МИНИСТЕРСТВО ПРОМЫШЛЕННОСТИ И ТОРГОВЛИ ТВЕРСКОЙ ОБЛАСТИ

Государственное бюджетное профессиональное образовательное учреждение

«Тверской полиграфический колледж»

Н. В. Гордеев

Основы теории вероятностей и математической статистики

Конспект лекций

Тверь, 2020

Гордеев Н. В. Основы теории вероятностей и математической статистики: Конспект лекций. – Тверь: ГБП ОУ «Тверской полиграфический колледж», 2020. – 41с.

Табл. 6. Ил. 8. Библ. 6

В конспекте лекций рассмотрены темы рабочей программы дисциплины «Математика», касающиеся основных вопросов теории вероятности и математической статистики. Представлены примеры решения задач, контрольные вопросы.

Конспект лекций предназначен для студентов специальностей: 46. 02. 01; 42. 02. 02.

Ответственный за выпуск: Г. Ю. Волкова

Организация-разработчик:

ГБП ОУ «Тверской полиграфический колледж» г. Твери

Введение

Многие вещи нам непонятны не
потому, что наши понятия слабы;
но потому, что сии вещи не входят
в круг наших понятий.
Козьма Прутков

Представителям многих специальностей часто приходится составлять из конечного числа элементов различные комбинации и производить подсчет числа всех возможных комбинаций, составленных по некоторому правилу. Такие задачи получили название комбинаторных, а раздел математики, занимающийся их решением комбинаторикой.

Комбинаторика играет важную роль при решении некоторых проблем теории вероятностей, кибернетики, вычислительной техники и т. д.

Теория вероятностей сформировалась в самостоятельную научную дисциплину в 17 – 18 веках. Приоритет в научном подходе к изучению случайных явлений и введение в науку понятие вероятности случайного события принадлежит двум французским математикам: Блезу Паскалю и Пьеру Ферма. Общепринятой в мире в настоящее время считается аксиоматика Колмогорова, русского ученого - математика, на которой строится теория вероятностей. В настоящее время особое значение приобрели вероятностно – статистические методы, которые находят широкое применение не только в физике, кибернетике, геодезии, строительстве, экономике, медицине, биологии и т.д., но и в теории стихосложения и лингвистике, психологии. Практически, пожалуй, нет области знаний, в которой не использовались бы методы теории вероятностей и математической статистики.

Данные конспекты лекций написаны с целью оказания помощи студентам в изучении раздела математики «Основы теории вероятностей и математической статистики». В лекциях последовательно вводятся все базовые понятия теории вероятности и математической статистики, формулируются основные теоремы, большая часть которых не доказывается. Рассматриваются основные задачи и методы их решения и технологии применения этих методов к решению практических задач. Изложение сопровождается подробными комментариями и многочисленными примерами.

Лекции могут быть использованы для первичного ознакомления с изучаемым материалом, при конспектировании лекций, для подготовки к практическим занятиям, для закрепления полученных знаний, умений и навыков. Кроме того, пособие будет полезно и студентам - старшекурсникам как справочное пособие, позволяющее быстро восстановить в памяти то, что было изучено ранее.

В конце каждой лекции приведены контрольные вопросы.

Нумерация тем в конспекте лекций приведена в соответствии с рабочей программой раздела «Основы теории вероятностей и математической статистики» дисциплины «Математика».

Лекция 1 Задачи теории вероятностей

Тема 1.1 Задачи теории вероятностей

Теория вероятностей — раздел математики, изучающий закономерности случайных явлений: случайные события, случайные величины, их свойства и операции над ними.

Теория вероятностей изучает объективные закономерности массовых случайных событий. Путем наблюдений (испытаний, экспериментов), т.е. опыта в широком смысле слова, происходит познание явлений действительного мира.

В своей практической деятельности мы часто встречаемся с явлениями, исход которых невозможно предсказать, результат которых зависит от случая.

Случайное явление можно охарактеризовать отношением числа его наступлений к числу испытаний, в каждом из которых при одинаковых условиях всех испытаний оно могло наступить или не наступить.

Возникновение теории вероятностей как науки относят к средним векам и первым попыткам математического анализа азартