 0$, $b = 1$, то есть $X \sim U[0,1]$, то такое непрерывное равномерное распределение называют стандартным. Имеет место элементарное утверждение:

Если случайная величина $X \sim U[0,1]$, и $Y = a + (b - a) \times X$, где $a < b$, то $Y \sim U[a,b]$.

Таким образом, имея генератор случайной выборки из стандартного непрерывного равномерного распределения, легко построить генератор выборки любого непрерывного равномерного распределения.

Более того, имея такой генератор и зная функцию обратную к функции распределения случайной величины, можно построить генератор выборки любого непрерывного распределения (не обязательно равномерного) с помощью метода обратного преобразования. Поэтому, стандартно равномерно распределённые случайные величины иногда называют базовыми случайными величинами.

Линейное преобразование $Y = \frac{X - a}{b - a}$ переводит СВ $X \sim R(a,b)$ в СВ $Y \sim R(0,1)$. Действительно,

$$F_Y(x) = P\{Y \leq y\} = P\{X \leq (b-a)y + a\} = \begin{cases} 0, & y < 0 \\ y, & 0 \leq y \leq 1 \\ 1, & y > 1 \end{cases}$$

Равномерное распределение является непрерывным аналогом дискретного распределения вероятностей для опытов с равновероятными исходами.

Значение

С помощью линейного преобразования приводится к равномерному распределению на отрезке $[0,1]$. Равномерное распределение является непрерывным аналогом распределений классической теории вероятностей, описывающих случайные эксперименты с равновероятными исходами.

Погрешность, происходящая от округления числа, удовлетворительно описывается равномерным распределением на отрезке $[-1/2, 1/2]$.

Если случайная величина ζ имеет непрерывную функцию распределения $F_\zeta(x)$, то случайная величина $\xi = F_\zeta(\zeta)$ имеет равномерное распределение на отрезке $[0,1]$. Этим объясняется широкое использование равномерного распределения в статистическом моделировании (методы Монте-Карло).

Нормальное распределение

Нормальное распределение (синоним - гауссово распределение) - распределение непрерывной случайной величины с плотностью

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

Здесь "мю" - математическое ожидание, "сигма" - среднеквадратическое отклонение (корень квадратный из дисперсии). Такой вариант задания плотности распределения позволяет легко отыскивать математическое ожидание и дисперсию в случае **нормального распределения**. Строго говоря, сейчас было задано множество различных распределений, элементы которого отличаются значениями двух параметров. При конкретных значениях этих параметров получаем вполне конкретное распределение. График плотности распределения для нормально распределённой случайной величины имеет вид, отдалённо напоминающий колокол:

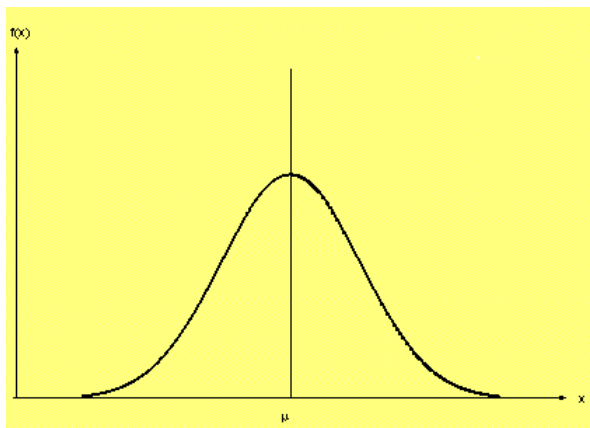


Рисунок 2- График плотности распределения для нормально распределённой случайной

Точка наивысшего подъёма данного графика - математическое ожидание случайной величины. Как нетрудно догадаться, при реальных испытаниях в большом количестве или крупных экспериментальных выборках очень высокий процент полученных случайных значений будет приходиться на узкий диапазон, включающий в себя математическое ожидание и некую "округу". Значения, находящиеся далеко от "высшей" точки, встречаются редко, являясь "аномалией". Функция распределения не может быть записана через элементарные функции, поскольку интеграл от плотности распределения "неберущийся". Поэтому её записывают вот так:

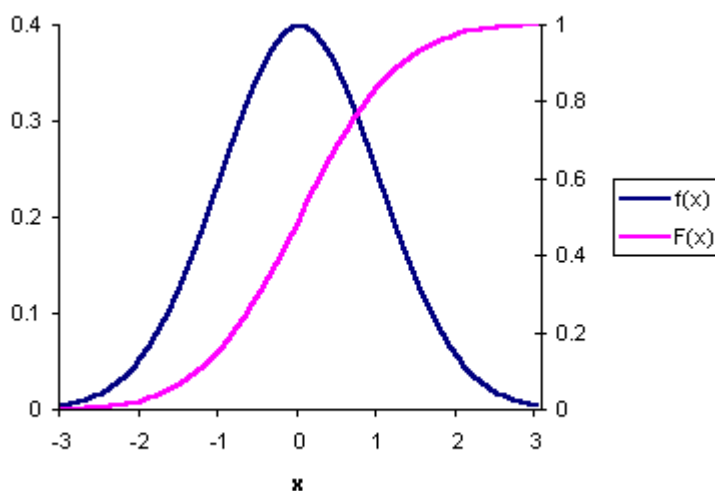
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

Вероятность того, что непрерывная случайная величина окажется каким-либо действительным числом, равна единице, поскольку полагается, что величина может принимать значение только на множестве действительных чисел. Поэтому

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt = 1$$

Стандартное нормальное распределение

Нормальное распределение с нулевым математическим ожиданием и единичной дисперсией называется **стандартным нормальным распределением**. Вот так в данном случае выглядят плотность и функция распределения.



Экспоненциальное распределение

Начнем с вероятностной модели, приводящей к таким распределениям. Для этого рассмотрим "поток событий", т.е. последовательность событий, происходящих одно за другим в какие-то моменты времени. Примерами могут служить: поток вызовов на телефонной станции; поток отказов оборудования в технологической цепочке; поток отказов изделий при испытаниях продукции; поток обращений клиентов в отделение банка; поток покупателей, обращающихся за товарами и услугами, и т.д. В теории потоков событий

справедлива теорема, аналогичная центральной предельной теореме, но в ней речь идет не о суммировании случайных величин, а о суммировании потоков событий. Рассматривается суммарный поток, составленный из большого числа независимых потоков, ни один из которых не оказывает преобладающего влияния на суммарный поток. Например, поток вызовов, поступающих на телефонную станцию, складывается из большого числа независимых потоков вызовов, исходящих от отдельных абонентов. Доказано, что в случае, когда характеристики потоков не зависят от времени, суммарный поток полностью описывается одним числом λ - интенсивностью потока. Для суммарного потока рассмотрим случайную величину X - длину промежутка времени между последовательными событиями. Ее функция распределения имеет вид

$$F(x, \lambda) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (10)$$

Это распределение называется экспоненциальным распределением, т.к. в формуле (10) участвует экспоненциальная функция $e^{-\lambda x}$. Величина $1/\lambda$ - масштабный параметр. Иногда вводят и параметр сдвига c , при этом экспоненциальным распределением называют распределение случайной величины $X + c$, где распределение X задается формулой (10).

В формуле (10) $e = 2,718281828\dots$ - основание натуральных логарифмов. Функция экспоненциального распределения $F(x, \lambda)$ и его плотность $f(x, \lambda)$ связаны простым соотношением

$$f(x, \lambda) = \lambda(1 - F(x, \lambda)).$$

Это соотношение имеет простую интерпретацию в терминах теории надежности технических изделий и устройств. Оно означает, что интенсивность отказов (т.е. интенсивность выхода изделий из строя) постоянна, другими словами, не зависит от того, сколько времени изделие уже проработало. Обычно интенсивность отказов постоянна на основном этапе эксплуатации, после того, как на начальном этапе выявлены скрытые дефекты, и до того, как из-за естественного старения материалов начинает происходить

ускоренный износ с резким возрастанием интенсивности выхода изделия из строя.

Логнормальное распределение

Говорят, что случайная величина имеет **логнормальное распределение** $\eta \in L_{a, \sigma^2}$, если ее логарифм имеет [нормальное распределение](#) $\ln \eta \in \Phi_{\alpha, \sigma^2}$.

Другими словами, $\eta = e^{\xi}$, где $\xi \in \Phi_{\alpha, \sigma^2}$.

Отсюда видно, что распределение L_{a, σ^2} сосредоточено на положительной полуоси.

Логнормальное распределение используется, например, при моделировании таких переменных, как доходы, допустимое отклонение от стандарта вредных веществ в продуктах питания и т.д.

Итак, если величина x имеет нормальное распределение, то величина $y = e^x$ имеет логнормальное распределение.

Если вы подставите нормальную величину в степень экспоненты, то легко поймете, что логнормальная величина получается в результате многократных умножений независимых величин, так же как нормальная случайная величина есть результат многократного суммирования.

Плотность $\eta \in L_{a, \sigma^2}$ в силу формул для плотности функций от случайной величины равна

$$\varphi_{\alpha, \sigma^2}(\ln x) x^{-1}.$$

Кроме того, находим

$$M\eta = \int e^y \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\alpha)^2}{2\sigma^2}} dy = \exp \frac{(\alpha + \sigma^2)^2 - \alpha^2}{2\sigma^2} \times$$

$$\times \int \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - \alpha - \sigma^2)^2}{2\sigma^2}\right) dy = e^{\alpha + \sigma^2/2},$$

$$M\eta^2 = \int e^{2y} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\alpha)^2}{2\sigma^2}} dy = e^{2\alpha + 2\sigma^2}.$$

Числовые характеристики одномерных случайных величин

Начальным моментом порядка K случайной величины X называют математическое ожидание величины X^k :

$$\nu_k = M(X^k)$$

В частности, $\nu_1 = M(X)$, $\nu_2 = M(X^2)$. Следовательно, $D(X) = \nu_2 - \nu_1^2$.

Центральным моментом порядка K случайной величины X называют математическое ожидание величины $(X - M(X))^K$:

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

В частности, $\mu_1 = 0$, $\mu_2 = D(X)$. Следовательно, $\mu_2 = \nu_2 - \nu_1^2$.

Математическим ожиданием дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности.

Пусть случайная величина X может принимать только значения $X_1, X_2, \dots, X_n$, вероятности которых соответственно равны $P_1, P_2, \dots, P_n$. Тогда математическое ожидание $M(X)$ случайной величины X определяется равенством

$$M(X) = X_1 P_1 + X_2 P_2 + \dots + X_n P_n.$$

Если дискретная случайная величина X принимает счетное множество возможных значений, то

$$M(X) = \sum_{i=1}^{\infty} x_i p_i,$$

Причем математическое ожидание существует, если ряд в правой части равенства сходится абсолютно.

Свойства математического ожидания

Свойство 1. Математическое ожидание суммы двух случайных величин равно сумме их математических ожиданий:

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

Свойство 2. Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:

$$M(XY)=M(X)M(Y).M(XY)=M(X)M(Y).$$

Свойство 3. Математическое ожидание постоянной величины равно самой постоянной:

$$M(c)=c.M(c)=c.$$

Свойство 4. Постоянный множитель случайной величины можно вынести за знак математического ожидания:

$$M(cX)=cM(X).M(cX)=cM(X).$$

Свойство 5. Математическое ожидание отклонения случайной величины от ее математического ожидания равно нулю:

$$M(X-M(X))=0.$$

Дисперсией (рассеянием) дискретной случайной величины называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

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

Пусть дискретная случайная величина задана рядом распределения

X	x1	X2	X3			Xn
P	p1	P2	P3			Pn

Тогда

$$D(X) = M[X - M(X)]^2 = [x1-M(X)]^2p1 + [x2-M(X)]^2p2 + ... + [Xn-M(X)]^2Pn.$$

Таким образом, *Чтобы найти дисперсию, достаточно вычислить сумму произведений возможных значений квадрата отклонения на их вероятности.*

Свойства дисперсии случайных величин

Свойство 1. Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин:

$$D[X+Y]=D[X]+D[Y].D[X+Y]=D[X]+D[Y].$$

Свойство 2. Дисперсия случайной величины равна разности между математическим ожиданием квадрата случайной величины X и квадратом ее математического ожидания:

$$D[X] = M(X^2) - (M(X))^2. \quad (4.3) \quad D[X] = M(X^2) - (M(X))^2.$$

Свойство 3. Дисперсия постоянной величины равна нулю:

$$D[c] = 0. D[c] = 0.$$

Свойство 4. Постоянный множитель случайной величины, можно выносить за знак дисперсии, предварительно возведя его в квадрат:

$$D[cX] = c^2 D[X]. D[cX] = c^2 D[X]$$

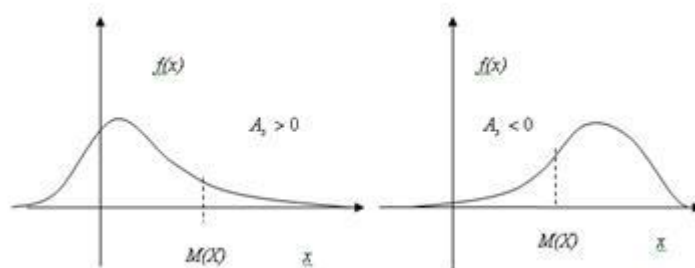
Свойство 5. Дисперсия произведения двух независимых случайных величин X и Y определяется по формуле

$$D[XY] = D[X]D[Y] + (M(X))^2 D[Y] + (M(Y))^2 D[X].$$

Асимметрией называют отношение центрального момента третьего порядка к кубу среднего квадратического отклонения:

$$A_3(X) = \frac{\mu_3(X)}{\sigma(X)^3}.$$

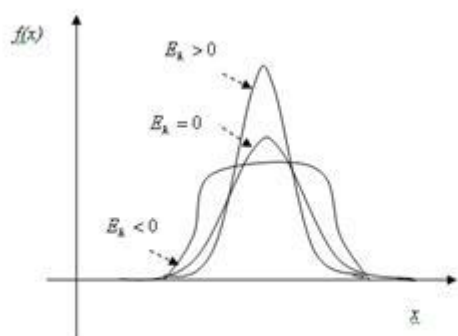
Асимметрия положительна, если “длинная часть” кривой распределения расположена справа от математического ожидания; асимметрия отрицательна, если “длинная часть” кривой расположена слева от математического ожидания. Практически определяют знак асимметрии по расположению кривой распределения относительно моды (см. ниже):



Экссессом называют характеристику, которая определяется равенством:

$$E_k(X) = \frac{\mu_4(X)}{\sigma^4(X)} - 3$$

Для нормального распределения эксцесс равен нулю. Следовательно, если эксцесс некоторого распределения отличен от нуля, то кривая этого распределения отличается от нормальной кривой: если эксцесс положительный, то кривая имеет более высокую и “острую” вершину, чем нормальная кривая; если эксцесс отрицательный, то сравниваемая кривая имеет более низкую и “плоскую” вершину, чем нормальная кривая:



Медианой случайной величины X называется число X_0 , которое удовлетворяет условию:

$$P(X < x_0) = P(X > x_0)$$

Модой непрерывной случайной величины называется точка локального максимума плотности распределения этой случайной величины.

Модой дискретной случайной величины называется такое значение этой случайной величины, вероятность принятия которой больше, чем вероятности принятия соседних с ним значений.

Квантилем уровня P называется число X_p , которое удовлетворяет условию:

$$F(x_p) = p$$

Производящая функция

При решении задач, связанных с ДСВ, часто используются методы комбинаторики. Одним из наиболее развитых теоретических методов

комбинаторного анализа является метод производящих функций, который является одним из самых сильных методов и в приложениях. Кратко познакомимся с ним.

Если принимает только целые неотрицательные значения, т.е.

$$\sum_{k=0}^{\infty} P(\xi = k) = \sum_{k=0}^{\infty} p_k = 1,$$

то **производящей функцией** распределения вероятностей случайной величины называется функция

$$\varphi_{\xi}(z) = M[z^{\xi}] = \sum_{k=0}^{\infty} z^k p_k, \quad (6.13)$$

где z — действительная или комплексная переменная. Отметим, что между множеством производящих функций $\varphi_{\xi}(x)$ и множеством распределений $\xi \{P(k)\}$ существует взаимно однозначное соответствие.

Пусть имеет случайная величина *биномиальное распределение*

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

Тогда, используя формулу бинома Ньютона, получим

$$\varphi_{\xi}(z) = \sum_{k=0}^n C_n^k p^k q^{n-k} z^k = (pz + q)^n,$$

т.е. **производящая функция биномиального распределения имеет вид**

$$\varphi_{\xi}(z) = (pz + q)^n. \quad (6.14)$$

Добавление. Производящая функция распределения Пуассона

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

имеет вид

$$\varphi_{\xi}(z) = e^{\lambda(z-1)}. \quad (6.15)$$

Производящая функция геометрического распределения

$$P(\xi = k) = pq^k$$

имеет вид

$$\varphi_{\xi}(z) = \frac{p}{1 - qz} . \quad (6.16)$$

При помощи производящих функций удобно находить основные числовые характеристики ДСВ. Например, первый и второй начальный моменты связаны с производящей функцией следующими равенствами:

$$\nu_1 = M\xi = \varphi'_{\xi}(1) , \quad (6.17)$$

$$\nu_2 = \varphi'_{\xi}(1) + \varphi''_{\xi}(1) . \quad (6.18)$$

Метод производящих функций часто бывает удобен тем, что в некоторых случаях функцию распределения ДСВ очень трудно определить, тогда как производящую функцию порой легко найти. Например, рассмотрим схему последовательных независимых испытаний Бернулли, но внесем в нее одно изменение. Пусть вероятность осуществления события A от испытания к испытанию меняется. Это означает, что формула Бернулли для такой схемы становится неприменимой. Задача нахождения функции распределения в таком случае представляет значительные трудности. Однако для данной схемы легко находится производящая функция, а, следовательно, легко находятся и соответствующие числовые характеристики.

Широкое применение производящих функций основано на том, что изучение сумм случайных величин можно заменить изучением произведений соответствующих производящих ξ функций. Так, если $\xi_1, \xi_2, \dots, \xi_n$ независимы, то

$$\varphi_{\xi_1 + \dots + \xi_n}(x) = \varphi_{\xi_1}(x) \dots \varphi_{\xi_n}(x) . \quad (6.19)$$

Пусть $p_k = P_k(A)$ – вероятность "успеха" в k -м испытании в схеме Бернулли (соответственно, $q_k = 1 - p_k$ – вероятность "неуспеха" в k -м испытании). Тогда, в соответствие с формулой (6.19), производящая функция будет иметь вид

$$\varphi_{\xi}(z) = \prod_{k=1}^n (q_k + p_k z) . \quad (6.20)$$

Пользуясь данной производящей функцией, можем написать

$$\nu_1 = M\xi = \left. \frac{d\varphi_{\xi}(z)}{dz} \right|_{z=1} = \sum_{k=1}^n p_k .$$

Здесь учтено, что $p_k + q_k = 1$. Теперь по формуле (6.1) найдем второй начальный момент. Для этого предварительно вычислим

$$\varphi_{\xi}''(1) = p_1 \sum_{k \neq 1}^n p_k + p_2 \sum_{k \neq 2}^n p_k + \dots + p_n \sum_{k \neq n}^n p_k = \left(\sum_{k=1}^n p_k \right)^2 - \sum_{k=1}^n p_k^2.$$

Тогда

$$v_2 = \sum_{k=1}^n p_k + \left(\sum_{k=1}^n p_k \right)^2 - \sum_{k=1}^n p_k^2 \quad \text{и} \quad D\xi = v_2 - v_1^2 = \sum_{k=1}^n p_k q_k.$$

В частном случае $p_1 = p_2 = \dots = p_n = p$ (т.е. в случае биномиального распределения) из полученных формул следует, что $Mnp = \xi$, $Dnpq$.

Предельные теоремы в схеме Бернулли.

Пуассоновское

приближение

Теорема Пуассона:

Пусть $n \rightarrow \infty$, $p \rightarrow 0$ таким образом, что $np \rightarrow a$, где $a > 0$ - заданное число. Тогда для любого фиксированного k

$$P_{n,p}(k) \rightarrow \frac{a^k}{k!} e^{-a}.$$

Другими словами, в описанном предельном переходе биномиальные вероятности $P_{n,p}(k)$ аппроксимируются пуассоновским распределением.

Доказательство:

Для краткости будем считать, что $n \rightarrow \infty$, $p = a / n$. Тогда

$$\begin{aligned} P_{n,a/n}(k) &= C_n^k \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n-k} = \\ &= \frac{a^k}{k!} \cdot \left(1 - \frac{a}{n}\right)^n \cdot \frac{n(n-1) \cdots (n-k+1)}{n^k} \left(1 - \frac{a}{n}\right)^{-k} = \frac{a^k}{k!} \cdot \left(1 - \frac{a}{n}\right)^n \cdot \left[1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k}{n}\right)\right] \end{aligned}$$

, поскольку выражение в квадратных скобках стремится к единице, если k фиксировано, а $n \rightarrow \infty$.

Нормальное приближение

Здесь мы рассмотрим случай, когда число испытаний в схеме Бернулли растет ($n \rightarrow \infty$), а вероятность успеха в единичном испытании p остается фиксированной. Верна так называемая интегральная теорема Муавра-Лапласа.

Теорема Муавра-Лапласа:

Пусть ν_n - число успехов в последовательности из n независимых испытаний Бернулли с вероятностью успеха в единичном испытании p . Пусть $-\infty \leq c < d \leq +\infty$. При $n \rightarrow \infty$

$$P\left\{c \leq \frac{\nu_n - np}{\sqrt{np(1-p)}} \leq d\right\} \rightarrow \Phi(d) - \Phi(c) \quad (2), \quad \text{где}$$

$$\Phi(t) = \int_{-\infty}^t \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Замечание 1.

Функция $\Phi(t)$, появившаяся в этой теореме, называется *функцией распределения стандартного нормального закона*. Для значений этой функции существуют подробные таблицы. Отметим, что она не зависит ни от каких параметров. Следовательно, предел в теореме Муавра-Лапласа является универсальным, так как он не зависит от параметра p , который имеется в допредельном выражении. На самом деле, эта теорема является частным случаем другой, еще более универсальной *центральной предельной теоремы*.

Замечание 2.

Чтобы понять смысл выражения $\frac{\nu_n - np}{\sqrt{np(1-p)}} \quad (3)$, необходимо вспомнить, что $E\nu_n = np$ и $D\nu_n = np(1-p)$. Таким образом, это выражение имеет вид $\eta_n = (\nu_n - E\nu_n)/\sqrt{D\nu_n}$. Легко видеть, что $E\eta_n = 0$, а $D\eta_n = 1$.

Преобразование (3) называется *центрированием и нормированием случайной величины ν_n* .

Замечание 3.

В предельном переходе " $n \rightarrow \infty$ ", " p фиксировано" каждая "индивидуальная" вероятность $P_{n,p}(k)$ стремится к нулю. Асимптотика этого стремления описывается так называемой *локальной предельной теоремой*. Что же касается интегральной предельной теоремы Муавра-Лапласа, то можно сказать, что она описывает предельное поведение сумм большого числа таких малых вероятностей. Действительно,

$$P\left\{c \leq \frac{\nu_n - np}{\sqrt{np(1-p)}} \leq d\right\} = \sum_{np+c\sqrt{np(1-p)} \leq k \leq np+d\sqrt{np(1-p)}} P_{n,p}(k),$$

таким образом, в последней сумме содержится много (порядка $\sqrt{n}$) слагаемых.

Замечание 4.

Скорость сходимости в (2) хорошо изучена. Имеет место так называемая оценка *Берри-Эссеена*: Существует такое $D > 0$, что

$$\sup_{c,d} \left| P\left\{c \leq \frac{\nu_n - np}{\sqrt{np(1-p)}} \leq d\right\} - (\Phi(d) - \Phi(c)) \right| \leq \frac{D}{\sqrt{np(1-p)}}.$$

Понятие системы случайных величин.

При теоретико-множественной трактовке любая СВ X есть функция элементарного события, входящего в пространство элементарных событий Ω :

$$X = \varphi(\omega), \omega \in \Omega,$$

т.е. каждому элементарному событию $\omega \in \Omega$ ставится в соответствие некоторое действительное число $x \in \Omega_X$, где Ω_X – множество возможных значений случайной величины X .

Теперь перейдем к рассмотрению системы случайных величин – двух и более.

Например: координаты падения снаряда X и Y ; набор оценок $X_1, X_2, \dots, X_n$, выставленных в приложении к диплому.

Будем обозначать систему нескольких случайных величин $X_1, \dots, X_n$ как $(X_1, \dots, X_n)$. Эта система есть функция элементарного события:

$$(X_1, \dots, X_n) = \varphi(\omega), \omega \in \Omega,$$

т.е. каждому элементарному событию $\omega \in \Omega$ ставится в соответствие упорядоченный набор действительных чисел $x_1, \dots, x_n \in \Omega_{X_1 \dots X_n}$, где $\Omega_{X_1 \dots X_n}$ – множество возможных значений n -мерной случайной величины $(X_1, \dots, X_n)$.

Т.о., каждому элементарному событию ω ставится в соответствие несколько действительных чисел – значения, принятые случайными величинами $X_1, \dots, X_n$ в результате опыта.

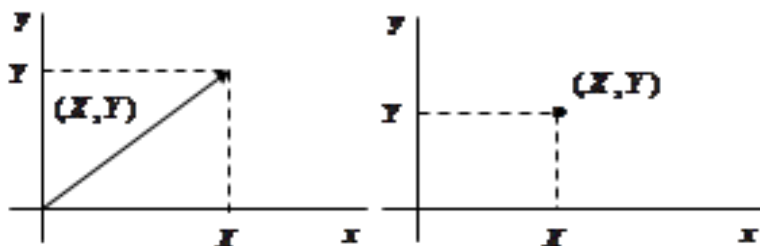


Рис. 1.2. Случайный вектор **Рис. 1.1. Случайная точка**

СВ, входящие в систему, могут быть как дискретными, так и непрерывными. Систему двух СВ (X, Y) можно изобразить случайной точкой на плоскости с координатами X и Y (см. рис. 1.1). Для системы из трех СВ (X, Y, Z) – случайной точкой в 3-х мерном пространстве, с координатами X, Y, Z . И то, и другое можно изобразить в виде вектора (см. рис. 1.2). Использование геометрической интерпретации удобно для любой системы СВ $(X_1, X_2, \dots, X_n)$, как вектора в n -мерном пространстве: $\vec{X} = (X_1, X_2, \dots, X_n)$.

Свойства системы случайных величин определяются как свойствами отдельных величин, входящих в систему, так и зависимостями между случайными величинами.

Полной характеристикой системы случайных величин является закон распределения, который может быть представлен в виде:

- функции распределения,
- плотности распределения,
- матрицы распределения (таблицы вероятностей отдельных значений случайного вектора) и т.д.

Матрица распределения двумерной дискретной случайной величины.

Двухмерная случайная величина (X, Y) является дискретной, если множества значений ее компонент Ω_X и Ω_Y представляют собой счетные множества. Для описания вероятностных характеристик таких величин используется двухмерная функция распределения и матрица распределения.

Матрица распределения представляет собой прямоугольную таблицу, которая содержит значения компоненты $X - \Omega_X = \{x_1, x_2, \dots, x_n\}$, значения компоненты $Y - \Omega_Y = \{y_1, y_2, \dots, y_m\}$ и вероятности всевозможных пар значений

$$p_{ij} = p(X = x_i, Y = y_j), i = 1, \dots, n, j = 1, \dots, m.$$

Свойства матрицы распределения вероятностей:

- Переход к ряду распределения вероятностей составляющей X :
- Переход к ряду распределения вероятностей составляющей Y :

Функция распределения системы случайных величин.

Функцией распределения системы случайных величин называется вероятность совместного выполнения неравенств $X_1 < x_1, \dots, X_n < x_n$:

$$F(x_1, \dots, x_n) = P\{(X_1 < x_1) \cap \dots \cap (X_n < x_n)\}.$$

Для случая **двумерной случайной величины**:

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

Геометрически функция распределения $F(x, y)$ это вероятность попадания случайной точки (X, Y) в бесконечный квадрант с вершиной в точке (x, y) , лежащей левее и ниже ее.

Свойства функции распределения.

1. Значения функции распределения удовлетворяют двойному неравенству:

$$0 \leq F(x, y) \leq 1.$$

Доказательство этого свойства вытекает из определения функции распределения как вероятности: вероятность есть неотрицательное число, не превышающее 1.

2. Функция распределения $F(x, y)$ есть **неубывающая функция** по каждому из аргументов т.е

$$x_1 < x_2 \Rightarrow F(x_1, y) \leq F(x_2, y)$$

$$y_1 < y_2 \Rightarrow F(x, y_1) \leq F(x, y_2)$$

Доказательство этого свойства вытекает из того, что при увеличении какого-нибудь из аргументов (x, y) квадрант, увеличивается; следовательно, вероятность попадания в него случайной точки (X, Y) уменьшаться не может.

3. Если хотя бы один из аргументов функции распределения обращается в $-\infty$, то функция распределения равна 0:

$$F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0.$$

Доказательство. По определению

$$F(-\infty, y) = P\{(X < -\infty) \cap (Y < y)\}.$$

Событие $\{(X < -\infty) \cap (Y < y)\}$ невозможное событие, т.к. невозможным является событие $(X < -\infty)$ событие; тогда

$$F(-\infty, y) = 0.$$

4. Если оба аргумента функции распределения $F(x,y)$ равны $+\infty$, то функция распределения равна 1.

Доказательство следует из определения функции распределения системы случайных величин:

$$\lim_{x \rightarrow \infty, y \rightarrow \infty} F(x,y) = P(X < \infty, Y < \infty) = 1$$

5. Если один из аргументов обращается в $+\infty$, то функция распределения $F(x,y)$ становится равной функции распределения случайной величины, соответствующей другому аргументу:

$$F(x, +\infty) = F_1(x); \quad F(+\infty, y) = F_1(y)$$

Доказательство. По определению функции распределения:

$$F(x, +\infty) = P\{(X < x) \cap (Y < +\infty)\}.$$

Событие $(Y < +\infty)$ является достоверным событием. Тогда

$$P\{(X < x) \cap \Omega\} = P\{X < x\} = F_1(x).$$

Точно так же доказывается, что

$$F(+\infty, y) = P\{Y < y\} = F_1(y).$$

6. Вероятность попадания в прямоугольную область

$$P(a \leq X \leq b; d \leq Y \leq g) = F(b,g) - F(b,d) - F(a,g) + F(a,d).$$

Плотность распределения вероятности и ее свойства

Функция распределения непрерывной случайной величины является ее вероятностной характеристикой. Но она имеет недостаток, заключающийся в том, что по ней трудно судить о характере распределения случайной величины в небольшой окрестности той или другой точки числовой оси. Более наглядное представление о характере распределения непрерывной случайной величины дает функция, которая называется плотностью распределения вероятности, или дифференциальной функцией распределения случайной величины.

Плотность распределения $f(x)$ равна производной от функции распределения $F(x)$, т. е.

$$f(x)=F'(x).$$

Смысл плотности распределения $f(x)$ состоит в том, что она указывает на то, как часто случайная величина X появляется в некоторой окрестности точки x при повторении опытов. Кривая, изображающая плотность распределения $f(x)$ случайной величины, называется **кривой распределения**.

Свойства плотности распределения.

Свойство 1. Плотность распределения неотрицательна, т. е.

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

Свойство 2. Функция распределения случайной величины равна интегралу от плотности в интервале от $-\infty$ до x , т. е.

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

Свойство 3. Вероятность попадания непрерывной случайной величины X на участок $(\alpha; \beta)$ равна интегралу от плотности распределения, взятому по этому участку, т. е.

$$P\{\alpha \leq X \leq \beta\} = \int_{\alpha}^{\beta} f(x) dx.$$

Свойство 4. Интеграл в бесконечных пределах от плотности распределения равен единице:

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

Условные законы распределения

Выразим теперь плотность распределения каждой из величин входящих в систему, через плотность распределения системы. Пользуясь формулой выражающей функцию распределения через плотность распределения, напомним:

$$F_1(x) = F(x, \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(x, y) dx dy,$$

откуда, дифференцируя по x , получим выражение для плотности распределения величины X :

$$f_1(x) = F_1'(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (8.4.2)$$

Аналогично

$$f_2(y) dy = F_2'(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (8.4.3)$$

Таким образом, для того чтобы получить плотность распределения одной из величин, входящих в систему, нужно плотность распределения системы проинтегрировать в бесконечных пределах по аргументу, соответствующему другой случайной величине.

Формулы (8.4.1), (8.4.2) и (8.4.3) дают возможность, зная закон распределения системы (заданный в виде функции распределения или плотности распределения), найти законы распределения отдельных величин, входящих в систему. Естественно, возникает вопрос об обратной задаче: нельзя ли по законам распределения отдельных величин, входящих в систему, восстановить закон распределения системы? Оказывается, что в общем случае этого сделать нельзя: зная только законы распределения отдельных величин, входящих в систему, не всегда можно найти закон распределения системы. Для того чтобы исчерпывающим образом охарактеризовать систему, недостаточно знать распределение каждой из величин, входящих в систему; нужно еще знать зависимость между величинами, входящими в систему. Эта зависимость может быть охарактеризована с помощью так называемых условных законов распределения.

Условным законом распределения величины X , входящей в систему (X, Y) , называется ее закон распределения, вычисленный при условии, что другая случайная величина Y приняла определенное значение y .

Условный закон распределения можно задавать как функцией распределения, так и плотностью. Условная функция распределения обозначается $F(x|y)$ условная плотность распределения $f(x|y)$. Так как системы непрерывных величин имеют основное практическое значение, мы в

данном курсе ограничимся рассмотрением условных законов, заданных плотностью распределения.

Чтобы нагляднее пояснить понятие условного закона распределения, рассмотрим пример. Система случайных величин L и Q представляет собой длину и вес осколка снаряда. Пусть нас интересует длина осколка L безотносительно к его весу; это есть случайная величина, подчиненная закону распределения с плотностью $f_1(l)$. Этот закон распределения мы можем исследовать, рассматривая все без исключения осколки и оценивая их только по длине; $f_1(l)$ есть безусловный закон распределения длины осколка. Однако нас может интересовать и закон распределения длины осколка вполне определенного веса, например 10 г. Для того чтобы его определить, мы будем исследовать не все осколки, а только определенную весовую группу, в которой вес приблизительно равен 10 г, и получим условный закон распределения длины осколка при весе 10 г с плотностью $f(l|q)$ при $q = 10$. Этот условный закон распределения вообще отличается от безусловного $f_1(l)$; очевидно, более тяжелые осколки должны в среднем обладать и большей длиной; следовательно, условный закон распределения длины существенно зависит от веса q .

Зная закон распределения одной из величин, входящих в систему, и условный закон распределения второй, можно составить закон распределения системы. Выведем формулу, выражающую это соотношение, для непрерывных случайных величин. Для этого воспользуемся понятием об элементе вероятности. Рассмотрим прилежащий к точке (x, y) элементарный прямоугольник R_d со сторонами dx , dy (рис. 8.4.1). Вероятность попадания в этот прямоугольник - элемент вероятности $f(x, y)dx dy$ - равна вероятности одновременного попадания случайной точки (X, Y) в элементарную полосу I, опирающуюся на отрезок dx , и в полосу II, опирающуюся на отрезок dy :

$$f(x, y) dx dy = P((X, Y) \in R_d) = P((x < X < x + dx)(y < Y < y + dy)).$$

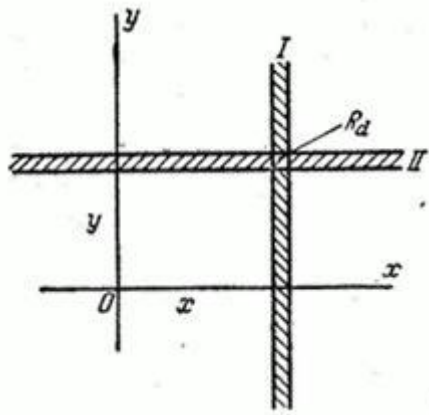


Рис.8.4.1

Вероятность произведения этих двух событий, по теореме умножения вероятностей, равна вероятности попадания в элементарную полосу I, умноженной на условную вероятность попадания в элементарную полосу II, вычисленную при условии, что первое событие имело место. Это условие в пределе равносильно условию $X = x$, следовательно,

$$f(x, y) dx dy = f_1(x) dx f(y|x) dy,$$

откуда

$$f(x, y) = f_1(x) f(y|x), \quad (8.4.4)$$

т.е. плотность распределения системы двух величин равна плотности распределения одной из величин, входящих в систему, умноженной на условную плотность распределения другой величины, вычисленную при условии, что первая величина приняла заданное значение.

Формулу (8.4.4) часто называют теоремой умножения законов распределения. Эта теорема в схеме случайных величин аналогична теореме умножения вероятностей в схеме событий.

Очевидно, формуле (8.4.4) можно придать другой вид, если задать значение не величины X , а величины Y :

$$f(x, y) = f_2(y) f(x|y). \quad (8.4.5)$$

Разрешая формулы (8.4.4) и (8.4.5) относительно $f(y|x)$ и $f(x|y)$, получим выражения условных законов распределения через безусловные:

$$\left. \begin{aligned} f(y|x) &= \frac{f(x,y)}{f_1(x)}, \\ f(x|y) &= \frac{f(x,y)}{f_2(y)}, \end{aligned} \right\} \quad (8.4.6)$$

или, применяя формулы (8.4.2) и (8.4.3),

$$\left. \begin{aligned} f(y|x) &= \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y) dy}, \\ f(x|y) &= \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y) dx} \end{aligned} \right\}$$

Числовые характеристики двумерной случайной величины.

Ковариация

Если между случайными величинами X и Y существует стохастическая связь, то одним из параметров, характеризующих меру этой связи является *ковариация* $\text{cov}(X, Y)$. Ковариацию вычисляют по формулам $\text{cov}(X, Y) = M[(X - M_X)(Y - M_Y)] = M(XY) - M_X M_Y$.

Если случайные величины X и Y *независимы*, то $\text{cov}(X, Y) = 0$.

Обратное, вообще говоря, неверно. Из равенства нулю ковариации не следует независимость случайных величин. Случайные величины могут быть зависимыми в то время как их ковариация нулевая! Но зато, *если ковариация случайных величин отлична от нуля, то между ними существует стохастическая связь, мерой которой и является величина ковариации.*

Свойства ковариации:

$$\text{cov}(X, X) = D_X;$$

$$\text{cov}(\xi + C_1, \eta + C_2) = \text{cov}(\xi, \eta);$$

$$\text{cov}(\xi, \eta) = \text{cov}(\eta, \xi);$$

$$\text{cov}(C_1\xi + C_2\eta, \zeta) = C_1 \text{cov}(\xi, \zeta) + C_2 \text{cov}(\eta, \zeta),$$

где C_1 и C_2 - произвольные константы.

Ковариационной матрицей случайного вектора (ξ, η) называется матрица вида

$$\text{Cov}(\xi, \eta) = \begin{bmatrix} \text{cov}(\xi, \xi) & \text{cov}(\xi, \eta) \\ \text{cov}(\eta, \xi) & \text{cov}(\eta, \eta) \end{bmatrix} \equiv \begin{bmatrix} \sigma_\xi^2 & \text{cov}(\xi, \eta) \\ \text{cov}(\eta, \xi) & \sigma_\eta^2 \end{bmatrix}.$$

Эта матрица симметрична и положительно определена. Ее определитель называется *обобщенной дисперсией* и может служить мерой рассеяния системы случайных величин (ξ, η) .

Как уже отмечалось ранее, дисперсия суммы *независимых* случайных величин равна сумме их дисперсий: $D(\xi + \eta) = D(\xi) + D(\eta)$. Если же случайные величины *зависимы*, то $D(\xi \pm \eta) = D(\xi) + D(\eta) \pm 2\text{cov}(\xi, \eta)$.

Корреляция

Понятно, что значение ковариации зависит не только от “тесноты” связи случайных величин, но и от самих значений этих величин, например, от единиц измерения этих значений. Для исключения этой зависимости вместо ковариации используется безразмерный *коэффициент корреляции*

$$k_{\xi\eta} = \frac{\text{cov}(\xi, \eta)}{\sqrt{D(\xi) \cdot D(\eta)}}.$$

Этот коэффициент обладает следующими свойствами:

он безразмерен;

его модуль не превосходит единицы, т.е. $|k_{\xi\eta}| \leq 1$;

если ξ и η независимы, то $k_{(\xi, \eta)} = 0$ (обратное неверно!);

если $|k_{\xi\eta}| = 1$, то случайные величины ξ и η связаны *функциональной зависимостью вида*

$$\xi = a\eta + b,$$

где a и b - некоторые числовые коэффициенты;

$$k_{\xi\xi} = k_{\eta\eta} = 1;$$

Корреляционной матрицей случайного вектора называется матрица

$$K = \begin{bmatrix} 1 & k_{\xi\eta} \\ k_{\xi\eta} & 1 \end{bmatrix}.$$

Если $\sigma_{\xi} = \sqrt{D(\xi)}$ и $\sigma_{\eta} = \sqrt{D(\eta)}$, то ковариационная и корреляционная матрицы случайного вектора (ξ, η) связаны соотношением $Cov(\xi, \eta) = \Sigma K \Sigma$, где $\Sigma = \begin{bmatrix} \sigma_{\xi} & 0 \\ 0 & \sigma_{\eta} \end{bmatrix}$.

Энтропия.

Энтропию (обозначение H_p), как меру неопределенности источника сообщений, осуществляющего выбор из m состояний, каждое из которых реализуется с вероятностью p_i , предложил Клод Шеннон.

Энтропия вычисляется по формуле:

$$H_p = - \sum_{i=1}^m p_i \log p_i, \text{ где } \sum_{i=1}^m p_i = 1$$

$-\log p_i$ – количество информации, которое несет i - тое состояние сигнала
 есть величина случайная. Тогда, энтропия – это математическое ожидание информации, которое несет сигнал с m состояниями, каждое из которых имеет вероятность появления p_i .

Энтропия позволяет сравнивать количества информации, приносимые различными сигналами.

Пример: В первой урне имеются 7 белых, 5 черных и 2 синих шара. Во второй урне- 4 белых, 6 черных и 2 синих шара. Наудачу вынимают один шар. Для какой из урн исход более определенный?

Вычислим энтропию для каждого из случаев: шар – из первой урны, шар – из второй урны. Менее определенному исходу соответствует большее значение энтропии.

Для первой урны возможны три состояния : шар может оказаться или белым, или черным, или синим.

Найдем вероятность каждого состояния: белый шар- $p = 7/14 = 1/2$;

черный шар – $p = 5/14$; синий шар- $p = 2/14 = 1/7$.

Энтропия для первой урны :

$$H_1 = -\left(\frac{1}{2} \log \frac{1}{2} + \frac{5}{14} \log \frac{5}{14} + \frac{1}{7} \log \frac{1}{7}\right) = 1,43 \text{ бит.}$$

Аналогичные расчеты выполним для второй урны. Вероятности состояний: белый шар- $p = 4/12 = 1/3$; черный шар – $p = 6/12 = 1/2$; синий шар- $p = 2/12 = 1/6$.

Энтропия для второй урны :

$$H_2 = -\left(\frac{1}{3} \log \frac{1}{3} + \frac{1}{2} \log \frac{1}{2} + \frac{1}{6} \log \frac{1}{6}\right) = 1,41 \text{ бит.}$$

Так как энтропия для второй урны меньше, значит исход для нее более определенный (получено меньшее количество информации).

Количество информации

На вход системы передачи информации (СПИ) от источника информации подается совокупность сообщений, выбранных из ансамбля сообщений (рис. 1).

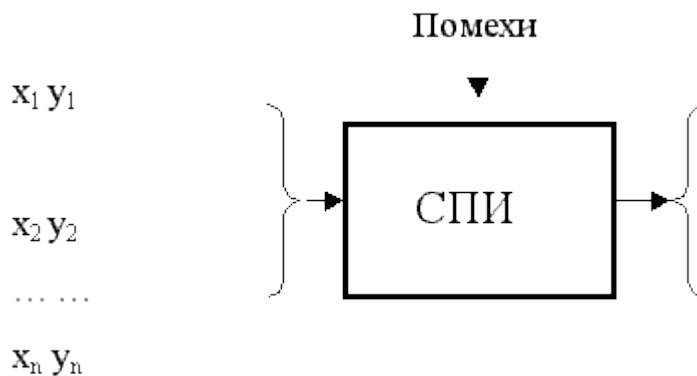


Рис. 1. Система передачи информации

Ансамбль сообщений – множество возможных сообщений с их вероятностными характеристиками – $\{X, p(x)\}$. При этом: $X = \{x_1, x_2, \dots, x_m\}$ – множество возможных сообщений источника; $i = 1, 2, \dots, m$, где m – объем алфавита; $p(x_i)$ – вероятности появления сообщений, причем $p(x_i) \geq 0$ и поскольку вероятности сообщений представляют собой полную группу событий, то их суммарная вероятность равна единице

$$\sum_{i=1}^m p(x_i) = 1.$$

Каждое сообщение несет в себе определенное количество информации. Определим количество информации, содержащееся в сообщении x_i , выбранном из ансамбля сообщений источника $\{X, p(x)\}$. Одним из параметров, характеризующих данное сообщение, является вероятность его появления – $p(x_i)$, поэтому естественно предположить, что количество информации $I(x_i)$ в сообщении x_i является функцией $p(x_i)$. Вероятность появления двух независимых сообщений x_1 и x_2 равна произведению вероятностей $p(x_1, x_2) = p(x_1) \cdot p(x_2)$, а содержащаяся в них информация должна обладать свойством аддитивности, т.е.:

$$I(x_1, x_2) = I(x_1) + I(x_2). \quad (1)$$

Поэтому для оценки количества информации предложена логарифмическая мера:

$$I(x_i) = \log \frac{1}{p(x_i)} = -\log p(x_i). \quad (2)$$

При этом наибольшее количество информации содержат наименее вероятные сообщения, а количество информации в сообщении о достоверном событии равно нулю. Т. к. все логарифмы пропорциональны, то выбор основания определяет единицу информации: $\log_a x = \log_b x / \log_b a$.

В зависимости от основания логарифма используют следующие единицы информации:

2 – [бит] (*binary digit* – двоичная единица), используется при анализе информационных процессов в ЭВМ и др. устройствах, функционирующих на основе двоичной системы счисления;

e – [нит] (*natural digit* – натуральная единица), используется в математических методах теории связи;

10 – [дит] (*decimal digit* – десятичная единица), используется при анализе процессов в приборах работающих с десятичной системой счисления.

Бит - двоичная единица статистической меры информации, количество информации, которое снимает неопределенность в отношении наступления одного из двух равновероятных ($p_0 = 0.5$, $p_1 = 0.5$), независимых событий.

Схожую меру другого иерархического уровня (комбинаторно-геометрического) - двоичный разряд часто путают с битом, считая, что байт (8 двоичных разрядов) = 8 битам, хотя это не так. (в двоичных разрядах не задана вероятность, а значит, не может быть определена мера от вероятности).

Среднее количество информации для всей совокупности сообщений можно получить путем усреднения по всем независимым событиям:

$$I(x) = - \sum_{i=1}^m p_i \log p_i \quad . (3)$$

Количество информации, в сообщении, состоящем из n не равновероятных его элементов равно (эта мера предложена в 1948 г. К. Шенноном):

$$I(x) = -n \sum_{i=1}^m p_i \log p_i \quad . (4)$$

Для случая независимых равновероятных событий количество информации определяется (эта мера предложена в 1928 г. Р. Хартли):

$$I(x) = -n \sum_{i=1}^m p_i \log p_i = -n \cdot m \left(\frac{1}{m} \log \frac{1}{m} \right) = n \cdot \log m \quad (5)$$

Кроме этих мер информации существуют и иные, производные от энтропии статистические меры:

- расстояние Кульбака как асимметричная информационная мера:

$$D_{KL}(p, q) = \sum_{x \in \mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}.$$

где $p(x)$, $q(x)$ - вероятности двух наборов случайных величин, а D - информационное расстояние от ансамбля $\{p\}$ до элементов другого ансамбля $\{q\}$.

- дивергенция Кульбака как симметричная информационная мера:

$$D_{KL}(p, q) = \sum_{x \in \mathcal{X}} (p(x) - q(x)) \ln \frac{p(x)}{q(x)}.$$

где $p(x)$, $q(x)$ - вероятности двух наборов случайных величин, а D - расстояние между ансамблями $\{p\}$ и $\{q\}$.

Свойства количества информации

1. Количество информации в сообщении обратно – пропорционально вероятности появления данного сообщения.

2. Свойство аддитивности – суммарное количество информации двух источников равно сумме информации источников.

3. Для события с одним исходом количество информации равно нулю.

4. Количество информации в дискретном сообщении растет в зависимости от увеличения объема алфавита – m .

Пример 1. Определить количество информации в сообщении из 8 двоичных символов ($n = 8$, $m = 2$), если вероятности равны: $p_{i0} = p_{i1} = 1/2$.

Количество информации равно:

$$I = n \log m = 8 \log_2 2 = 8 \text{ бит}.$$

Пример 2. Определить количество информации в сообщении из 8 двоичных символов ($n = 8$, $m = 2$), если вероятности равны:

$$p_{i0} = 3/4; p_{i1} = 1/4.$$

Количество информации равно:

$$\begin{aligned} I(x) &= -n \sum_{i=1}^m p_i \log p_i = -8 \cdot \left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) = \\ &= -8 \left[\frac{3}{4} (\log_2 3 - \log_2 4) + \frac{1}{4} (\log_2 1 - \log_2 4) \right] = 5.67 \text{ бит}. \end{aligned}$$

Многомерный нормальный закон распределения.

Многомерное нормальное распределение строится из распределения одномерных гауссовских величин.

Лучше всего представить процесс построения по шагам.

Рассмотрим n независимых случайных величин $\xi_{01}, \dots, \xi_{0n}$, имеющих одно и то же (стандартное) гауссовское распределение со средним 0 и стандартным отклонением 1.

Эти величины назовем исходными.

Каждая величина из исходных величин имеет плотность

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

Так как величины независимые, то совместная плотность вероятности этих величин есть

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{k=1}^n x_k^2}, \quad -\infty < x_1, \dots, x_n < \infty$$

Мы просто перемножили плотности между собой и получили написанное выражение.

Основываясь на исходных величинах, определим далее величины

$$\xi_i = \sum_{k=1}^n \sigma_{ik} \xi_{0k} + \alpha_i, \quad i = 1, \dots, n \quad (*)$$

Это линейное невырожденное преобразование величин ξ_{0k} , $k=1, \dots, n$.

Очевидно, среднее этих новых величин есть: $M \xi_i = \alpha_i$

Ковариация равна:

$$M(\xi_i - \alpha_i)(\xi_j - \alpha_j) = \sum_{k=1}^n \sigma_{ik} \sigma_{jk}, \quad i, j = 1, \dots, n$$

Формула для ковариаций верна (вы можете проверить это непосредственным вычислением), поскольку исходные величины ξ_{0k} , $k=1, \dots, n$ с нулевыми средними значениями являются *некоррелированными*:

$$M \xi_{0k} = 0, \quad M(\xi_{0k} \xi_{0l}) = M \xi_{0k} \cdot M \xi_{0l} = 0 \quad \text{при } k \neq l, k, l=1, \dots, n.$$

Собственно, именно эти величины и имеют совместное многомерное нормальное распределение в общем случае.

Самое главное для них – корреляционная матрица, корреляция измеряет меру связи.

Построенные новые случайные величины уже зависимы (связаны) между собой.

Не нужно запоминать формулу плотности многомерного нормального распределения, но нужно понять ее смысл.

Определим корреляционную матрицу величин $\xi_1, \dots, \xi_n$.

Матрица $R = \{R_{ij}\}$ с элементами $R_{ij} = M(\xi_i - a_i)(\xi_j - a_j)$, $i, j = 1, \dots, n$

называется корреляционной матрицей случайных величин $\xi_1, \dots, \xi_n$

($r_{ij} = \frac{R_{ij}}{\sqrt{R_{ii}R_{jj}}}$ есть коэффициент корреляции между величинами ξ_i, ξ_j).

Легко показать, что корреляционная матрица R имеет также вид $R = \sigma\sigma^*$

где $\sigma = \{\sigma_{ij}\}$ – матрица линейного преобразования (*), а σ^* – матрица, сопряженная к σ (с элементами $\sigma_{ij}^* = \sigma_{ji}$, $i, j = 1, \dots, n$).

Определители $|R|$ и $|\sigma|$ указанных матриц связаны равенством $|R| = |\sigma|^2$; при этом $|\sigma|$ есть якобиан преобразования (*), и, следовательно, совместная плотность распределения случайных величин $\xi_1, \dots, \xi_n$ может быть описана формулой

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |R|} \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^n b_{ij} (x_i - a_i)(x_j - a_j) \right\},$$

где $\sum_{i,j=1}^n b_{ij} (x_i - a_i)(x_j - a_j)$ – та квадратичная форма, в которую переходит

сумма квадратов $\sum_{k=1}^n x_k^2$ при линейном преобразовании, обратном к (*).

Легко видеть, что матрица $\{b_{ij}\}$ этой квадратичной формы является обратной к корреляционной матрице $\{R_{ij}\}$, поскольку сумма квадратов переходит в квадратичную форму

$$\sum_{k=1}^n \left[\sum_{j=1}^n c_{kj} (x_j - a_j) \right]^2 = \sum_{i,j=1}^n \left(\sum_{k=1}^n c_{ki} c_{kj} \right) (x_i - a_i)(x_j - a_j), \quad (**)$$

где $\{c_{kj}\}$ есть матрица, обратная к $\{\sigma_{ij}\}$, так что

$$\{b_{ij}\} = \{c_{ik}\}^* \{c_{kj}\} = [\{c_{ik}\} \{c_{kj}\}^*]^{-1} = \{R_{ij}\}^{-1}.$$

Как уже отмечалось ранее, корреляционная матрица $\{R_{ij}\}$ является положительно определенной: при любых действительных $c_1, \dots, c_n$

$$\sum_{i,j=1}^n R_{ij} c_i c_j = M \left[\sum_{j=1}^n c_j (\xi_j - a_j) \right]^2 \geq 0;$$

Свойством положительной определенности обладает, очевидно, и обратная матрица $\{b_{ij}\} = R^{-1}$.

Для любой невырожденной положительно определенной матрицы $\{b_{ij}\}$ формула (**), где $|R|^{-1}$ означает определитель матрицы $\{b_{ij}\}$ и $a_1, \dots, a_n$ – произвольные постоянные, задает некоторую плотность вероятности (именно, такой вид имеет плотность вероятности величин $\xi_1, \dots, \xi_n$, которые связаны со «стандартными» гауссовскими величинами $\xi_{01}, \dots, \xi_{0n}$ линейным преобразованием (*) с матрицей $\{\sigma_{ij}\}$, являющейся квадратным корнем из матрицы $R = \{b_{ij}\}^{-1}$).

Распределение вероятностей с плотностью (**), так же как и случайные величины $\xi_1, \dots, \xi_n$ с таким распределением вероятностей, называется нормальным (или гауссовским).

Комплексные случайные величины.

Комплексной случайной величиной называется случайная величина вида:

$$, \quad (15.9.1)$$

где ξ_i – действительные случайные величины; j – мнимая единица.

Математическим ожиданием комплексной случайной величины называется комплексное число (15.9.2)

Дисперсией комплексной случайной величины называется математическое ожидание квадрата модуля соответствующей центрированной величины: (15.9.3), где. Выразим дисперсию комплексной случайной величины через дисперсии ее действительной и мнимой частей. Очевидно,;

Отсюда

Или, (15.9.4) Выразим корреляционный момент двух комплексных случайных величин через корреляционные моменты их действительных и мнимых частей. Имеем:

$$(15.9.9)$$

где $\mu_{11}, \mu_{22}, \mu_{12}$ - соответственно корреляционные моменты величин X_1, X_2 .

Очевидно, в случае, когда все эти величины между собой не коррелированы, корреляционный момент величин X_1, X_2 также равен нулю.

Примеры построения законов распределения.

Рассмотрим примеры нахождения законов распределения случайных величин и их числовых характеристик.

Пример 1. 1). Составить закон распределения числа попаданий в цель при трех выстрелах по мишени, если вероятность попадания при каждом выстреле равна 0,4.

Решение. Дискретная случайная величина X - число попаданий в цель при трех выстрелах может принимать четыре значения: 0, 1, 2, 3. Вероятность того, что она примет каждое из них, найдем по формуле Бернулли при $n=3$, $p=0,4$, $q=1-p=0,6$ и $m=0, 1, 2, 3$:

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

Получим вероятности возможных значений X :

$$P(X=0) = C_3^0 \cdot 0,4^0 \cdot 0,6^3 = 0,216,$$

$$P(X=1) = C_3^1 \cdot 0,4^1 \cdot 0,6^2 = 0,432, \quad P(X=2) = 0,288, \quad P(X=3) = 0,064.$$

Составим искомый закон распределения случайной величины X :

X	0	1	2	3
P	0,216	0,432	0,288	0,064

Контроль: $0,216+0,432+0,288+0,064=1$.

Построим многоугольник распределения полученной случайной величины X . Для этого в прямоугольной системе координат отметим точки $(0;0,216)$, $(1;0,432)$, $(2; 0,288)$, $(3;0,064)$. Соединим эти точки отрезками

прямых, полученная ломаная и есть искомый многоугольник распределения. См. рисунок в п.2).

2). Найти интегральную функцию $F(x)$ для полученного распределения дискретной случайной величины X и начертить ее график.

Решение. 1. Если $x \leq 0$, то $F(x)=0$. Действительно, значений, меньших нуля, величина X не принимает. Следовательно, при всех $x \leq 0$, пользуясь определением $F(x)$, получим $F(x)=P(X < x)=0$ (как вероятность невозможного события).

2. Если $0 < x \leq 1$, то $F(x)=0,216$.

Действительно, в этом случае $F(x)=P(X < x)=P(-\infty < X \leq 0)+P(0 < X < x)=0,216+0=0,216$.

Если взять, например, $x=0,2$, то $F(0,2)=P(X < 0,2)$. Но вероятность события $X < 0,2$ равна 0,216, так как случайная величина X лишь в одном случае принимает значение меньшее 0,2, а именно 0 с вероятностью 0,216.

3. Если $1 < x \leq 2$, то $F(x)=P(X < x)=P(-\infty < X \leq 0)+P(0 < X \leq 1)+P(1 < X < x)=0,216+0,432+0=0,648$. Действительно, X может принять значение 0 с вероятностью 0,216 и значение 1 с вероятностью 0,432; следовательно, одно из этих значений, безразлично какое, X может принять (по теореме сложения вероятностей несовместных событий) с вероятностью 0,648.

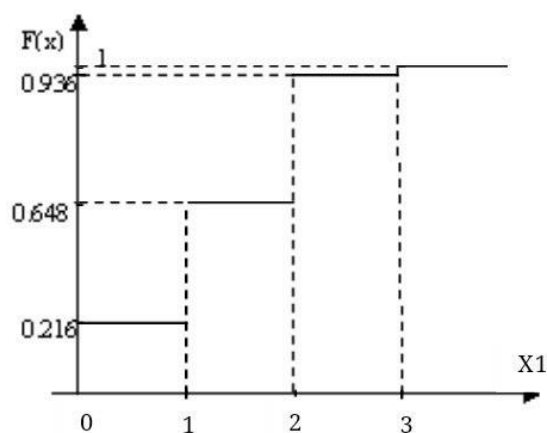
4. Если $2 < x \leq 3$, то рассуждая аналогично, получим $F(x)=0,216+0,432+0,288=0,936$. Действительно, пусть, например, $x=3$. Тогда $F(3)=P(X < 3)$ выражает вероятность события $X < 3$ – стрелок сделает меньше трех попаданий, т.е. ноль, один или два. Применяя теорему сложения вероятностей, получим указанное значение функции $F(x)$.

5. Если $x > 3$, то $F(x)=0,216+0,432+0,288+0,064=1$. Действительно, событие $X \leq 3$ является достоверным и вероятность его равна единице, а $X > 3$ – невозможным. Учитывая, что

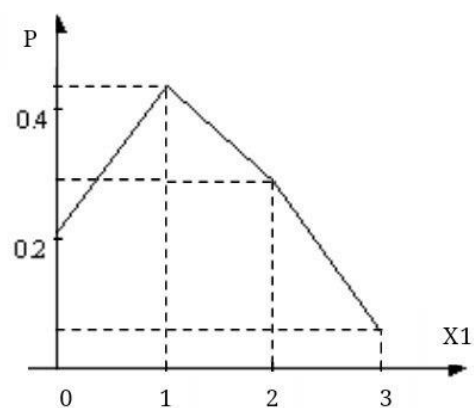
$F(x)=P(X < x)=P(X \leq 3) + P(3 < X < x)$, получим указанный результат.

Итак, получена искомая интегральная функция распределения случайной величины X :

$$F(x) = \begin{cases} 0, & \text{если } -\infty < x \leq 0, \\ 0,216, & " \quad 0 < x \leq 1, \\ 0,648, & " \quad 1 < x \leq 2, \\ 0,936, & " \quad 2 < x \leq 3, \\ 1, & " \quad 3 < x < \infty. \end{cases}$$



Интегральная функция F(X)



Многоугольник

распределения

3). Найти математическое ожидание $M(x)$, дисперсию $D(x)$ и среднее квадратическое отклонение $\sigma(X)$ случайной величины X , закон распределения которой получен в п.1).

Решение. Математическое ожидание дискретной случайной величины равно сумме произведений всех возможных значений X на их вероятности:

$$M(x) = 0 \cdot 0,216 + 1 \cdot 0,432 + 2 \cdot 0,288 + 3 \cdot 0,064 = 1,2.$$

То есть, в среднем происходит одно попадание в цель при трех выстрелах.

Дисперсию можно вычислить, исходя из определения дисперсии $D(X) = M(X - M(X))^2$ или воспользоваться формулой $D(X) = M(X^2) - (M(X))^2$, которая ведет к цели быстрее.

Напишем закон распределения случайной величины X^2 :

X^2	0	1	4	9
P	0.216	0.432	0.228	0.064

Найдем математическое ожидание для X^2 :

$$M(X^2) = 0 \cdot 0,216 + 1 \cdot 0,432 + 4 \cdot 0,228 + 9 \cdot 0,064 = 2,16.$$

Вычислим искомую дисперсию:

$$D(X) = M(X^2) - (M(X))^2 = 2,16 - (1,2)^2 = 0,72.$$

Среднее квадратическое отклонение найдем по формуле

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0,72} = 0,848.$$

Интервал $(M - \sigma; M + \sigma) = (1,2 - 0,85; 1,2 + 0,85) = (0,35; 2,05)$ - интервал наиболее вероятных значений случайной величины X , в него попадают значения 1 и 2.

Пример 2. Дана дифференциальная функция распределения (функция плотности) непрерывной случайной величины X :

$$f(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ \frac{a(2x-1)}{2} & \text{при } 1 < x \leq 2, \\ 0 & \text{при } x > 2. \end{cases}$$

- 1). Определить постоянный параметр a .
- 2). Найти интегральную функцию $F(x)$.
- 3). Начертить графики функций $f(x)$ и $F(x)$.
- 4). Найти двумя способами вероятности $P(0,5 < X \leq 1,5)$ и $P(1,5 < X < 3,5)$.

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

Решение.

1). Дифференциальная функция по свойству $f(x)$ должна удовлетворять условию

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Вычислим этот несобственный интеграл для данной функции $f(x)$:

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^1 0 dx + \int_1^2 \frac{a(2x-1)}{2} dx + \int_2^{\infty} 0 dx = 0 + \int_1^2 \frac{a(2x-1)}{2} dx + 0 = \frac{a}{2} (x^2 - x) \Big|_1^2 = \frac{a}{2} 2 = a.$$

Подставляя этот результат в левую часть равенства, получим, что $a=1$. В условии для $f(x)$ заменим параметр a на 1:

$$f(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ \frac{2x-1}{2} & \text{при } 1 < x \leq 2, \\ 0 & \text{при } x > 2. \end{cases}$$

2). Для нахождения $F(x)$ воспользуемся формулой

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

Если $x \leq 1$, то $f(x) = 0$, следовательно $F(x) = \int_{-\infty}^x 0 dx = 0$.

Если $1 < x \leq 2$,

то
$$F(x) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^1 0 dx + \int_1^x \frac{2x-1}{2} dx = \frac{x^2 - x}{2} \Big|_1^x = \frac{x^2 - x}{2}.$$

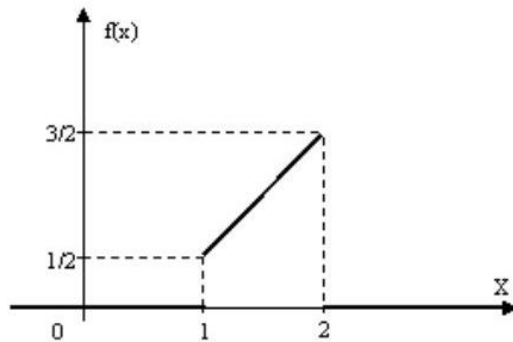
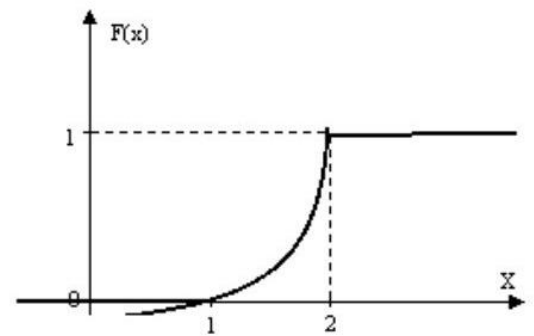
Если $x > 2$, то

$$F(x) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^1 0 dx + \int_1^2 \frac{2x-1}{2} dx + \int_2^x 0 dx = 0 + \frac{x^2 - x}{2} \Big|_1^2 + 0 = \frac{4-2}{2} = 1.$$

Итак, искомая интегральная функция $F(x)$ имеет вид

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

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

График $F(x)$ График $f(x)$ 

4). Вероятность попадания случайной величины в заданный

интервал (a, b) вычисляется по формуле а)
$$P(a < X < b) = \int_a^b f(x) dx$$
, если известна функция $f(x)$

и по формуле б) $P(a < X < b) = F(b) - F(a)$, если известна функция $F(x)$.

Найдем $P(0,5 < X \leq 1,5)$ по двум формулам и сравним результаты. По условию $a=0,5$; $b=1,5$; функция $f(x)$ задана в п.1). Следовательно, искомая вероятность по формуле а) равна:

$$P(0,5 < X \leq 1,5) = \int_{0,5}^{1,5} f(x) dx = \int_{0,5}^1 0 dx + \int_1^{1,5} \frac{2x-1}{2} dx = 0 + \frac{(2x-1)^2}{8} \Big|_1^{1,5} = \frac{3}{8}.$$

Та же вероятность может быть вычислена по формуле б) через приращение полученной в п.2). интегральной функции $F(x)$ на этом интервале:

$$P(0,5 < X \leq 1,5) = F(1,5) - F(0,5) = \left(\frac{x^2 - x}{2} \right)_{x=1,5} - 0 = \frac{1}{2} \left(\frac{9}{4} - \frac{3}{2} \right) = \frac{3}{8}, \text{ т.к. } F(0,5) =$$

0.

Аналогично находим

$$\text{а) } P(1,5 < X < 3,5) = \int_{1,5}^{3,5} f(x) dx = \int_{1,5}^2 \frac{2x-1}{2} dx + \int_2^{3,5} 0 dx = \frac{(2x-1)^2}{8} \Big|_{1,5}^2 + 0 = \frac{5}{8}$$

$$b) \quad P(1,5 < X < 3,5) = F(3,5) - F(1,5) = 1 - \left(\frac{x^2 - x}{2} \right)_{x=1,5} = 1 - \frac{3}{8} = \frac{5}{8}. \quad \text{т.к. } F(3,5)=1.$$

5). Для нахождения математического ожидания $M(X)$ воспользуемся

формулой $M(X) = \int_{-\infty}^{\infty} xf(x)dx$. Функция $f(x)$ задана в п.1), она равна нулю вне интервала $(1,2]$.

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^1 x \cdot 0dx + \int_1^2 x \cdot \frac{2x-1}{2} dx + \int_2^{\infty} x \cdot 0dx = \frac{1}{2} \int_1^2 (2x^2 - x) dx = \frac{19}{12} = 1,583 \approx 1,58.$$

Дисперсия непрерывной случайной величины $D(X)$ определяется равенством

$$D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 f(x) dx, \quad \text{или равносильным равенством}$$

$$D(X) = M(X^2) - (M(X))^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - (M(X))^2.$$

Для нахождения $D(X)$ воспользуемся последней формулой и учтем, что все возможные значения $f(x)$ принадлежат интервалу $(1,2]$.

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\frac{19}{12} \right)^2 = \int_1^2 x^2 \frac{2x-1}{2} dx - \left(\frac{19}{12} \right)^2 = \frac{31}{12} - \left(\frac{19}{12} \right)^2 = 0,076 \approx 0,08.$$

Среднее квадратическое отклонение $\sigma(X) = \sqrt{D(X)} = 0,276$.

Интервал наиболее вероятных значений случайной величины X равен $(M - \sigma, M + \sigma) = (1,58 - 0,28; 1,58 + 0,28) = (1,3; 1,86)$.

Распределение Пирсона.

Распределение Пирсона χ^2 (хи - квадрат) – распределение случайной величины

$$X = X_1^2 + X_2^2 + \dots + X_n^2,$$

где случайные величины $X_1, X_2, \dots, X_n$ независимы и имеют одно и тоже распределение $N(0,1)$. При этом число слагаемых, т.е. n , называется «числом степеней свободы» распределения хи -квадрат.

Распределение хи-квадрат используют при оценивании дисперсии (с помощью доверительного интервала), при проверке гипотез согласия, однородности, независимости, прежде всего для качественных (категоризованных) переменных, принимающих конечное число значений, и во многих других задачах статистического анализа данных.

Распределение Стьюдента.

Распределение t Стьюдента – это распределение случайной величины

$$T = \frac{U\sqrt{n}}{\sqrt{X}},$$

где случайные величины U и X независимы, U имеет распределение стандартное нормальное распределение $N(0,1)$, а X – распределение хи – квадрат с n степенями свободы. При этом n называется «числом степеней свободы» распределения Стьюдента.

Распределение Стьюдента было введено в 1908 г. английским статистиком В. Госсетом, работавшем на фабрике, выпускающей пиво. Вероятностно-статистические методы использовались для принятия экономических и технических решений на этой фабрике, поэтому ее руководство запрещало В. Госсету публиковать научные статьи под своим именем. Таким способом охранялась коммерческая тайна, «ноу-хау» в виде вероятностно-статистических методов, разработанных В. Госсетом. Однако он имел возможность публиковаться под псевдонимом «Стьюдент». История Госсета - Стьюдента показывает, что еще сто лет назад менеджерам Великобритании была очевидна большая экономическая эффективность вероятностно-статистических методов.

В настоящее время распределение Стьюдента – одно из наиболее известных распределений среди используемых при анализе реальных данных. Его применяют при оценивании математического ожидания, прогнозного

значения и других характеристик с помощью доверительных интервалов, по проверке гипотез о значениях математических ожиданий, коэффициентов регрессионной зависимости, гипотез однородности выборок и т.д.

Распределение Фишера.

Распределение Фишера – это распределение случайной величины

$$F = \frac{\frac{1}{k_1} X_1}{\frac{1}{k_2} X_2},$$

где случайные величины X_1 и X_2 независимы и имеют распределения хи – квадрат с числом степеней свободы k_1 и k_2 соответственно. При этом пара (k_1, k_2) – пара «чисел степеней свободы» распределения Фишера, а именно, k_1 – число степеней свободы числителя, а k_2 – число степеней свободы знаменателя. Распределение случайной величины F названо в честь великого английского статистика Р.Фишера (1890-1962), активно использовавшего его в своих работах.

Распределение Фишера используют при проверке гипотез об адекватности модели в регрессионном анализе, о равенстве дисперсий и в других задачах прикладной статистики.

Характеристическая функция.

Плотность вероятности нормально распределённой случайной величины с параметром смещения μ и масштаба σ (или, что тоже самое, дисперсией σ^2) имеет следующий вид:

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

Функция распределения такой величины не выражается через элементарные функции и записывается по определению через интеграл Римана как

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt.$$

Функция распределения стандартной нормальной случайной величины (т. е. при $\mu = 0$, $\sigma = 1$) часто обозначают как $\Phi(\cdot)$:

$$\Phi(x) = F(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

Функцию распределения нормальной случайной величины с любыми параметрами легко выразить через $\Phi(\cdot)$:

$$F(x; \mu, \sigma) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Характеристическая функция нормального распределения имеет вид

$$f(t) = E\{e^{it\xi}\} = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right),$$

где $\xi \sim N(\mu, \sigma^2)$ — нормально распределённая с параметрами μ и σ случайная величина.

Значение предельных теорем.

Теория вероятностей изучает закономерности, свойственные массовым случайным явлениям. Предельные теоремы теории вероятностей устанавливают зависимость между случайностью и необходимостью. Изучение закономерностей, проявляющихся в массовых случайных явлениях, позволяет научно предсказывать результаты будущих испытаний.

Предельные теоремы теории вероятностей делятся на две группы, одна из которых получила название **закона больших чисел**, а другая — **центральной предельной теоремы**.

Рассмотрим теоремы, относящихся к закону больших чисел: неравенство Чебышева, теоремы Чебышева и Бернулли.

Закон больших чисел состоит из нескольких теорем, в которых доказывается приближение средних характеристик при соблюдении определённых условий к некоторым постоянным значениям.

Сходимость по вероятности

Как и в математическом анализе, в теории вероятностей приходится иметь дело с разными видами сходимости случайных величин. Рассмотрим

основные виды сходимости: по вероятности; с вероятностью 1; в среднем порядка s ; по распределению.

Пусть $X_1, X_2, \dots$ - последовательность случайных величин задана на вероятностном пространстве $\langle W, F, P \rangle$. Пусть на этом пространстве задана также и случайная величина X .

Определение 1. Последовательность случайных величин $X_1, \dots, X_n, \dots$ называется сходящейся **по вероятности** к случайной величине X (обозначение: $X_n \xrightarrow{P} X$), если для любого $\varepsilon > 0$ ($\varepsilon > 0$) существует:

$$\lim_{n \rightarrow \infty} \{ |X_n - X| < \varepsilon \} = 1 \quad P\{|X_n - X| > \varepsilon\} = 0, n \in \mathbb{N} \quad (1)$$

Определение 2. Последовательность случайных величин $X_1, X_2, \dots$ называется сходящейся **с вероятностью единица** (почти наверное, почти всюду) к случайной величине X , если

$$P\{\omega : X_n \not\rightarrow X\} = 0 \quad (2)$$

т.е. если множество исходов ω , для которых $X_n(\omega)$ не сходятся к $X(\omega)$, имеет нулевую вероятность. Этот вид сходимости обозначают следующим

образом: $X_n \xrightarrow{п.н.} X; X_n \xrightarrow{п.в.} X$

Определение 3. Последовательность случайных величин $X_1, X_2, \dots$ называется сходящейся **в среднем порядка s** , $0 < s < \infty$, к случайной величине X , если

$$M |X_n - X|^s \rightarrow 0, n \in \mathbb{N} \text{ или } \lim_{n \rightarrow \infty} M |X_n - X|^s = 0 \quad (3)$$

Если $s = 2$, то сходимость (3) называют сходимостью **в среднем квадратическом** и используют формулу.

$$\lim_{n \rightarrow \infty} M |X_n - X|^2 = 0 \quad (4)$$

Определение 4. Последовательность случайных величин $X_1, X_2, \dots$ называется сходящейся **по распределению** к случайной величине, если во всех точках x , где функция распределения $F_X(x)$ - непрерывна, выполняется условие

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (5)$$

Это обозначается так: $X_n \xrightarrow{d} X; F_{X_n}(x) \Rightarrow F_X(x)$.

Замечание 1. О сходимости по распределению можно говорить и в том случае, если $\{X_n\}$ и X заданы на разных вероятностных пространствах.

Замечание 2. Сходимость по вероятности, почти всюду, в среднем порядка s применима и к неслучайным величинам.

Неравенство Чебышева

Математические законы теории вероятностей получены абстрагированием реальных статистических закономерностей, свойственных массовым случайным явлениям. Наличие этих закономерностей связано именно с массовостью явлений, то есть с большим числом выполняемых однородных опытов или с большим числом складывающихся случайных воздействий, порождающих в своей совокупности случайную величину, подчиненную вполне определенному закону. Свойство устойчивости массовых случайных явлений известно человечеству еще с глубокой древности. В какой бы области оно ни проявлялось, суть его сводится к следующему: конкретные особенности каждого отдельного случайного явления почти не сказываются на среднем результате массы таких явлений; случайные отклонения от среднего, неизбежные в каждом отдельном явлении, в массе взаимно погашаются, нивелируются, выравниваются. Именно эта **устойчивость средних** и представляет собой физическое содержание «закона

больших чисел», понимаемого в широком смысле слова: при очень большом числе случайных явлений средний их результат практически перестает быть случайным и может быть предсказан с большой степенью определенности.

В узком смысле слова под «законом больших чисел» в теории вероятностей понимается ряд математических теорем, в каждой из которых для тех или иных условий устанавливается факт приближения средних характеристик большого числа опытов к некоторым определенным постоянным. С общими формами закона больших чисел мы познакомимся в данной главе. Все они устанавливают факт и условия сходимости по вероятности тех или иных случайных величин к постоянным, не случайным величинам.

Закон больших чисел играет важную роль в практических применениях теории вероятностей. Свойство случайных величин при определенных условиях вести себя практически как не случайные позволяет уверенно оперировать с этими величинами, предсказывать результаты массовых случайных явлений почти с полной определенностью.

Возможности таких предсказаний в области массовых случайных явлений еще больше расширяются наличием другой группы предельных теорем, касающихся уже не предельных значений случайных величин, а **предельных законов распределения**. Речь идет о группе теорем, известных под названием «центральной предельной теоремы». Мы уже говорили о том, что при суммировании достаточно большого числа случайных величин закон распределения суммы неограниченно приближается к нормальному при соблюдении некоторых условий. Эти условия, которые математически можно формулировать различным образом - в более или менее общем виде, - по существу сводятся к требованию, чтобы влияние на сумму отдельных слагаемых было **равномерно малым**, т. е. чтобы в состав суммы не входили члены, явно преобладающие над совокупностью остальных по своему влиянию на рассеивание суммы. Различные формы центральной предельной

теоремы различаются между собой теми условиями, для которых устанавливается это предельное свойство суммы случайных величин.

Различные формы закона больших чисел вместе с различными формами центральной предельной теоремы образуют совокупность так называемых **предельных теорем** теории вероятностей. Предельные теоремы дают возможность не только осуществлять научные прогнозы в области случайных явлений, но и оценивать точность этих прогнозов.

Теоремы Чебышева и Бернулли

Теорема 1: (теорема Чебышева). Если $X_1, X_2, \dots, X_n$ – попарно независимые случайные величины, дисперсии которых равномерно ограничены ($D(X_i) \leq C$), то для сколь угодно малого числа ε вероятность неравенства

$$\left| \frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n} \right| < \varepsilon$$

будет сколь угодно близка к 1, если число случайных величин достаточно велико.

Замечание. Иначе говоря, при выполнении этих условий

$$\lim_{n \rightarrow \infty} P\left(\left| \frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n} \right| < \varepsilon \right) = 1.$$

Доказательство. Рассмотрим новую случайную величину $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ и найдем ее математическое ожидание. Используя свойства математического

ожидания, получим, что $M\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}$.

Применим к $\bar{X}$ неравенство Чебышева:

$$p\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| < \varepsilon\right) \geq 1 - \frac{D\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)}{\varepsilon^2}.$$

Так как рассматриваемые случайные величины независимы, то, учитывая условие

$$\text{теоремы, имеем: } D\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{D(X_1) + D(X_2) + \dots + D(X_n)}{n^2} \leq \frac{Cn}{n^2} = \frac{C}{n}.$$

Используя этот результат, представим предыдущее неравенство в виде:

$$p\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| < \varepsilon\right) \geq 1 - \frac{C}{n\varepsilon^2}.$$

Перейдем к пределу при $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} p\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| < \varepsilon\right) \geq 1.$$

Поскольку вероятность не может быть больше 1, можно утверждать, что

$$\lim_{n \rightarrow \infty} p\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{M(X_1) + M(X_2) + \dots + M(X_n)}{n}\right| < \varepsilon\right) = 1.$$

Теорема доказана.

Следствие.

Если $X_1, X_2, \dots, X_n$ – попарно независимые случайные величины с равномерно ограниченными дисперсиями, имеющие одинаковое математическое ожидание, равное a , то для любого сколь угодно малого $\varepsilon > 0$ вероятность

неравенства $\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < \varepsilon$ будет как угодно близка к 1, если число случайных величин достаточно велико. Иначе говоря,

$$\lim_{n \rightarrow \infty} p\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < \varepsilon\right) = 1.$$

Вывод: среднее арифметическое достаточно большого числа случайных величин принимает значения, близкие к сумме их математических ожиданий, то есть утрачивает характер случайной величины. Например, если проводится серия измерений какой-либо физической величины, причем: а) результат

каждого измерения не зависит от результатов остальных, то есть все результаты представляют собой попарно независимые случайные величины; б) измерения производятся без систематических ошибок (их математические ожидания равны между собой и равны истинному значению a измеряемой величины); в) обеспечена определенная точность измерений, следовательно, дисперсии рассматриваемых случайных величин равномерно ограничены; то при достаточно большом числе измерений их среднее арифметическое окажется сколь угодно близким к истинному значению измеряемой величины.

Теорема 2: (теорема Бернулли). Если в каждом из n независимых опытов вероятность p появления события A постоянна, то при достаточно большом числе испытаний вероятность того, что модуль отклонения относительной частоты появлений A в n опытах от p будет сколь угодно малым, как угодно близка к 1:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1.$$

Доказательство. Введем случайные величины $X_1, X_2, \dots, X_n$, где X_i – число появлений A в i -м опыте. При этом X_i могут принимать только два значения: 1 (с вероятностью p) и 0 (с вероятностью $q = 1 - p$). Кроме того, рассматриваемые случайные величины попарно независимы и их дисперсии равномерно ограничены (так как $D(X_i) = pq$, $p + q = 1$, откуда $pq \leq 1/4$). Следовательно, к ним можно применить теорему Чебышева при $M_i = p$:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - p\right| < \varepsilon\right) = 1.$$

Но $\frac{X_1 + X_2 + \dots + X_n}{n} = \frac{m}{n}$, так как X_i принимает значение, равное 1, при появлении A в данном опыте, и значение, равное 0, если A не произошло. Таким образом,

$$\lim_{n \rightarrow \infty} p \left(\left| \frac{m}{n} - p \right| < \varepsilon \right) = 1,$$

что и требовалось доказать.

Замечание. Из теоремы Бернулли *не следует*, что $\lim_{n \rightarrow \infty} \frac{m}{n} = p$. Речь идет лишь о *вероятности* того, что разность относительной частоты и вероятности по модулю может стать сколь угодно малой. Разница заключается в следующем: при обычной сходимости, рассматриваемой в математическом анализе, для всех n , начиная с некоторого значения, неравенство $\left| \frac{m}{n} - p \right| < \varepsilon$ выполняется всегда; в нашем случае могут найтись такие значения n , при которых это неравенство неверно. Этот вид сходимости называют **сходимостью по вероятности**.

Центральная предельная теорема.

Центральная предельная теорема Ляпунова.

Закон больших чисел не исследует вид предельного закона распределения суммы случайных величин. Этот вопрос рассмотрен в группе теорем, называемых **центральной предельной теоремой**. Они утверждают, что закон распределения суммы случайных величин, каждая из которых может иметь различные распределения, приближается к нормальному при достаточно большом числе слагаемых. Этим объясняется важность нормального закона для практических приложений. Характеристические функции.

Для доказательства центральной предельной теоремы используется метод характеристических функций.

Определение: **Характеристической функцией** случайной величины X называется функция

$$g(t) = M(e^{itX}) \quad (14.1)$$

Таким образом, $g(t)$ представляет собой математическое ожидание некоторой комплексной случайной величины $U = e^{itX}$, связанной с величиной X . В частности, если X – дискретная случайная величина, заданная рядом распределения, то

$$g(t) = \sum_{k=1}^n e^{itx_k} p_k. \quad (14.2)$$

Для непрерывной случайной величины с плотностью распределения $f(x)$

$$g(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx. \quad (14.3)$$

Пример 1. Пусть X – число выпадений 6 очков при одном броске игральной кости. Тогда по формуле (14.2) $g(t) = e^{it \cdot 0} \cdot \frac{5}{6} + e^{it \cdot 1} \cdot \frac{1}{6} = \frac{5 + e^{it}}{6}$.

Пример 2. Найдем характеристическую функцию для нормированной непрерывной случайной величины, распределенной по нормальному закону

$$\left(f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right). \text{ По формуле (14.3) } g(t) = \int_{-\infty}^{+\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx - \frac{x^2}{2}} dx = e^{-\frac{t^2}{2}} \quad ($$

использовалась формула $\int_{-\infty}^{+\infty} e^{-Ax^2 \pm 2Bx - C} dx = \sqrt{\frac{\pi}{A}} e^{-\frac{AC - B^2}{A}}$ и то, что $i^2 = -1$).

Свойства характеристических функций:

1. Функцию $f(x)$ можно найти по известной функции $g(t)$ по формуле

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} g(t) dt. \quad (14.4)$$

(преобразование (14.3) называется *преобразованием Фурье*, а преобразование (14.4) – *обратным преобразованием Фурье*).

2. Если случайные величины X и Y связаны соотношением $Y = aX$, то их характеристические функции связаны соотношением

$$g_y(t) = g_x(at). \quad (14.5)$$

3. Характеристическая функция суммы независимых случайных величин

равна произведению характеристических функций слагаемых: для $Y = \sum_{k=1}^n X_k$

$$g_y(t) = g_{x_1}(t) \cdot g_{x_2}(t) \cdot \dots \cdot g_{x_n}(t) \quad (14.6)$$

Центральная предельная теорема для одинаково распределенных слагаемых.

Если $X_1, X_2, \dots, X_n, \dots$ - независимые случайные величины с одинаковым законом распределения, математическим ожиданием m и дисперсией σ^2 , то при

неограниченном увеличении n закон распределения суммы $Y_n = \sum_{k=1}^n X_k$ неограниченно приближается к нормальному.

Доказательство.

Докажем теорему для непрерывных случайных величин $X_1, X_2, \dots, X_n$ (доказательство для дискретных величин аналогично). Согласно условию теоремы, характеристические функции слагаемых одинаковы:

$g_x(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx$. Тогда по свойству 3 характеристическая функция суммы Y_n будет $g_{y_n}(t) = g_x^n(t)$. Разложим функцию $g_x(t)$ в ряд Маклорена:

$$g_x(t) = g_x(0) + g'_x(0)t + \left(\frac{g''_x(0)}{2} + \alpha(t) \right) t^2, \quad \text{где} \quad \alpha(t) \rightarrow 0 \text{ при } t \rightarrow 0.$$

Найдем

$$g_x(0) = \int_{-\infty}^{+\infty} f(x) dx = 1, \quad g'_x(0) = \left. \int_{-\infty}^{+\infty} ixe^{itx} f(x) dx \right|_{t=0} = i \left. \int_{-\infty}^{+\infty} xe^{itx} f(x) dx \right|_{t=0} = i \int_{-\infty}^{+\infty} xf(x) dx = im.$$

Если предположить, что $m = 0$ (то есть перенести начало отсчета в точку m), то $g'_x(0) = 0$.

$g''_x(0) = - \left. \int_{-\infty}^{+\infty} x^2 e^{itx} f(x) dx \right|_{t=0} = - \int_{-\infty}^{+\infty} x^2 f(x) dx = -\sigma^2$ (так как $m = 0$). Подставив полученные результаты в формулу Маклорена, найдем, что

$$g_x(t) = 1 - \left(\frac{\sigma^2}{2} - \alpha(t) \right) t^2.$$

Рассмотрим новую случайную величину $Z_n = \frac{Y_n}{\sigma\sqrt{n}}$, отличающуюся от Y_n тем, что ее дисперсия при любом n равна 0. Так как Y_n и Z_n связаны линейной зависимостью, достаточно доказать, что Z_n распределена по нормальному закону, или, что то же самое, что ее характеристическая функция приближается к характеристической функции нормального закона (см. пример 2). По свойству характеристических функций

$$g_{z_n}(t) = g_{y_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(g_x\left(\frac{t}{\sigma\sqrt{n}}\right) \right)^n = \left(1 - \left(\frac{\sigma^2}{2} - \alpha\left(\frac{t}{\sigma\sqrt{n}}\right) \right) \frac{t^2}{n\sigma^2} \right)^n.$$

Прологарифмируем

полученное

выражение:

$$\ln g_{z_n}(t) = n \ln(1 - k), \text{ где } k = \left(\frac{\sigma^2}{2} - \alpha \left(\frac{t}{\sigma \sqrt{n}} \right) \right) \frac{t^2}{n \sigma^2}, \quad \lim_{n \rightarrow \infty} k = 0.$$

Разложим $\ln(1 - k)$ в ряд при $n \rightarrow \infty$, ограничившись двумя членами разложения, тогда $\ln(1 - k) \approx -k$. Отсюда

$$\lim_{n \rightarrow \infty} \ln g_{z_n}(t) = \lim_{n \rightarrow \infty} n \cdot (-k) = \lim_{n \rightarrow \infty} \left(-\frac{t^2}{2} + \alpha \left(\frac{t}{\sigma \sqrt{n}} \right) \frac{t^2}{\sigma^2} \right) = -\frac{t^2}{2} + \lim_{n \rightarrow \infty} \frac{t^2}{\sigma^2} \alpha \left(\frac{t}{\sigma \sqrt{n}} \right), \text{ где последний}$$

предел равен 0, так как $\alpha(t) \rightarrow 0$ при $t \rightarrow 0$. Следовательно, $\lim_{n \rightarrow \infty} \ln g_{z_n}(t) = -\frac{t^2}{2}$, то

есть $\lim_{n \rightarrow \infty} g_{z_n}(t) = e^{-\frac{t^2}{2}}$ - характеристическая функция нормального распределения.

Итак, при неограниченном увеличении числа слагаемых характеристическая функция величины Z_n неограниченно приближается к характеристической функции нормального закона; следовательно, закон распределения Z_n (и Y_n) неограниченно приближается к нормальному. Теорема доказана.

Центральная предельная теорема

Нормально распределенные величины широко распространены на практике. Чем это объясняется? Ответ на этот вопрос был дан выдающимся русским математиком А.М. Ляпуновым (центральная предельная теорема): *если случайная величина X представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то X имеет распределение, близкое к нормальному.*

Пример: Пусть производится измерение некоторой физической величины. Любое измерение дает лишь приближенное значение измеряемой величины, так как на результат измерения влияют очень многие независимые случайные факторы (температура, колебание прибора, влажность и др.).

Каждый из этих факторов порождает ничтожную “частную ошибку”. Однако их совокупное действие порождает уже заметную “суммарную ошибку”, которая имеет распределение, близкое к нормальному.

Таким образом, *центральная предельная теорема устанавливает условия, при которых сумма большого числа независимых слагаемых имеет нормальное распределение.*

Центральная предельная теорема:

Пусть $X_1, X_2, \dots, X_n$ – последовательность независимых случайных величин, каждая из которых имеет конечные математическое ожидание и дисперсию: $M(X_k) = a_k, D(X_k) = b_k^2$.

Обозначим $S_n = \sum_{k=1}^n X_k, A_n = \sum_{k=1}^n a_k, B_n^2 = \sum_{k=1}^n b_k^2$, а функцию распределения

нормированной суммы через $F_n(x) = P\left[\frac{S_n - A_n}{B_n} < x\right]$.

Тогда при любом x функция распределения нормированной суммы стремится к нормальной функции распределения при $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} P\left[\frac{S_n - A_n}{B_n} < x\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz.$$

В частности, если все случайные величины $X_1, X_2, \dots$ одинаково распределены, то к этой последовательности применима центральная предельная теорема, если дисперсии всех величин X_i ($i=1,2,\dots$) конечны и отличны от нуля.

Условие Ляпунова: Если для $\delta > 0$ при $n \rightarrow \infty$ отношение Ляпунова

$$L_n = C_n / B_n^{2+\delta}, \text{ где } C_n = \sum_{k=1}^n M|X_k - a_k|^{2+\delta}$$

стремится к нулю, то к последовательности X_1, X_2 применима центральная предельная теорема.

Сущность условия Ляпунова состоит в требовании, чтобы каждое

слагаемое суммы $\frac{S_n - A_n}{B_n}$ оказывало на сумму ничтожное влияние.

Роль нормального распределения в приложениях.

Пример, характеризующий роль нормального распределения в приложениях теории вероятностей, дает массовое производство, существующее во многих отраслях современной промышленности. В процессе массового производства изготавливаются большие партии однотипных изделий. Все наиболее существенные характеристики выпускаемых изделий должны, естественно, соответствовать определенному стандарту однако в действительности наблюдаются отклонения от стандарта, которые порождаются причинами случайного характера (следует, учесть, что выпуск изделия связан, как правило, с большим числом операций, некоторые из них не могут быть выполнены абсолютно точно). Каждая из этих причин сама по себе порождает ничтожную ошибку ξ_i (это и есть, в каком то смысле, условие Ляпунова, что «достаточно вероятные» значения слагаемых ξ_k должны быть при больших n малы по сравнению со всей суммой v_n), но, складываясь, такие ошибки могут давать вполне ощутимые отклонения от стандарта. И здесь, так же как в случае ошибок измерения, имеются все основания считать, что суммарное отклонение от стандарта подчиняется нормальному распределению.

Раздел 2. Математическая статистика.

Тема 2. 1. Предмет и задачи математической статистики.

Предмет математической статистики.

Математическая статистика – это наука, изучающая случайные явления посредством обработки и анализа результатов наблюдений и измерений.

Первая задача математической статистики – указать способы получения, группировки и обработки статистических данных, собранных в результате наблюдений, специально поставленных опытов или произведённых измерений.

Вторая задача математической статистики – разработка методов анализа статистических сведений в зависимости от целей исследования. Например, целью исследования может быть:

- оценка неизвестной вероятности события;
- оценка параметров распределения случайной величины;
- оценка неизвестной функции распределения случайной величины;
- проверка гипотез о параметрах распределения или о виде неизвестного распределения;
- оценка зависимости случайной величины от одной или нескольких случайных величин и т.д.

Случайную величину X будем называть *генеральной совокупностью* X .

Исходным материалом для изучения свойств генеральной совокупности X являются статистические данные, т.е. значения X , полученные в результате повторения случайного опыта (измерения случайной величины X). Предполагается, что опыт может быть повторён сколько угодно раз в неизменных условиях. Это означает, что распределение случайной величины X_i , $i = 1, 2, \dots$, заданной на множестве исходов i -го опыта, не зависит от i и совпадает с распределением генеральной совокупности X .

Набор n независимых в совокупности случайных величин $(X_1, X_2, \dots, X_n)$, где X_i соответствует i -му опыту, называют *случайной выборкой* из генеральной совокупности X . Число n называется *объёмом выборки*.

Совокупность чисел $(x_1, x_2, \dots, x_n)$, полученных в результате n -кратного повторения опыта по измерению генеральной совокупности X , называется *реализацией случайной выборки* или просто *выборкой объёма n* .

В основе большинства результатов математической статистики лежит *выборочный метод*, состоящий в том, что свойства генеральной совокупности X устанавливаются путём изучения тех же свойств на случайной выборке.

Генеральная и выборочная совокупности.

Пусть требуется изучить совокупность однородных объектов относительно некоторого качественного или количественного признака, характеризующего эти объекты. Например, если имеется партия деталей, то качественным признаком может служить стандартность детали, а количественным—контролируемый размер детали.

Иногда проводят сплошное обследование, т. е. обследуют каждый из объектов совокупности относительно признака, которым интересуются. На практике, однако, сплошное обследование применяют сравнительно редко. Например, если совокупность содержит очень большое число объектов, то провести сплошное обследование физически невозможно. Если обследование объекта связано с его уничтожением или требует больших материальных затрат, то проводить сплошное обследование практически не имеет смысла. В таких случаях случайно отбирают из всей совокупности ограниченное число объектов и подвергают их изучению. Различают генеральную и выборочную совокупности:

Генеральной совокупностью называют совокупность всех мысленно возможных объектов данного вида, над которыми проводятся наблюдения с целью получения конкретных значений случайной величины, или совокупность результатов всех мыслимых наблюдений, проводимых в неизменных условиях над одной из случайных величин, связанных с данным видом объектов.

Замечание: Часто генеральная совокупность содержит конечное число объектов. Однако если это число достаточно велико, то иногда в целях

упрощения вычислений допускают, что генеральная совокупность состоит из бесчисленного множества объектов. Такое допущение оправдывается тем, что увеличение объема генеральной совокупности (достаточно большого объема) практически не сказывается на результатах обработки данных выборки.

Выборочной совокупностью называют часть отобранных объектов из генеральной совокупности.

Объемом совокупности (выборочной или генеральной) называют число объектов этой совокупности. Например, если из 1000 деталей отобрано для обследования 100 деталей, то объем генеральной совокупности $N = 1000$, а объем выборки $n = 100$.

Число объектов генеральной совокупности N значительно превосходит объем выборки n .

Понятие выборки.

Выборкой называется любая подгруппа элементов (испытуемых, респондентов), выделенная из генеральной совокупности для проведения эксперимента. При этом отдельный индивид из выборки, с которым работает психолог, называется испытуемым (респондентом).

Полное или сплошное исследование всей генеральной совокупности – задача нереальная. Поэтому исследования проводятся на репрезентативных выборках.

Варианта (x) – это единица выборки, каждое отдельное x – результат отдельного измерения.

Объем выборки (n) – общее число вариантов в выборке. Объем выборки может быть любым, но не меньшим чем два респондента. В статистике различают малую ($n < 30$), среднюю ($30 < n < 100$) и большую выборку $n > 100$.

Частота (f) – число, показывающее сколько раз встречается в выборке каждая варианта x .

Частость (ω) – это доля каждой частоты в общем объеме выборки, т.е. $\omega = f/n$.

Выборки могут быть независимыми (несвязными) и зависимыми (связными).

Выборки называются независимыми (несвязными), если процедура эксперимента и полученные результаты измерения некоторого свойства у испытуемых одной выборки не оказывают влияния на особенности протекания этого же эксперимента и результаты измерения этого же свойства у испытуемых (респондентов) другой выборки.

Выборки называются зависимыми (связными), если процедура эксперимента и полученные результаты измерения некоторого свойства у испытуемых одной выборки оказывают влияния на особенности протекания этого же эксперимента и результаты измерения этого же свойства у испытуемых (респондентов) другой выборки.

К выборке применяется ряд обязательных требований, определенных прежде всего целями и задачами исследования. Одним из важных требований является требование однородности выборки. Оно означает, что психолог, изучая, например, подростков, не может включить в эту же выборку взрослых людей.

Все требования, предъявляемые к любой выборке, сводятся к тому, что на ее основе психологом должна быть получена наиболее полная, неискаженная информация об особенностях генеральной совокупности, из которой взята эта выборка. Иными словами, выборка должна быть репрезентативной. Репрезентативная выборка, или представительная выборка, - это такая выборка, в которой все основные признаки генеральной совокупности представлены приблизительно в той же пропорции и той же частотой, с которой данный признак выступает в данной генеральной совокупности. Репрезентативная выборка представляет собой меньшую по размеру, но точную модель той генеральной совокупности, которую она должна отражать. Репрезентативность выборки позволяет распространить полученные на ней выводы на всю генеральную совокупность.

Методы, обеспечивающие репрезентативность выборки:

Первый метод формирования **простой случайной** выборки. Получить простую случайную выборку можно путем обычной жеребьевки (по аналогии с лотореей) или с помощью специальных таблиц случайных чисел. В последнем случае элементы генеральной совокупности перенумеровываются и из таблицы случайных чисел выписываются номера элементов, которые должны быть взяты в выборку. Данная процедура трудно осуществима, поскольку для ее реализации необходимо учитывать каждого представителя генеральной совокупности.

Второй метод основывается на понятии **стратифицированной случайной выборки**. Для этого необходимо разбить элементы генеральной совокупности на страты (группы) в соответствии с некоторыми характеристиками (возраст, пол, социальная принадлежность, национальность, место жительства (город, деревня)). Случайная выборка производится отдельно из каждой группы (страты).

Эмпирическая функция распределения.

Чтобы от отдельных событий перейти к одновременному рассмотрению многих событий, используют накопленную частоту. Так называется отношение числа единиц, для которых результаты наблюдения меньше заданного значения, к общему числу наблюдений. (Это понятие используется, если результаты наблюдения – действительные числа, а не вектора, функции или объекты нечисловой природы.) Функция, которая выражает зависимость между значениями количественного признака и накопленной частотой, называется эмпирической функцией распределения. Итак, эмпирической функцией распределения $F_n(x)$ называется доля элементов выборки, меньших x . Эмпирическая функция распределения содержит всю информацию о результатах наблюдений.

Чтобы записать выражение для эмпирической функции распределения в виде формулы, введем функцию $c(x, y)$ двух переменных:

$$c(x, y) = \begin{cases} 0, & x \leq y, \\ 1, & x > y. \end{cases}$$

Случайные величины, моделирующие результаты наблюдений, обозначим $X_1(\omega), X_2(\omega), \dots, X_n(\omega), \omega \in \Omega$. Тогда эмпирическая функция распределения $F_n(x)$ имеет вид

$$F_n(x) = F_n(x, \omega) = \frac{1}{n} \sum_{1 \leq i \leq n} c(x, X_i(\omega)).$$

Из закона больших чисел следует, что для каждого действительного числа x эмпирическая функция распределения $F_n(x)$ сходится к функции распределения $F(x)$ результатов наблюдений, т.е.

$$F_n(x) \rightarrow F(x) \quad (1)$$

при $n \rightarrow \infty$. Советский математик В.И. Гливенко (1897-1940) доказал в 1933 г. более сильное утверждение: сходимость в (1) равномерна по x , т.е.

$$\sup_x |F_n(x) - F(x)| \rightarrow 0 \quad (2)$$

при $n \rightarrow \infty$ (сходимость по вероятности).

В (2) использовано обозначение $\sup$ (читается как «супремум»). Для функции $g(x)$ под $\sup_x g(x)$ понимают наименьшее из чисел a таких, что $g(x) \leq a$ при всех x . Если функция $g(x)$ достигает максимума в точке x_0 , то $\sup_x g(x) = g(x_0)$. В таком случае вместо $\sup$ пишут $\max$. Хорошо известно, что не все функции достигают максимума.

В том же 1933 г. А.Н. Колмогоров усилил результат В.И. Гливенко для непрерывных функций распределения $F(x)$. Рассмотрим случайную величину

$$D_n = \sqrt{n} \sup_x |F_n(x) - F_0(x)|$$

и ее функцию распределения

$$K_n(x) = P(D_n \leq x).$$

По теореме А.Н. Колмогорова

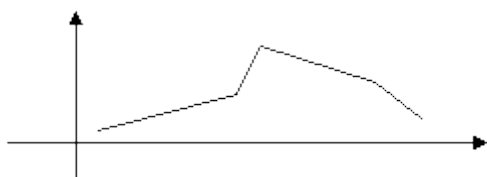
$$\lim_{n \rightarrow \infty} K_n(x) = K(x)$$

при каждом x , где $K(x)$ – т.н. функция распределения Колмогорова.

Рассматриваемая работа А.Н. Колмогорова породила одно из основных направлений математической статистики – т.н. непараметрическую статистику. И в настоящее время непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат широко используются. Они были разработаны для проверки согласия с *полностью известным* теоретическим распределением, т.е. предназначены для проверки гипотезы $H_0 : F(x) \equiv F_0(x)$. Основная идея критериев Колмогорова, омега-квадрат и аналогичных им состоит в измерении расстояния между функцией эмпирического распределения и функцией теоретического распределения. Различаются эти критерии видом расстояний в пространстве функций распределения. Аналитические выражения для предельных распределений статистик, расчетные формулы, таблицы распределений и критических значений широко распространены, поэтому не будем их приводить.

Полигон частот, гистограмма.

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – **полигон частот**: ломаная, отрезки которой соединяют точки с координатами $(x_1, n_1), (x_2, n_2), \dots, (x_k, n_k)$, где x_i откладываются на оси абсцисс, а n_i – на оси ординат. Если на оси ординат откладывать не абсолютные (n_i) , а относительные (w_i) частоты, то получим **полигон относительных частот**



(рис.2).

Рис.

2.

По аналогии с функцией распределения случайной величины можно задать некоторую функцию, относительную частоту события $X < x$.

Точечные оценки и их свойства.

Получив статистические оценки параметров распределения (выборочное среднее, выборочную дисперсию и т.д.), нужно убедиться, что они в достаточной степени служат приближением соответствующих характеристик генеральной совокупности. Определим требования, которые должны при этом выполняться.

Пусть Θ^* - статистическая оценка неизвестного параметра Θ теоретического распределения. Извлечем из генеральной совокупности несколько выборок одного и того же объема n и вычислим для каждой из них оценку параметра Θ : $\Theta_1^*, \Theta_2^*, \dots, \Theta_k^*$. Тогда оценку Θ^* можно рассматривать как случайную величину, принимающую возможные значения $\Theta_1^*, \Theta_2^*, \dots, \Theta_k^*$. Если математическое ожидание Θ^* не равно оцениваемому параметру, мы будем получать при вычислении оценок систематические ошибки одного знака (с избытком, если $M(\Theta^*) > \Theta$, и с недостатком, если $M(\Theta^*) < \Theta$). Следовательно, необходимым условием отсутствия систематических ошибок является требование $M(\Theta^*) = \Theta$.

Статистическая оценка Θ^* называется **несмещенной**, если ее математическое ожидание равно оцениваемому параметру Θ при любом объеме выборки:

$$M(\Theta^*) = \Theta. \quad (1)$$

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Однако **несмещенность** не является достаточным условием хорошего приближения к истинному значению оцениваемого параметра. Если при этом

возможные значения Θ^* могут значительно отклоняться от среднего значения, то есть дисперсия Θ^* велика, то значение, найденное по данным одной выборки, может значительно отличаться от оцениваемого параметра. Следовательно, требуется наложить ограничения на дисперсию.

Статистическая оценка называется **эффективной**, если она при заданном объеме выборки n имеет наименьшую возможную дисперсию. При рассмотрении выборок большого объема к статистическим оценкам предъявляется еще и требование состоятельности.

Состоятельной называется статистическая оценка, которая при $n \rightarrow \infty$ стремится по вероятности к оцениваемому параметру (если эта оценка несмещенная, то она будет состоятельной, если при $n \rightarrow \infty$ ее дисперсия стремится к 0).

Убедимся, что $\bar{X}_B$ представляет собой несмещенную оценку математического ожидания $M(X)$.

Будем рассматривать $\bar{X}_B$ как случайную величину, а $x_1, x_2, \dots, x_n$, то есть значения исследуемой случайной величины, составляющие выборку, – как независимые, одинаково распределенные случайные величины $X_1, X_2, \dots, X_n$, имеющие математическое ожидание a . Из свойств математического ожидания следует, что

$$M(\bar{X}_B) = M\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = a.$$

Но, поскольку каждая из величин $X_1, X_2, \dots, X_n$ имеет такое же распределение, что и генеральная совокупность, $a = M(X)$, то есть $M(\bar{X}_B) = M(X)$, что и требовалось доказать. Выборочное среднее является не только несмещенной, но и состоятельной оценкой математического ожидания. Если предположить,

что $X_1, X_2, \dots, X_n$ имеют ограниченные дисперсии, то из теоремы Чебышева следует, что их среднее арифметическое, то есть $\bar{X}_B$, при увеличении n стремится по вероятности к математическому ожиданию a каждой из величин, то есть к $M(X)$. Следовательно, выборочное среднее есть состоятельная оценка математического ожидания.

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Можно доказать, что

$$M(D_B) = \frac{n-1}{n} D_G, \quad (2)$$

где D_G – истинное значение дисперсии генеральной совокупности. Можно предложить другую оценку дисперсии – **исправленную дисперсию s^2** , вычисляемую по формуле

$$s^2 = \frac{n}{n-1} D_B = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n-1}. \quad (3)$$

Такая оценка будет являться несмещенной. Ей соответствует **исправленное среднее квадратическое отклонение**

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n-1}}. \quad (4)$$

Неравенство Крамера-Рао.

Дисперсия любой несмещенной оценки одного параметра θ удовлетворяет неравенству Рао-Крамера:

$$D(\theta^*) \geq -\frac{1}{nM\left(\frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2}\right)} \quad (16.7)$$

где $f(x, \theta)$ – плотность распределения вероятностей случайной величины;
 n – объем выборки.

Оценка θ^* , для которой в неравенстве Рао-Крамера достигается знак равенства, будет эффективной. В математической статистике применяются также асимптотически эффективные оценки, дисперсия которых стремится к нижней границе неравенства Рао-Крамера при $n \rightarrow \infty$.

Докажем с помощью неравенства (16.7), что выборочная средняя является эффективной оценкой для генеральной средней.

$$\sigma^2(\bar{X}) = \sigma^2\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \left(\frac{1}{n}\right)^2 \left(\sum_{i=1}^n \sigma_i^2\right) = \frac{1}{n^2} n \sigma_{\text{ген}}^2 = \frac{\sigma_{\text{ген}}^2}{n}.$$

$$\begin{aligned} M\left(\frac{\partial^2}{\partial a^2} \ln f(x, a, \sigma_{\text{ген}})\right) &= M\left(\frac{\partial^2}{\partial a^2} \ln \frac{1}{\sigma_{\text{ген}} \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma_{\text{ген}}^2}}\right) = \\ &= M\left(\frac{\partial^2}{\partial a^2} \ln \left[\frac{1}{\sigma_{\text{ген}} \sqrt{2\pi}} - \frac{(x-a)^2}{2\sigma_{\text{ген}}^2}\right]\right) = \frac{1}{\sigma_{\text{ген}}^2}. \end{aligned}$$

Подставляем найденное математическое ожидание в неравенство Рао-Крамера:

$$D(\bar{X}) \geq -\frac{1}{-n/\sigma_{\text{ген}}^2} = \frac{\sigma_{\text{ген}}^2}{n}$$

Таким образом, нижняя граница неравенства Рао-Крамера совпадает с дисперсией выборочной средней, что свидетельствует об эффективности выборочной средней.

Оценка неизвестной вероятности. Оценка математического ожидания.

При проведении экспериментов часто приходится оценивать неизвестную вероятность события P по его частоте P^* в N независимых экспериментах. Частота некоторого события в N независимых экспериментах есть не что иное, как среднее арифметическое наблюдаемых значений

величины X , которая в каждом отдельном опыте принимает значение 1 (если событие совершилось), или значение 0 (если событие не произошло):

$$P^* = \frac{\sum_{i=1}^N X_i}{N}.$$

Показано, что математическое ожидание величины X равно P , а ее дисперсия равна $P(1-P)$. Математическое ожидание выборочного среднего равно P :

$$M[P^*] = P,$$

т.е. оценка P^* является несмещенной. Дисперсия величины P^* равна:

$$D[P^*] = P(1-P)/N.$$

Специфика этой задачи в том, что X в данном случае - дискретная случайная величина только с двумя возможными значениями: 0 и 1. Сделаем ограничение практически всегда выполняемым - число экспериментов достаточно велико, так что выполняются условия:

$$N(1-P) > 4, NP > 4.$$

Если эти условия выполнены, то частоту P^* можно считать распределенной по гауссовскому закону. Тогда параметры этого закона:

$$m^* = P, \sigma^* = [(1-P)P]^{0.5}$$

Приведена методика оценки доверительного интервала, которую мы приведем далее. Границы интервала, в котором заключено истинное значение вероятности события, определяются следующим образом:

$$P_1 = P^* - t[(1-P^*)P^*/N]^{0.5},$$

$$P_2 = P^* + t[(1-P^*)P^*/N]^{0.5}.$$

Здесь P^* - конкретная оценка вероятности (частоты события), а t находится, исходя из заданной доверительной вероятности β :

$$t = F^{-1}[(1+\beta)/2].$$

Пример. Производится серия из N экспериментов с целью оценки вероятности некоторого события. В результате этой серии экспериментов получено значение $P^* = 0.34$. Построить доверительный интервал, в котором с вероятностью 0.85 вычисляется истинная вероятность рассматриваемого события.

По табл. 1 при $\beta = 0.85$ находим: $t = 1.439$. Умножая это значение на величину

$$\sqrt{\frac{P^*(1-P^*)}{N}} \approx 0.0335,$$

получим 0.048. Откуда искомый доверительный интервал $I \approx (0.292; 0.388)$.

Методы нахождения точечных оценок.

1. Метод наибольшего правдоподобия.

Пусть X – дискретная случайная величина, которая в результате n испытаний приняла значения $x_1, x_2, \dots, x_n$. Предположим, что нам известен закон распределения этой величины, определяемый параметром Θ , но неизвестно численное значение этого параметра. Найдем его точечную оценку.

Пусть $p(x_i, \Theta)$ – вероятность того, что в результате испытания величина X примет значение x_i . Назовем **функцией правдоподобия** дискретной случайной величины X функцию аргумента Θ , определяемую по формуле:

$$L(x_1, x_2, \dots, x_n; \Theta) = p(x_1, \Theta)p(x_2, \Theta) \dots p(x_n, \Theta).$$

Тогда в качестве точечной оценки параметра Θ принимают такое его значение $\Theta^* = \Theta(x_1, x_2, \dots, x_n)$, при котором функция правдоподобия достигает максимума. Оценку Θ^* называют **оценкой наибольшего правдоподобия**.

Поскольку функции L и $\ln L$ достигают максимума при одном и том же значении Θ , удобнее искать максимум $\ln L$ – **логарифмической функции правдоподобия**. Для этого нужно:

1.

найти производную $\frac{d \ln L}{d \Theta}$;

2.

приравнять ее нулю (получим так называемое *уравнение правдоподобия*) и найти критическую точку;

3.

найти вторую производную $\frac{d^2 \ln L}{d \Theta^2}$; если она отрицательна в критической точке, то это – точка максимума.

Достоинства метода наибольшего правдоподобия: полученные оценки состоятельны (хотя могут быть смещенными), распределены асимптотически нормально при больших значениях n и имеют наименьшую дисперсию по сравнению с другими асимптотически нормальными оценками; если для оцениваемого параметра Θ существует эффективная оценка Θ^* , то уравнение правдоподобия имеет единственное решение Θ^* ; метод наиболее полно использует данные выборки и поэтому особенно полезен в случае малых выборок.

Недостаток метода наибольшего правдоподобия: сложность вычислений. Для непрерывной случайной величины с известным видом плотности распределения $f(x)$ и неизвестным параметром Θ функция правдоподобия имеет

вид:

$$L(x_1, x_2, \dots, x_n; \Theta) = f(x_1, \Theta)f(x_2, \Theta) \dots f(x_n, \Theta).$$

Оценка наибольшего правдоподобия неизвестного параметра проводится так же, как для дискретной случайной величины.

2. **Метод моментов.**

Метод моментов основан на том, что начальные и центральные эмпирические моменты являются состоятельными оценками соответственно начальных и центральных теоретических моментов, поэтому можно приравнять теоретические моменты соответствующим эмпирическим моментам того же порядка.

Если задан вид плотности распределения $f(x, \Theta)$, определяемой одним неизвестным параметром Θ , то для оценки этого параметра достаточно иметь одно уравнение. Например, можно приравнять начальные моменты первого порядка:

$$\bar{x}_B = M(X) = \int_{-\infty}^{\infty} xf(x; \Theta)dx = \varphi(\Theta),$$

получив тем самым уравнение для определения Θ . Его решение Θ^* будет точечной оценкой параметра, которая является функцией от выборочного среднего и, следовательно, и от вариантов выборки:

$$\Theta = \psi(x_1, x_2, \dots, x_n).$$

Если известный вид плотности распределения $f(x, \Theta_1, \Theta_2)$ определяется двумя неизвестными параметрами Θ_1 и Θ_2 , то требуется составить два уравнения, например

$$v_1 = M_1, \quad \mu_2 = m_2.$$

Отсюда $\begin{cases} M(X) = \bar{x}_B \\ D(X) = D_B \end{cases}$ - система двух уравнений с двумя неизвестными Θ_1 и Θ_2 .
Ее решениями будут точечные оценки Θ_1^* и Θ_2^* - функции вариант выборки:

$$\Theta_1 = \psi_1(x_1, x_2, \dots, x_n),$$

$$\Theta_2 = \psi_2(x_1, x_2, \dots, x_n).$$

3. Метод наименьших квадратов.

Если требуется оценить зависимость величин y и x , причем известен вид связывающей их функции, но неизвестны значения входящих в нее коэффициентов, их величины можно оценить по имеющейся выборке с помощью метода наименьших квадратов. Для этого функция $y = \varphi(x)$ выбирается так, чтобы сумма квадратов отклонений наблюдаемых значений $y_1, y_2, \dots, y_n$ от $\varphi(x_i)$ была минимальной:

$$\sum_{i=1}^n (y_i - \varphi(x_i))^2 = \min.$$

При этом требуется найти стационарную точку функции $\varphi(x; a, b, c \dots)$, то есть решить систему:

$$\begin{cases} \sum_{i=1}^n (y_i - \varphi(x_i; a, b, c \dots)) \left(\frac{\partial \varphi}{\partial a} \right)_i = 0 \\ \sum_{i=1}^n (y_i - \varphi(x_i; a, b, c \dots)) \left(\frac{\partial \varphi}{\partial b} \right)_i = 0 \\ \sum_{i=1}^n (y_i - \varphi(x_i; a, b, c \dots)) \left(\frac{\partial \varphi}{\partial c} \right)_i = 0 \\ \dots \end{cases}$$

(решение, конечно, возможно только в случае, когда известен конкретный вид функции φ).

Рассмотрим в качестве примера подбор параметров линейной функции методом наименьших квадратов.

Для того, чтобы оценить параметры a и b в функции $y = ax + b$, найдем

$$\left(\frac{\partial \varphi}{\partial a}\right)_i = x_i; \left(\frac{\partial \varphi}{\partial b}\right)_i = 1. \quad \text{Тогда} \quad \begin{cases} \sum_{i=1}^n (y_i - (ax_i + b))x_i = 0 \\ \sum_{i=1}^n (y_i - (ax_i + b)) = 0 \end{cases} \quad \text{Отсюда}$$

$$\begin{cases} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - bn = 0 \end{cases} \quad \text{Разделив оба полученных уравнения на } n \text{ и}$$

вспомнив определения эмпирических моментов, можно получить выражения для a и b в виде:

$$a = \frac{(K_{xy})_B}{(D_x)_B}, \quad b = \bar{y}_B - \frac{(K_{xy})_B}{(D_x)_B} \bar{x}_B. \quad \text{Следовательно, связь между } x \text{ и } y \text{ можно задать в виде:}$$

$$y - \bar{y}_B = \frac{(K_{xy})_B}{(D_x)_B} (x - \bar{x}_B).$$

Интервальное оценивание. Доверительная вероятность и доверительный интервал.

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому в таком случае лучше пользоваться **интервальными оценками**, то есть указывать интервал, в который с заданной вероятностью попадает

истинное значение оцениваемого параметра. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра. Поэтому, если для оценки Θ^* некоторого параметра Θ справедливо неравенство $|\Theta^* - \Theta| < \delta$, число $\delta > 0$ характеризует **точность оценки** (чем меньше δ , тем точнее оценка). Но статистические методы позволяют говорить только о том, что это неравенство выполняется с некоторой вероятностью.

Надежностью (доверительной вероятностью) оценки Θ^* параметра Θ называется вероятность γ того, что выполняется неравенство $|\Theta^* - \Theta| < \delta$. Если заменить это неравенство двойным неравенством $-\delta < \Theta^* - \Theta < \delta$, то получим:

$$P(-\delta < \Theta^* - \Theta < \delta) = \gamma.$$

Таким образом, γ есть вероятность того, что Θ попадает в интервал $(\Theta^* - \delta, \Theta^* + \delta)$.

Доверительным называется интервал, в который попадает неизвестный параметр с заданной надежностью γ .

Построение доверительных интервалов для оценки математического ожидания нормального распределения.

Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии.

Пусть исследуемая случайная величина X распределена по нормальному закону с известным средним квадратическим σ , и требуется по значению выборочного среднего $\bar{x}_B$ оценить ее математическое ожидание a . Будем рассматривать выборочное среднее $\bar{x}_B$ как случайную величину $\bar{X}$, а значения вариант выборки $x_1, x_2, \dots, x_n$ как одинаково распределенные независимые случайные величины $X_1, X_2, \dots, X_n$, каждая из которых имеет математическое

ожидание a и среднее квадратическое отклонение σ . При этом $M(\bar{X}) = a$, $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ (используем свойства математического ожидания и дисперсии суммы независимых случайных величин). Оценим вероятность выполнения неравенства $|\bar{X} - a| < \delta$. Применим формулу для вероятности попадания нормально распределенной случайной величины в заданный интервал:

$$p(|\bar{X} - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma(\bar{X})}\right). \text{ Тогда, с учетом того, что } \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}, p(|\bar{X} - a| < \delta) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) =$$

$= 2\Phi(t)$, где $t = \frac{\delta\sqrt{n}}{\sigma}$. Отсюда $\delta = \frac{t\sigma}{\sqrt{n}}$, и предыдущее равенство можно переписать так:

$$p\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}} < a < \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma. \quad (1)$$

Итак, значение математического ожидания a с вероятностью (надежностью) γ

попадает в интервал $\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}}; \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right)$, где значение t определяется из таблиц для функции Лапласа так, чтобы выполнялось равенство $2\Phi(t) = \gamma$.

Пример. Найдём доверительный интервал для математического ожидания нормально распределенной случайной величины, если объем выборки $n = 49$, $\bar{x}_B = 2,8$, $\sigma = 1,4$, а доверительная вероятность $\gamma = 0,9$.

Определим t , при котором $\Phi(t) = 0,9:2 = 0,45$: $t = 1,645$. Тогда

$2,8 - \frac{1,645 \cdot 1,4}{\sqrt{49}} < a < 2,8 + \frac{1,645 \cdot 1,4}{\sqrt{14}}$, или $2,471 < a < 3,129$. Найден доверительный интервал, в который попадает a с надежностью 0,9.

Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии.

Если известно, что исследуемая случайная величина X распределена по нормальному закону с неизвестным средним квадратическим отклонением, то для поиска доверительного интервала для ее математического ожидания построим новую случайную величину

$$T = \frac{\bar{x}_B - a}{\frac{s}{\sqrt{n}}}, \quad (2)$$

где $\bar{x}_B$ - выборочное среднее, s – исправленная дисперсия, n – объем выборки. Эта случайная величина, возможные значения которой будем обозначать t , имеет распределение Стьюдента (см. лекцию 12) с $k = n - 1$ степенями свободы.

Поскольку плотность распределения Стьюдента $s(t, n) = B_n \left(1 + \frac{t^2}{n-1} \right)^{-\frac{n}{2}}$, где

$$B_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)},$$

явным образом не зависит от a и σ , можно задать

вероятность ее попадания в некоторый интервал $(-t_\gamma, t_\gamma)$, учитывая четность

плотности распределения, следующим образом:

$$p\left(\left|\frac{\bar{x}_B - a}{\frac{s}{\sqrt{n}}}\right| < t_\gamma\right) = 2 \int_0^{t_\gamma} s(t, n) dt = \gamma.$$

Отсюда получаем:

$$p\left(\bar{x}_B - \frac{t_\gamma s}{\sqrt{n}} < a < \bar{x}_B + \frac{t_\gamma s}{\sqrt{n}}\right) = \gamma. \quad (3)$$

Таким образом, получен доверительный интервал для a , где t_γ можно найти по соответствующей таблице при заданных n и γ .

Пример. Пусть объем выборки $n = 25$, $\bar{x}_B = 3$, $s = 1,5$. Найдём доверительный интервал для a при $\gamma = 0,99$. Из таблицы находим, что t_γ ($n = 25$, $\gamma = 0,99$) = 2,797. Тогда $3 - \frac{2,797 \cdot 1,5}{\sqrt{25}} < a < 3 + \frac{2,797 \cdot 1,5}{\sqrt{25}}$, или $2,161 < a < 3,839$ – доверительный интервал, в который попадает a с вероятностью 0,99.

Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.

Будем искать для среднего квадратического отклонения нормально распределенной случайной величины доверительный интервал вида $(s - \delta, s + \delta)$, где s – исправленное выборочное среднее квадратическое отклонение, а для δ выполняется условие: $p(|\sigma - s| < \delta) = \gamma$.

Запишем это неравенство в виде: $s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right)$ или, обозначив $q = \frac{\delta}{s}$,

$$s(1 - q) < \sigma < s(1 + q). \quad (1)$$

Рассмотрим случайную величину χ , определяемую по формуле

$$\chi = \frac{s}{\sigma} \sqrt{n-1},$$

которая распределена по закону «хи-квадрат» с $n-1$ степенями свободы.

Плотность ее распределения

$$R(\chi, n) = \frac{\chi^{n-2} e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right)}$$

не зависит от оцениваемого параметра σ , а зависит только от объема выборки n . Преобразуем неравенство (1) так, чтобы оно приняло вид $\chi_1 < \chi < \chi_2$. Вероятность выполнения этого неравенства равна доверительной вероятности

γ , следовательно, $\int_{\chi_1}^{\chi_2} R(\chi, n) d\chi = \gamma$. Предположим, что $q < 1$, тогда неравенство (1) можно записать так:

$$\frac{1}{s(1+q)} < \frac{1}{\sigma} < \frac{1}{s(1-q)},$$

или, после умножения на $s\sqrt{n-1}$, $\frac{\sqrt{n-1}}{1+q} < \frac{s\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q}$. Следовательно,

$\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}$. Тогда $\int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} R(\chi, n) d\chi = \gamma$. Существуют таблицы для

распределения «хи-квадрат», из которых можно найти q по заданным n и γ , не решая этого уравнения. Таким образом, вычислив по выборке значение s и определив по таблице значение q , можно найти доверительный интервал (1), в который значение σ попадает с заданной вероятностью γ .

Замечание. Если $q > 1$, то с учетом условия $\sigma > 0$ доверительный интервал для σ будет иметь границы

$$0 < \sigma < s(1+q). \quad (2)$$

Пример.

Пусть $n = 20$, $s = 1,3$. Найдем доверительный интервал для σ при заданной надежности $\gamma = 0,95$. Из соответствующей таблицы находим $q(n = 20, \gamma = 0,95) = 0,37$. Следовательно, границы доверительного интервала: $1,3(1-0,37) = 0,819$ и $1,3(1+0,37) = 1,781$. Итак, $0,819 < \sigma < 1,781$ с вероятностью 0,95.

Тема 2.2. Точечные оценки параметров генеральной совокупности.

Эмпирические моменты.

Обычным эмпирическим моментом порядка k называют среднее значение k -х степеней разностей $x_i - C$

$$M'_k = \frac{\sum n_i (x_i - C)^k}{n},$$

где x_i — наблюдаемая варианта,

n_i — частота варианты,

$n = \sum n_i$ — объем выборки,

C — произвольное постоянное число (ложный нуль).

Начальным эмпирическим моментом порядка k называется обычный эмпирический момент порядка k при $C = 0$

$$M'_k = \frac{\sum n_i x_i^k}{n}.$$

В частности

$$M'_1 = \frac{\sum n_i x_i}{n} = \bar{x}_s.$$

Центральным эмпирическим моментом порядка k называется обычный момент порядка k при $C = \bar{x}_s$

$$m_k = \frac{\sum n_i (x_i - \bar{x}_s)^k}{n}.$$

В частности

$$m_2 = \frac{\sum n_i (x_i - \bar{x}_s)^2}{n} = D_s. \quad (1)$$

Легко выразить центральные моменты через обычные:

$$m_2 = M'_2 - (M'_1)^2; \quad (\text{IV.2})$$

$$m_3 = M'_3 - 3M'_2 M'_1 + 2(M'_1)^3;$$

$$m_4 = M'_4 - 4M'_3 M'_1 + 6M'_2 (M'_1)^2 - 3(M'_1)^4. \quad (3)$$

Условным эмпирическим моментом порядка k называется начальный момент порядка k , вычисленный для условных вариантов

$$M_k^* = \frac{\sum n_i u_i^k}{n} = \frac{\sum n_i \left(\frac{x_i - C}{h}\right)^k}{n}.$$

В частности

$$\begin{aligned} M_1^* &= \frac{\sum n_i u_i}{n} = \frac{1}{h} \cdot \left(\frac{\sum n_i x_i}{n} - C \frac{\sum n_i}{n} \right) = \\ &= \frac{1}{h} \cdot (x_i - C). \end{aligned}$$

Откуда: $\bar{x}_s = M_1^* h + C$.

Выразим обычные моменты через условные:

$$M_k^* = \frac{1}{h^k} \cdot \frac{\sum n_i (x_i - C)^k}{n} = \frac{M'_k}{h^k}.$$

Отсюда: $M'_k = M_k^* h^k$.

Найдя **обычные моменты**, легко найти **центральные моменты** по соотношениям (2) и (3), тогда будем иметь удобные формулы для вычислений

$$m_2 = \left[M_2^* - (M_1^*)^2 \right] \cdot h^2; \quad (4)$$

$$m_3 = \left[M_3^* - 3 M_2^* M_1^* + 2 (M_1^*)^3 \right] \cdot h^3;$$

$$m_4 = \left[M_4^* - 4 M_3^* M_1^* + 6 M_2^* (M_1^*)^2 - 3 (M_1^*)^4 \right] \cdot h^4. \quad (5)$$

Метод произведений для вычисления выборочной средней и дисперсии.

Метод произведений – это удобная техника вычислений условных моментов различных порядков вариационного ряда с равноотстоящими вариантами. Зная условные моменты, нетрудно найти начальные и центральные эмпирические моменты. В частности, методом произведений удобно вычислять выборочную среднюю и выборочную дисперсии. При этом рекомендуем пользоваться расчетной таблицей, которая составляется следующим образом:

- в первый столбец таблицы записывают выборочные (первоначальные) варианты, располагая их в возрастающем порядке;
- во второй столбец записывают частоты вариантов; складывают все частоты и их сумму (объем выборки n) помещают в нижнюю клетку столбца;

$$u_i = \frac{x_i - C}{h},$$

- в третий столбец записывают условные варианты
- причем в качестве ложного нуля C выбирают варианту с наибольшей

частотой и полагают h равным разности между любыми двумя соседними вариантами;

- умножают частоты на условные варианты и записывают их произведения $n_i u_i$ в четвертый столбец; сложив все полученные числа, их сумму $\sum n_i u_i$ помещают в нижнюю клетку столбца;

- умножают частоты на квадраты условных вариантов и записывают их выражения $n_i u_i^2$ в пятый столбец; сложив все полученные числа, их сумму $\sum n_i u_i^2$, помещают в нижнюю клетку столбца;

- умножают частоты на квадраты условных вариантов, увеличенных каждая на единицу, и записывают произведения $n_i (u_i + 1)^2$ в шестой контрольный столбец; сложив все полученные числа, их сумму $\sum n_i (u_i + 1)^2$ помещают в нижнюю клетку столбца.

Далее вычисляются условные моменты:

$$M_1^* = \frac{\sum n_i u_i}{n}; \quad M_2^* = \frac{\sum n_i u_i^2}{n}.$$

И, наконец, вычисляются:

выборочная средняя $\bar{x}_s = M_1^* h + C;$

и выборочная дисперсия $D_s = \left[M_2^* - (M_1^*)^2 \right] \cdot h^2.$

Сведение первоначальных вариантов к равноостоящим.

Интервал, в котором заключены все наблюдаемые значения признака (первоначальные варианты), делят на несколько равных частичных интервалов. (Практически в каждый частичный интервал должно попасть не

менее 8-10 первоначальных вариантов.) Затем находятся середины частичных интервалов, которые и образуют последовательность равноотстоящих вариантов.

В качестве частоты «новой» варианты (середины частичного интервала) принимают общее число первоначальных вариантов, попавших в соответствующий частичный интервал.

Ясно, что замена первоначальных вариантов серединами частичных интервалов сопровождается ошибками (первоначальные варианты левой половины частичного интервала будут увеличены, а варианты правой половины уменьшены), однако эти ошибки будут в основном погашаться, поскольку они имеют разные знаки.

x_i	n_i	x_i	n_i		f	x	f	x	
1						1			
,00		1		1		,32		1	
1		,12		,23		1		,44	
,03	1	1		1		,33	4	1	
1	3	,15		,25	4	1	5	,45	3
,05	6	1		1	8	,37	6	1	3
1	4	,16		,26	4	1	2	,46	2
,06	2	1		1	4	,38	1	1	4
1	4	,19		,29	6	1	2	,49	2
,08		1		1		,39		1	
1		,20		,30		1		,50	
,10						,40			
1	1	1	1	1					
,00	-,10	-,20	-,30	-,40	-				
1,10	1,20	1,30	1,40	1,50					

Приняв середины частичных интервалов в качестве новых вариантов y_i , получим равноотстоящие варианты: $y_1 = 1,05; y_2 = 1,15; y_3 = 1,25;$

$y_4 = 1,35; y_5 = 1,45$. Найдем частоту варианты y_i :

$$n_1 = 1 + 3 + 6 + 4 + 2 + \frac{4}{2} = 18$$
. (Поскольку первоначальная варианта 1,10 одновременно является концом первого частичного интервала и началом второго частичного интервала, частота 4 этой варианты поровну распределяется между обоими частичными интервалами.) Аналогично вычисляются частоты остальных вариантов. Таким образом, для полученных новых равноотстоящих вариантов будем иметь следующий статистический ряд:

y_i	1,05	1,15	1,25	1,35	1,45
n_i	18	20	25	22	15

Эмпирические и выравнивающие частоты.

Дискретное распределение.

Рассмотрим дискретную случайную величину X , закон распределения которой неизвестен. Пусть произведено n испытаний, в которых величина X приняла n_1 раз значение x_1 , n_2 раза - значение $x_2, \dots, n_k$ раз - значение x_k , причем $\sum n_i = n$.

Эмпирическими частотами называют фактически наблюдаемые частоты n_i .

Предположим, что у нас имеются основания предположить, что изучаемая величина X распределена по некоторому определенному закону. Для того, чтобы проверить, согласуется ли это предположение с данными наблюдений, вычисляют частоты наблюдаемых значений, то есть находят

теоретически сколько раз величина X должна была принять каждое из наблюдаемых значений, если она распределена по наблюдаемому закону.

Выравнивающими (теоретическими), в отличие от фактически наблюдаемых эмпирических частот, называют частоты n'_i , найденные теоретически (вычислениями). Их находят по соотношению

$$n'_i = n \cdot P_i,$$

где: n — число испытаний;

P_i — вероятность наблюдаемого значения x_i , вычисленная при допущении, что X имеет предполагаемое распределение.

Непрерывное распределение.

В случае непрерывного распределения, вероятности отдельных возможных значений равны нулю. Поэтому весь интервал возможных значений делят на k непересекающихся интервалов и вычисляют вероятности P_i попадания X в i — й частичный интервал, а затем, как и для дискретного распределения, умножают число испытаний на эти вероятности.

Итак, выравнивающие частоты непрерывного распределения находят по соотношению:

$$n'_i = n \cdot P_i,$$

где: n — число испытаний;

P_i — вероятность наблюдаемого значения x_i , вычисленная при допущении, что X имеет предполагаемое распределение.

В частности, если имеются основания предположить, что случайная величина X (генеральная совокупность) распределена нормально, то выравнивающие частоты могут быть найдены по формуле

$$n'_i = \frac{nh}{\sigma_\varepsilon} \varphi(u_i), \quad (\text{IV.6})$$

где: n — число испытаний (объем выборки);

h — длина частичного интервала;

σ_ε — выборочное среднее квадратическое отклонение;

$$u_i = \frac{x_i - \bar{x}_\varepsilon}{\sigma_\varepsilon} \quad (x_i - \text{середина } i\text{-го частичного интервала}),$$

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

Замечание: Как известно, дифференциальная функция (функция плотности распределения вероятностей) общего нормального распределения имеет следующий вид

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (\text{IV.7})$$

При $a = 0$ и $\sigma = 1$ получим дифференциальную функцию нормированного распределения

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

или, заменив обозначение аргумента

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

Далее, положив, $u = \frac{x-a}{\sigma}$, имеем

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (\text{IV.8})$$

Сравнивая (IV.7) и (IV.8), можно сделать заключение, что

$$p_X(x) = \frac{1}{\sigma} \varphi(u).$$

Если математическое ожидание α и среднее квадратическое отклонение σ неизвестны, то в качестве оценок этих параметров принимают соответственно выборочную среднюю $\bar{x}_s$ и выборочное среднее квадратическое отклонение σ_s . Тогда

$$p_X(x) = \frac{1}{\sigma_s} \varphi(u), \quad \text{где} \quad u_i = \frac{x_i - \bar{x}_s}{\sigma_s}.$$

Пусть x_i — середина i — го частичного интервала (на которые разбита совокупность всех наблюдаемых значений нормально распределенной случайной величины X) длиной h . Тогда вероятность попадания X в этот интервал приблизительно равна произведению длины интервала на значение дифференциальной функции $p_X(x)$ в любой точке интервала и, в частности, при $x = x_i$

$$P_i = h p_X(x) = h \cdot \frac{1}{\sigma_s} \varphi(u_i).$$

Следовательно, выравнивающая частота

$$n'_i = n P_i = \frac{nh}{\sigma_s} \cdot \varphi(u_i), \quad \text{где} \quad u_i = \frac{x_i - \bar{x}_s}{\sigma_s}.$$

Построение нормальной кривой распределения по опытным данным.

Один из способов построения нормальной кривой по опытным данным наблюдений (либо экспериментов) заключается в следующем:

* находят $\bar{x}_s$ и σ_s , например, по методу произведений;

* находят ординаты y_i (выравнивающие частоты) теоретической

$$y_i = \frac{nh}{\sigma_{\varepsilon}} \cdot \varphi(u_i),$$

кривой по формуле где n – сумма наблюдаемых частот,

h – разность между двумя соседними вариантами

$$u_i = \frac{x_i - \bar{x}_{\varepsilon}}{\sigma_{\varepsilon}} \quad \text{и} \quad \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

* строят точки (x_i, y_i) в прямоугольной системе координат и соединяют их плавной линией.

Близость выравнивающих частот к наблюдаемым подтверждает правильность допущения о том, что обследуемый признак распределен нормально.

В качестве иллюстрации, рассмотрим следующий пример.

Пример 8: Построим нормальную кривую по данным

варианта x_i	5	10	15	20	25	30	35	40	45	50
частота n_i	6	13	8	3	4	7	1	8	3	1

Используя метод произведений, нетрудно получить:

$$\bar{x}_{\varepsilon} = 34,7; \quad \sigma_{\varepsilon} = 7,38.$$

Расчеты выравнивающих частот осуществлены в таблице:

x	n_i	$x_i - \bar{x}_{\varepsilon}$	$u_i = \frac{x_i - \bar{x}_{\varepsilon}}{\sigma_{\varepsilon}}$	$\varphi(u_i)$	$y_i = \frac{nh}{\sigma_{\varepsilon}} \cdot \varphi(u_i)$
5	6	-19,7	-2,67	0,01	3

0	2	13	-14,7	-1,99	51	0,05	14
5	2	38	-9,7	-1,31	91	0,16	42
0	3	74	-4,7	-0,63	71	0,32	82
5	3	106	0,3	0,05	84	0,39	99
0	4	85	5,3	0,73	56	0,30	76
5	4	30	10,3	1,41	76	0,14	37
0	5	10	15,3	2,09	49	0,04	11
5	5	4	20,3	2,77	86	0,00	2
		$n = 36$					$\sum y_i = 366$

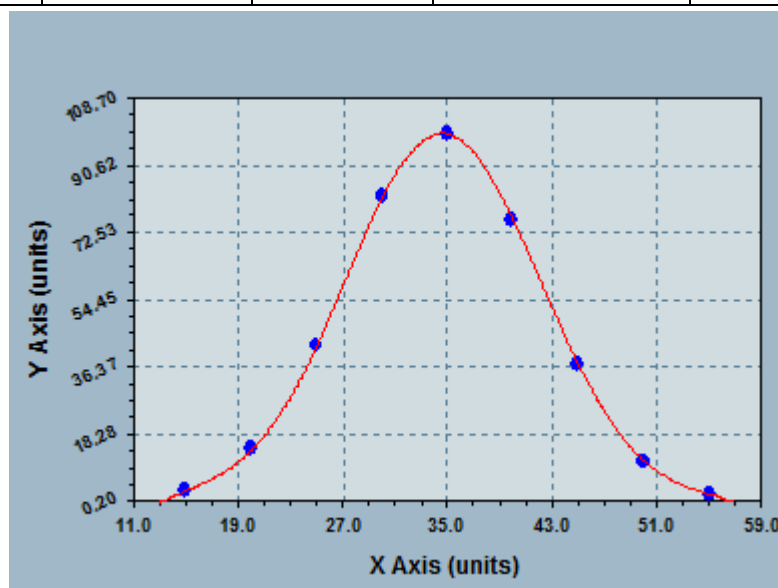


Рис. 4.

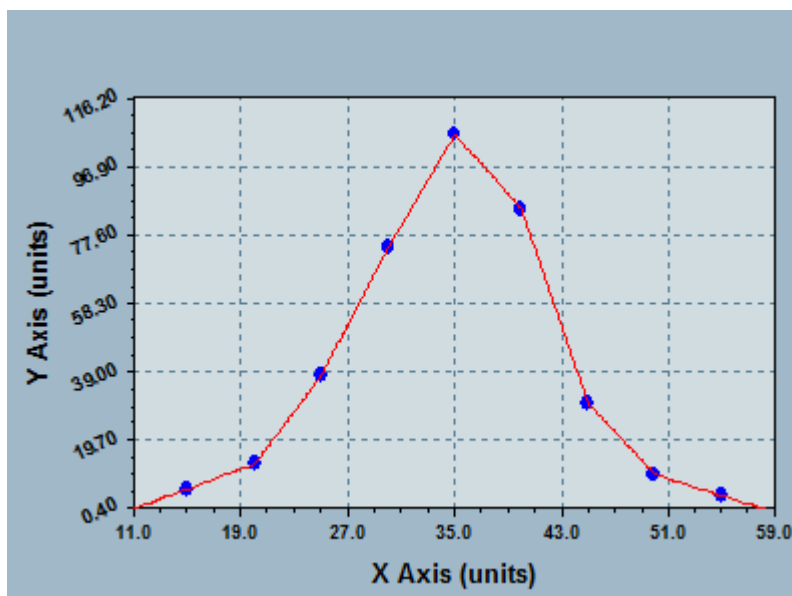


Рис. 5.

На рис. 4 построена нормальная (теоретическая) кривая по выравнивающим частотам (см. рис. 4) и полигон наблюдаемых частот (они изображены на рис. 5).

Сравнение графиков визуально показывает, что построенная теоретическая кривая удовлетворительно отражает данные наблюдений.

Для того, чтобы более уверенно считать, что данные наблюдений свидетельствуют о нормальном распределении признака, необходимо использовать специальные правила (их называют **критериями согласия**).

Оценка отклонений эмпирического распределения от нормального. Асимметрия и эксцесс.

Для оценки отклонения эмпирического распределения от нормального используются различные характеристики, к числу которых относятся *асимметрия* и *эксцесс*. Их определения аналогичны определениям асимметрии и эксцесса теоретических распределений.

Асимметрия эмпирического распределения определяется соотношением:

$$A_s = \frac{m_3}{\sigma_s^3} \quad (\text{здесь } m_3 - \text{центральный эмпирический момент третьего порядка}).$$

$$E_k = \frac{m_4}{\sigma_s^4} - 3$$

Экссесс эмпирического распределения определяется соотношением:

(здесь m_4 – центральный эмпирический момент четвертого порядка).

Указанные в соответствующих формулах моменты удобно вычислять методом произведений.

Пример 9: Найдем асимметрию и эксцесс эмпирического распределения:

вар	1	1	1	1	1	1	1	1	1	1	1
ианта x_i	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	1,8	2,0	
час					1	2	2	1	1		
тота n_i	2	3	8	3	5	0	2	0	6	1	

Последний столбец, приведенной ниже таблицы расчетов, служит для контроля вычислений по тождеству:

$$\sum n_i (u_i + 1)^4 = \sum n_i u_i^4 + 4 \cdot \sum n_i u_i^3 + 6 \cdot \sum n_i u_i^2 + 4 \cdot \sum n_i u_i + n$$

Расчеты приведены в таблице:

	n_i		$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	$n_i (u_i + 1)^4$
0,2	2	4	-8	32	128	512	162
0,4	3	3	-9	27	-81	243	48
0,6	8	2	-16	32	-64	128	8

0,8	1	13	1	-13	13	-13	13	-
1,0	1	25		$A_1 =$		-		25
1,2	1	20		20	20	20	20	320
1,4	1	12		24	48	96	192	972
1,6	1	10		30	90	270	810	2560
1,8	1	6		24	96	384	1536	3750
2,0	1	1		5	25	125	625	1296
				$A_2 =$		895		
		$n =$		$\sum n$	$\sum n$	$\sum n$	$\sum n$	$\sum n_i(u_i + 1)^4$
				$= 5$	$=$	$=$	$= 40$	$= 9141$

Контроль:

$$\sum n_i(u_i + 1)^4 = 9141;$$

$$\begin{aligned} \sum n_i(u_i + 1)^4 &= \sum n_i u_i^4 + 4 \cdot \sum n_i u_i^3 + 6 \cdot \sum n_i u_i^2 + 4 \cdot \sum n_i u_i + n \\ &= 4079 + 4 \cdot 609 + 6 \cdot 383 + 4 \cdot 57 + 100 = 9141. \end{aligned}$$

Совпадение сумм свидетельствует о том, что вычисления произведены верно.

Ранее было найдено: $M_1^* = 0,57$; $M_2^* = 3,83$; $D_s = 0,14$,

следовательно $\sigma_s = \sqrt{0,14}$.

Далее находим условные моменты третьего и четвертого порядков:

$$M_3^* = \frac{\sum n_i u_i^3}{n} = \frac{609}{100} = 6,09;$$

$$M_4^* = \frac{\sum n_i u_i^4}{n} = \frac{4079}{100} = 40,79.$$

Находим центральные эмпирические моменты третьего и четвертого порядков:

$$\begin{aligned} m_3 &= \left[M_3^* - 3M_1^* M_2^* + 2(M_1^*)^3 \right] \cdot h^3 = \\ &= \left[6,09 - 3 \cdot 0,57 \cdot 3,83 + 2 \cdot (0,57)^3 \right] \cdot (0,2)^3 = \\ &= -0,0007; \end{aligned}$$

$$\begin{aligned} m_4 &= \left[M_4^* - 4 \cdot M_1^* M_3^* + 6 \cdot (M_1^*)^2 M_2^* - 3 \cdot (M_1^*)^4 \right] \cdot h^4 = \\ &= \left[40,79 - 4 \cdot 0,57 \cdot 6,09 + 6 \cdot (0,57)^2 \cdot 3,83 - 3 \cdot (0,57)^4 \right] \cdot (0,2)^4 = \\ &= 0,054. \end{aligned}$$

$$A_s = \frac{m_3}{\sigma_s^3} = -\frac{0,0007}{(\sqrt{0,14})^3} = -0,01$$

И, наконец, находим асимметрию:

$$E_k = \frac{m_4}{\sigma_s^4} - 3 = \frac{0,054}{(\sqrt{0,14})^4} - 3 = -0,24.$$

и эксцесс:

Замечание: В случае малых выборок для оценок асимметрии и эксцесса необходимо находить дополнительно и точность этих оценок.

Тема 2.3. Методы расчета сводных характеристик выборки.

Функциональная и статистическая зависимость.

По характеру зависимости признаков различают функциональную (полную) связь и корреляционную (статистическую, неполную) связь.

Функциональная зависимость (жестко детерминированная $y=2x$) – величине факторного признака строго соответствует одно или несколько значений результативного признака.

Статистическая – определенному значению факторного признака соответствует лишь среднее значение результативного признака. Стат. Связь не имеет ограничений и условий как в функциональной. Корреляционная связь явл. Частным случаем стат. Связи, состоит в том, что разным значениям одной переменной соответствуют различные средние значения др.

Зависимость между случайными величинами может иметь функциональный характер, т.е. быть строгим функциональным отношением, связывающим их значения. Однако при обработке экспериментальных данных гораздо чаще встречаются зависимости другого рода: статистические зависимости. Различие между двумя видами зависимостей состоит в том, что функциональная зависимость устанавливает строгую взаимосвязь между переменными, а статистическая зависимость лишь говорит о том, что распределение случайной величины Y зависит от того, какое значение принимает случайная величина X .

Условное среднее. Выборочные уравнения регрессии.

При рассмотрении многомерных случайных величин рассматривались условные законы распределения и их числовые характеристики: математическое ожидание, дисперсия и различные моменты. Оценками этих величин служат их выборочные аналоги. Наиболее важными являются условные математические ожидания, вычисленные по выборке – **условные средние**.

Условное среднее $\bar{y}_x$ – среднее арифметическое значений случайной величины Y , наблюдавшихся при фиксированном значении случайной величины $X=x$.

Условное среднее $\bar{X}_y$ – среднее арифметическое значений случайной величины X , наблюдавшихся при фиксированном значении случайной величины $Y = y$.

Напомним определение уравнения регрессии:

$$M\left(\frac{Y}{x}\right) = f(x)$$

условное математическое ожидание $M\left(\frac{Y}{x}\right)$ является функцией x .

Эта функция $f(x)$ называется **функцией регрессии** Y на X , а ее график – **линией регрессии**.

Выборочный аналог этого уравнения, $\bar{y}_x = f^*(x)$, называется **выборочным уравнением регрессии** Y на X , функция $f^*(x)$ – **выборочной функцией регрессии** Y на X , ее график – **выборочной линией регрессии** Y на X .

Аналогично определяются выборочные характеристики и для регрессии X на Y .

Корреляционная таблица.

Пусть в результате эксперимента для системы (X, Y) получена выборка значений $x_1, x_2, \dots, x_n$.

Если значения x и y повторяются, то их группируют

$$(x_i, y_j, n_{ij}), i = 1, 2, \dots, l$$

$$j = 1, 2, 3, \dots, k$$

$$\sum_{i,j} n_{ij} = n$$

Здесь x_i и y_j – наблюдаемые значения X и Y , а n_{ij} – частота появления пары значений (x_i, y_j) .

Чаще всего в этом случае данные организуют в виде **корреляционной таблицы**:

Y	X	x_1	x_2	...	x_i	m_j
	y_1	n_{11}	n_{21}	..	n_{i1}	$m_1 = \sum_i n_{i1}$
	y_2	n_{12}	n_{22}	..	n_{i2}	$m_2 = \sum_i n_{i2}$
	...	...	...	..		
	y_k	n_{1k}	n_{2k}	..	n_{ik}	$m_k = \sum_i n_{ik}$
	n_x	$n_1 = \sum_j n_{1j}$	$n_2 = \sum_j n_{2j}$	..	$n_i = \sum_j n_{ij}$	$n = \sum_i n_i = \sum_j n_j$

Группируя данные по значениям x_i или y_j :

$$\sum_{j=1}^k n_{i,j}; i = 1, 2, \dots, l; (n_x)$$

$$\sum_{i=1}^l m_{i,j}; j = 1, 2, \dots, k; (m_y)$$

$$\sum_{i=1}^l n_i = \sum_{j=1}^k m_j = n$$

по данным корреляционной таблицы можно составить законы распределения составляющих (последняя строка и последний столбец таблицы) и их средние по выборки $\overline{X_s}$ и $\overline{Y_s}$.

$$\overline{X_s} = \frac{1}{n} \sum_i n_i x_i \quad \text{и} \quad \overline{Y_s} = \frac{1}{n} \sum_j m_j y_j.$$

Для наглядности данные таблицы изображают графически. Каждую пару (x_i, y_j) изображают точкой в системе координат (XOY) . Частоту n_{ij} , с которой данная пара встречается в таблице, изображают соответствующим числом близко расположенных точек либо пишут число n_{ij} возле одной точки. Построенное таким образом в системе координат изображение

корреляционной таблицы называют **полем корреляции**. Также возможно изображать данные таблицы кругами, центр которых находится в точке (x_i, y_j) , а диаметр (или площадь) пропорционален m_{ij} .

Точка в системе координат (XOY) с координатами $(\bar{X}, \bar{Y})$ называется **центром рассеивания**.

Можно также составить условные законы распределения, например Y при $X = x_j$ или X при $Y = y_i$.

$\frac{Y}{X = x_j}$	y_1	y_2		y_i
m	m_{1j}	m_{2j}		m_{ij}

Зная условные законы распределения, можно найти условные средние:

$\frac{\bar{Y}}{X = x_1}, \frac{\bar{Y}}{X = x_2}, \dots, \frac{\bar{Y}}{X = x_j}$ и т.п. Построим в системе координат (XOY) точки

$(x_j, \frac{\bar{Y}}{X = x_j})$ и соединим их отрезками прямых. Полученную ломаную называют

выборочной линией регрессии Y на X .

Если распределения случайных величин X и (или) Y заданы **интервальным** вариационным рядом, то удобно перейти к вспомогательным переменным, значения которых совпадают с серединами интервалов.

Кроме того, если **варианты** (значения вариационного ряда) являются равноотстоящими, т.е., образуют арифметическую прогрессию с разностью h , бывает удобно перейти к **условным вариантам**:

$$u_i = \frac{x_i - C}{h},$$

где C **ложный нуль** (новое начало отсчета),

h – **шаг**, т.е. разность между двумя соседними первоначальными вариантами (новая единица масштаба).

Если в качестве ложного нуля взята какая-то из вариант x_m , то условные варианты – целые числа, что упрощает вычисления

$$u_i = \frac{x_i + (i-1)h - [x_1 + (m-1)h]}{h} = i - m$$

Отыскание параметров выборочного уравнения прямой линии среднеквадратичной регрессии по несгруппированным данным.

Допустим, что количественные признаки X и Y связаны линейной корреляционной зависимостью. В этом случае обе линии регрессии будут прямыми.

Предположим, что для отыскания уравнений этих прямых проведено n независимых испытаний, в результате которых получены n пар чисел:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

Поскольку наблюдаемые пары чисел можно рассматривать как случайную выборку из генеральной совокупности всех возможных значений случайной величины (X, Y) , то величины и уравнения, найденные по этим данным, называют *выборочными*.

Для определенности будем искать выборочное уравнение прямой линии регрессии Y на X .

Рассмотрим простейший случай: различные значения x признака X и соответствующие им значения y признака Y наблюдались по одному разу. Очевидно, что группировать данные нет необходимости. Также нет надобности использовать понятие условной средней, поэтому искомое уравнение

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

Угловым коэффициентом прямой линии регрессии Y на X принято называть *выборочным коэффициентом регрессии* Y на X и обозначать через ρ_{yx}

$$\overline{y}_x = \rho_{yx}x + b$$

Выборочное корреляционное отношение и его свойства.

Выборочный коэффициент корреляции вводится для оценки тесноты линейной корреляционной связи. Для оценки тесноты любой (в том числе и нелинейной) корреляционной связи вводятся новые характеристики:

η_{yx} - выборочное корреляционное отношение Y к X ;

η_{xy} - выборочное корреляционное отношение X к Y .

Выборочным корреляционным отношением Y к X называют отношение межгруппового среднего квадратичного отклонения к общему среднему квадратичному отклонению признака Y :

$$\eta_{yx} = \frac{\sigma_{\text{межгр}}}{\sigma_{\text{общ}}} , (4.17)$$

где

$$\sigma_{\text{межгр}} = \sqrt{D_{\text{межгр}}} = \sqrt{\frac{\sum n_x (\overline{y}_x - \overline{y})^2}{n}} ;$$

$$\sigma_{\text{общ}} = \sqrt{D_{\text{общ}}} = \sqrt{\frac{\sum n_y (y - \overline{y})^2}{n}} ;$$

n – объем выборки; n_x – число появлений (частота) значения x признака X ; n_y - частота значения y признака Y ; $\overline{y}$ - общая средняя признака Y ; $\overline{y}_x$ - условная средняя признака Y .

Выборочное корреляционное отношение X к Y определяется аналогично. Чтобы получить определение , нужно в определении везде поменять местами x и y , а также X и Y .

Пример. Найти по данным наблюдений, сведенным в корреляционную таблицу.

X Y	3	5	7	9	n_y
2	1	7	8	3	19
4	7	3	1	6	17
6	2	0	1	1	4
n_x	10	10	10	10	n = 40
	4,2	2,6	2,6	3,6	

Решение. Условные средние можно понимать как групповые средние.

Групповая средняя группы, соответствующей $x = 3$, равна

$$\bar{y}_3 = \frac{1 \cdot 2 + 7 \cdot 4 + 2 \cdot 6}{10} = 4,2;$$

Другие групповые средние равны:

$$\bar{y}_5 = \frac{7 \cdot 2 + 3 \cdot 4 + 0 \cdot 6}{10} = 2,6;$$

$$\bar{y}_7 = \frac{8 \cdot 2 + 1 \cdot 4 + 1 \cdot 6}{10} = 2,6;$$

$$\bar{y}_9 = \frac{3 \cdot 2 + 6 \cdot 4 + 1 \cdot 6}{10} = 3,6.$$

Общая средняя равна

$$\bar{y} = \frac{\sum n_y y}{n} = \frac{19 \cdot 2 + 17 \cdot 4 + 4 \cdot 6}{40} = 3,25.$$

Межгрупповое среднее квадратичное отклонение равно

$$\begin{aligned} \sigma_{\text{межгр}} &= \sqrt{\frac{\sum n_x (\bar{y}_x - \bar{y})^2}{n}} = \\ &= \sqrt{\frac{10(4,2 - 3,25)^2 + 10(2,6 - 3,25)^2 + 10(2,6 - 3,25)^2 + 10(3,6 - 3,25)^2}{40}} = 0,68. \end{aligned}$$

Общее среднее квадратичное отклонение равно

$$\sigma_{\text{общ}} = \sqrt{\frac{\sum n_y (y - \bar{y})^2}{n}} =$$

$$= \sqrt{\frac{19(2 - 3,25)^2 + 17(4 - 3,25)^2 + 4(6 - 3,25)^2}{40}} = 1,32.$$

Следовательно, выборочное корреляционное отношение равно

$$\eta_{yx} = \frac{\sigma_{\text{межгр}}}{\sigma_{\text{общ}}} = \frac{0,68}{1,32} = 0,515.$$

Свойства выборочного корреляционного отношения:

Выборочные корреляционные отношения и обладают одинаковыми свойствами. Поэтому далее перечислены без доказательств свойства выборочного корреляционного отношения (для упрощения записи оно обозначено η).

1. Выборочное корреляционное отношение удовлетворяет двойному неравенству

$$0 \leq \eta \leq 1.$$

2. Если $\eta = 0$, то признак Y не связан с признаком X корреляционной зависимостью.

3. Если $\eta = 1$, то признак Y связан с признаком X функциональной зависимостью. Справедливо и обратное утверждение: если признак Y связан с признаком X функциональной зависимостью, то $\eta = 1$.

При возрастании значения признака Y внутри всех групп, соответствующих различным наблюдаемым значениям x признака X , будут все меньше различаться между собой, корреляционная связь между Y и X будет становиться все теснее и при $\eta = 1$ превратится в функциональную.

4. Выборочное корреляционное отношение не меньше абсолютной величины выборочного коэффициента корреляции:

$$\eta \geq |r_s|.$$

5. Если выборочное корреляционное отношение равно абсолютной величине выборочного коэффициента корреляции, то между признаками Y и X имеет место точная линейная корреляционная зависимость. Верно и

обратное утверждение: если между признаками Y и X имеет место точная линейная корреляционная зависимость, то $\eta = |r_a|$.

Достоинством выборочного корреляционного отношения является то, что оно служит мерой тесноты связи любой формы, в том числе и нелинейной. Таким образом, является более общей характеристикой связи, чем выборочный коэффициент корреляции, который служит мерой тесноты лишь линейной связи. Недостатком выборочного корреляционного отношения может считаться то, что оно никак не указывает, какая именно (линейная, параболическая, гиперболическая, ..) форма связи имеет место между признаками.

Вычислять корреляционное отношение имеет смысл, если связь между признаками далека от линейной ($|r_a| \leq 0,7$). Более строгая оценка линейности

связи может быть получена с помощью критерия $t = \frac{k}{\delta_k}$, где $k = \eta^2 - r_a^2$ - мера криволинейности, δ_k - основная ошибка меры k , вычисляемая по формуле $\delta_k = \frac{2\sqrt{k + k^2(2 - \eta^2 - r_a^2)}}{\sqrt{n}}$. Если

$t_3 < t_\alpha$, связь может быть признана линейной.

Значимость корреляционного отношения определяется критерием $t = \frac{\eta}{\delta_\eta}$, где $\delta_\eta = \frac{1 - \eta^2}{\sqrt{n}}$. Если вычисленное t_3 больше табличного (Приложение 4) при $f = n - 2$, то найденный значимый.

Доверительный интервал для корреляционного отношения η_A , соответствующего генеральной совокупности, находят в виде $\eta - t\delta_\eta \leq \eta_A \leq \eta + t\delta_\eta$.

Проверка статистических гипотез.

Статистическая гипотеза - всякое предположение о виде закона распределения исследуемой переменной или параметрах известного распределения.

Так, например, можно предположить (выдвинуть гипотезу), что изучаемая переменная X распределена по нормальному закону. В этой гипотезе речь идет о виде предполагаемого закона распределения. Достаточно типична и такая ситуация: закон распределения изучаемой переменной известен, но неизвестны параметры этого распределения. Тогда естественно выдвинуть гипотезу о том, что неизвестный параметр θ принадлежит, например, заданному интервалу.

Таким образом, статистические гипотезы подразделяются на две группы:

- гипотезы о виде закона распределения;
- гипотезы о параметрах известного закона распределения (параметрические гипотезы).

Выдвигаемую гипотезу называют нулевой (основной) и обозначают через H_0 . Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу $\bar{H}_0$. Гипотезу, которая противоречит нулевой, называют конкурирующей (альтернативной) и обозначают через H_1 ($H_1 = \bar{H}_0$).

Выдвинутая гипотеза H_0 , как и всякое предположение, в действительности может быть либо верной, либо неверной; поэтому возникает необходимость ее проверки.

Исходным материалом для проверки выдвинутой гипотезы H_0 служат выборочные данные (выборка).

Задача проверки гипотезы H_0 описательно заключается в следующем: на заданном уровне значимости α требуется установить, согласуется ли выдвинутая гипотеза H_0 с выборочными данными или противоречит им.

Уровень значимости $\alpha = 1 - \gamma$ – вероятность совершить ошибку первого рода ("степень риска"), т.е. вероятность ошибочно отвергнуть верную гипотезу. Уровень значимости α назначается исследователем; наиболее часто

α принимают равным 0,05 (5%) или 0,01 (1%), что соответствует практически ничтожному риску, и тем самым обеспечивают высокую надежность правильного решения задачи.

Основные принципы и необходимые этапы проверки статистической гипотезы

Для проверки выдвинутой гипотезы H_0 используется статистический критерий K (разрешающее правило), согласно которому на основании данных выборки принимается решение сохранить либо отвергнуть нулевую гипотезу H_0 .

В основе критерия K лежит его статистика Z – специально подбираемая для выдвинутой гипотезы H_0 случайная величина, закон распределения которой достаточно хорошо изучен (имеется таблица квантилей этого распределения).

Обозначим через Δ множество всех возможных значений статистики Z . Это множество Δ разбивается на два непересекающихся подмножества Δ_0 и Δ_1 :

$$\Delta_0 \cup \Delta_1 = \Delta, \Delta_0 \cap \Delta_1 = \emptyset,$$

где Δ_0 – область допустимых значений статистики Z ;

Δ_1 – критическая область статистики Z .

Точки, отделяющие Δ_0 от Δ_1 , называются критическими точками статистики Z . Вопрос построения критической области Δ_1 мы здесь рассматривать не будем, отметим лишь только, что $\alpha = P(Z \in \Delta_1 | H_0)$.

По выборочным данным (выборке) вычисляется наблюдаемое значение статистики: $Z_{\text{набл}}$.

Критерий K (разрешающее правило) проверки выдвинутой гипотезы H_0 заключается в следующем:

1. Если $Z_{набл} \in \Delta_1$, то гипотеза H_0 отвергается.
2. Если $Z_{набл} \in \Delta_0$, то гипотеза H_0 сохраняется (т.е. она согласуется с выборочными данными).

Заметим, что отвергают гипотезу более решительно, чем принимают. Принимают гипотезу весьма осторожно. Дело в том, что в случае $Z_{набл} \in \Delta_0$ выдвинутая гипотеза H_0 еще не доказана (по данным одной ограниченной выборки). На практике для большей уверенности принятия гипотезы повторяют эксперимент, увеличив объем выборки, и еще раз проверяют гипотезу (может быть другими способами).

Итак, необходимыми этапами проверки статистической гипотезы являются:

- формирование выборки;
- выдвижение гипотез H_0 и H_1 ;
- назначение уровня значимости α ;
- выбор подходящей статистики Z для проверки H_0 ;
- вычисление по выборке наблюдаемого значения статистики $Z_{набл}$;
- определение по таблице критических точек $Z_{кр}$ статистики Z и построение критической области Δ_1 ;
- принятие решения согласно критерию K проверки гипотезы H_0 .

Байесовский подход к принятию решений.

Предположим, предприниматель раздумывает над выбросом на рынок нового перспективного товара. Но он не знает, “пойдет” ли товар. Для уточнения ситуации он производит пробную партию и смотрит, как он раскупается. После этого ситуация становится более определенной, более прогнозируемой. Для уточнения этой ситуации можно выпустить еще одну пробную партию и проанализировать какие-нибудь другие моменты.

В общем, байесовский подход выглядит следующим образом. Предположим, мы имеем вероятностный прогноз ситуации S : $P(S=H_i)=p_i$. Имея такой прогноз, можно найти средний ожидаемый доход $\bar{Q}$ или средний ожидаемый риск $\bar{R}$. Рассмотрим возможность проведения пробной операции, которая уточнит $\{p_i\}$. Новое распределение вероятностей есть $\{p_i'\}$. Новому распределению вероятностей соответствуют новые характеристики: средний ожидаемый доход $\bar{Q}'$, средний ожидаемый риск $\bar{R}'$. Если ЛПР решит, что при уточнении пробная операция оправдывается (например, если увеличение среднего ожидаемого дохода превышает затраты на проведение пробной операции), то он ее проводит.

$$Q = \begin{vmatrix} 0 & 6 & 5 & 2 \\ 6 & 2 & 8 & 22 \\ 9 & 4 & 3 & 32 \\ -6 & -4 & -12 & 10 \end{vmatrix}$$

$$p_j' = (1/6 \ 1/6 \ 1/3 \ 1/3)$$

$$\bar{Q}_1' : \begin{vmatrix} 0 & 6 & 5 & 2 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{vmatrix}$$

$$\bar{Q}_2' : \begin{vmatrix} 6 & 2 & 8 & 22 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{vmatrix}$$

$$\bar{Q}_3' : \begin{vmatrix} 9 & 4 & 3 & 32 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{vmatrix}$$

$$\bar{Q}_4' : \begin{vmatrix} -6 & -4 & -12 & 10 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{vmatrix}$$

$$\bar{Q}_1' = 6/6 + 5/3 + 2/3 = 20/6$$

$$\bar{Q}_2' = 6/6 + 2/6 + 8/3 + 22/3 = 68/6$$

$$\bar{Q}_3' = 9/6 + 4/6 + 3/3 + 32/3 = 83/6$$

$$\bar{Q}_4' = -6/2 - 4/4 - 12/5 + 10/20 = -14/6$$

Наибольший доход при пробной операции будет получен при 3-ем решении. Теперь выясним, стоит ли производить пробную операцию, т.е. найдем разность между средним ожидаемым доходом от основной операции

(см. Правило максимизации среднего ожидаемого дохода) и полученными в результате пробной операции данными, $(83/6 - 7,7 = 184/30 = 92/15 \approx 6,13)$. В итоге можно сказать, что стоимость пробной операции в данном примере не должна превышать

Для нахождения лучших операций иногда применяют подходящую взвешивающую формулу, которая для пар (Q, r) дает одно число, по которому и определяют лучшую операцию.

Для анализа ситуаций можно применить взвешивающую формулу $E(Q, r) = 4Q - r$. Данная формула говорит, что доход ценится в четыре раза больше, чем риск, т.е. увеличение риска на 4 компенсируется увеличением дохода на единицу.

$$E_1 = 4 \cdot 2.6 - 6.6 = 3.8$$

$$E_2 = 4 \cdot 6.2 - 3 = 21.8$$

$$E_3 = 4 \cdot 7.7 - 1.5 = 29.3$$

$$E_4 = 4 \cdot (-5.9) - 25.1 = -48.7$$

Согласно этой формуле лучшей операцией считается операция № 3, а худшей — операция № 4.

Сложные гипотезы.

Сложной называют гипотезу, которая состоит из конечного или бесконечного множества простых гипотез. Сложная гипотеза H_0 о неравенстве $\square > 10$ состоит из бесконечного множества простых гипотез H_0 о равенстве $\square = b_i$, где b_i – любое число, большее 10. Гипотеза H_0 о том, что математическое ожидание нормального распределения равно двум при неизвестной дисперсии, тоже является сложной. Сложной гипотезой будет предположение о распределении случайной величины X по нормальному закону, если не фиксируются конкретные значения математического ожидания и дисперсии.

Проверка гипотезы о законе распределения.

1.Критерий

Хи-квадрат.

$$x_{cp} := \frac{104}{30} \quad x_{cp} = 3.467 \quad D := \frac{456 - 30 \cdot x_{cp}^2}{30 - 1} \quad D = 3.292$$

$$\lambda := \frac{D + x_{cp}}{2} \quad \lambda = 3.379 \quad i := 0..7$$

$$N := \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{pmatrix}$$

- количество происшествий

Распределение Пуассона.
Вероятность наступления N происшествий:

$$P := \text{dpois}(N, \lambda)$$

$$P = \begin{pmatrix} 0.034 \\ 0.115 \\ 0.195 \\ 0.219 \\ 0.185 \\ 0.125 \\ 0.07 \\ 0.034 \\ 0.014 \end{pmatrix}$$

$$\sum P = 0.992$$

Наблюдаемая частота

$$ob := \begin{pmatrix} 0 \\ 4 \\ 6 \\ 7 \\ 5 \\ 4 \\ 2 \\ 1 \\ 1 \end{pmatrix}$$

$$\sum ob = 30$$

Ожидаемая частота

$$ex := \text{round}(P \cdot 30)$$

$$ex = \begin{pmatrix} 1 \\ 3 \\ 6 \\ 7 \\ 6 \\ 4 \\ 2 \\ 1 \\ 0 \end{pmatrix}$$

$$\sum ex = 30$$

$$\text{obs} := \begin{pmatrix} 10 \\ 7 \\ 5 \\ 8 \end{pmatrix} \quad \text{exp} := \begin{pmatrix} \text{ex}_0 + \text{ex}_1 + \text{ex}_2 \\ \text{ex}_3 \\ \text{ex}_4 \\ \text{ex}_5 + \text{ex}_6 + \text{ex}_7 + \text{ex}_8 \end{pmatrix} \quad \text{exp} = \begin{pmatrix} 10 \\ 7 \\ 6 \\ 7 \end{pmatrix}$$

$\alpha := 0.05$ - уровень значимости

Степень свободы: $v := \text{length}(\text{obs}) - 1 \xrightarrow{v = 3}$

Хи-квадрат статистика: $\chi^2 := \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} \quad \chi^2 = 0.31$

$X2 := \text{qchisq}(1 - \alpha, v) \quad X2 = 7.815 \quad X2 < \chi^2 = 0$

2. Критерий Колмогорова-Смирнова.

Наблюдаемая вероятность	Теоретическая вероятность	Интегральные вероятности (распределение)	
$PN := \frac{ob}{30}$			
$PN = \begin{pmatrix} 0 \\ 0.133 \\ 0.2 \\ 0.233 \\ 0.167 \\ 0.133 \\ 0.067 \\ 0.033 \\ 0.033 \end{pmatrix}$	$P = \begin{pmatrix} 0.034 \\ 0.115 \\ 0.195 \\ 0.219 \\ 0.185 \\ 0.125 \\ 0.07 \\ 0.034 \\ 0.014 \end{pmatrix}$	$PNI = \begin{pmatrix} 0 \\ 0.1333 \\ 0.3333 \\ 0.5667 \\ 0.7333 \\ 0.8667 \\ 0.9333 \\ 0.9667 \\ 1 \end{pmatrix}$	$PI = \begin{pmatrix} 0.034 \\ 0.149 \\ 0.344 \\ 0.563 \\ 0.748 \\ 0.873 \\ 0.944 \\ 0.978 \\ 0.992 \end{pmatrix}$
Абсолютная разность	$AR_i := PNI_i - PI_i $		
$AR = \begin{pmatrix} 0.03407 \\ 0.01587 \\ 0.01041 \\ 0.00378 \\ 0.01469 \\ 0.00648 \\ 0.01028 \\ 0.01097 \end{pmatrix}$	$\max(AR) = 0.034$	$D_{kp} := \frac{1.36}{\sqrt{30}} \quad D_{kp} = 0.248$	
	$\max(AR) < D_{kp} = 1$	гипотеза о том, что экспериментальное распределение пуассоновское не отвергается	

Практические задания

Модуль 1. Теория вероятностей

1.1. Занятие №1

Задача 1. В кабинете работают 6 мужчин и 4 женщины. Для переезда наудачу отобраны 7 человек. Найти вероятность того, что среди отобранных лиц три женщины.

Решение. Общее число возможных исходов равно числу способов, которыми можно отобрать 7 человек из 10, т.е.
 $n = C_{10}^7 = C_{10}^3$.

Найдем число исходов, благоприятствующих интересующему нас событию: трех женщин можно выбрать из четырех C_4^3 способами; при этом остальные четыре человека должны быть мужчинами, их можно отобрать C_6^4 способами. Следовательно, число благоприятствующих исходов равно $C_4^3 C_6^4 = C_4^1 C_6^2$.

Искомая вероятность

$$P = \frac{C_4^1 C_6^2}{C_{10}^3} = \frac{4 \cdot \frac{6 \cdot 5}{2}}{\frac{10 \cdot 9 \cdot 8}{2 \cdot 3}} = \frac{1}{2}.$$

Задача 2. В урне 10 шаров: 6 белых и 4 черных. Вынули два шара. Какова вероятность, что оба шара белые?

Задача 3. Случайным образом одновременно выбирают две буквы из 33 букв русского алфавита. Найдите вероятность того, что обе они согласные;

Задача 4. Два нумизмата обмениваются коллекционными монетами. Найти число способов обмена, если первый нумизмат обменивает 5 монет, а второй — 8 монет.

Задача 5. Абонент забыл последнюю цифру телефонного номера. Найти вероятность того, что при наборе номера наугад он наберет его правильно не более чем с четырех попыток.

Задача 6. В лотерее разыгрывается 150 вещевых и 50 денежных выигрышей на каждые 5000 билетов. Найти вероятность выигрыша вообще.

Задача 7. В партии из 20 изделий 5 изделий имеют скрытый дефект. Какова вероятность того, что из взятых наугад 4 изделий 2 изделия являются дефектными.

Задача 8. На склад поступило 1500 изделий с первой фабрики и 2000 изделий со второй. Известно, что средний процент нестандартных изделий среди продукции первой фабрики равен 3%, второй – равен 2%. Найти вероятность того, что наудачу взятое со склада изделие будет нестандартным.

Задача 9. Бросают игральный кубик. Какова вероятность того, что выпало четное число очков, если известно, что число выпавших очков меньше пяти?

Задача 10. Была проведена одна и та же контрольная работа в трех группах. В первой группе из 30 студентов 8 выполнили работу на «отлично», во второй, где 28 студентов, – 6 «отличных» работ, в третьей, где 27 студентов, – 9 работ выполнены на «отлично». Найти вероятность того, что первая выбранная наудачу работа из работ, принадлежащих группе, которая также выбрана наудачу, окажется «отличной».

Задача 11. В правом и левом карманах имеются по три монетки в 10 коп и по четыре монетки в 5 коп. Из правого кармана в левый наудачу перекладывается 5 монет. Определить вероятность извлечения из левого кармана после перекладывания монеты достоинством в 10 коп.

Задача 12. В розыгрыше кубка страны по футболу берут участие 17 команд. Сколько существует способов распределить золотую, серебряную и бронзовую медали?

Задача 13. Игральный кубик бросают два раза. Описать пространство элементарных событий. Описать события: А - сумма появившихся очков равна 8; В - по крайней мере один раз появится 6.

Задача 14. В вазе с цветами 15 гвоздик: 5 белых и 10 красных. Из вазы наугад вынимают 2 цветка. Какова вероятность того, что эти цветки: а) оба белые; б) оба красные; в) разного цвета; г) одного цвета.

Задача 15. Вероятность того, что в течении одной смены возникнет поломка станка равна 0,05. Какова вероятность того, что не возникнет ни одной поломки за три смены?

Задача 16. С помощью наблюдений установлено, что в некоторой местности в сентябре в среднем бывает 25 дней без дождя. Какова вероятность того, что 1-го и 2-го сентября дождя не будет?

Задача 17. В белом ящике 12 красных и 6 синих шаров. В черном - 15 красных и 10 синих шаров. Бросают игральный кубик. Если выпадет количество очков, кратное 3, то наугад берут шар из белого ящика. Если

выпадет любое другое количество очков, то наугад берут шар из черного ящика. Какова вероятность появления красного шара?

Задача 18. Из каждых 40-ка изделий, изготовленных станком-автоматом 4 бракованных. Наугад взяли 400 изделий. Найти вероятность того, что среди них 350 без дефекта.

Задача 19. Ткач обслуживает 1000 веретен. Вероятность обрыва нитки на одном из веретен в течении одной минуты равна 0,005. Найти вероятность того, что в течении одной минуты обрыв произойдет на 7 веретенах.

Задача 20. Из 20 экзаменационных билетов 3 содержат простые вопросы. Пять студентов по очереди берут билеты. Найти вероятность того, что хотя бы одному из них достанется билет с простыми вопросами.

Занятие №2

Задача 1. Магазин получил продукцию в ящиках с четырех оптовых складов: четыре с 1-го, пять со 2-го, семь с 3-го и четыре с 4-го. Случайным образом выбран ящик для продажи. Какова вероятность того, что это будет ящик с первого или третьего склада.

Решение: всего получено магазином: $4 + 5 + 7 + 4 = 20$ ящиков.

В данной задаче удобнее воспользоваться «быстрым» способом оформления без расписывания событий большими латинскими буквами. По классическому определению:

$p_1 = \frac{4}{20} = 0,2$ – вероятность того, что для продажи будет выбран ящик с 1-го склада;

$p_3 = \frac{7}{20} = 0,35$ – вероятность того, что для продажи будет выбран ящик с 3-го склада.

Бесконечных «хвостов» после запятой тут нет и не ожидается, поэтому можно работать с десятичными дробями – компактнее будет запись.

По теореме сложения несовместных событий:

$p = p_1 + p_3 = 0,2 + 0,35 = 0,55$ – вероятность того, что для продажи будет выбран ящик с первого или третьего склада.

Ответ: 0,55

Задача 2. Два стрелка сделали по одному выстрелу в мишень. Вероятность попадания для первого стрелка равна 0,8, для второго – 0,6. Найти вероятность того, что:

- а) только один стрелок попадёт в мишень;
- б) хотя бы один из стрелков попадёт в мишень.

Задача 3. Для сигнализации о возгорании установлены два независимо работающих датчика. Вероятности того, что при возгорании датчик сработает, для первого и второго датчиков соответственно равны 0,5 и 0,7. Найти вероятность того, что при пожаре:

- а) оба датчика откажут;
- б) оба датчика сработают.

Задача 4. Рабочий обслуживает три станка. Вероятность того, что в течение смены первый станок потребует настройки, равна 0,3, второй – 0,75, третий – 0,4. Найти вероятность того, что в течение смены:

- а) все станки потребуют настройки;
- б) только один станок потребует настройки;
- в) хотя бы один станок потребует настройки.

Задача 5. На станции отправления имеется 8 заказов на отправку товара: пять – внутри страны, а три – на экспорт. Какова вероятность того, что два выбранных наугад заказа окажутся предназначенными для потребления внутри страны?

Задача 6. Из партии изделий товаровед наудачу отбирает изделия высшего сорта. Вероятность того, что выбранная вещь окажется высшего сорта равна, 0,8; первого сорта – 0,7; второго сорта – 0,5. Найти вероятность того, что из трех наудачу отобранных изделий будут:

- а) только два высшего сорта;
- б) все разные.

Задача 7. В типографии имеется 4 печатных машины. Для каждой машины вероятность того, что она работает в данный момент, равна 0,9. Найти вероятность того, что в данный момент работает хотя бы одна машина (событие A).

Задача 8. В читальном зале имеется 6 учебников по теории вероятностей, из которых три в переплете. Библиотекарь наудачу взял два учебника. Найти вероятность того, что оба учебника окажутся в переплете.

Задача 9. Вероятность появления простого самостоятельного предложения в текстах Н.М. Карамзина равна 0,065, а в текстах А.С. Пушкина – 0,132. Из текстов каждого автора извлекается по одному предложению. Найти вероятность событий: а) «оба предложения простые»; б) «хотя бы одно предложение простое»; в) «одно из предложений простое»; г) «оба предложения не являются простыми».

Задача 10. Слово «МАТЕМАТИКА» составлено из букв разрезной азбуки. Карточки с буквами этого слова положены в урну.

Найти вероятность события A = «Получится слово МАТЕМАТИКА», если:
а) последовательно извлекается карточка с буквой и возвращается обратно (безусловная вероятность); б) карточка с буквой извлекается и не возвращается обратно (условная вероятность).

Задача 11. Для разрушения моста достаточно попадания одной авиационной бомбы. Найдите вероятность того, что мост будет разрушен, если на него сбросили три бомбы, вероятности попадания которых в цель соответственно равны 0,8; 0,9 и 0,85.

Задача 12. Третья часть одной из трех партий изделий второсортна, другие изделия во всех партиях первого сорта. Изделие, взятое из одной партии, оказалось первосортным. Найдите вероятность того, что изделие было взято из партии, которая содержала второсортные изделия.

Задача 13. Вероятность выхода изделия из строя при сроке его эксплуатации до одного года равна 0,13, а при эксплуатации сроком до 3 лет – 0,36. Найти вероятность выхода изделия из строя при эксплуатации от 1 года до 3 лет.

Задача 14. Покупатель ищет необходимую вещь, обходя два магазина. Вероятность наличия ее в каждом магазине равна 50%. Что вероятнее – найдет он искомую вещь или нет?

Задача 15. Вероятность получить высокие дивиденды по акциям на первом предприятии – 0,2, на втором – 0,32, на третьем – 0,15. Определить вероятность того, что акционер, имеющий акции всех трех предприятий, получит высокие дивиденды хотя бы на одном предприятии.

Задача 16. На дактилоскопическую экспертизу поступило 3 отпечатка пальцев рук. Вероятность непригодности к работе каждого отпечатка соответственно составляет 0,1; 0,15; 0,2. Найти вероятность того, что все три отпечатка будут обработаны.

Задача 17. Только 1 из 9 ключей подходит к данному замку. Какова вероятность того, что придется опробовать 2 ключа для открывания замка?

Задача 18. Трое охотников одновременно выстрелили по медведю, который был убит одной пулей. Определить вероятность того, что медведь был убит первым стрелком, если вероятности попадания для этих стрелков равны соответственно 0,3, 0,4, 0,5

Задача 19. Задача о переливании крови. При переливании крови надо учитывать группы крови донора и больного. Человеку, имеющему четвертую группу крови, можно перелить кровь любой группы; человеку со второй или третьей группой можно перелить кровь либо той же группы, либо первой; человеку с первой группой крови можно перелить кровь только первой группы. Среди населения 33,7% имеют первую группу, 37,5% - вторую, 20,9% - третью, 7,9 % - четвертую группы крови. Найти вероятность того, что случайно взятому пациенту можно перелить кровь случайного донора.

Задача 20. Подбросили 2 монеты. Появление герба на второй монете не зависит от того, что выпало на первой и наоборот. Это два независимых друг от друга события.

Задача 21. Из множества всех последовательностей длины 10, состоящих из цифр 0; 1; 2; 3, наудачу выбирается одна. Какова вероятность того, что выбранная последовательность содержит ровно 5 нулей, причем два из них находятся на концах последовательности.

Занятие №3

Задача 1. В лотерее выпущено 100 билетов. Разыгрывался один выигрыш в 50 у.е. и десять выигрышей по 10 у.е. Найти закон распределения величины X – стоимости возможного выигрыша.

Решение. Возможные значения величины X : $x_1 = 0$; $x_2 = 10$ и $x_3 = 50$. Так как «пустых» билетов – 89, то $p_1 = 0,89$, вероятность выигрыша 10 у.е. (10 билетов) – $p_2 = 0,10$ и для выигрыша 50 у.е. – $p_3 = 0,01$. Таким образом:

X	0	10	50
P	0,89	0,10	0,01

Легко проконтролировать: $p_1 + p_2 + p_3 = 0,89 + 0,10 + 0,01 = 1$.

Задача 2. Произведено 5000 патронов. Вероятность того, что один патрон бракованный $p = 0,0002$. Какова вероятность того, что во всей партии будет ровно 3 бракованных патрона?

Задача 3. При стрельбе до первого попадания с вероятностью попадания $p = 0,6$ при выстреле надо найти вероятность того, что попадание произойдет при третьем выстреле.

Задача 4. Выпущено 1000 лотерейных билетов: на 5 из них выпадает выигрыш в сумме 500 рублей, на 10 – выигрыш в 100 рублей, на 20 – выигрыш

в 50 рублей, на 50 – выигрыш в 10 рублей. Определить закон распределения вероятностей случайной величины X – выигрыша на один билет.

Задача 5. На пути движения автомашины 4 светофора, каждый из которых запрещает дальнейшее движение автомашины с вероятностью 0,5. Найти ряд распределения числа светофоров, пройденных машиной до первой остановки. Чему равны математическое ожидание и дисперсия этой случайной величины?

Задача 6. В магазине имеется 15 автомобилей определенной марки. Среди них 7 черного цвета, 6 серого и 2 белого. Представители фирмы обратились в магазин с предложением о продаже им 3 автомобилей этой марки, безразлично какого цвета. Составьте ряд распределения числа проданных автомобилей черного цвета при условии, что автомобили отбирались случайно.

Задача 7. В городе 4 коммерческих банка. У каждого риск банкротства в течение года составляет 20%. Составьте ряд распределения числа банков, которые могут обанкротиться в течение следующего года.

Задача 8. В магазине продаются 5 отечественных и 3 импортных телевизора. Составить закон распределения случайной величины – числа импортных из четырех наудачу выбранных телевизоров. Найти функцию распределения этой случайной величины и построить ее график.

Задача 9. Из ящика с семью деталями, среди которых имеется 5 стандартных, наудачу взяты четыре детали. Составить закон распределения дискретной случайной величины X числа стандартных деталей среди отобранных.

Задача 10. Тираж календаря 50 тыс. экземпляров. Вероятность брака в одном календаре равна 0,0003. Найти вероятность содержания в тираже ровно 10 бракованных календарей.

Задача 11. Случайная величина X задана на интервале $(0, 2)$ плотностью распределения $f(x) = x/8$; вне этого интервала $f(x) = 0$. Найти функцию распределения и дисперсию величины X .

Задача 12. Размер женской обуви является случайной величиной с нормальным законом распределения, математическим ожиданием 37 и дисперсией 4. Какой процент от общего объема закупок следует предусмотреть магазину для обуви 38 размера, если этот размер находится в интервале $(37,5, 38,5)$?

Задача 13. Книга в 1000 страниц имеет 100 опечаток. Какова вероятность того, что на случайно выбранной странице не менее четырех опечаток?

Задача 14. В лотереи из ста билетов выигрышными являются один билет в 50 рублей, три билета в 25 рублей и пять билетов по 10 рублей. Для случайной величины X – стоимости возможного выигрыша для владельца одного лотерейного билета найти:

- а) закон распределения;
- б) функцию распределения;
- в) математическое ожидание, дисперсию и среднее квадратическое отклонение.

Задача 15. Частота вспышек средней мощности не стационарной звезды составляет в среднем 10 вспышек в сутки. Какова вероятность того, что в течение минуты будет зафиксировано

- а) ровно две вспышки, б) не более двух вспышек?

Задача 16. Хоккеист, исполняя штрафной юросок, забрасывает шайбу с вероятностью 0.9. По какому закону распределено число заброшенный шайб при 10 бросках? Каково среднее число заброшенных шайб и какова дисперсия этой случайной величины?

Задача 17. Автоматическая телефонная станция получает в среднем 240 вызовов в час. Какова вероятность того что в данную минуту она получит 5 вызовов?

Задача 18. Поезда метро идут строго по расписанию. Интервал движения – 5 минут. Составить $f(x)$ и $F(x)$ случайной величины X – времени ожидания очередного поезда и построить их графики. Найти $M(X)$, $D(X)$.

Задача 19. Непрерывная случайная величина распределена по показательному закону с параметром $\lambda = 2$. Составить функцию распределения, функцию плотности этой случайной величины. Найти числовые характеристики и вероятность того, что случайная величина попадет в интервал $(0,3;1)$.

Задача 20. Детали, выпускаемые цехом, по размеру диаметра распределены по нормальному закону. Стандартная длина диаметра детали равна $a=35$, среднее квадратическое отклонение $\sigma=4$. Требуется:

а) составить функцию плотности вероятностей;

б) найти вероятность того, что диаметр наудачу взятой детали будет больше $\alpha=34$ и меньше $\beta=40$;

в) найти вероятность того, что диаметр детали отклонится от стандартной длины не больше чем на $\delta=2$.

Занятие №4

Задача 1. Имеется выборка из 10 наблюдений экономических показателей (X) и (Y).

x_i	180	172	173	169	175	170	179	170	167	174
y_i	186	180	176	171	182	166	182	172	169	177

Требуется найти выборочное уравнение регрессии Y на X. Построить выборочную линию регрессии Y на X.

Решение: Случайная величина X принимает значения 0,1000,10000 с вероятностями:

$P(X=0)=\frac{89}{100}=0,89$, $P(X=1000)=\frac{10}{100}=0,1$, $P(X=10000)=\frac{1}{100}=0,01$. Таким образом, ряд распределения имеет следующий вид:

X	0	1000	10000
P	0,89	0,1	0,01

Условие нормировки выполняется: $\sum_{i=1}^3 p_i = 0,89 + 0,1 + 0,01 = 1$.

Найдем функцию распределения данной случайной величины X :

Если $x \leq 0$, то $F(x) = 0$ (третье свойство). Если $0 < x \leq 1000$, то $F(x) = 0,89$.

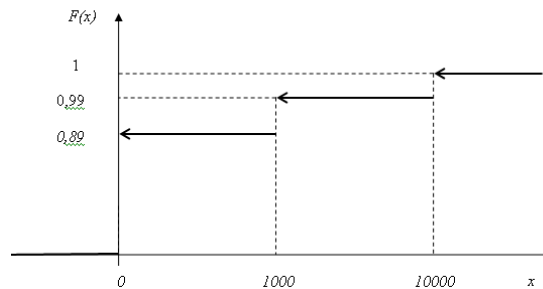
Действительно, X может принять значение 0 с вероятностью 0,89. Если $1000 < x \leq 10000$, то $F(x) = 0,99$.

Действительно, если x_1 удовлетворяет неравенству $1000 < x_1 \leq 10000$, то $F(x_1)$ равно вероятности события $X < x_1$, которое может быть осуществлено, когда X примет значение 0 с вероятностью 0,89 или 1000 с вероятностью 0,1. Поскольку эти два события несовместны, то по теореме сложения вероятность события $X < x_1$ равна сумме вероятностей $0,89 + 0,1 = 0,99$.

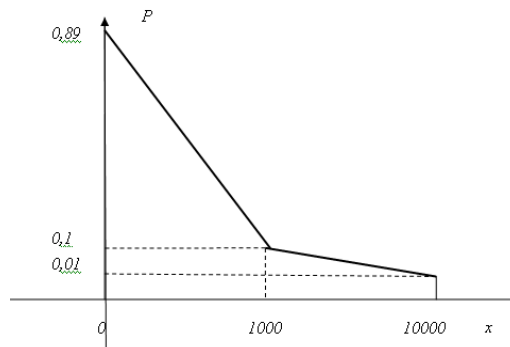
Если $x > 10000$, то $F(x) = 1$ (третье свойство). Итак, функция распределения аналитически может быть записана следующим образом:

$$F(x) = \begin{cases} 0, & \text{если } x \leq 0, \\ 0,89 & \text{если } 0 < x \leq 1000, \\ 0,99 & \text{если } 1000 < x \leq 10000, \\ 1 & \text{если } x > 10000. \end{cases}$$

График функции распределения:



Многоугольник распределения имеет следующий вид:



Задача 2. Задана плотность вероятности случайной величины X

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ 2x & \text{при } 0 < x \leq 1, \\ 0 & \text{при } x > 1. \end{cases}$$

Найти вероятность того, что в результате испытания X примет значение, принадлежащее интервалу $(0,5; 1)$.

Задача 3. Число телефонных звонков, поступивших в справочное бюро от абонентов между полуднем и часом дня в любой день недели, есть случайная величина X , заданная таблицей:

XX	0	11	22	33	44	55
P	0.3	0,2	0,2	0,1	0,1	0.1

- а) Убедитесь, что задан ряд распределения.
- б) Найти функцию распределения $F(x)$ и построить график.
- в) Используя функцию распределения $F(x)$, определите вероятность того, что между 12 час.34 мин. и 12 час. 35 мин. в справочное бюро поступит больше двух звонков

Задача 4. В пяти ВУЗах случайным образом отобрали по одному студенту для проведения тестирования на употребление наркотиков. По статистике, вероятность того, что студент ВУЗа употребляет наркотики равна 0,4.

Пусть $A = \{\text{студент является наркоманом}\}$. $P(A) = 0,4 = p$, $q = 0,6$.

X - число наркоманов.

- 1) Составить ряд распределения X .

Задача 5. Закон распределения случайной величины X имеет вид:

X	-2	0	2	3
P	0,2	0,3	0,1	0,4

Рассмотрим две числовые функции $\varphi_1(x) = x^3$ и $\varphi_2(x) = x^2$. Подставляя вместо аргумента x случайную величину X , получим новые случайные величины $Y = \varphi_1(X) = X^3$ и $Z = \varphi_2(X) = X^2$. Построить ряды распределений случайных величин: 1) $Y = X^3$, 2) $Z = X^2$. Составить их функции распределения.

Задача 6. Случайная величина X распределена нормально с параметрами m и

σ ($X \sim N(m; \sigma)$). Доказать, что линейная функция $Y = aX + b$, где $a \neq 0$, также распределена нормально, причем $M[Y] = am + b$, $\sigma[Y] = |a| \sigma$.

Задача 7. Случайная величина X распределена по закону Коши

$$f_X(x) = \frac{1}{\pi(1+x^2)}.$$

Найти плотность распределения случайной величины $Y = X^3 + 2$.

Задача 8. Случайная величина X распределена равномерно в интервале $(0; 2\pi)$ ($X \sim R(0; 2\pi)$). Найти плотность распределения случайной величины $Y = \cos X$.

Задача 9. Задана функция распределения $F_X(x)$ случайной величины X . Найти функцию распределения $F_Y(y)$ случайной величины $Y = \varphi(X)$, если:

1) $Y = 3X + 2$; 2) $Y = -\frac{2}{3}X + 2$.

Задача 10. Закон распределения случайной величины X имеет вид:

X	-1	0	1	2
P	0,1	0,2	0,3	0,4

Вычислить m_Y и D_Y , если $Y = X^2$

Задача 11. Случайная величина X задана плотностью распределения

$$f_X(x) = \begin{cases} \cos x, & x \in (0; \pi/2), \\ 0, & x \notin (0; \pi/2). \end{cases}$$

Найти математическое ожидание и дисперсию функции $Y = \varphi(X) = X^2$.

Задача 12. Закон распределения случайной величины X имеет вид:

X	$\pi/4$	$\pi/2$	$3\pi/4$
P	0,2	0,7	0,1

Найти закон распределения случайной величины $Y = \sin X$.

Задача 13. Задана плотность распределения $f_X(x)$ случайной величины X , возможные значения которой заключены в интервале $(0; +\infty)$. Найти плотность распределения $f_Y(y)$ случайной величины Y , если:

1) $Y = e^{-X}$; 2) $Y = 1/X^2$; 3) $Y = \sqrt{X}$.

Задача 14. Задана плотность распределения $f_X(x)$ случайной величины X , возможные значения которой заключены в интервале $(-\infty; +\infty)$. Найти плотность распределения $f_Y(y)$ случайной величины Y , если:

1) $Y = e^{-X^2}$; 2) $Y = |X|$; 3) $Y = \arctg X$.

Задача 15. Число X неисправностей на участке высоковольтной линии в течение года имеет распределение Пуассона с параметром a ($X \sim P(a)$). Общий материальный ущерб Y от этих неисправностей пропорционален

квадрату их числа: $Y = kX^2$, где $k > 0$ – неслучайная величина. Найти закон распределения этого ущерба.

Модуль 2. Математическая статистика

Занятие №5

Задача 1. В процессе исследования среднедушевого дохода (в руб.) обследовано 100 семей. Выявлены оценки: $\bar{X} = 1800$, $s = 230$. В предположении о нормальном законе:

а) найти долю семей, чей среднедушевой доход находится в пределах от 1200 до 1800;

Решение

$$P = \Phi\left(\frac{1800-1800}{230}\right) - \Phi\left(\frac{1200-1800}{230}\right) = \Phi(0) - \Phi\left(-\frac{600}{230}\right) = \Phi(0) + \Phi(2,61) = \\ = 0 + 0,495 = 0,495$$

Для того, чтобы найти долю семей, умножим полученную вероятность на 100.

Т.е. 49,5 семей из 100.

Ответ: 49,5 из 100.

Задача 2.

В 2009 г. Среднегодовая численность населения города составляла 1018,3 тыс. чел, в 2010 г. – 1025, 8 тыс. чел, в 2011 г. – 1030, 6 тыс. чел.

1) Определить в % базисные относительные величины динамики (с точностью до 0,1 %):

А) 102,8 и 101,2; в) 100,7 и 101,2 с) 101,8 и 102,3.

Задача 3.

Цены на бензин в 4м квартале текущего года по сравнению с 4-м кварталом предыдущего года возросли на 1 литр с 23 до 27 руб. Каков ежеквартальный темп прироста цен на бензин (в %):

А) 8,3 с) 2,9

В) 4,1 d) 10,4

Задача 4. Измерение роста детей младшей группы детского сада представлено выборкой:

92, 96, 95, 96, 94, 97, 98, 94, 95, 96.

Найдем некоторые характеристики этой выборки.

Задача 5. Из генеральной совокупности извлечена выборка объема N. Оценить с надежностью $\gamma = 0,95$ математическое ожидание A нормально распределенного признака X генеральной совокупности по выборочной средней с помощью доверительного интервала.

Значение признака XI	2	1	1	3	4	5
Частота Ni	2	1	2	2	2	1

Задача 6. Найти методом сумм асимметрию и эксцесс по заданному распределению выборки объема N=100:

Варианта XI	4	8	5	2	5	6	6	0	6	4	6	8	7	2	7	6	8	0	8	4
Частота Ni	2	4	6	8	12	30	18	8	7	5										

Задача 7. Администрацию универсама интересует оптимальный уровень запасов продуктов в торговом зале, а также среднемесячный объем продаж товаров, не являющихся предметом ежедневного потребления в семье (таких, например, как сода). Для выяснения этого вопроса менеджер универсама в течение месяца регистрировал продажи соды и представил результаты в виде дискретного вариационного ряда

x_i 1 2 3 4 5 6 7
 m_i 2 3 1 2 5 5 2 1

Требуется:

а) построить полигон относительных частот $w_i = \frac{m_i}{n}$;

Задача 8. Построить полигон частот и эмпирическую функцию по данному распределению выборки:

Варианты XI	3	0	1	4	6	7
Частоты N_i	3	6	1	2	5	1

Задача 9. Имеется информация о среднедушевых доходах на душу населения по РФ за 2010 год, руб.:

Год	Среднедушевой доход на душу населения, руб./чел.
2006	10000
2007	12315
2008	14826
2009	16739
2010	18565

Для анализа динамики изучаемого показателя определите:

- 1) а) в соответствии с классификацией – вид ряда динамики;
- Б) средний уровень ряда;
- В) цепные и базисные показатели по среднедушевому доходу: абсолютные приросты, темпы (коэффициенты) роста и прироста;
- Г) абсолютное значение 1% прироста.

Задача 10. Имеются данные о работы предприятия (таблицы 1 и 2). Провести анализ его деятельности. Определить для отчетного и базисного периода:

- 1) Среднесписочную численность рабочих, показатели движения персонала (коэффициенты оборота рабочей силы по приему, увольнению, коэффициент текучести рабочих).
- 2) Часовую, дневную и годовую производительность труда.

Задача 11. Общий индекс себестоимости продукции (в %) ...при условии:

Вид продукции	Общая сумма затрат на производство продукции	Изменение себестоимости продукции в отчетном периоде по сравнению с базисным, %	
Базисный период	Отчетный период		
Обувь муж.	230	240	+6
Обувь жен.	290	300	-2

A) 103,4; b) 101,4; c) 104,3; d) 98,5.

Задача 12.

Средняя списочная численность работников за сентябрь ...чел. (с точностью до 1 чел.) при условии:

Списочная численность на 1 сентября	Принято на работу	Уволено
С 5 сентября	С 19 сентября	С 8 сентября
24 чел.	5 чел.	2 чел.

A) 26; b) 28; c) 25; d) 27.

Задача 13. При изучении структуры коммерческих банков по объявленному уставному фонду из трех тысяч банков страны было отобрано по схеме собственно случайной бесповторной выборки сто. Данные о распределении банков по этому признаку представлены в таблице:

Размер уставного фонда	x_{min}	До 10	10	20	30	40	Свыше 50	Итого
	x_{max}		20	30	40	50		
Число банков	n_i	7	9	18	34	22	10	100

Найти:

А) вероятность того, что средний размер уставного фонда всех коммерческих банков отличается от среднего размера его в выборке не более чем на пять миллионов рублей (по абсолютной величине);

Задача 14. С целью размещения рекламы опрошено 400 телезрителей, из которых данную передачу смотрят 150 человек. С доверительной вероятностью 0,95 найти долю телезрителей, охваченных рекламой в лучшем случае. Случайны ли результаты опроса, если согласно статистике доля телезрителей, охваченных рекламой составляет 0,41 при уровне значимости $\alpha=0,05$?

Задача 15. Заданы среднее квадратическое отклонение s нормально распределенной случайной величины, выборочное среднее $\bar{x}$. И объем выборки N . Найти доверительный интервал для математического ожидания с доверительной вероятностью g .

$S = 4$; $\bar{x} = 24$; $N = 36$; $g = 0,87$.

Задача 16. По выборке объема $N=41$ найдена смещенная оценка генеральной дисперсии $D_B = 3$. Найти несмещенную оценку дисперсии генеральной совокупности.

Задача 17. Найти доверительный интервал для оценки с надежностью $P=0,95$ неизвестного математического ожидания A нормально распределенного признака X генеральной совокупности, если даны генеральное среднее квадратическое отклонение $S=5$, выборочная средняя $\bar{x}_B = 14$, а объем выборки $N=25$.

Задача 18. Задано распределение частот выборки:

x_i	2	6	12	15
m_i	3	10	7	5

Составить распределение относительных частот.

Задача 19. Случайная величина X распределена нормально. По выборке объема $n = 16$ найдены $\bar{X} = 20,2$ и $\bar{\sigma} = 0,8$. Оценить неизвестное математическое ожидание $\bar{a}$ (при неизвестной дисперсии) с надежностью $\gamma = 0,975$.

Задача 20. Проведено выборочное обследование магазинов города. Имеются следующие данные о величине товарооборота для 50 магазинов города (x_i – товарооборот, млн. руб.; n_i – число магазинов).

x_i 25-75 75-125 125-175 175-225 225-275 275-325

n_i 12 15 9 7 4 3

Найти

- среднее, среднее квадратическое отклонение S и коэффициент V ;
- построить гистограмму и полигон частот.

2.2 Занятие №6

Задача 1. По результатам анализа темпов роста стоимости акций 30 компаний выявлено, что средний темп роста составил $X=104\%$.

Определить:

- границы доверительного интервала для математического ожидания при известном генеральном среднеквадратическом отклонении $\alpha = 5\%$ с надежностью $\gamma = 0.97$, и с надежностью $\gamma = 0.95$ при не известном α , если выборочное среднеквадратическое отклонение $S = 5.2\%$;

Решение.

- При нахождении доверительного интервала для математического ожидания а при известном среднеквадратическом отклонении σ и нормальном законе распределения исследуемой генеральной совокупности используется статистика

$$t = \frac{\bar{X} - a}{\sigma} \sqrt{n} \in N(0,1).$$

При заданной величине γ : $P(|t| < \delta) = \gamma = 2\Phi(\delta)$,

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

где

Из уравнения $2\Phi(\delta) = \gamma/2$ находят величину $\delta\gamma = \Phi^{-1}(\gamma/2)$ по таблице для $\Phi(x)$, а из неравенства

$$\frac{|\bar{X} - a|}{\sigma} \sqrt{n} < \delta_\gamma$$

найдем доверительный интервал для а:

$$\begin{aligned} \frac{-\delta_\gamma \sigma}{\sqrt{n}} + \bar{X} < a < \frac{-\delta_\gamma \sigma}{\sqrt{n}} + \bar{X} \\ \frac{-2.17 * 5}{\sqrt{30}} + 104 < a < 104 + \frac{2.17 * 5}{\sqrt{30}} \end{aligned}$$

Окончательно $102.06 \leq a \leq 105.94$.

$$t = \frac{\bar{X} - a}{S} \sqrt{n-1}$$

Если σ неизвестно воспользуемся статистикой, имеющей распределение Стьюдента с $v=n-1$ степенями свободы. По определению доверительного интервала $P(|t| < \delta) = \gamma$, $1-\gamma = \alpha$. Для случайной величины t , имеющей распределение Стьюдента затабулированы значения функции $St(x, v) = P(|t| > x)$.

В силу свойств функции $St(x)$: $P(-\delta \leq t \leq \delta) = 1 - 2St(\delta) = \gamma$, $St(\delta) = 1 - \gamma/2 = \alpha/2$. По таблице найдем δ_α :

$$\bar{X} - \frac{S\delta_\alpha}{\sqrt{n-1}} \leq a \leq \bar{X} + \frac{S\delta_\alpha}{\sqrt{n-1}}$$

$v=n-1=30-1=29$ и $\alpha=1-\gamma=1-0.95=0.05$. $\delta_\alpha=2.045$.

Доверительный интервал имеет вид:

$$104 - \frac{2.045 * 5.2}{\sqrt{30-1}} \leq a \leq 104 + \frac{2.045 * 5.2}{\sqrt{30-1}},$$

$102.03 \leq a \leq 105.97$.

Задача 2. Из генеральной совокупности извлечена выборка объема $n=60$:

x	1	3	6	26
n	8	40	10	2

Найти несмещенную оценку генеральной средней.

Задача 3. Найти выборочную среднюю по данному распределению выборки объема $n=20$:

X	2560	2600	2620	2650	2700
n	2	3	10	4	1

Задача 4. По выборке объема $n=51$ найдена смещенная оценка $=5$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

Задача 5. В итоге четырех измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: 8; 9; 11;

Найти:

- а) выборочную среднюю результатов измерений;
- б) выборочную и исправленную дисперсии ошибок прибора.

Задача 6. Найти выборочную дисперсию по данному распределению выборки объема $n=100$:

X	340	360	375	380
n	20	50	18	12

Задача 7. Найти выборочную дисперсию по данному распределению выборки объема $n = 50$:

x_i	0.1	0.5	0.6	0.8
n	5	15	20	10

Задача 8. Найти исправленную выборочную дисперсию по данному распределению выборки объема $n= 20$:

X_i	0.1	0.5	0.7	0.9
n	6	12	1	1

Задача 9. Найти исправленную выборочную дисперсию по данному распределению выборки объема $n= 10$:

x	23.5	26.1	28.2	30.4
n	2	3	4	1

Задача 10. Дана выборка объема n из нормально распределенной генеральной совокупности Y с известным $\sigma=3$. Точность оценки неизвестного математического ожидания μ_y при уровне доверительной вероятности $\gamma=0.95$ равна 1. Определить минимальный объем выборки, обеспечивающий заданную точность оценки математического ожидания.

Задача 11. По данным торгов на российских биржах 8 банков определили средний курс $X=31.5$ рубля за один доллар. Определить с надежностью $\gamma=0.95$ границы доверительного интервала для генеральной средней, в предположении, что курс доллара распределен по нормальному закону с $\sigma=2$ рубля.

Задача 12. Случайная величина, распределенная по нормальному закону, имеет дисперсию $\sigma^2=9$. Сделана случайная выборка объемом $n=25$, для которой выборочная средняя $X=8$. С надежностью $\gamma=0.95$ указать интервальную оценку для неизвестного математического ожидания.

Задача 13. Случайная величина X имеет нормальное распределение с известным среднеквадратическим отклонением $\sigma=2$. Определить доверительный интервал для оценки неизвестного математического ожидания по выборочной средней $X=5$, если объем выборки $n=16$ и доверительная вероятность $\gamma=0.90$.

Задача 14. На контрольных испытаниях 16 осветительных ламп были определены: средняя продолжительность работы ламп $X=3000$ ч и оценка среднеквадратического отклонения $S=20$ ч. Считая, что срок службы каждой лампы является нормально распределенной случайной величиной, указать вычисленный с надежностью 0.9 интервал для математического ожидания.

Задача 15. Определить доверительный интервал для дисперсии σ^2 генеральной совокупности X по выборке объемом $n=20$, если $X=67.6$, $S=29.20$ и заданной доверительной вероятности $\gamma=0.95$.

Задача 16. Средний размер основных фондов у 30 предприятий отрасли, вошедших в выборку, составил $X=85$ млн. руб., а $S=0.09$. В предположении о нормальном распределении определить с надежностью 0.9 границы доверительного интервала для σ генеральной совокупности.

Задача 17. Средний размер основных фондов у 30 предприятий отрасли, вошедших в выборку, составил $X=85$ млн.руб., а $S=0.09$. В предположении о нормальном распределении определить с надежностью 0.95 границы доверительного интервала для генеральной средней.

Задача 18. На основании выборочных наблюдений над производительностью труда 42 работниц было установлено, что средняя суточная выработка работницы составляет $X=608$ м ткани в час, S

=15м/час. Определить доверительную вероятность того, что генеральная средняя попадает в интервал $(0.992X; 1.008X)$ дисперсии в предположении, что производительность работниц имеет нормальное распределение.

Задача 19. На основании выборочных наблюдений над производительностью труда 42 работниц было установлено, что средняя суточная выработка работницы составляет $X=608$ м ткани в час, $S=15$ м/час. Определить с надежностью 0.8 границы доверительного интервала для генеральной дисперсии в предположении, что производительность работниц имеет нормальное распределение.

Задача 20. По результатам 10 наблюдений установлено, что среднегодовой темп роста производства продукции предприятия равен в среднем 104.4%. В предположении, что ошибки наблюдений распределены по нормальному закону со среднеквадратическим отклонением $S=1\%$ определить с надежностью 0.95 интервальную оценку для генерального среднеквадратического отклонения.

2.3 Занятие №7

Задача 1. Из большой партии электроламп случайным образом отобрано 100. Средняя продолжительность горения ламп в выборке оказалась равной 1000 ч. Найти с надежностью $\gamma = 0,95$ доверительный интервал для средней продолжительности a горения ламп во всей партии, если известно, что среднее квадратическое отклонение продолжительности горения лампы $\sigma = 40$ ч и продолжительность горения ламп распределена по нормальному закону.

Решение. По условию $\bar{x}_B = 1000$, $\gamma = 0,95$, $\sigma = 40$. Для решения воспользуемся формулой

$$\bar{x}_B - \frac{\sigma}{\sqrt{n}} < a < \bar{x}_B + \frac{\sigma}{\sqrt{n}}$$

По приложению 3 находим t из условия:

$$\Phi(t) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475; \Rightarrow t = 1,96.$$

Тогда доверительный интервал:

$$1000 - \frac{1,96 \cdot 40}{\sqrt{100}} < a < 1000 + \frac{1,96 \cdot 40}{\sqrt{100}} \\ 992,16 < a < 1007,84.$$

Задача 2.

Выборочно обследование 30 предприятий машиностроительной

промышленности по валовой продукции и получены следующие данные, в млн. руб.:

18,0; 12,0; 11,9; 1,9; 5,5; 14,6; 4,8; 5,6; 4,8; 10,9; 9,7; 7,2; 12,4; 7,6;
9,7; 11,2; 4,2; 4,9; 9,6; 3,2; 8,6; 4,6; 6,7; 8,4; 6,8; 6,9; 17,9; 9,6;
14,8; 15,8.

Составить интервальное распределение выборки с началом $x_0 = 1$ и длиной частичного интервала $h = 3$. Построить гистограмму частот.

Задача 3.

Из большой группы предприятий одной из отраслей промышленности случайным образом отобрано 30, по которым получены показатели основных фондов в млн. руб.: 2; 3; 2; 4; 5; 2; 3; 3; 6; 4; 5; 4; 6; 5; 3; 4; 2; 4; 3; 3; 5; 4; 6; 4; 5; 3; 4; 3; 2; 4.

1. Составить дискретное статистическое распределение выборки.
2. Найти объем выборки.
3. Составить распределение относительных частот.

Задача 4. Выборка задана таблицей распределения

X_i	1	2	3	5
n_i	15	20	10	5

Найти выборочные характеристики: среднюю, дисперсию и среднее квадратическое отклонение.

Задача 5.

На экзамене по истории студенты получили оценки:

3	4	4	4	3	4
3	4	3	5	4	4
5	5	2	3	2	3
3	4	4	5	3	3
5	4	5	4	4	4

Построить дискретный вариационный ряд распределения студентов по баллам и изобразить его графически.

Задача 6.

Во время выборочной проверки было установлено, что продолжительность одной покупки в кондитерском отделе магазина была такой: (секунды).

77	70	82	81	81
82	75	80	71	80
81	89	75	67	78
73	76	78	73	76
82	69	61	66	84
72	74	82	82	76

Построить интервальный вариационный ряд распределения покупок по продолжительности, создав 4 группы с одинаковыми интервалами. Обозначить элементы ряда. Изобразить его графически, сделать вывод.

Задача 7. В результате эксперимента получены следующие значения случайной величины X :

3; 6; 8; 11; 6; 10; 7; 9; 7; 3; 4; 8;

7; 9; 4; 9; 11; 7; 8; 4; 10; 5; 6; 7; 2.

Требуется: а) составить статистический ряд;

б) построить статистическое распределение;

Задача 8. В результате эксперимента получены следующие значения случайной величины X :

16; 17; 9; 13; 21; 11; 7; 7; 19; 5; 17; 5; 20;

18; 11; 4; 6; 22; 21; 15; 15; 23; 19; 25; 1.

Требуется: а) построить интервальный статистический ряд, разбив промежуток $[0; 25]$ на 5 промежутков равной длины;

б) построить гистограмму относительных частот.

Задача 9. Имеются данные о количестве студентов в 30 группах физико-математического факультета:

26	25	25	26	25	23
23	24	19	23	20	19
22	24	24	23	20	23
24	19	21	18	21	18
20	18	18	21	15	15

Найти вариационный ряд количества студентов в группах и размах варьирования. Построить полигон частот.

Задача 10. В результате испытаний получены следующие статистические значения случайной величины X :

-0,9; 0,1; -2,9; 1,1; 5,1; 0,1; -6,9; 1,1; -3,9; -0,9;

5,1; -8,9; -2,9; 0,1; 1,1; 6,1; 3,1; 0,1; 1,1; -2,9;

-2,9; 0,1; -3,9; -0,9; 0,1; 3,1; 0,1; 1,1; 3,1; 0,1.

Требуется: а) составить статистический ряд;

б) найти статистическую функцию распределения $\tilde{F}(x)$;

в) изобразить полигон относительных частот.

Задача 11. В итоге пяти измерений длины стержня одним прибором (без систематических ошибок) получены следующие результаты (в мм): 92; 94; 103; 105; 106. Найти: а) выборочную среднюю длину стержня; б) выборочную и исправленную дисперсии ошибок прибора.

Задача 12. Произведено 10 фиксаций курса валюты X и валюты Y . Результаты (в условных единицах) сведены в таблицу:

i	1	2	3	4	5	6	7	8	9	10
-----	---	---	---	---	---	---	---	---	---	----

x_i	1.8	1.85	1.85	1.7	1.72	1.77	1.8	1.83	1.89	1.89
y_i	1.5	1.5	1.45	1.5	1.6	1.6	1.55	1.5	1.55	1.55

Найти оценки для числовых характеристик системы случайных величин (X,Y).

Задача 13. Вычислить несмещенные оценки среднего значения μ , дисперсии $2\sigma^2$ и стандартного отклонения σ генеральной совокупности x , $2S^2$, S по выборкам (43, 51, 44, 47, 34) и (52, 42, 40, 38, 37).

Задача 14. Вычислить несмещенные оценки параметров генеральной совокупности x , $2S^2$, S по трем выборкам:

$x = (82, 70, 83, 80, 76, 81, 80, 76, 70, 79, 74, 77, 65, 80, 79)$, $y = (-237, -209, -243, -239, -221, -238, -238, -223, -207, -237, -216, -228, -193, -234, -229)$, $z = (730, 626, 737, 716, 681, 721, 710, 680, 626, 708, 658, 690, 579, 714, 704)$.

Задача 15. Вычислить несмещенные оценки параметров генеральной совокупности x , $2S^2$, S по выборкам А и В, используя результаты, полученные в примере 3.

Выборка А: 0 4 2 0 5 1 1 3 0 2 2 4 3 2 3 3 0 4 5 1 3 1 5 2 0 2 2 3 2 2 2 6 2 1 3 1 3 1 5 4 5 5 3 2 2 0 2 1 1 3 2 3 5 3 5 2 5 2 1 1 2 3 4 3 2 3 2 4 2. $N=69$.

Начало первого интервала: 0. Длина интервала: 1.

Выборка В: 135 133 124 132 104 152 134 130 129 120 122 124 117 123 123 129 121 122 125 131 147 124 137 112 126 128 111 129 115 147 131 132 137 119 125 120 129 125 123 127 132 118 133 132 132 134 131 120 135 132 125 132 108 114 121 133 133 135 131 125 114 115 122 131 125 132 120 126 115 117 118 118 132 134 127 127 124 135 128 127 115 144 129 120 137 127 125 116 132 120 117 127 118 109 127 122 120 135 116 118 133 136 125 126 119 126 129 127 129 124 127 132 126 131 127 130 126 124 135 127 124 123 123 130 132 143 122 138 120 134 108 132 121 111 123 140 137 120 125 131 118 120 120 136 129 127 116 138 128 133 122 131 128 140 138 134 120 126 109 137 111 115 117 130 113 126 115 124 125 118 115 128 123 129 128 120 115 134 118 135 134. $N=181$. Начало первого интервала: 102. Длина интервала: 4

Задача 16. Случайная величина X имеет нормальное распределение с известным средним квадратическим отклонением $\sigma=3$. Найти доверительные интервалы для оценки неизвестной генеральной средней по выборочным средним, если объем выборок $n=36$ и надежность оценки $p=0,95$

Задача 17. Количественный признак X генеральной совокупности распределен нормально. По выборке объема $n=16$ найдены выборочная средняя $\bar{x}=20,2$ и исправленное среднеквадратическое отклонение $s=0,8$. Оценить неизвестную генеральную среднюю с помощью доверительного интервала с надежностью $p=0,95$

Задача 18. Дана выборка объема $n=30$. Сделать интервальную группировку этой выборки.

20,3; 15,4; 17,2; 19,2; 23,1; 18,1; 21,9; 15,3; 16,8; 13,2; 20,4; 16,5; 19,7; 20,5; 14,3; 20,1; 16,8; 14,7; 20,8; 19,5; 15,4; 19,3; 17,8; 16,2; 15,7; 22,8; 21,9; 12,5; 10,1; 21,1

Занятие №8

Задача 1. Исследование зависимости между среднемесячными доходами X на семью (в тыс. у.е.) и расходами Y на покупку кондитерских изделий (в у.е.) представлено в таблице:

X	4,8	3,8	5,4	4,2	3,4	4,6	3,4	4,8	5,0	3,8	5,2	4,0	3,8	4,6	4,4
Y	75	68	78	71	64	73	66	75	75	65	77	69	67	72	70

Построить корреляционное поле и сделать предварительный вывод о форме зависимости случайных величин.

Решение. Корреляционное поле, построенное по статистическим данным, приведено на рис. 6.1.

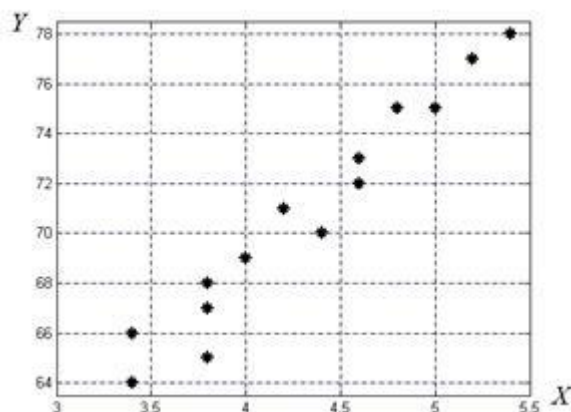


Рис. 6.1.

Анализ рис. 6.1 позволяет сделать вывод о наличии сильной линейной статистической связи между среднемесячными доходами семьи и затратами на приобретение ею кондитерских изделий. При этом связь имеет положительную тенденцию, т.е. с ростом переменной X наблюдается увеличение отклика Y .

При большом объеме выборки результаты группируются и представляются в виде корреляционной таблицы.

Задача 2. Имеются данные средней выработки на одного рабочего Y (тыс. руб.) и товарооборота X (тыс. руб.) в 20 магазинах за квартал. На основе указанных данных требуется:

- 1) определить зависимость (коэффициент корреляции) средней выработки на одного рабочего от товарооборота,
- 2) составить уравнение прямой регрессии этой зависимости.

Задача 3. С целью анализа взаимного влияния зарплаты и текучести рабочей силы на пяти однотипных фирмах с одинаковым числом работников проведены измерения уровня месячной зарплаты X и числа уволившихся за год рабочих Y :

X	100	150	200	250	300
Y	60	35	20	20	15

Найти линейную регрессию Y на X , выборочный коэффициент корреляции.

Задача 4. В таблице представлены результаты измерений роста X (см) и веса Y (кг) 50 мужчин – слушателей военной академии:

Y	[55; 65)	[65; 75)	[75; 85)	[85; 95)	$n_{\cdot j}$
X					
[155; 165)	2	5	4	1	12
[165; 175)	2	8	9	4	23
[175; 185)	0	4	6	5	15
$n_{\cdot j}$	4	17	19	10	50

Вычислить выборочный коэффициент корреляции и проверить гипотезу о значимости корреляционной связи.

Задача 5. Имеется связанная выборка из 11 пар значений (x_k, y_k):

k	1	2	3	4	5	6	7	8	9
x_k	51.00000	50.00000	48.00000	51.00000	46.00000	47.00000	49.00000	60.00000	51.00000
y_k	13.00000	15.00000	13.00000	16.00000	12.00000	14.00000	12.00000	10.00000	18.00000

Требуется вычислить/построить:

- коэффициент корреляции;
- коэффициенты уравнения линейной регрессии;

Задача 6. По данным корреляционной таблицы найти условные средние $\bar{Y}_x$ и $\bar{X}_y$. Оценить тесноту линейной связи между признаками X и Y и составить уравнение линейной регрессии Y по X и X по Y. Сделать чертеж, нанеся на него условные средние и найденные прямые регрессии. Оценить силу связи между признаками с помощью корреляционного отношения.

Y\X	5	10	15	20	25	30	N _y
35	4	2					6
45		5	3				8
55			5	45	5		55
65			2	8	7		17
75				4	7	3	14
N _x	4	7	10	57	19	3	n=100

Задача 7. Используя метод сопоставления параллельных рядов, метод построения корреляционной таблицы и поля корреляции, установите направление и характер связи между прожиточным минимумом и средней заработной платой населения по 10 районам РФ:

Таблица 1.

№ района	Средняя заработная плата, тыс. руб.	Прожиточный минимум на душу населения, тыс. руб./месяц
	1,08	0,49

	1,63	0,49
	1,04	0,46
	1,49	0,52
	0,97	0,38
	0,90	0,33
	0,77	0,34
	0,69	0,34
	0,57	0,33
	0,52	0,28

По данным таблицы вычислить линейный коэффициент корреляции и коэффициент детерминации. Охарактеризуйте тесноту и направление связи между признаками.

Задача 8. По 20 туристическим фирмам были установлены затраты X на рекламу и количества туристов Y, воспользовавшихся услугами каждой фирмы. В таблице фирмы ранжированы по величине затрат на рекламу:

Порядковый номер фирмы	Затраты на рекламу, усл. ден. ед.	Количество туристов, воспользовавшихся услугами фирмы, чел.
1	8	800
2	8	850
3	8	720
4	9	850
5	9	800
6	9	880
7	9	950
8	9	820
9	10	900
10	10	1000
11	10	920
12	10	1060
13	10	950
14	11	900
15	11	1200
16	11	1150
17	11	1000

18	12	1200
19	12	1100
20	12	1000

Построить корреляционную таблицу и сделать предварительный вывод о форме зависимости случайных величин.

Задача 9. Найти выборочное уравнение прямой линии регрессии по данным $n=8$ наблюдений, которые получены при изучении зависимости количества продаж товара y от затрат на рекламу этого товара x :

x	1,5	4,0	5,0	7,0	8,5	10,0	11,0	12,5
y	5,0	4,5	7,0	6,5	9,5	9,0	11,0	9,0

Задача 10. установить взаимосвязь балансовой прибыли со стоимостным выпуском продукции по данным интервальной группировки 40 фирм, приведенным ниже:

Таблица II.3

Выпуск продукции, X, млн. руб.	Балансовая прибыль, Y, млн.руб.	Число фирм
300-400 400-500 500-600 600-700 700-800	10-20, 20-30, 30-40 20-30, 30-40, 40-50 30-40, 40-50, 50-60 40-50, 50-60 50-60	2; 4; 2 1; 5; 3 4; 8; 2 2; 4 3

ЗАДАЧА 11. По данной корреляционной таблице построить прямые регрессии с X на Y и с Y на X . Найти соответствующие коэффициенты регрессии и коэффициент корреляции между X и Y .

y/x	15	20	25	30	35	40
100					2	2
120			4	3	10	3
140		2	50	7	10	
160	1	4		3		
180	1	1				

Задача 12. Дана корреляционная таблица для величин X и Y , X - срок службы колеса вагона в годах, а Y - усредненное значение износа по толщине обода колеса в миллиметрах. Определить коэффициент корреляции и уравнения регрессий.

X / Y	0	2	7	12	17	22	27	32	37	42
0	3	6	0	0	0	0	0	0	0	0
1	25	108	44	8	2	0	0	0	0	0
2	30	50	60	21	5	5	0	0	0	0
3	1	11	33	32	13	2	3	1	0	0
4	0	5	5	13	13	7	2	0	0	0
5	0	0	1	2	12	6	3	2	1	0
6	0	1	0	1	0	0	2	1	0	1
7	0	0	1	1	0	0	0	1	0	0

Задача 13. Среднемесячная заработная плата (тыс. руб.) в Ярославской области в 2001-2002 годах составила по отраслям:

отрасль	ЖКХ	здравоохранение	наука	образование	транспорт	промышленность
2001 год	2	1,5	2,7	1,3	3,2	3,2
2002 год	3	2,8	3,6	2,4	4,9	4,5

Найдите выборочный коэффициент корреляции для заработной платы в указанные годы.

Задача 14. Имеется выборка из 10 наблюдений экономических показателей (X) и (Y).

x_i	180	172	173	169	175	170	179	170	167	174
y_i	186	180	176	171	182	166	182	172	169	177

Требуется найти выборочное уравнение регрессии Y на X . Построить выборочную линию регрессии Y на X .

Лабораторные работы

Лабораторная работа №1.

Тема: Элементы комбинаторики.

Цель работы: Закрепить навыки и умения решать комбинаторные задачи на «перестановки», «сочетания», «размещения» по формулам, развивать аналитические способности, логическое мышление.

Теоретическое обоснование:

Комбинаторика - это раздел математики, изучающий дискретные объекты, множества (сочетания, перестановки, размещения и перечисления элементов) и отношения на них (например, частичного порядка).

Перестановками называют комбинации, состоящие из одних тех же n различных элементов и отличающиеся только порядком их расположения.

Число всех возможных перестановок находится по формуле:

$$P_n = 1 \cdot 2 \cdot \dots \cdot n = n!$$

при этом, по определению, полагают, что $0! = 1$.

Размещениями называют комбинации, составленные из n различных элементов по m элементов, которые отличаются либо составом элементов, либо их порядком.

Число всех возможных размещений вычисляется по формуле:

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

Сочетаниями называют комбинации, составленные из n различных элементов по m элементов, которые отличаются хотя бы одним элементом.

Число всех возможных сочетаний находится по формуле:

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

Число размещений, число перестановок и число сочетаний связаны формулой:

$$A_n^m = P_m \cdot C_n^m.$$

При решении задач комбинаторики используют правило суммы и произведения.

Правило суммы. Если некоторый объект A может быть выбран из совокупности объектов m способами, а другой объект B может быть n способами, то выбрать либо A , либо B можно $m + n$ способами.

Правило произведения. Если некоторый объект A может быть выбран из совокупности объектов m способами и после каждого такого выбора другой объект B может быть выбран n способами, то пара объектов (A, B) в указанном порядке может быть выбрана $m \cdot n$ способами.

Порядок выполнения работы:

Пример.

Задача 1. Пусть даны шесть цифр: 1; 2; 3; 4; 5; 6. Определить сколько трехзначных чисел можно составить из этих цифр.

- Рассмотрим некоторое множество X , состоящее из n элементов $X = \{x_1, x_2, \dots, x_n\}$.
- Будем выбирать из этого множества различные упорядоченные подмножества Y из k элементов.

Размещением из n элементов множества X по k элементам назовем любой упорядоченный набор $(x_{i_1}, x_{i_2}, \dots, x_{i_k})$ элементов множества X .

Если выбор элементов множества Y из X происходит с возвращением, т.е. каждый элемент множества X может быть выбран несколько раз, то число размещений из n по k находится по формуле n^k (**размещения с повторениями**).

Если же выбор делается без возвращения, т.е. каждый элемент множества X можно выбирать только один раз, то количество размещений из n по k обозначается A_n^k и определяется равенством

$$A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

(размещения без повторений).

Решение. Если цифры могут повторяться, то количество трехзначных чисел будет $m = n^k = 6^3 = 216$. Если цифры не повторяются, то $m = A_6^3 = 6 \cdot 5 \cdot 4 = 120$.

Задача 2. Предприятие может предоставить работу по одной специальности 4 женщинами, по другой - 6 мужчинам, по третьей - 3 работникам независимо от пола. Сколькими способами можно заполнить вакантные места, если имеются 14 претендентов: 6 женщин и 8 мужчин?

Решение. Имеем 14 претендентов и 13 рабочих мест. Сначала выберем работников на первую специальность, то есть 4 женщин из 6:

$$C_4^6 = 6! / 4! \cdot 2! = 15.$$

Далее независимо аналогичным образом выберем мужчин на вторую специальность:

$$C_6^8 = 8! / 6! \cdot 2! = 28.$$

Осталось 2 женщины, 2 мужчин и 3 вакантных места, которые, по условию, могут занять любые из четырех оставшихся человек. Это может быть сделано 2 вариантами: 1. 1 женщина и 2 мужчин (выбираем женщину $C_1^2 = 2$ способами) 2. 1 мужчина и 2 женщины (выбираем мужчину $C_1^2 = 2$ способами). В итоге получаем $15 \cdot 28(2 + 2) = 1680$ способов.

Ответ: 1680 способов.

Контрольные задания:

- В пассажирском поезде 9 вагонов. Сколькими способами можно рассадить в поезде 4 человека, при условии, что все они должны ехать в различных вагонах?
- Группу из 20 студентов нужно разделить на 3 бригады, причем в первую бригаду должны входить 3 человека, во вторую — 5 и в третью — 12. Сколькими способами это можно сделать.
- Каких чисел от 1 до 1 000 000 больше: тех, в записи которых встречается единица, или тех, в которых она не встречается?

- Сколько различных дробей можно составить из чисел 3, 5, 7, 11, 13, 17 так, чтобы в каждую дробь входили 2 различных числа? Сколько среди них будет правильных дробей?
- Сколько слов можно получить, переставляя буквы в слове Гора и Институт?
- В шахматном турнире принимали участие 15 шахматистов, причем каждый из них сыграл только одну партию с каждым из остальных. Сколько всего партий было сыграно в этом турнире?

Контрольные вопросы:

- Что изучает комбинаторика?
- Сформулируйте правила суммы?
- Сформулируйте правила произведения?
- Какой формулой связаны: число размещений, число перестановок и число сочетаний?

Лабораторная работа №2.

Тема: Операции над событиями.

Цель работы: Рассмотреть события и операции над ними в теории вероятностей.

Теоретическое обоснование:

Достоверное событие - в теории вероятности называется событие U , которое в результате опыта или наблюдения непременно должно произойти.

Для достоверного события $P(U) = 1$

Т.е. вероятность события U равна единице.

Но, не всякое событие, вероятность которого равна 1, является достоверным, если оговорена некоторая допустимая погрешность (например, 10-50), то событие, вероятность которого не более чем на значение погрешности отличается от 1, называется практически достоверным.

Невозможное событие - событие V , которое в результате опыта произойти не может. Очевидно, что вероятность невозможного события равна нулю.

Однако, не всякое событие, вероятность которого равна нулю, является невозможным событием. Пример: событие, состоящее в том, что нормально распределенная случайная величина примет некоторое конкретное значение. Для любой непрерывной случайной величины верно утверждение: вероятность того, что случайная величина примет определенное, наперед заданное значение равна нулю. $P\{x=x_0\}=0$. Другой пример события с нулевой вероятностью: эксперимент состоит в том, что монета подбрасывается бесконечное число раз. Событие "Монета бесконечное число раз упадет цифрой вверх" имеет нулевую вероятность, но оно может произойти.

Если оговорена некоторая допустимая погрешность (например, 10^{-50}), то событие, вероятность которого не больше значения этой погрешности, называют практически невозможным.

Сложение вероятностей – вероятность появления одного из двух несовместимых событий равна сумме вероятностей этих событий:

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

Вероятность суммы совместных событий вычисляется по формуле:

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

Сумма вероятностей событий $A_1, A_2, \dots, A_n$, образующие полную группу, равна единице

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

Произведением событий A и B называется событие AB , которое наступает тогда и только тогда, когда наступают оба события A и B одновременно.

Случайные события A и B называются совместными, если при данном испытании могут произойти оба события.

События событий A и B называются независимыми, если появление одного события не мешает появлению другого. Событие A называется зависимым от события B , если вероятность события A меняется в зависимости от того, произошло ли событие B или нет.

Теорема об умножении вероятностей. Вероятность произведения независимых событий A и B вычисляется по формуле:

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

Противоположными называются два единственно возможных события, образующих полную группу.

События называются **несовместными**, если появление одного из них исключает появление других. То есть, может произойти только одно определённое событие, либо другое.

События называются **совместными**, если наступление одного из них не исключает наступления другого.

Порядок выполнения работы:

Пример 1. Записать пространство элементарных событий для следующих экспериментов:

- а) бросается 1 раз правильная симметричная монета,
- б) бросается 1 раз правильная игральная кость,
- в) выбирается наугад любое число, заключенное между числами a и b ,
- г) бросаются правильная игральная кость и правильная монета.

Решение:

а) Очевидно, что в этом случае пространство элементарных событий состоит из двух событий: пусть A – событие, состоящее в том, что при подбрасывании монеты она упала гербом вверх, B – упала гербом вниз. (Строго говоря, при бросании монеты не обязательно выпадает герб или решка, монета может куда-нибудь укатиться или стать на ребро. Тем не менее, мы условимся рассматривать герб или решку как единственно возможные исходы бросания монеты. Это соглашение не сказывается на возможности применения теории, но заметно упрощает ее). Следовательно, $\Omega = \{A, B\}$.

б) В этом эксперименте возможны следующие исходы: кость может упасть гранью с шестью очками вверх – событие обозначим A_6 , может упасть пятью очками вверх – событие A_5 и т. д. Т. е. пространство элементарных событий состоит из шести событий $\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\}$, где A_i ($i = 1, 2, \dots, 6$) – событие, состоящее в том, что кость упала гранью с i очками вверх.

в) Если выбирается одно число, заключенное между a и b , то пространство элементарных событий состоит из событий вида A_t ($a < t < b$), где A_t – событие, состоящее в том, что выбрано число t . Очевидно, что событий вида A_t ($a < t < b$) бесконечно много.

г) В этом случае могут быть следующие исходы опыта. Введем обозначения: $B1j, j = 1, \dots, 6$ – «кость выпала j очками вверх и монета гербом вверх», $B2j, j = 1, \dots, 6$ – «кость выпала j очками вверх и монета упала решкой вверх».

Пример 4. Опыт – подбрасывание 1 раз игрального кубика. Обозначим элементарные события A_k ($k = 1, 2, 3, 4, 5, 6$) – выпадение k очков, сложные события:

A – выпадение четного числа очков,

B – выпадение нечетного числа очков,

C – выпадение числа очков, кратного трем,

D – выпадение числа очков, большего трех,

E – число очков дробно,

F – число очков в промежутке $(0,5; 1,5)$.

Выразить события A, B, C, D, E, F через события A_k . Выявить пары совместных событий.

Решение:

Пространство элементарных исходов для данного опыта – это возможное число очков, выпавших на верхней грани, т. е. $\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\}$. Событие A наступает тогда и только тогда, когда наступает A_2 , или A_4 , или A_6 . Это означает, что $A = A_2 + A_4 + A_6$, где события A_2, A_4, A_6 – несовместные события в данном опыте. Рассуждая аналогичным образом, получаем $B = A_1 + A_3 + A_5$, $C = A_3 + A_6$, $D = A_4 + A_5 + A_6$, $E = \Omega$, $F = A_1$. Сопоставляя попарно события и проверяя наличие общих элементов, находим пары совместных событий A и C , A и D , B и C , B и D , B и F , C и D .

Контрольные задания:

- Описать пространство элементарных событий для следующих экспериментов:
 - а) подбрасывание монеты до тех пор, пока не выпадет герб;

б) два шара – красный и синий – помещают в два занумерованных ящика, причем оба шара можно положить в один ящик.

- При наличии 4-х патронов стрелок производит стрельбу до первого попадания. Описать пространство .
- Рабочий обслуживает четыре станка, каждый из которых в любой момент времени может остановиться независимо от других. В момент t рабочий проводит осмотр (смотрит, какие станки работают, а какие нет). Описать пространство элементарных событий. Сделать то же, если известно, что при остановке второго станка обязательно остановится и четвертый.
- Два парохода должны подойти к одному и тому же причалу в любой момент времени в течение данных суток. Время стоянки первого парохода один час, а второго – два часа. Описать пространство элементарных событий и событие A – одному из пароходов придется ожидать освобождения причала.
- Опыт – производится два выстрела по мишени. Описать множество элементарных исходов опыта и случайные события:

A – два попадания,

B – ровно одно попадание,

C – хотя бы одно попадание.

Отметить совместные из них.

Контрольные вопросы:

- Что такое событие?
- Что такое достоверное событие?
- Что такое невозможное событие?
- Какие события называются несовместными, а какие совместными?
- Дайте определение операции произведению двух событий?

Лабораторная работа №3.

Тема: Изучение законов распределения непрерывных случайных величин.

Цель работы: Понять, что такое непрерывные случайные величины, научиться строить графики случайных величин и искать функции их распределения.

Теоретическое обоснование:

Случайная величина - одно из основных понятий теории вероятностей. В самом общем смысле случайная величина - это некоторая переменная, принимающая те или иные значения с определенными вероятностями. Так, например, число очков, выпадающее на верхней грани игральной кости, может оказаться равным 1, 2, 3, 4, 5, 6 с вероятностями $1/6$ для каждого значения. Строгое математическое определение случайной величины как измеримой функции, заданной на некотором вероятностном пространстве, дается в рамках общепринятой аксиоматики теории вероятностей.

Дискретные и непрерывные случайные величины.

Случайная величина называется **дискретной**, если она принимает отдельные, изолированные возможные значения с определенными вероятностями. Число возможных значений дискретной случайной величины может быть конечным или бесконечным (счетным).

Случайная величина называется **непрерывной**, если она принимает все значения из некоторого конечного или бесконечного промежутка. Число возможных значений непрерывной случайной величины бесконечно.

Для описания закона распределения непрерывной случайной величины X предлагается другой подход: рассматривать не вероятности событий $X=x$ для разных x , а вероятности события $X < x$. При этом вероятность $P(X < x)$ зависит от текущей переменной, т. е. является некоторой функцией от x .

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 **интегральной функцией** распределения или **интегральным законом** распределения.

Способ задания непрерывной случайной величины с помощью функции распределения не является единственным. Необходимо определить некоторую функцию, отражающую вероятности попадания случайной точки в различные участки области возможных значений непрерывной случайной величины. Т. е. представить некоторую замену вероятностям p_i для дискретной случайной величины в непрерывном случае.

Такой функцией является плотность распределения вероятностей. Плотностью вероятности (**плотностью распределения, дифференциальной функцией**) случайной величины X называется функция $f(x)$, являющаяся первой производной интегральной функции распределения:

$$f(x)=F'(x)$$

Про случайную величину X говорят, что она имеет распределение (распределена) с плотностью $f(x)$ на определенном участке оси абсцисс.

Свойства функции распределения.

Свойство1: Значения функции распределения принадлежат отрезку $[0,1]$:

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

Свойство 2: $F(x)$ – неубывающая функция, то есть

$$F(x_2) \geq F(x_1), \text{ если } x_2 > x_1.$$

Свойство3: Если возможные значения случайной величины принадлежат интервалу (a,b) , то

$$1) F(x) = 0 \text{ при } x \leq a; 2) F(x) = 1 \text{ при } x \geq b.$$

Свойство4: Функция распределения непрерывна слева, то есть

$$F(x_0) = F(x_0 - 0).$$

Порядок выполнения работы:

Пример. Известна плотность вероятности случайной величины:

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ a \sin x & \text{при } 0 < x \leq \pi, \\ 0 & \text{при } x \geq \pi. \end{cases}$$

а) параметр а;

б) функцию распределения F(x);

в) вероятность попадания X в интервал $(-\pi/4; \pi/4)$.

Построить графики f(x), F(x).

Решение:

1. Зная свойства плотности вероятности - функции f(x), найдем неизвестный параметр а. Из неравенства $f(x) \geq 0$, делаем вывод, что $a \geq 0$.

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^0 0 \cdot dx + \int_0^{\pi} a \cdot \sin x dx + \int_{\pi}^{\infty} 0 \cdot dx = \int_0^{\pi} a \cdot \sin x dx$$

2. Вычислим данный интеграл. Зная, что его значение должно быть равно единице, выразим а.

$$\int_0^{\pi} a \cdot \sin x dx = -a \cos x \Big|_0^{\pi}$$

= $a - (-a) = 2a$. Зная, что

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

получаем $2a=1$, отсюда $a=1/2$.

3. Для нахождения функции F(x) используем формулу, определяющую интегральную функцию распределения. Так как f(x) задана различным образом на трех разных интервалах, то выражение для F(x) находим отдельно для каждого интервала.

Если $x \leq 0$

$$F(x) = \int_{-\infty}^x 0 \cdot dz = 0$$

Если $0 < x \leq \pi$, то

$$F(x) = \int_{-\infty}^0 0 \cdot dz + \int_0^x \frac{1}{2} \cdot \sin z dz = 0 + \frac{1}{2} (-\cos z) \Big|_0^x$$

Если $x > \pi$, то

$$F(x) = \int_{-\infty}^0 0 \cdot dz + \int_0^{\pi} \frac{1}{2} \cdot \sin z \, dz + \int_{\pi}^x 0 \cdot dz = \frac{1}{2} (-\cos z) \Big|_0^{\pi} = 1$$

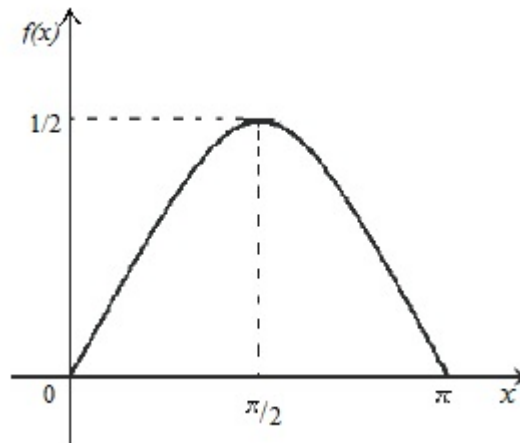
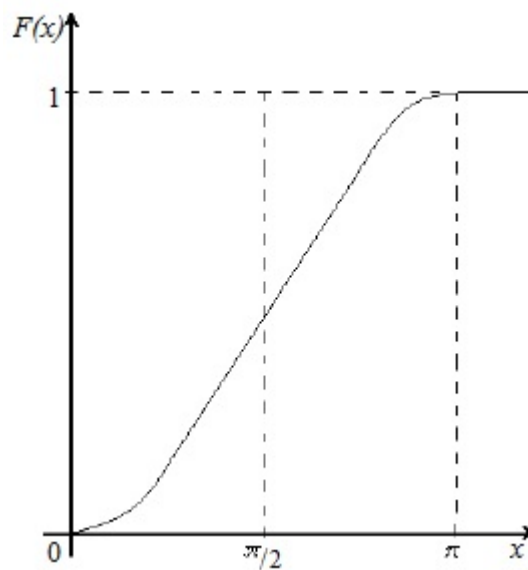


График функции $f(x)$ - плотности распределения вероятностей случайной величины.

Искомая интегральная функция принимает окончательный вид:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \frac{1}{2}(1 - \cos x) & \text{при } 0 < x \leq \pi, \\ 1 & \text{при } x > \pi; \end{cases}$$



4. Вероятность попадания случайной величины X в интервал $(-\pi/4; \pi/4)$ найдем по формуле: $P(a < x < b) = F(b) - F(a)$.

$$P(-\pi/4 < x < \pi/4) = F(\pi/4) - F(-\pi/4) = \frac{1}{2} (1 - \cos \pi/4) - 0 = \frac{1}{2} (1 - \frac{1}{2}\sqrt{2}).$$

Контрольные задания:

1. Из урны, содержащей три шара с номерами 1, 2 и 3, последовательно достают два шара. Найти математические ожидания сумм номеров извлеченных шаров, если шары достают с возвратом и без возврата.

2. Найти математические ожидания сумм выпавших очков при подбрасывании костей, если:

а) бросают один раз одну кость;

б) бросают один раз две кости;

в) бросают вместе две кости сто раз.

3. Биатлонист поражает мишень из положения лежа с вероятностью 0,9, а из положения стоя с вероятностью - 0,7. Найти математическое ожидание суммарного числа пораженных мишеней при десяти выстрелах из каждого положения.

4. Найти дисперсию для числа выпавших "гербов" при подбрасывании одной монеты; двух и ста монет.

5. Найти средние квадратичные отклонения для сумм выпавших очков при двух и десяти подбрасываниях игральной кости.

Контрольные вопросы:

1. Дайте определение случайной величины?

2. Случайная величина называется дискретной, если?

3. Случайная величина называется, если?

4. Формула распределения функции $F(X)$?

5. О чем гласит четвертое свойство функции распределения?

Лабораторная работа №4.

Тема: Изучение законов распределения непрерывных случайных величин.

Цель работы: Моделирование непрерывных случайных величин, получение случайных выборок заданного объема, оценка точности моделирования.

Теоретическое обоснование:

Характеристическая функция.

Характеристическая функция случайной величины — один из способов задания распределения. Характеристические функции могут быть удобнее в тех случаях, когда, например, плотность или функция распределения имеют очень сложный вид. Также характеристические функции являются удобным инструментом для изучения вопросов слабой сходимости (сходимости по распределению).

- Пусть есть случайная величина X с распределением P^X . Тогда характеристическая функция задаётся формулой:

$$\phi_X(t) = \mathbb{E} [e^{itX}]$$

Пользуясь формулами для вычисления математического ожидания, определение характеристической функции можно переписать в виде:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} P^X(dx)$$

то есть характеристическая функция — это преобразование Фурье распределения случайной величины.

Если случайная величина X принимает значения в произвольном гильбертовом пространстве H , то её характеристическая функция имеет вид:

$$\phi_X(t) = \mathbb{E} [e^{i\langle t, X \rangle}] , \quad \forall t \in H$$

Свойства характеристических функций.

- Характеристическая функция однозначно определяет распределение. Пусть X, Y — две случайные величины, и

$$\phi_X(t) = \phi_Y(t), \quad \forall t, \text{ тогда}$$

$$P^X = P^Y.$$

В частности, если обе величины абсолютно непрерывны, то совпадение характеристических функций влечёт совпадение плотностей. Если обе случайные величины дискретны, то совпадение характеристических функций влечёт совпадение функций вероятности.

- Характеристическая функция всегда ограничена:

$$|\phi_X(t)| \leq 1$$

- Характеристическая функция в нуле равна единице:

$$\phi_X(0) = 1$$

- Характеристическая функция всегда непрерывна:

$$\phi_X \in C(\mathbb{R})$$

- Характеристическая функция как функция случайной величины однородна:

$$\phi_{aX}(t) = \phi_X(at), \quad \forall a \in \mathbb{R}.$$

Нормальный закон распределения(закон Гаусса).

Нормальный закон распределения (часто называемый законом Гаусса) играет исключительно важную роль в теории вероятностей и занимает среди других законов распределения особое положение. Это – наиболее часто встречающийся на практике закон распределения. Главная особенность, выделяющая нормальный закон среди других законов, состоит в том, что он является предельным законом, к которому приближаются другие законы распределения при весьма часто встречающихся типичных условиях.

Можно доказать, что сумма достаточно большого числа независимых (или слабо зависимых) случайных величин, подчиненных каким угодно законам распределения (при соблюдении некоторых весьма нежестких ограничений), приближенно подчиняется нормальному закону, и это выполняется тем точнее, чем большее количество случайных величин суммируется. Большинство встречающихся на практике случайных величин, таких, например, как ошибки измерений, ошибки стрельбы и т.д., могут быть представлены как суммы весьма большого числа сравнительно малых слагаемых – элементарных ошибок, каждая из которых вызвана действием отдельной причины, не зависящей от остальных. Каким бы законам распределения ни были подчинены отдельные элементарные ошибки, особенности этих распределений в сумме большого числа слагаемых нивелируются, и сумма оказывается подчиненной закону, близкому к нормальному. Основное ограничение, налагаемое на суммируемые ошибки, состоит в том, чтобы они все равномерно играли в общей сумме относительно малую роль. Если это условие не выполняется и, например,

одна из случайных ошибок окажется по своему влиянию на сумму резко преобладающей над всеми другими, то закон распределения этой преобладающей ошибки наложит свое влияние на сумму и определит в основных чертах её закон распределения.

- Плотность вероятности имеет следующий вид:

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

- Функция распределения такой величины не выражается через элементарные функции и записывается по определению через интеграл Римана как:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt.$$

- Функция распределения стандартной нормальной случайной величины:

$$\Phi(x) = F(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

- Характеристическая функция нормального распределения имеет вид:

$$f(t) = E\{e^{it\xi}\} = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right),$$

- Производящая функция моментов задается формулой:

$$M(t) = E\{e^{t\xi}\} = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

Логнормальное распределение.

Логнормальное распределение в теории вероятностей — это двухпараметрическое семейство абсолютно непрерывных распределений. Если случайная величина имеет логнормальное распределение, то её логарифм имеет нормальное распределение.

- Функция распределения:

$$\frac{1}{2} + \frac{1}{2} \operatorname{Erf}\left[\frac{\ln(x) - \mu}{\sigma\sqrt{2}}\right]$$

- Плотность вероятности:

$$\exp \left(- \left[\frac{\ln(x) - \mu}{\sigma} \right]^2 / 2 \right) / (x\sigma\sqrt{2\pi})$$

Свойства логнормального распределения.

Если $X_1, \dots, X_n$ — независимые логнормальные случайные величины, такие что $X_i \sim \text{LogN}(\mu, \sigma_i^2)$, то их произведение также логнормально:

$$Y = \prod_{i=1}^n X_i \sim \text{LogN} \left(n\mu, \sum_{i=1}^n \sigma_i^2 \right)$$

Гамма-распределение.

Гамма распределение в теории вероятностей — это двухпараметрическое семейство абсолютно непрерывных распределений. Если параметр k принимает целое значение, то такое гамма-распределение также называется распределением Эрланга.

- Характеристическая функция:

$$(1 - \theta i t)^{-k}$$

- Производящая функция моментов:

$$(1 - \theta t)^{-k}, \text{ когда } t < 1/\theta$$

- Функция распределения:

$$\frac{\gamma(k, x/\theta)}{\Gamma(k)}$$

- Плотность вероятности:

$$x^{k-1} \frac{\exp(-x/\theta)}{\Gamma(k) \theta^k}$$

Свойства гамма-распределения.

- Если $X_1, \dots, X_n$ — независимые случайные величины, такие что $X_i \sim \Gamma(k_i, \theta)$, $i = 1, \dots, n$, то

$$Y = \sum_{i=1}^n X_i \sim \Gamma \left(\sum_{i=1}^n k_i, \theta \right)$$

- Если $X \sim \Gamma(k, \theta)$ и $a < 0$ — произвольная константа, то

$$aX \sim \Gamma(k, a\theta)$$

- Гамма-распределение бесконечно делимо.

Экспоненциальное распределение.

Экспоненциальное распределение — абсолютно непрерывное распределение, моделирующее время между двумя последовательными свершениями одного и того же события.

- Функция экспоненциального распределения:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & , x \geq 0, \\ 0 & , x < 0. \end{cases}$$

- Производящая функция моментов:

$$\left(1 - \frac{t}{\lambda}\right)^{-1}$$

Распределение Вейбулла.

Распределение Вейбулла (иначе — распределение Вейбулла-Гнеденко) в теории вероятностей — двухпараметрическое семейство абсолютно непрерывных распределений. Названо в честь шведского инженера Валодди Вейбулла (Waloddi Weibull, 1887—1979).

- Функция распределения:

$$1 - e^{-(x/\lambda)^k}$$

- Коэффициент асимметрии:

$$\frac{\Gamma(1 + \frac{3}{k})\lambda^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

- Дисперсия:

$$\lambda^2\Gamma\left(1 + \frac{2}{k}\right) - \mu^2$$

Равномерный закон распределения.

Равномерным называется распределение вероятностей непрерывной случайной величины X , плотность которого сохраняет постоянное значение на отрезке $[a;b]$ и имеет вид

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

- Функция распределения:

$$F(x) = \begin{cases} 0 & \text{при } x \leq a, \\ \frac{x-a}{b-a} & \text{при } a < x \leq b, \\ 1 & \text{при } x > b. \end{cases}$$

- Дисперсия:

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

- Вероятность попадания в интервал:

$$P(x_1 < X < x_2) = \frac{x_2 - x_1}{b - a}$$

Показательный закон распределения.

Показательным (экспоненциальным) называется распределение вероятностей непрерывной случайной величины X , которое описывается плотностью, имеющей вид

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

- Функция распределения:

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

- Дисперсия:

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

- Вероятность попадания в интервал:

$$P(a < X < b) = \int_a^b \lambda e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}$$

Порядок выполнения работы:

Пример. Автобусы идут строго по расписанию. Интервал движения 7 мин.

Найти: а) вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее двух минут; б) вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус не менее трех минут; в) математическое ожидание и среднее квадратическое отклонение случайной величины X – времени ожидания пассажира.

Решение:

1. По условию задачи непрерывная случайная величина $X = \{\text{время ожидания пассажира}\}$ равномерно распределена между приходами двух автобусов. Длина интервала распределения случайной величины X равна $b-a=7$, где $a=0$, $b=7$.

2. Время ожидания будет менее двух минут, если случайная величина X попадает в интервал $(5;7)$. Вероятность попадания в заданный интервал найдем по формуле: $P(x_1 < X < x_2) = (x_2 - x_1) / (b - a)$.

$$P(5 < X < 7) = (7-5)/(7-0) = 2/7 \approx 0,286.$$

3. Время ожидания будет не менее трех минут (т.е. от трех до семи мин.), если случайная величина X попадает в интервал $(0;4)$. Вероятность попадания в заданный интервал найдем по формуле: $P(x_1 < X < x_2) = (x_2 - x_1) / (b - a)$.

$$P(0 < X < 4) = (4-0)/(7-0) = 4/7 \approx 0,571.$$

4. Математическое ожидание непрерывной, равномерно распределенной случайной величины X – времени ожидания пассажира, найдем по формуле: $M(X) = (a+b)/2$. $M(X) = (0+7)/2 = 7/2 = 3,5$.

5. Среднее квадратическое отклонение непрерывной, равномерно распределенной случайной величины X – времени ожидания пассажира, найдем по формуле: $\sigma(X) = \sqrt{D} = (b-a)/2\sqrt{3}$. $\sigma(X) = (7-0)/2\sqrt{3} = 7/2\sqrt{3} \approx 2,02$.

Контрольные задания:

1. Показательное распределение задано при $x \geq 0$ плотностью $f(x) = 5e^{-5x}$. Требуется: а) записать выражение для функции распределения; б) найти вероятность того, что в результате испытания X попадает в интервал $(1;4)$; в) найти вероятность того, что в результате испытания $X \geq 2$; г) вычислить $M(X)$, $D(X)$, $\sigma(X)$.

2. Найти вероятность попадания случайной величины, имеющей равномерное распределение на отрезке $[a;b]$, на участок $(\alpha;\beta)$, представляющий собой часть отрезка $[a;b]$.

3. Случайная величина X — время работы радиолампы — имеет показательное распределение. Определить вероятность того, что время работы лампы будет не меньше 600 ч, если среднее время работы

радиолампы 400 ч.

4. Найти вероятность того, что нормально распределённая случайная величина X удовлетворяет неравенству $a < X < b$.

5. Компьютерная система содержит 50 одинаковых микрочипов. Вероятность того, что любой микрочип будет работать в заданное время, равна 0,9. Для выполнения некоторой операции требуется, чтобы, по крайней мере, 30 микрочипов было в рабочем состоянии.

а) Чему равна вероятность того, что операция будет выполнена успешно?

б) Чему равна вероятность того, что будут работать 47 микрочипов?

6. Экзамен по математической статистике успешно сдают 75% студентов дневного отделения. Если на втором курсе факультета обучается 250 студентов, то какова вероятность того, что 203 студента сдадут экзамен успешно?

Контрольные вопросы:

1. Определение характеристической функции?
2. Свойства характеристических функций?
3. Особенность закона Гаусса?
4. Формула плотности вероятности закона Гаусса?
5. определение логнормального распределения?
6. Функция логнормального распределения?
7. Определения гамма-распределения?
8. Определение экспоненциального распределения?
9. Формула функции распределения показательного закона?

.

Лабораторная работа №5.

Тема: Генеральная и выборочная совокупности.

Цель работы: Освоение методики организации и проведения выборочного наблюдения; статистических методов, методов оценки параметров генеральной совокупности на основе выборочных данных.

Теоретическое обоснование:

Выборочное наблюдение - важнейший вид не сплошного наблюдения. Теория выборочного наблюдения, т.н. выборочный метод, - совокупность принципов и способов отбора единиц совокупности, а также способов и методов оценки параметров генеральной совокупности на основе выборочных единиц. Выборочный метод в настоящее время получил широкое практическое применение, поскольку обладает целым рядом преимуществ по сравнению со сплошным наблюдением и иными видами несплошного наблюдения.

Преимущества выборочного наблюдения по сравнению со сплошным:

1. Экономия времени, финансовых, трудовых, материальных ресурсов.
2. Возможность расширить программу наблюдения.

С другими видами не сплошного наблюдения:

3. Благодаря хорошо разработанной теории выборки и используемых при выборочных наблюдениях способах формирования выборки появляется возможность дать вероятностную оценку параметров генеральной совокупности.

Генеральная совокупность - совокупность, которая собственно интересует исследователя и из которой отбираются единицы в выборочную совокупность. Выборочная совокупность - совокупность отобранных единиц, по которым будут фиксироваться значения тех или иных признаков.

Основной принцип формирования выборочной совокупности - случайность отбора, т.е. всем единицам генеральной совокупности должна быть обеспечена равная вероятность попадания в выборку. Этот принцип обеспечивает объективность выборочного наблюдения, поскольку позволяет сформировать репрезентативную выборку. Репрезентативность способствует получению несмещённой выборки, т.е. структура или закономерность распределения в выборочной совокупности соответствует распределению единиц в генеральной совокупности.

Способы отбора единиц в выборочную совокупность:

1. Случайный отбор. Реализуют методом жеребьёвки или с использованием таблиц случайных чисел.
2. Механический отбор - частный случай случайного отбора.

Рассчитывается шаг отбора, который равен отношению объёма совокупности к объёму выборки.

Отбор может проводиться по принципу бесповторного отбора, когда, извлекаемая из генеральной совокупности, единица назад не возвращается, и повторного отбора.

Основу статистического исследования составляет множество данных, полученных в результате измерения одного или нескольких признаков. Реально наблюдаемая совокупность объектов, статистически представленная рядом наблюдений случайной величины, является **выборкой**, а гипотетически существующая (домысливаемая) — генеральной совокупностью. **Генеральная совокупность** может быть конечной (число наблюдений $N = \text{const}$) или бесконечной ($N = \infty$), а выборка из генеральной совокупности — это всегда результат ограниченного ряда наблюдений. Число наблюдений, образующих выборку, называется **объемом выборки**. Если объем выборки достаточно велик ($n \rightarrow \infty$) выборка считается **большой**, в противном случае она называется выборкой ограниченного объема. Выборка считается **малой**, если при измерении одномерной случайной величины объем выборки не превышает 30 ($n \leq 30$), а при измерении одновременно нескольких (**k**) признаков в многомерном пространстве отношение n к k не превышает 10 ($n/k < 10$). Выборка образует вариационный ряд, если ее члены являются порядковыми статистиками, т. е. выборочные значения случайной величины **X** упорядочены по возрастанию (ранжированы), значения же признака называются вариантами.

Пример. Практически одна и та же случайно отобранная совокупность объектов — коммерческих банков одного административного округа Москвы, может рассматриваться как выборка из генеральной совокупности всех коммерческих банков этого округа, и как выборка из генеральной совокупности всех коммерческих банков Москвы, а также как выборка из коммерческих банков страны и т.д.

Основные способы организации выборки.

Достоверность статистических выводов и содержательная интерпретация результатов зависит от репрезентативности выборки, т.е. полноты и адекватности представления свойств генеральной совокупности, по отношению к которой эту выборку можно считать представительной.

Изучение статистических свойств совокупности можно организовать двумя способами: с помощью сплошного и несплошного наблюдения . Сплошное наблюдение предусматривает обследование всех единиц изучаемой совокупности, а несплошное (выборочное) наблюдение — только его части.

Существуют пять основных способов организации выборочного наблюдения:

1. Простой случайный отбор, при котором объектов случайно извлекаются из генеральной совокупности объектов (например с помощью таблицы или датчика случайных чисел), причем каждая из возможных выборок имеют равную вероятность. Такие выборки называются собственно-случайными;

2. Простой отбор с помощью регулярной процедуры осуществляется с помощью механической составляющей (например, даты, дня недели, номера квартиры, буквы алфавита и др.) и полученные таким способом выборки называются механическими;

3. Стратифицированный отбор заключается в том, что генеральная совокупность объема подразделяется на подсовкупности или слои (страты) объема так что . Страты представляют собой однородные объекты с точки зрения статистических характеристик (например, население делится на страты по возрастным группам или социальной принадлежности; предприятия — по отраслям). В этом случае выборки называются стратифицированными (иначе, расслоенными, типическими, районированными);

4. Методы серийного отбора используются для формирования серийных или гнездовых выборок. Они удобны в том случае, если необходимо обследовать сразу "блок" или серию объектов (например, партию товара, продукцию определенной серии или население при территориально-административном делении страны). Отбор серий можно осуществить собственно-случайным или механическим способом. При этом проводится сплошное обследование определенной партии товара, или целой территориальной единицы (жилого дома или квартала);

5. Комбинированный (ступенчатый) отбор может сочетать в себе сразу несколько способов отбора (например, стратифицированный и случайный или случайный и механический); такая выборка называется комбинированной.

Виды отбора.

По виду различаются индивидуальный, групповой и комбинированный отбор. При **индивидуальном отборе** в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при **групповом отборе** — качественно однородные группы (серии) единиц, а комбинированный отбор предполагает сочетание первого и второго видов.

По методу отбора различают **повторную и бесповторную** выборку.

Бесповторным называется отбор, при котором попавшая в выборку единица не возвращается в исходную совокупность и в дальнейшем выборе не участвует; при этом численность единиц генеральной совокупности N сокращается в процессе отбора. При повторном отборе попавшая в выборку единица после регистрации возвращается в генеральную совокупность и таким образом сохраняет равную возможность наряду с другими единицами быть использованной в дальнейшей процедуре отбора; при этом численность единиц генеральной совокупности N остается неизменной (метод в социально-экономических исследованиях применяется редко). Однако, при большом N ($N \rightarrow \infty$) формулы для бесповторного отбора приближаются к аналогичным для повторного отбора и практически чаще используются последние ($N = \text{const}$).

Основные характеристики параметров основной и выборочной совокупности.

В основе статистических выводов проведенного исследования лежит распределение случайной величины, наблюдаемые же значения ($x_1, x_2, \dots, x_n$) называются **реализациями случайной величины X (n — объем выборки)**. Распределение случайной величины в генеральной совокупности носит теоретический, идеальный характер, а ее выборочный аналог является эмпирическим распределением. Некоторые теоретические распределения заданы аналитически, т.е. их параметры определяют значение функции распределения в каждой точке пространства возможных значений случайной величины. Для выборки же функцию распределения определить трудно, а иногда невозможно, поэтому параметры оценивают по эмпирическим данным, а затем их подставляют в аналитическое выражение, описывающее теоретическое распределение. При этом предположение (или гипотеза) о виде распределения может быть как статистически верным, так и ошибочным. Но в любом случае восстановленное по выборке эмпирическое распределение лишь грубо характеризует истинное. Важнейшими

параметрами распределений являются математическое ожидание и дисперсия .

По своей природе распределения бывают **непрерывными и дискретными**. Наиболее известным непрерывным распределением является **нормальное**. Выборочными аналогами параметров и для него являются: **среднее значение и эмпирическая дисперсия** . Среди дискретных в социально-экономических исследованиях наиболее часто применяется альтернативное **(дихотомическое) распределение**. Параметр математического ожидания этого распределения выражает относительную величину (или долю) единиц совокупности, которые обладают изучаемым признаком (она обозначена буквой **p**); доля совокупности, не обладающая этим признаком, обозначается буквой **q** (**q = 1 — p**). Дисперсия же альтернативного распределения также имеет эмпирический аналог .

В зависимости от вида распределения и от способа отбора единиц совокупности по-разному вычисляются характеристики параметров распределения. Основные из них для теоретического и эмпирического распределений приведены в таблице 1.

Долей выборки **kn** называется отношение числа единиц выборочной совокупности к числу единиц генеральной совокупности:

$$kn = n/N.$$

Выборочная доля **w** — это отношение единиц, обладающих изучаемым признаком **x** к объему выборки **n**:

$$w = nn/n.$$

Характеристика параметров распределения	Совокупность	
	генеральная	выборочная
Объем выборки	N	n
Альтернативный признак		
Численность единиц совокупности, обладающих признаком x	N_x	n_x
Доля единиц, обладающих изучаемым признаком x	$p = \frac{N_x}{N}$	$w = \frac{n_x}{n}$
Дисперсия	$\sigma^2 = p(1 - p)$	$\sigma^2 = w(1 - w)$
Среднее квадратическое отклонение	$\sigma = \sqrt{p(1 - p)}$	$\sigma = \sqrt{w(1 - w)}$
Количественный признак		
Среднее значение признака	$\mu = \frac{\sum x_i}{N}$	$\bar{x} = \frac{\sum x_i}{n}$
Дисперсия	$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N - 1}$	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$
Среднее квадратическое отклонение	$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N - 1}}$	$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$

Основные параметры генеральной и выборочной совокупности (таблица 1).

Ошибки выборки.

При любом статистическом наблюдении (сплошном и выборочном) могут встретиться ошибки двух видов: **регистрации и репрезентативности**. Ошибки **регистрации** могут иметь случайный и систематический характер. Случайные ошибки складываются из множества различных неконтролируемых причин, носят непреднамеренный характер и обычно по совокупности уравнивают друг друга (например, изменения показателей прибора при температурных колебаниях в помещении).

Систематические ошибки тенденциозны, так как нарушают правила отбора объектов в выборку (например, отклонения в измерениях при изменении настройки измерительного прибора).

Пример. Для оценки социального положения населения в городе предусмотрено обследовать 25% семей. Если при этом выбор каждой четвертой квартиры основан на ее номере, то существует опасность отобрать все квартиры только одного типа (например, однокомнатные), что обеспечит систематическую ошибку и исказит результаты; выбор же номера квартиры

по жребию более предпочтителен, так как ошибка будет случайной.

Ошибки **репрезентативности** присущи только выборочному наблюдению, их невозможно избежать и они возникают в результате того, что выборочная совокупность не полностью воспроизводит генеральную. Значения показателей, получаемых по выборке, отличаются от показателей этих же величин в генеральной совокупности (или получаемых при сплошном наблюдении).

Ошибка выборочного наблюдения есть разность между значением параметра в генеральной совокупности и ее выборочным значением. Для среднего значения количественного признака она равна: $\varepsilon_{\bar{x}} = |\mu - \bar{x}|$, а для доли (альтернативного признака) $\varepsilon_w = |p - w|$.

Ошибки выборки свойственны только выборочным наблюдениям. Чем больше эти ошибки, тем больше эмпирическое распределение отличается от теоретического. Параметры эмпирического распределения и являются случайными величинами, следовательно, ошибки выборки также являются случайными величинами, могут принимать для разных выборок разные значения и поэтому принято вычислять **среднюю ошибку**.

Средняя ошибка выборки есть величина $m = \sqrt{\frac{\sigma^2}{n}}$, выражающая среднее квадратическое отклонение выборочной средней от математического ожидания. Эта величина при соблюдении принципа случайного отбора зависит прежде всего от объема выборки и от степени варьирования признака: чем больше n и чем меньше вариация признака (следовательно, и значение), тем меньше величина средней ошибки выборки m . Соотношение между дисперсиями генеральной и выборочной совокупностей выражается формулой: $\sigma^2 = s^2 * \frac{n}{n-1}$, т.е. при достаточно больших n можно считать, что $Q=S$. Средняя ошибка выборки показывает возможные отклонения параметра выборочной совокупности от параметра генеральной. В таблице 2 приведены выражения для вычисления средней ошибки m выборки при разных методах организации наблюдения.

Вид выборки	Отбор	
	Повторный	Бесповторный
Количественный признак		
Собственно-случайная	$m_{\bar{x}} = \sqrt{\frac{s^2}{n}}$	$m_{\bar{x}} = \sqrt{\frac{s^2(1-n/N)}{n}}$
Механическая	–	– ^н –
Типическая (стратифицированная)	$m_{\bar{x}} = \sqrt{\frac{s_i^2}{n}}$	$m_{\bar{x}} = \sqrt{\frac{s_i^2(1-n/N)}{n}}$
Серийная	$m_{\bar{x}} = \sqrt{\frac{\delta_x^2}{r}}$	$m_{\bar{x}} = \sqrt{\frac{\delta_x^2(1-r/R)}{r}}$
Альтернативный признак		
Собственно-случайная	$m_p = \sqrt{\frac{w(1-w)}{n}}$	$m_p = \sqrt{\frac{w(1-w)(1-n/N)}{n}}$
Механическая	–	– ^н –
Типическая (стратифицированная)	$m_p = \sqrt{\frac{w_i(1-w_i)}{n}}$	$m_p = \sqrt{\frac{w_i(1-w_i)(1-n/N)}{n}}$
Серийная	$m_{\bar{x}} = \sqrt{\frac{\delta_w^2}{r}}$	$m_{\bar{x}} = \sqrt{\frac{\delta_w^2(1-r/R)}{r}}$

Средняя ошибка (m) выборочных средних и доли для разных видов выборки(таблица 2).

Порядок работы:

Пример. Вычисление среднего значения и доверительного интервала для непрерывного количественного признака.

Для оценки скорости расчета с кредиторами в банке проведена случайная выборка 10 платежных документов. Их значения оказались равными (в днях): 10; 3; 15; 15; 22; 7; 8; 1; 19; 20.

Необходимо с вероятностью $P = 0,954$ определить предельную ошибку Δ выборочной средней и доверительные пределы среднего времени расчетов.

Решение.

1. Среднее значение вычисляется по формуле из таблицы 1 для выборочной совокупности

$$\bar{x} = \frac{\sum x_i}{n} = 12 \text{ дней.}$$

2. Дисперсия вычисляется по формуле из таблицы 1

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = 53 \text{ дня.}$$

3. Средняя квадратическая погрешность $Q=7.3$ дня.

4. Ошибка средней вычисляется по формуле:

$$m = \frac{\sigma}{\sqrt{n}} = 2,3, \text{ т.е. среднее значение равно } \bar{x} \pm m = 12,0 \pm 2,3 \text{ дней.}$$

5. Достоверность среднего составила

$$t = \frac{\bar{x}}{m} = \frac{12}{2,3} = 5,21.$$

6. Предельную ошибку вычислим по формуле для повторного отбора, так как численность генеральной совокупности N неизвестна, и для $P = 0,954$ уровня достоверности.

$$\Delta_{\bar{x}} = \pm t m_{\bar{x}} = 2 \times 2,3 = 4,6 \text{ дней,}$$

Таким образом, среднее значение равно $\bar{x} \pm D = \bar{x} \pm 2m = 12,0 \pm 4,6$, т.е. его истинное значение лежит в пределах от 7,4 до 16,6 дней.

Контрольные задания:

1. При механическом выборочном способе обследования социального положения 1000 семей выявлено, что доля малообеспеченных семей составила $w = 0,3$ (30%) (выборка была 2%, т.е. $n/N = 0,02$). Необходимо с уровнем достоверности $p = 0,997$ определить показатель p малообеспеченных семей во всем регионе.

2. На основе случайного повторного выборочного обследования в отделении связи города предполагается определить долю писем частных лиц в общем объеме отправляемой корреспонденции. Никаких предварительных данных об удельном весе этих писем в общей массе отправляемой корреспонденции не имеется. Определите численность выборки, если результаты выборки необходимо дать с точностью до 1% и гарантировать это с вероятностью 0,95.

3. В городе 500 тыс. жителей. По материалам учета городского населения было обследовано 50 тыс. жителей методом случайного бесповторного отбора. В результате обследования установлено, что в городе 15% жителей старше 60 лет. Определите с вероятностью 0,683 пределы, в которых находится доля жителей в городе в возрасте старше 60 лет.

4. В области, состоящей из 20 районов, проводилось выборочное обследование урожайности на основе бесповторного отбора серий (районов). Выборочные средние по районам составили соответственно 14,5 ц/га; 16,0; 15,5; 15,0 и 14,0 ц/га. Определите с вероятностью 0,954 пределы урожайности во всей области.

5. При проверке веса импортируемого груза на таможне методом случайной

повторной выборки отобрано 200 изделий. В результате был установлен средний вес изделия 30 г при среднем квадратическом отклонении 4 г. Определите с вероятностью 0,9973 пределы, в которых находится средний вес изделий в генеральной совокупности.

6. В 100 туристических агентствах города предполагается провести обследование среднемесячного количества реализованных путевок методом механического отбора. Какова должна быть численность выборки, чтобы с вероятностью 0,683 ошибка не превышала 3 путевок, если по данным пробного обследования дисперсия составляет 225?

7. Произведено выборочное наблюдение партии однородной продукции для определения процента изделий высшего сорта. При механическом способе отбора из партии готовых изделий в 20000 единиц было обследовано 800 единиц, из которых 640 изделий отнесены к высшему сорту. Определите с вероятностью 0,9973 возможный процент изделий высшего сорта во всей партии.

8. При обследовании 100 образцов изделий, отобранных из партии в случайном порядке, оказалось 20 нестандартных. Определите с вероятностью 0,954 пределы, в которых находится доля нестандартной продукции в партии.

Контрольные вопросы:

1. Что такое генеральная совокупность?
2. Дайте определение выборочного наблюдения?
3. Преимущества выборочного наблюдения по сравнению со сплошным?
4. Перечислите основные способы организации выборки?
5. Какие ошибки встречаются при статистическом наблюдении?
6. Что такое средняя ошибка выборки?
7. Какими бывают распределения?

Лабораторная работа №6.

Тема: Вычисление основных характеристик выборки.

Цель работы: Усвоить основные понятия, изучить основные статистические характеристики ряда результатов измерений, приобрести практические знания расчета указанных характеристик.

Теоретическое обоснование:

После установления основной структуры статистических данных с помощью таблиц распределения частот вариационного ряда на следующем этапе статистического анализа необходимо указать небольшое число показателей, при помощи которых можно в сжатом виде охарактеризовать все распределение частот наблюдений случайной выборки.

Такими основными показателями являются:

- выборочные функции распределения,
- выборочные меры центра распределения наблюдений,
- выборочные меры рассеяния (или изменчивости) наблюдений,
- выборочная мера асимметрии,
- выборочная мера эксцесса (островершинности) наблюдений.

Выборочные меры центра распределения наблюдений

К выборочным мерам центра распределения наблюдений случайной выборки принято относить выборочное среднее значение, выборочную медиану и выборочную моду.

Выборочное среднее значение – наиболее часто используемый показатель центра распределения наблюдений. Для наблюдений $x(1), x(2), \dots, x(n)$ его обычно обозначают $\bar{x}$ (или $\bar{x}(n)$) и определяют формулой:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Если воспользоваться знаком Σ , то $\bar{x}$ записывают в виде:

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

Для дискретного вариационного ряда можно использовать такую формулу:

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

Для выборок большого объема выборочное среднее значение находят лишь после построения интервального вариационного ряда по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i \cdot \bar{x}_i$$

где n – общее число наблюдений, k – число интервалов, $\bar{x}_i$ – середина i -го интервала, n_i – частота в i -ом интервале.

Следует заметить, что формула $\bar{x}$ для интервального вариационного ряда рекомендуется лишь для симметричного или умеренно симметричного распределения частот по интервалам.

Отметим некоторые свойства выборочных средних значений, применяемых на практике.

1. Если наблюдения выборки имеют вид $sx_1, sx_2, \dots, sx_n$, где s – заданное число, то среднее значение имеет вид $s\bar{x}$, где $\bar{x}$ – среднее значение для $x_1, x_2, \dots, x_n$.

2. Если наблюдения выборки имеют вид $x_1+y_1, x_2+y_2, \dots, x_n+y_n$, то среднее значение имеет вид $\bar{x}+\bar{y}$, где $\bar{x}$ – среднее значение для наблюдений $x_1, x_2, \dots, x_n$, а $\bar{y}$ – среднее значение для наблюдений $y_1, y_2, \dots, y_n$.

3. Если наблюдения выборки объема n являются объединением наблюдений двух выборок объемов n_1 и n_2 , то выборочное среднее объединенной выборки выражается через выборочные средние $\bar{x}_1, \bar{x}_2$ ее составных частей по формуле:

$$\bar{x} = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2.$$

4. Сумма отклонений результатов наблюдений от среднего значения равно нулю, т.е.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

5. Сумма квадратов отклонений результатов наблюдений от среднего значения меньше, чем сумма квадратов отклонений от любого другого значения, т.е.

$$\sum_{i=1}^n (x_i - \bar{x})^2 < \sum_{i=1}^n (x_i - x)^2 \quad \text{при всех } x \neq \bar{x}.$$

Выборочное среднее значение является наиболее важной характеристикой наблюдений выборки. Оно является центром рассеяния данных выборки. При сравнении двух выборок значений изучаемого признака, прежде всего, сравнивают между собой средние значения для этих выборок.

Если для наблюдений выборки $x_1, x_2, \dots, x_n$ построен дискретный вариационный ряд $x(1), x(2), \dots, x(n)$, где $x(1), x(2), \dots, x(n)$, то выборочной медианой x_{med} или $x_{med}(n)$ называется такое число, которое делит вариационный ряд на две равные части. При этом x_{med} может быть наблюдением выборки, а может им не быть. Если объем n выборки нечетное число, т.е. $n=2k+1$, то $x_{med}=x(k+1)$. Если же объем n выборки четное число, т.е. $n=2k$,

$$x_{med} = \frac{1}{2} (x_{(k)} + x_{(k+1)})$$

Выборочная медиана в ряде случаев имеет некоторые преимущества по сравнению с выборочным средним. Она может быть вычислена в случаях, когда вычисление выборочного среднего не возможно, например, вследствие неопределенности первого или последнего интервала группировки данных выборки в интервальном вариационном ряде.

Кроме того, выборочную медиану легко вычислять и она более устойчива по сравнению с выборочным средним при резких выбросах в наблюдениях выборки. Однако выборочная медиана имеет и свои недостатки по сравнению с выборочным средним. Например, для выборочной медианы невозможны аналоги свойств 1)-5) для выборочного среднего. Другие различия между $\bar{x}$ и x_{med} будут ясны далее.

К выборочным характеристикам центра группирования, условно относят и выборочную моду x_{mod} . Выборочной модой x_{mod} называется наиболее часто встречающееся значение среди наблюдений выборки.

Например, для выборки 1, 2, 5, 5, 5, 4, 4 мода $x_{mod}=5$, а для выборки 1, 2, 5, 5, 5, 4, 4, 4 имеем две моды $x(1)_{mod}=5$, $x(2)_{mod}=4$. Таким образом, в отличие от $\bar{x}$ и x_{med} выборочная мода x_{mod} может быть не единственной. В таких случаях говорят об одномодальном, двухмодальном и т.д. распределении наблюдений выборки.

Выбор , x_{med} и x_{mod} зависит, с одной стороны, от природы данных, а с другой – от того, как этот показатель будет использоваться. Для дальнейшего теоретиковероятностного и статистического анализа наиболее полезно.

Выборочные меры рассеяния наблюдений.

К выборочным мерам рассеяния относятся: выборочная дисперсия, выборочное среднее квадратическое отклонение, выборочный коэффициент вариации, выборочное квартильное отклонение.

Наиболее часто употребляемыми мерами рассеяния наблюдений выборки относительно центра группирования являются выборочная дисперсия, обозначаемая s^2 (или $s^2(n)$) и выборочное среднее квадратическое (или стандартное) отклонение s (или $s(n)$). Если наблюдения $x(1), x(2), \dots, x(n)$ имеют выборочное среднее , то по определению

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} > 0.$$

Для данных, сгруппированных в интервальный вариационный ряд формула выборочной дисперсии выглядит так:

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

Здесь n – общее число наблюдений, k – число интервалов, $\bar{x}$ – выборочное среднее всех данных, n_i – частота в i -ом интервале, x_i – центр i -го интервала.

Выборочные дисперсия и стандартное отклонение выражаются в единицах измерения рассматриваемого количественного признака и непригодны, например, при сравнении степеней рассеяния наблюдений двух выборок различной природы. Для сравнения таких степеней рассеяния используется выборочный коэффициент вариации:

$$\hat{V} = \frac{s}{\bar{x}} \cdot 100\%$$

Квартильное отклонение Q определяют формулой:

$$Q = \frac{1}{2} (Q_3 - Q_1).$$

Положение Q_1 определяется номером, ближайшим к числу:

$$\frac{n+1}{4},$$

А положение Q3 определяется номером, ближайшим к числу:

$$\frac{3(n+1)}{4}.$$

Выборочные коэффициенты асимметрии и эксцесса.

На практике также значительный интерес представляет собой степень отклонения одномодального распределения частот данных выборки от симметричного распределения. Если имеется асимметрия (т.е. несимметричность) распределения частот, то в этом наглядно можно убедиться по гистограмме и полигону частот. В этом случае одна из их ветвей, начиная от вершины, более длинная, чем другая. При симметричном распределении частот выборочные среднее значение, медиана и мода имеют приблизительно одинаковое значение. В случае длинной правой ветви гистограммы и полигона говорят о положительной асимметрии распределения наблюдений, а в случае более длинной левой ветви говорят об отрицательной асимметрии наблюдений. Для положительной асимметрии среднее значение больше медианы, которая в свою очередь больше моды. Для отрицательной асимметрии среднее значение меньше медианы, которая в свою очередь меньше моды.

Существует несколько коэффициентов для расчета асимметрии наблюдений выборки. Наиболее точным из них является **выборочный коэффициент асимметрии** следующего вида:

$$\hat{A} = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3,$$

где $x_1, x_2, \dots, x_n$ – рассматриваемые наблюдения, $\bar{x}$ – выборочное среднее значение, s – выборочное стандартное отклонение.

Существует еще одна характеристика одновершинного распределения частот наблюдений выборки, указывающая на степень концентрации частот около выборочного среднего значения и на концах интервала наблюдений выборки. Это **выборочный коэффициент эксцесса**. Он задается формулой следующего вида:

$$\hat{E} = \frac{1}{n \cdot s^4} \sum_{i=1}^n (x_i - \bar{x})^4 - 3,$$

где $x(1), x(2), \dots, x(n)$ – наблюдения выборки, $\bar{x}$ – выборочное среднее значение, s – выборочное стандартное отклонение.

Выборочная ковариация и выборочный коэффициент корреляции Пирсона.

Если в каждом наблюдении определяются значения не одной случайной величины, а значения двух (или нескольких) случайных величин одновременно, то получаем двумерную (или многомерную) выборку. Пусть, например, получена случайная выборка значений двух случайных величин X и Y : $(x_1, y_1); (x_2, y_2); \dots; (x_n, y_n)$.

Для характеристики такой выборки вводятся понятия выборочной ковариации и выборочного коэффициента корреляции Пирсона.

Пусть $\bar{x}$ – выборочное среднее значение и s_x – выборочное стандартное отклонение наблюдений $x_1, x_2, \dots, x_n$, а $\bar{y}$ – выборочное среднее значение и s_y – выборочное стандартное отклонение наблюдений $y_1, y_2, \dots, y_n$. Тогда выборочная ковариация определяется формулой:

$$\hat{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

а **выборочный коэффициент корреляции Пирсона** определяется формулой:

$$\hat{r} = \frac{\hat{cov}(X, Y)}{s_x \cdot s_y}.$$

Свойства выборочных характеристик.

Теперь о свойствах выборочных характеристик, построенных по наблюдениям случайной выборки. Предположим, что наблюдалась некоторая случайная величина ξ для которой $F(x)$ – функция распределения (ее называют теоретической функцией распределения) и для которой существуют основные числовые характеристики (их называют теоретическими характеристиками): математическое ожидание, дисперсия, стандартное отклонение, коэффициенты асимметрии и эксцесса и т.п.

На практике точный вид функции $F(x)$ и точные значения основных теоретических числовых характеристик случайной величины ξ бывают неизвестны. Исследователь вынужден строить свои выводы о них лишь на основании ограниченного ряда выборочных характеристик, полученных для наблюдений ξ случайной выборки из интересующей генеральной совокупности. К основным выборочным характеристикам относятся: выборочная функция распределения $F_n(x)$, выборочная относительная частота n_i/n появления i -го возможного значения в n наблюдениях,

выборочное среднее значение, выборочная дисперсия $s^2(n)$, выборочный коэффициент асимметрии $A(n)$, выборочный коэффициент эксцесса $E(n)$, выборочный коэффициент корреляции $r(n)$.

Из закона больших чисел теории вероятностей следует, что при неограниченном увеличении объема выборки (т.е. при $n \rightarrow \infty$) с вероятностью, близкой к единице, все основные выборочные характеристики стремятся к соответствующим теоретическим характеристикам исследуемой случайной величины ξ . Этот факт позволяет использовать выборочные характеристики для приближенного описания свойств случайной величины ξ для всей генеральной совокупности.

Все выборочные характеристики являются случайными величинами и поэтому возникает вопрос о законе распределения вероятностей каждой из них.

Из центральной предельной теоремы теории вероятностей следует, что асимптотически (при $n \rightarrow \infty$) практически независимо от типа случайной величины все основные выборочные характеристики за исключением r ведут себя как нормально распределенные случайные величины. При этом, разумеется, параметры нормального закона, т.е. математическое ожидание и дисперсия, различные для разных выборочных характеристик. Так, например,

$$M[\bar{x}(n)] = \alpha$$

- математическому ожиданию ξ ,

$$D[\bar{x}(n)] = \frac{\sigma^2}{n}$$

- где σ^2 - дисперсия ξ ,

$$M[s^2(n)] = (1 - \frac{1}{n})\sigma^2, \quad M[F_n(x)] = F(x), \quad M\left[\frac{n_i}{n}\right] = p_i$$

- вероятности i -го значения соответствующей дискретной случайной величины ξ и т.д.

Из центральной предельной теоремы также следует, что асимптотически (при $n \rightarrow \infty$) случайная величина

$$\sqrt{n}[\hat{r}(n) - r(X, Y)]$$

распределена по нормальному закону с математическим ожиданием равным нулю. При конечных (ограниченных) объемах случайной выборки поведение

основных выборочных характеристик существенно зависит от закона распределения случайной величины ξ .

Пусть, например, случайная величина $\xi \sim (a, \sigma^2)$ Тогда оказывается (теорема Фишера), что при любом конечном объеме $n >$ выборки (а не только при $n \rightarrow \infty$) случайная величина

$$\bar{x}(n) \sim N\left(a, \frac{\sigma^2}{n}\right),$$

т.е. распределена по нормальному закону с параметрами a и σ^2/n . Кроме того, в этом случае случайные величины $\bar{x}(n)$ и s^2 независимы, причем

Если совместное распределение пары случайных величин (X, Y) является нормальным, то асимптотически (при $n \rightarrow \infty$) распределение $r(n)$ тоже является нормальным с

$$M[\hat{r}(n)] = r(X, Y), D[\hat{r}(n)] = \frac{(1-r^2)^2}{n}.$$

При малых объемах n выборки или при значениях $|r(n)|$, близких к единице, это приближение является достаточно грубым.

Заметим в заключение, что каждый член дискретного вариационного ряда $x(1), x(2), \dots, x(n)$, построенного для наблюдений $x_1, x_2, \dots, x_n$, называется порядковой статистикой. Порядковые статистики тоже относятся к основным выборочным характеристикам. Они являются выборочными аналогами квантилей случайной величины.

Порядок выполнения работы:

Пример. По результатам контрольной проверки налоговыми службами 400 бизнес-структур, у 140 из них в налоговых декларациях не полностью указаны доходы, подлежащие налогообложению. Определите в генеральной совокупности (по всему району) долю бизнес-структур, скрывших часть доходов от уплаты налогов, с вероятностью 0,954.

Решение:

По условию задачи число единиц в выборочной совокупности $n=400$, число единиц, обладающих рассматриваемым признаком $m=140$, вероятность $P=0,954$.

Из теории вероятностей известно, что при вероятности $P=0,954$ коэффициент доверия $t=2$.

Долю единиц, обладающих указанным признаком, определим по формуле: $p=w+\Delta p$, где $w = m/n=140/400=0,35=35\%$,

а предельную ошибку признака Δp получим из формулы: $\Delta p = t \sqrt{w(1-w)/n} = 2\sqrt{0,35 \times 0,65/400} \approx 0,5 = 5\%$

Тогда $p = 35 \pm 5\%$.

Ответ: Доля бизнес-структур, скрывших часть доходов от уплаты налогов с вероятностью 0,954 равна $35 \pm 5\%$.

Контрольные задания:

1. При обследовании 100 семей по числу детей были получены следующие результаты: 20 семей не имеют детей, 40 семей имеют по одному ребенку, 25 семей имеют по два ребенка, 10 семей имеют по три ребенка и 5 семей имеют по четыре ребенка.

а) Построить таблицу частот дискретного вариационного ряда, полигон частот, выборочную функцию распределения и ее график.

б) Найти выборочное среднее значение, выборочную дисперсию и выборочный коэффициент вариации.

2. На экзамене по математике 25 студентов группы по пятибалльной шкале были оценены следующим образом: 5, 5, 4, 4, 4, 3, 3, 2, 5, 3, 4, 4, 4, 3, 2, 5, 2, 3, 4, 5, 5, 2, 3, 3, 3.

а) Построить: таблицу частот дискретного вариационного ряда, полигон частот, выборочную функцию распределения и ее график.

б) Найти выборочное среднее значение, выборочную дисперсию и выборочный коэффициент вариации.

3. При тестировании общего интеллекта 20 школьников были получены следующие баллы: 11, 17, 17, 12, 13, 12, 15, 15, 14, 16, 13, 14, 13, 15, 16, 16, 15, 15, 14, 14.

а) Построить: таблицу частот дискретного вариационного ряда, полигон частот, выборочную функцию распределения и ее график.

б) Найти выборочное среднее значение, выборочную дисперсию и выборочный коэффициент вариации.

4. Известны следующие результаты (в баллах) сдачи вступительных экзаменов в университет группы абитуриентов: 18, 16, 20, 17, 19, 20, 17, 17, 12, 15, 20, 18, 19, 18, 18, 16, 18, 14, 17, 19, 16, 14, 19, 12, 15, 16, 20, 14.

а) Построить таблицу частот интервального вариационного ряда, разбив данные на 4 равных интервала, и гистограмму частот.

б) Найти выборочное среднее значение, выборочную дисперсию и выборочный коэффициент вариации.

5. Имеются следующие данные выборочного обследования затраты времени (в часах) 100 студентов на дорогу до университета:

Затраты времени	Число студентов
$(0;0,5]$	7
$(0,5;1]$	18
$(1;1,5;]$	32
$(1,5;2]$	37
$(2;2,5]$	6

- а) Построить гистограмму частот и гистограмму накопленных частостей.
- б) Найти выборочное среднее значение, выборочную дисперсию и выборочный коэффициент вариации.

Контрольные вопросы:

1. Что такое репрезентативная случайная выборка?
2. Привести примеры применения каждой из четырех шкал измерения.
3. Что такое таблица частот наблюдений?
4. Как строятся гистограмма и полигон частот наблюдений?
5. Что принято относить к выборочным мерам центра распределения наблюдений случайной выборки?
6. Что относится к выборочным мерам рассеяния наблюдений случайной выборки?
7. Назовите основные методы исследования, применяемые в методологии науки?

Лабораторная работа №7.

Тема: Расчет и построение нормальной кривой по опытным данным.

Цель: Научить пользоваться методами математической статистики: оценкой параметров распределения случайных величин, построению гистограммы распределения по опытным данным и др.

Теоретическое обоснование:

Точечные оценки параметров распределения.

Пусть требуется изучить количественный признак генеральной совокупности. Допустим, что из теоретических соображений удалось установить, какое именно распределение имеет признак. Возникает задача оценки параметров, которыми определяется это распределение.

Обычно в распоряжении исследователя имеются лишь данные выборки, полученные в результате n наблюдений (здесь и далее наблюдения предполагаются независимыми). Через эти данные и выражают оцениваемый параметр. Рассматривая значения количественного признака как независимые случайные величины, можно сказать, что найти статистическую оценку неизвестного параметра теоретического распределения - это значит найти функцию от наблюдаемых случайных величин, которая и дает приближенное значение оцениваемого параметра.

Итак, **статистической оценкой неизвестного параметра** теоретического распределения называют функцию от наблюдаемых случайных величин.

Для того чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определенным требованиям: оценка должна быть несмещенной, эффективной и состоятельной.

Несмещенной называют статистическую оценку, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки.

$$\bar{x} = \frac{\sum_{i=1}^n (x_i - \bar{x}_{\text{нп}})^2 n_i}{n-1} \approx \frac{\sum_{i=1}^n x_i^2}{n-1} - \frac{n}{n-1} \bar{x}_{\text{нп}}^2$$

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

$$\sigma_{est} = \sqrt{\sigma_{est}^2}$$

Эффективной называют статистическую оценку, которая (при заданном объеме выборки n) имеет наименьшую возможную дисперсию.

Состоятельной называют статистическую оценку, которая при $n \rightarrow \infty$ стремится по вероятности к оцениваемому параметру. Например, если дисперсия несмещенной оценки при $n \rightarrow \infty$ стремится к нулю, то такая оценка оказывается и состоятельной.

Генеральная средняя.

Генеральной средней называют среднее арифметическое значений признака генеральной совокупности.

Выборочная средняя.

Выборочной средней называют среднее арифметическое значение признака выборочной совокупности.

Выборочная средняя является несмещенной и состоятельной оценкой генеральной средней.

Если выборка представлена интервальным вариационным рядом, то за принимают середины частичных интервалов.

Генеральная дисперсия.

Для того чтобы охарактеризовать рассеяние значений количественного признака X генеральной совокупности вокруг своего среднего значения, вводят сводную характеристику - **генеральную дисперсию**.

Генеральной дисперсией называют среднее арифметическое квадратов отклонений значений признака генеральной совокупности от их среднего значения.

Кроме дисперсии для характеристики рассеяния значений признака генеральной совокупности вокруг своего среднего значения пользуются сводной характеристикой - **средним квадратическим отклонением**.

Генеральным средним квадратическим отклонением (стандартом) называют квадратный корень из генеральной дисперсии.

Выборочная дисперсия.

Для того чтобы наблюдать рассеяние количественного признака значений выборки вокруг своего среднего значения, вводят сводную характеристику - **выборочную дисперсию**.

Выборочной дисперсией называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения.

Для характеристики рассеивания значений признака выборки вокруг своего среднего значения пользуются сводной характеристикой - **средним квадратическим отклонением**.

Выборочным средним квадратическим отклонением называют квадратный корень из выборочной дисперсии.

Исправленная дисперсия.

В качестве оценки генеральной дисперсии принимают исправленную дисперсию.

Для оценки среднего квадратического генеральной совокупности используют исправленное среднее квадратическое отклонение

Интервальные оценки параметров распределения.

Интервальной называют оценку, которая определяется двумя числами - концами интервала. Интервальные оценки позволяют установить точность и надежность оценок.

Пусть найденная по данным выборки статистическая характеристика служит оценкой неизвестного параметра. Будем считать постоянным числом (может быть и случайной величиной). Ясно, что тем точнее определяет параметр, чем меньше абсолютная величина разности.

Однако статистические методы не позволяют категорически утверждать, что оценка удовлетворяет неравенству; можно лишь говорить о вероятности, с которой это неравенство осуществляется.

Надежностью (доверительной вероятностью) оценки называют вероятность, с которой осуществляется неравенство.

Обычно надежность оценки задается наперед, причем в качестве γ берут число, близкое к единице. Наиболее часто задают надежность, равную 0,95; 0,99 и 0,999.

Виды интервальных оценок параметра:

1. Доверительный интервал для оценки математического ожидания при известном s .
2. Доверительный интервал для оценки математического ожидания при неизвестном s .
3. Доверительный интервал для оценки дисперсии и среднего квадратического отклонения.

Контрольные вопросы:

1. Что такое статистическая оценка?
2. Запишите формулу несмешанной статистической оценки?
3. Запишите формулу смешанной статистической оценки?
4. Какие бывают статистические оценки?
5. Что такое генеральная дисперсия?
6. Что такое выборочная дисперсия?
7. Что такое интервальная оценка?

Лабораторная работа №8.

Тема: Построение корреляционной таблицы.

Цель работы: Освоение методики построения корреляционного поля, а также корреляционной таблицы.

Теоретическое обоснование:

Корреляция изучается на основании экспериментальных данных, представляющих собой измеренные значения (x_i, y_i) двух признаков. Если экспериментальных данных немного, то двумерное эмпирическое распределение представляется в виде двойного ряда значений x_i и y_i . При этом корреляционную зависимость между признаками можно описывать разными способами. Соответствие между аргументом и функцией может быть задано таблицей, формулой, графиком и т. д.

Корреляционный анализ, как и другие статистические методы, основан на использовании вероятностных моделей, описывающих поведение исследуемых признаков в некоторой генеральной совокупности, из которой получены экспериментальные значения x_i и y_i .

Когда исследуется корреляция между количественными признаками, значения которых можно точно измерить в единицах метрических шкал (метры, секунды, килограммы и т.д.), то очень часто принимается модель двумерной нормально распределенной генеральной совокупности. Такая модель отображает зависимость между переменными величинами x_i и y_i графически в виде геометрического места точек в системе прямоугольных координат. Эту графическую зависимость называют также диаграммой рассеивания или корреляционным полем.

Данная модель двумерного нормального распределения (корреляционное поле) позволяет дать наглядную графическую интерпретацию коэффициента корреляции, т.к. распределение в совокупности зависит от пяти параметров: m_x, m_y – средние значения (математические ожидания); s_x, s_y – стандартные отклонения случайных величин X и Y и r – коэффициент корреляции, который является мерой связи между случайными величинами X и Y .

Если $r = 0$, то значения, x_i, y_i , полученные из двумерной нормальной совокупности, располагаются на графике в координатах x, y в пределах области, ограниченной окружностью. В этом случае между случайными величинами X и Y отсутствует корреляция и они называются некоррелированными. Для двумерного нормального распределения некоррелированность означает одновременно и независимость случайных величин X и Y .

Если $r = 1$ или $r = -1$, то между случайными величинами X и Y существует линейная функциональная зависимость ($Y = c + dX$). В этом случае говорят о полной корреляции. При $r = 1$ значения x_i, y_i определяют точки, лежащие на прямой линии, имеющей положительный наклон (c увеличением x_i значения y_i также увеличиваются), при $r = -1$ прямая имеет отрицательный наклон.

В промежуточных случаях ($-1 < r < 1$) точки, соответствующие значениям x_i, y_i , попадают в область, ограниченную некоторым эллипсом (рис.1, в. г), причем при $r > 0$ имеет место положительная корреляция (c увеличением x_i значения y_i имеют тенденцию к возрастанию), при $r < 0$ корреляция отрицательная. Чем ближе r к ± 1 , тем уже эллипс и тем теснее экспериментальные значения группируются около прямой линии.

Здесь же следует обратить внимание на то, что линия, вдоль которой группируются точки, может быть не только прямой, а иметь любую другую форму: парабола, гипербола и т. д. В этих случаях мы рассматривали бы так называемую, нелинейную (или криволинейную) корреляцию.

Таким образом, визуальный анализ корреляционного поля помогает выявить не только наличия статистической зависимости (линейную или нелинейную) между исследуемыми признаками, но и ее тесноту и форму. Это имеет существенное значение для следующего шага в анализе – выбора и вычисления соответствующего коэффициента корреляции.

Корреляционную зависимость между признаками можно описывать разными способами. В частности, любая форма связи может быть выражена уравнением общего вида $Y = f(X)$, где признак Y – зависимая переменная, или функция от независимой переменной X , называемой аргументом. Соответствие между аргументом и функцией может быть задано таблицей, формулой, графиком и т. д.

На практике в результате независимых наблюдений над величинами X и Y , как правило, имеют дело не со всей совокупностью всех возможных пар значений этих величин, а лишь с ограниченной выборкой из генеральной совокупности, причем объем n выборочной совокупности определяется как количество имеющихся в выборке пар. Первоочередной задачей статистической обработки экспериментального материала является систематизация полученных данных и выяснение формы соответствующей генеральной совокупности.

Пусть величина X в выборке принимает значения $x_1, x_2, \dots, x_m$, где количество различающихся между собой значений этой величины, причем в общем случае каждое из них в выборке может повторяться. Пусть величина Y в выборке принимает значения $y_1, y_2, \dots, y_k$, где k – количество различающихся между собой значений этой величины, причем в общем случае каждое из них в выборке также может повторяться. В этом случае данные заносят в таблицу с учетом частот встречаемости. Такую таблицу с группированными данными называют корреляционной.

Первым этапом статистической обработки результатов является составление корреляционной таблицы (таблица 1).

$Y \backslash X$	x_1	x_2	...	x_m	$n_{y\cdot}$
y_1	n_{12}	n_{21}		n_{m1}	n_{y1}
y_2		n_{22}		n_{m2}	n_{y2}
...					
y_k	n_{1k}	n_{2k}		n_{mk}	n_{yk}
$n_{x\cdot}$	n_{x1}	n_{x2}		n_{xm}	n

В первой строке основной части таблицы в порядке возрастания перечисляются все встречающиеся в выборке значения величины X . В первом столбце также в порядке возрастания перечисляются все встречающиеся в выборке значения величины Y . На пересечении соответствующих строк и столбцов указываются частоты n_{ij} ($i=1,2,\dots,m$; $j=1,2,\dots,k$) равные количеству появлений пары $(x_i; y_j)$ в выборке. Например, частота n_{12} представляет собой количество появлений в выборке пары $(x_1; y_1)$.

Так же $\sum_{i=1}^m, 1 \leq i \leq m$, сумма элементов i -го столбца, $\sum_{j=1}^k, 1 \leq j \leq k$, - сумма элементов j -ой строки и $\sum_{i=1}^m \sum_{j=1}^k p_{ij} = n$

Аналоги формул, полученные по данным корреляционной таблицы, имеют вид:

$$F = \frac{b^2 \sum_{i=1}^n (x_i - \bar{x})^2}{S^2}$$

Порядок выполнения работы:

Пример. Изучалась зависимость между качеством стандартности товаров $Y(\%)$ и количеством товаров (X) шт. Результаты наблюдений приведены в виде корреляционной таблицы.

$Y \backslash X$	18	22	26	30	$\sum y_i$
70	5				5
75	7	46	1		54
80		29	72		101
85			29	8	
37					
90				3	3
$\sum x_j$	12	75	102	11	200

Требуется:

1. Найти выборочное уравнение прямой регрессии Y на X .
2. Определить выборочные аналоги функции регрессии.
3. Сравнить между собой при каждом значении X приближения средних значений Y , полученные по функции регрессии и по уравнению прямой регрессии.

Решение. Пользуясь данными, приведенными в этой таблице, по формулам, находим:

$$S^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-2}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{и} \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\text{Следовательно, } r_B = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_X \sigma_Y}$$

$$a = 79.475 - 1.111 \cdot 24.24 = 79.475 - 26.930 = 52.544$$

Таким образом, выборочное уравнение прямой регрессии Y на X выражается формулой:

$$Y = 79.475 + 1.111(x - 24.24) = 79.475 + 1.111x - 26.930 = 52.545 + 1.111x, \text{ откуда:}$$

X	18	22	26	30
Y _{лин}	72.5	76.98	81.45	85.92
Y _x	72.91	76.93	81.37	86.36

где Y_{лин}(x=x₁) = 52.545 + 1.111 · 18 = 72.5 и т.д. y_{x1} = (5 · 70 + 7 · 75) / 12 = 72.91 и т.д.

Сопоставляя полученные результаты, приходим к выводу, что значения, вычисленные по уравнению выборочной регрессии и по линейной зависимости хорошо согласуются.

Заключение. Величины, вычисленные путем подстановки возможных значений X в уравнение прямой регрессии и в функцию регрессии, практически совпадают.

Замечание. Для упрощения вычислений в корреляционной табл. удобно от (x_i; y_i) перейти к новым переменным (u_i; v_i), положив u_i = (x_i - x₀) / h₁; v_j = (y_j - y₀) / h₂ (*)

где x₀ и y₀ варианты соответствующие наибольшим частотам соответственно x_i и y_i. h_i = x_{i+1} - x_i.

Обратный пересчет осуществляется по формулам:

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{X^2 - (\bar{X})^2}, \sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{Y^2 - (\bar{Y})^2}, \bar{Y}^2 = \frac{1}{n} \sum_{j=1}^k n_{y_j} y_j^2$$

Контрольные задания:

1. По 10 банкам изучается зависимость прибыли (y – млн. руб.) от вложений в уставные капиталы предприятий (x – млн. руб.):

Прибыль млн,руб .	Вложения в уставные капиталы предприятий млн,руб.
55,3	20
50,2	25

60,9	35
62,8	42
63,9	47
64,5	50
65,5	55
66,8	63
67,9	70
69,3	80

а) Построить поле корреляции рассматриваемой зависимости.

б) Определить уравнение регрессии полулогарифметрической модели: $y = a + b \cdot \ln x$.

в) Найти индекс корреляции и сравнить его с линейным коэффициентом корреляции. Пояснить причины различий.

г) Найти среднюю ошибку аппроксимации.

д) Стоит сказать, что рассчитать стандартную ошибку регрессии.

е) С вероятностью 0,95 оценить статистическую значимость уравнения и коэффициента регрессии. Сделать выводы.

ж) С вероятностью 0,95 оценить доверительный интервал ожидаемого размера прибыли, в случае если вложения в уставные капиталы предприятий составят 45 млн. руб.

Контрольные вопросы:

1. На каких основаниях изучается корреляция?
2. На чем основан корреляционный анализ?
3. В каких случаях принимается модель двумерной нормально распределенной генеральной совокупности?

Варианты контрольных заданий

Из каждой серии заданий (1-10, 11-20 и т.д.) выбрать и решить задания своего варианта. В общей сложности получится 8 заданий. Например, для варианта №7 это будут задания: 7, 17, 27, 37, 47 и т.п.

Задания 1-10

1. В коробке 6 чайных чашек, 4 из них синего цвета. Наудачу отбирают 3 чашки. Найти вероятность того, что среди них будут 2 синие чашки.
2. В группе 5 пловцов и 3 бегуна. Наудачу отбирают 5 спортсменов. Найти вероятность того, что среди них будут 3 пловца.
3. Совет директоров состоит из 3 бухгалтеров, 3 менеджеров и 2 инженеров. Планируется создать подкомитет из 4 его членов. Какова вероятность того, что в подкомитете будет 2 бухгалтера?
4. На полке 7 книг, 5 из них в жестком переплете. Наудачу отбирают 3 книги. Какова вероятность того, что 2 книги будут в жестком переплете?
5. В коробке 20 катушек ниток, 14 красных и 6 белых. Покупатель просит продать 3 красных катушки и 2 белых. Какова вероятность того, что продавец не глядя сразу подаст ему требуемое?
6. В производственном отделе компании работают 26 человек, среди них 6 женщин. Наудачу выбрано 3 служащих. Найти вероятность того, что все они мужчины.
7. Имеется 8 билетов в театр, 5 из которых на места в партер. Какова вероятность того, что среди 3 наудачу взятых билетов 2 билета будут в партер?
8. В пачке 15 накладных, 5 из них оформлены неправильно. Наудачу отбирают 4 штуки. Какова вероятность того, что одну из них надо переоформлять?
9. В коробке 15 конфет, 7 из них с коньяком. Найти вероятность того, что среди 5 наудачу взятых конфет с коньяком будут 3.
10. Из колоды в 36 карт наудачу отбирают 5. Найти вероятность того, что все они красные.

Задания 11-20

11. В двух коробках лежат яблоки. В 1-ой - 50 штук, 4 из них с червоточиной; во 2-ой - 60 штук и 5 из них с червоточиной. Из коробок наудачу отбирают по яблоку. Найти вероятность того, что среди них окажется:
 - а) только одно яблоко с червоточиной;
 - б) хотя бы одно яблоко с червоточиной.
12. На полке стоят 15 книг, 7 из них по математике. Наудачу отбирают 2 книги. Найти вероятность того, что среди них по математике окажется:

- а) только одна;

- б) ни одной;
- в) хотя бы одна книга.

13. В отделе реализации компании 10 мужчин и 15 женщин. Наудачу отбирают 2 человека. Найти вероятность того, что среди них:

- а) только одна женщина;
- б) оба мужчины;
- в) хотя бы один мужчина.

14. Иванов попадает в цель 70 раз из 100 выстрелов, Петров - 80 раз из 100. Оба делают по выстрелу. Найти вероятность того, что цель будет поражена.

15. На подносе 20 бокалов с шампанским, 8 из них с сухим, остальные с полусладким. Наудачу берут 2 бокала. Найти вероятность того, что среди отобранных:

- а) только один бокал с сухим шампанским;
- б) оба с полусладким;
- в) хотя бы один бокал с полусладким.

16. В двух коробках по 20 кубиков, в одной 7 кубиков из 20 - зеленого цвета, в другой - 9 из 20. Наудачу берут по одному кубику из каждой коробки. Найти вероятность того, что среди отобранных кубиков:

- а) только один зеленого цвета;
- б) оба зеленые;
- в) оба кубика другого цвета.

17. В магазине на полке стоят 12 чайных пар, 3 из них со скрытым дефектом. Покупатель просит продать 2 пары. Найти вероятность того, что:

- а) продавец сразу подаст 2 хороших пары;
- б) продавцу придется менять одну пару;
- в) продавцу придется менять обе пары.

18. На подставке 15 кассет, на 8 из них есть запись. Наудачу берут 2 кассеты. Найти вероятность того, что среди них:

- а) обе кассеты чистые;
- б) хотя бы одна кассета чистая.

19. В двух коробках по 20 пар перчаток. В первой - 6 пар размера L , во второй - 11 пар размера L . Наудачу отбирают по паре из каждой коробки. Найти вероятность того, что среди отобранных:

- а) только одна пара перчаток размера L ;
- б) хотя бы одна пара перчаток размера X .

20. В двух ящиках холодильной камеры по 30 упаковок мороженого, в первом ящике 16 с орехами, во втором - 18 с орехами. Наудачу берут по одной упаковке из каждого ящика. Найти вероятность того, что среди них окажется:

- а) только одно мороженое с орехами;
- б) обе упаковки мороженого без орехов.

Задания 21 - 30

21. Фирма получает комплектующие от трех поставщиков A , B и C . Поставщик A поставяет 40% общего объема поставок, B - 35%, C - 25%. Из практики известно, что 1% поставляемых A деталей с браком, для B - 2%, для

С -2,5%. Какова вероятность поступления бракованной детали на сборку и от какого поставщика она выше - А, В или С?

22. Покупатель посещает 1-ю булочную в два раза чаще, чем 2-ю и 3-ю. Вероятность того, что нужный сорт хлеба есть в булочной, для 1-й равна 0,6; для 2-й - 0,7 и для 3-й - 0,75. Покупатель купил хлеб. Что вероятнее: он купил его в 1-й, 2-й или 3-й булочной?

23. В 1-й партии микросхем 14 штук, во 2-й - 16 и в 3-й - 20 штук. Доля микросхем с браком, равна соответственно 2%, 1% и 1,5%. Выбранная наугад микросхема оказалась качественной. Что вероятнее: она из 1-й, 2-й или 3-й партии?

24. Над аэродромом в 35% случаев облачная погода. Вероятность взлета без задержек при хорошей погоде равна 0,9; при облачной - 0,6. Самолет взлетел вовремя. Сравнить вероятность того, что он сделал это при хорошей погоде и облачной.

25. При установившемся технологическом процессе $\frac{2}{3}$ всей продукции станок-автомат выпускает 1-го сорта, а остальную - 2-го. В среднем 1% продукции 1-го сорта имеет скрытые дефекты и 2,2% продукции 2-го сорта имеет скрытые дефекты. Наудачу взятое изделие оказалось со скрытым дефектом. Что вероятнее: оно 1-го сорта или 2-го сорта?

26. В центральную бухгалтерию фирмы поступают для проверки пачки накладных из трех цехов. Из 1-го цеха поступает 40% всех накладных, из 2-го - 25%, из 3-го - 35%. Доля накладных, оформленных неправильно, равна соответственно 1%, 1,5% и 2%. Взятую наугад накладную пришлось переоформлять. Что вероятнее: она из 1-го, 2-го или 3-го цеха?

27. В производственном отделе компании 20 мужчин и 6 женщин, в отделе реализации 10 мужчин и 15 женщин, на автобазе 10 мужчин и 4 женщины. Из всех служащих отбирается наугад один человек, и это женщина. Что вероятнее: она из производственного отдела, из отдела реализации или с автобазы?

28. В трехъярусной коробке конфет в 1-ом слое лежат 25 конфет, из них с помадкой 10; во 2-ом слое - 30 конфет, с помадкой 15; в 3-м слое - 20 конфет, с помадкой 12. Взятая конфета оказалась с помадкой. Что вероятнее: она из 1-го, 2-го или 3-го слоя?

29. Куплено 3 коробки мандаринов, в которых содержится 60, 65 и 55 штук соответственно. Вероятность того, что хотя бы один мандарин испортился, для 1-ой коробки равна 0,6; для 2-ой - 0,5; для 3-ей - 0,7. Взятый наугад мандарин испорчен. Сравните вероятности, что он из 1-ой, 2-ой или 3-ей коробки.

30. В команде 22 лыжника и 8 слаломистов. Вероятность получить травму для лыжника равна 0,1 и для слаломиста - 0,3. Спортсмен получил травму. Что вероятнее: он лыжник или слаломист?

Задания 31 - 40

31. Средний процент нарушения работы телевизора в течение года равен 22%. Найти вероятность того, что из пяти проданных в течение дня телевизоров менее трех придется ремонтировать.

32. Вероятность прибытия поезда без опоздания равна 0,7. Найти вероятность того, что из шести поездов опоздают не более двух.

33. Сидоров попадает в цель 80 раз из 100. Какова вероятность того, что при четырех контрольных выстрелах он поразит цель не менее двух раз?

34. Два равносильных шахматиста играют партии. Что вероятнее: выиграть не менее двух партий из четырех или не менее трех партий из пяти?

35. Из 25 студентов группы двадцать сдают экзамен с первого раза. Какова вероятность того, что следующий экзамен из 8 человек сдадут больше половины?

36. Найти вероятность того, что в семье с тремя детьми будет хотя бы один мальчик.

37. В коробке одна треть конфет с мармеладом. Найти вероятность того, что среди шести взятых конфет хотя бы одна будет с мармеладом.

38. Хозяйка купила дорогой пакет с десятью семенами помидоров. На пакете указана всхожесть 85%. Какова вероятность того, что у нее взойдет не менее семи семян?

39. Найти вероятность того, что при восьми бросаниях игральной кости шестерка выпадет менее трех раз.

40. Из ста билетов лотереи выиграют 15. Какова вероятность того, что из шести купленных билетов хотя бы один выиграет?

Задания 41 - 50

41. В здании института 6000 электроламп, вероятность включения каждой из которых равна 0,7. Найти наивероятнейшее число включенных ламп и соответствующую вероятность. Какова вероятность того, что будет включено от 4400 до 4600 ламп?

42. Средний процент отдельных дефектов в обуви, выпускаемой фабрикой, равен 20. Найти наивероятнейшее число пар качественной обуви среди 200 пар поступившей в магазин обуви и вероятность того, что обуви с дефектом будет от 30 до 50 пар.

43. К магистрали сжатого воздуха подключено 100 пневмомолотков, каждый из которых работает в данный момент с вероятностью 0,4. Магистраль не перегружена, если число одновременно работающих молотков не превышает 50. Найти вероятность того, что магистраль в данный момент не перегружена.

44. Вероятность того, что саженец сосны погибнет, равна 0,2. Найти наивероятнейшее число прижившихся саженцев среди 300 посаженных и соответствующую вероятность. Какова вероятность того, что погибнет от 40 до 70 саженцев?

45. В маршрутном такси едут 40 пассажиров. Вероятность выхода каждого из них на следующей остановке равна 0,25. Найти наивероятнейшее число пассажиров, которые выйдут на следующей остановке, и

соответствующую вероятность и вероятность того, что выйдут от 8 до 10 человек.

46. На складе 70% коробок с китайскими яблоками сорта «Фун-ши». Найти наивероятнейшее число коробок с «Фун-ши» среди 200 купленных и соответствующую вероятность и вероятность того, что яблок другого сорта будет от 60 до 80.

47. В книжном отделе 80% книг дороже 30 рублей. Покупают 20 книг, найти наивероятнейшее число книг дороже 30 рублей соответствующую вероятность.

48. Найти наивероятнейшее число выпадений «Герба» при 180 бросаниях монеты и соответствующую вероятность, а также вероятность того, что «решка» появится от 80 до 100 раз.

49. «Пикра» производит в среднем 10% кваса «Кузнецовский» (из всех сортов кваса). Найти наивероятнейшее число бутылок с этим квасом среди 500 привезенных в магазин, соответствующую вероятность и вероятность того, что бутылок с квасом другого сорта будет от 450 до 460.

50. При штамповке получается в среднем 90% годных изделий. Найти вероятнейшее число годных изделий из 900 полученных и соответствующую вероятность.

Задания 51 – 60

51. Производители карманных калькуляторов знают из опыта работы, что 1% произведенных калькуляторов имеет дефекты и их должны заменить по гарантии. Аудиторская фирма купила 500 калькуляторов. Какова вероятность того, что пять или более калькуляторов нужно заменить?

52. Полпроцента ксерокопий обычно испорчены. Найти вероятность того, что из 200 сделанных копий будет испорчено не более трех.

53. При перевозке повреждается в среднем 2 лампы из тысячи. Найти вероятность того, что из двух тысяч полученных ламп будет повреждено менее четырех.

54. Вероятность того, что на странице книги может оказаться опечатка, равна четверти процента. Проверяется из 800 страниц. Найти вероятность того, что с опечатками окажется больше четырех страниц.

55. В среднем из 100 бутылок некачественно закрыта одна. За смену автомат закрывает 3000 бутылок. Найти вероятность того, что некачественно будет закрыта хотя бы одна бутылка.

56. Из тысячи лотерейных билетов выигрывают два. Найти вероятность того, что из 500 купленных билетов выигрывает хотя бы один.

57. При укупорке взрывается в среднем две бутылки шампанского из тысячи. Какова вероятность того, что при укупорке трех тысяч бутылок взорвется не более пяти?

58. Доля брака при укупорке сока в пакеты «Тетрапак» 0,2%. Найти вероятность того, что из 1500 упаковок с дефектом будут пять.

59. Подошва туфель проверяется по качеству на изгиб. Найти вероятность того, что из 600 подошв не выдержит испытания более двух?

60. При пяти тысячах обращений к программе компьютера сбой происходит в одном случае. Найти вероятность того, что при десяти тысячах обращений к этой программе сбой произойдет в двух случаях.

Задания 61 – 70

Найти числовые характеристики дискретной случайной величины X , заданной законом распределения.

61.

X	-5	2	3	4
P	0,4	0,3	0,1	0,2

62.

x	0	1	2	4	6
P	0,1	0,2	0,3	0,15	0,25

63.

x	-3	0,5	1	3	4
P	0,05	0,1	0,15	0,5	0,2

64.

X	1,2	3,4	4,5	6,2
P	0,1	0,2	0,4	0,3

65.

X	0,21	0,35	0,65	1
P	0,3	0,2	0,4	0,1

66.

x	-1	0	1	2	3
P	0,3	0,2	0,4	0,05	0,05

67.

x	-2	-1	0	2	3
P	0,1	0,1	0,2	0,35	0,25

68.

x	1,5	2,1	4,2	5,3
P	0,2	0,3	0,4	0,1

69.

x	-0,5	0	1	2
P	0,1	0,25	0,35	0,3

70.

x	4,3	5,1	10,6	12,3
P	0,2	0,3	0,4	0,1

Задания 71 – 80

Непрерывная случайная величина задана функцией распределения вероятностей $F(x)$. Найти числовые характеристики.

$$71. F(x) = \begin{cases} 0, & x < 2 \\ \frac{x^2 - 2x}{8}, & 2 \leq x \leq 4 \\ 1, & x > 4 \end{cases}$$

$$72. F(x) = \begin{cases} 0, & x \leq -2 \\ \frac{x+2}{4}, & -2 < x \leq 2 \\ 1, & x > 2 \end{cases}$$

$$73. F(x) = \begin{cases} 0, & x < 3 \\ \frac{x^2 - 9}{7}, & 3 \leq x \leq 4 \\ 1, & x > 4 \end{cases}$$

$$74. F(x) = \begin{cases} 0, & x < \sqrt{5} \\ \frac{x^2 - 5}{4}, & \sqrt{5} \leq x \leq 3 \\ 1, & x > 3 \end{cases}$$

$$75. F(x) = \begin{cases} 0, & x < 2 \\ \frac{x^2 - 2x}{3}, & 2 \leq x \leq 3 \\ 1, & x > 3 \end{cases}$$

$$76. F(x) = \begin{cases} 0, & x < 1 \\ \frac{x^2 - x}{2}, & 1 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$$

$$77. F(x) = \begin{cases} 0, & x < 2 \\ \frac{x^2 - 4}{12}, & 2 \leq x \leq 4 \\ 1, & x > 4 \end{cases}$$

$$78. F(x) = \begin{cases} 0, & x \leq 1 \\ x^2, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

$$79. F(x) = \begin{cases} 0, & x \leq 0 \\ 3x^2 + 2x, & 0 \leq x \leq 1/3 \\ 1, & x > 1/3 \end{cases}$$

$$80. F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x^2}{4}, & 0 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$$

ВОПРОСЫ ДЛЯ ПОДГОТОВКИ К ЭКЗАМЕНУ

1. Что называется испытанием, событием? Приведите примеры испытаний, событий.
2. Какие события называются достоверными, невозможными, случайными. Приведите примеры этих событий.
3. Какие события называются несовместными, совместными? Приведите примеры.
4. Какие события называются противоположными? Приведите примеры.
5. Сформулируйте классическое определение вероятности. Укажите возможные границы вероятности.
6. Что понимается под суммой двух событий? Приведите примеры.
7. Сформулируйте теорему сложения вероятностей для несовместных событий.
8. Какие события называют независимыми, зависимыми? Приведите примеры.
9. Что понимается под произведением двух событий? Приведите примеры.
10. Сформулируйте теоремы умножения вероятностей для независимых и зависимых событий.
11. Сформулируйте теорему сложения вероятностей для совместных событий.
12. Что понимается под полной группой событий? Чему равна сумма вероятностей событий, составляющих полную группу?
13. Приведите формулу полной вероятности, формулу Байеса. При каких условиях они применяются?
14. Приведите формулу Бернулли. При решении какого типа задач она применяется?
15. Сформулируйте локальную теорему Лапласа. При каких условиях она применяется?

16. Сформулируйте интегральную теорему Лапласа.
17. Приведите формулу Пуассона. При каких условиях она применяется?
18. Какие случайные величины называются дискретными? непрерывными?
Приведите примеры.
19. Что называется законом распределения дискретной случайной величины? Как он задается?
20. Что называется многоугольником распределения дискретной случайной величины?
21. Дайте определение математического ожидания дискретной случайной величины.
22. Перечислите основные свойства математического ожидания.
23. Дайте определение дисперсии и среднего квадратического отклонения дискретной случайной величины.
24. Перечислите свойства дисперсии.
25. Дайте определение интегральной функции распределения. Перечислите ее свойства.
26. Дайте определение дифференциальной функции распределения. Перечислите ее свойства.
27. Как вычисляются числовые характеристики непрерывной случайной величины.
28. Какое распределение дискретной случайной величины называется биномиальным?
29. Чему равны числовые характеристики случайной величины, распределенной по биномиальному закону?
30. Какое распределение непрерывной случайной величины называется равномерным? Какой параметр характеризует равномерное распределение и как найти его значение?
31. Какое распределение непрерывной случайной величины называется нормальным? Какие параметры характеризуют нормальное распределение?

32. Начертите кривую нормального распределения. Как меняется кривая при изменении параметров нормального распределения?
33. Перечислите свойства нормального распределения.
34. Сформулируйте правило трех сигм.
35. Что понимается под генеральной совокупностью?
36. Что такое выборка? Что называется вариантами выборки и вариационным рядом?
37. Что такое частота появления варианты в выборке?
38. Как получают относительную частоту появления варианты в выборке?
39. Как построить полигоны частот и относительных частот?
40. Как построить гистограммы частот и относительных частот?
41. В чем сущность задачи по определению параметров генеральной совокупности?
42. Какую величину принимают за среднюю генеральной совокупности? Как она вычисляется?
43. Какую величину принимают за дисперсию генеральной совокупности? Как она вычисляется?
44. Как вычисляется среднее квадратическое отклонение средней выборки?
45. Что понимают под доверительным интервалом и доверительной вероятностью?
46. Как вычислить доверительный интервал для математического ожидания нормально распределенной случайной величины в случае, когда среднее квадратическое отклонение известно; когда среднее квадратическое неизвестно?
47. Дайте определение корреляционной зависимости.
48. В чем состоят две основные задачи теории корреляции?
49. Какую корреляционную зависимость называют линейной?
50. Дайте определение выборочного коэффициента корреляции и перечислите его свойства.

51. Что можно сказать о зависимости двух случайных величин, если коэффициент корреляции $r_B = 0$, $r_B = 1$, $r_B = -1$?
52. Запишите выборочные уравнения прямых регрессий.

ПРИЛОЖЕНИЯ

Приложение 1

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3865	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3104	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1513
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363

2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0043
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0006	0006	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Приложение 2

Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2} dz$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,0000	0,25	0,0987	0,50	0,1915	0,75	0,2734
0,01	0,0040	0,26	0,1026	0,51	0,1950	0,76	0,2764
0,02	0,0080	0,27	0,1064	0,52	0,1985	0,77	0,2794
0,03	0,0120	0,28	0,1103	0,53	0,2019	0,78	0,2823
0,04	0,0160	0,29	0,1141	0,54	0,2054	0,79	0,2852
0,05	0,0199	0,30	0,1179	0,55	0,2088	0,80	0,2881
0,06	0,0239	0,31	0,1217	0,56	0,2123	0,81	0,2910
0,07	0,0279	0,32	0,1255	0,57	0,2157	0,82	0,2939
0,08	0,0319	0,33	0,1293	0,58	0,2190	0,83	0,2967
0,09	0,0359	0,34	0,1331	0,59	0,2224	0,84	0,2995
0,10	0,0398	0,35	0,1368	0,60	0,2257	0,85	0,3023
0,11	0,0438	0,36	0,1406	0,61	0,2291	0,86	0,3051
0,12	0,0478	0,37	0,1443	0,62	0,2324	0,87	0,3078
0,13	0,0517	0,38	0,1480	0,63	0,2357	0,88	0,3106
0,14	0,0557	0,39	0,1517	0,64	0,2389	0,89	0,3133
0,15	0,0596	0,40	0,1554	0,65	0,2422	0,90	0,3159
0,16	0,0636	0,41	0,1591	0,66	0,2454	0,91	0,3186
0,17	0,0675	0,42	0,1628	0,67	0,2486	0,92	0,3212
0,18	0,0714	0,43	0,1664	0,68	0,2517	0,93	0,3238
0,19	0,0753	0,44	0,1700	0,69	0,2549	0,94	0,3264
0,20	0,0793	0,45	0,1736	0,70	0,2580	0,95	0,3289
0,21	0,0832	0,46	0,1772	0,71	0,2611	0,96	0,3315
0,22	0,0871	0,47	0,1808	0,72	0,2642	0,97	0,3340
0,23	0,0910	0,48	0,1844	0,73	0,2673	0,98	0,3365
0,24	0,0948	0,49	0,1879	0,74	0,2703	0,99	0,3389

1,00	0,3413	1,28	0,3997	1,56	0,4406	1,84	0, 4671
1,01	0,3438	1,29	0,4015	1,57	0,4418	1,85	0, 4678
1,02	0,3461	1,30	0, 4032	1,58	0,4429	1,86	0,4686
1,03	0,3485	1,31	0,4049	1,59	0,4441	1,87	0,4693
1,04	0,3508	1,32	0,4066	1,60	0,4452	1,88	0,4699
1,05	0,3531	1,33	0,4082	1,61	0,4463	1,89	0,4706
1,06	0,3554	1,34	0,4099	1,62	0,4474	1,90	0,4713
1,07	0,3577	1,35	0,4115	1,63	0,4484	1,91	0,4719
1,08	0,3599	1,36	0,4131	1,64	0,4495	1,92	0,4726
1,09	0,3621	1,37	0,4147	1,65	0,4505	1,93	0,4732
1,10	0,3643	1,38	0,4162	1,66	0,4515	1,94	0,4738
1,11	0,3665	1,39	0,4177	1,67	0,4525	1,95	0,4744
1,12	0,3686	1,40	0,4192	1,68	0,4535	1,96	0,4750
1,13	0,3708	1,41	0,4207	1,69	0,4545	1,97	0,4756
1,14	0,3729	1,42	0,4222	1,70	0,4554	1,98	0,4761
1,15	0,3749	1,43	0,4236	1,71	0,4564	1,99	0,4767
1,16	0,3770	1,44	0,4251	1,72	0,4573	2,00	0,4772
1,17	0,3790	1,45	0,4265	1,73	0,4582	2,02	0,4783
1,18	0,3810	1,46	0,4279	1,74	0,4591	2,04	0,4793
1,19	0,3830	1,47	0,4292	1,75	0,4599	2,06	0,4803
1,20	0,3849	1,48	0,4306	1,76	0,4608	2,08	0,4812
1,21	0,3869	1,49	0,4319	1,77	0,4616	2,10	0,4821
1,22	0,3883	1,50	0,4332	1,78	0,4525	2,12	0,4830
1,23	0,3907	1,51	0,4345	1,79	0,4633	2,14	0,4838
1,24	0,3925	1,52	0,4357	1,80	0,4641	2,16	0,4846
1,25	0,3944	1,53	0,4370	1,81	0,4649	2,18	0,4854
1,26	0,3962	1,54	0,4382	1,82	0,4656	2,20	0,4861
1,27	0,3980	1,55	0,4394	1,83	0,4664	2,22	0,4868
2,24	0,4875	2,48	0, 4934	2,72	0,4967	2,96	0,4985

2,26	0,4881	2,50	0,4938	2,74	0,4969	2,98	0,4986
2,28	0,4887	2,52	0,4941	2,76	0,4971	3,00	0,49865
2,30	0,4893	2,54	0,4945	2,78	0,4973	3,20	0,49931
2,32	0,4898	2,56	0,4948	2,80	0,4974	3,40	0,49966
2,34	0,4904	2,58	0,4951	2,82	0,4976	3,60	0,499841
2,36	0,4909	2,60	0,4953	2,84	0,4977	3,80	0,499928
2,38	0,4913	2,62	0,4956	2,86	0,4979	4,00	0,499968
2,40	0,4918	2,64	0,4959	2,88	0,4980	4,50	0,499997
2,42	0,4922	2,66	0,4961	2,90	0,4981	5,00	0, 499997
2,44	0,4927	2,68	0,4963	2,92	0,4982		
2,46	0,4931	2,70	0,4965	2,94	0,4984		

Приложение 3

Таблица значений $t_\gamma = t(\gamma, n)$

$\gamma \backslash n$	0,95	0,99	0,999	$\gamma \backslash n$	0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392

17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Приложение 4

Таблица значений $q = q(\gamma, n)$

$\gamma \backslash n$	0,95	0,99	0,999	$\gamma \backslash n$	0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	0,98	50	0,21	0,30	0,43
12	0,55	0,90	0,90	60	0,188	0,269	0,38
13	0,52	0,83	0,83	70	0,174	0,245	0,34
14	0,48	0,78	0,78	80	0,161	0,226	0,31
15	0,46	0,73	0,73	90	0,151	0,211	0,29
16	0,44	0,70	0,70	100	0,143	0,198	0,27
17	0,42	0,66	0,66	150	0,115	0,160	0,211
18	0,40	0,63	0,63	200	0,099	0,136	0,185
19	0,39	0,60	0,60	250	0,089	0,120	0,162

Заключение

Данное учебное пособие предназначено для студентов направления 38.03.01 «Экономика». Пособие содержит теоретический материал, практические задания, лабораторные работы и задания для выполнения контрольных работ. В каждом параграфе приведены примеры, пояснение изучаемого вопроса. Надеемся, что рассмотрение этих примеров поможет студентам–заочникам при решении контрольных заданий.

Задачи одного варианта оканчиваются на одну и ту же цифру, совпадающую с последней цифрой номера зачетной книжки. Например, если номер зачетной книжки оканчивается на 6, то все задачи № 06, 16, 26,... входят в контрольную работу варианта № 6.

При оформлении контрольной работы решение задач следует излагать по порядку, подробно, предварительно полностью переписав задание. Работа оформляется на листах формата А4 с одной стороны.

В некоторых случаях (в зависимости от специальности) преподаватель может исключить из контрольной работы те или иные задачи для каждого варианта.

Библиографический список

1. Маркович Э.С. Курс высшей математики с элементами теории вероятностей и математической статистики/ Маркович Э.С. – М.: Высшая школа, 1972 – 112с.
2. Гмурман Теория вероятностей и математическая статистика/ Гмурман – М.:
3. Карасев А. И. Теория вероятностей и математическая статистика. «Статистика», 1970.
4. Румшиский Л. З. Элементы теории вероятностей. Физматгиз, 1963.
5. Красс М.С. Основы математики и ее приложения в экономическом образовании, «Дело», 2000.
6. Гаврилец Т.Н. Теория вероятности и математическая статистика, 1997.
7. Вайнштейн И.И. Высшая математика: теория вероятности и математическая статистика. Красноярск: 2004

Министерство сельского хозяйства Российской Федерации
Красноярский Государственный Аграрный Университет
Ачинский филиал

Фёдорова И. А.

**«ТЕОРИЯ ВЕРОЯТНОСТЕЙ И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»**

Учебное пособие

Ачинск 2016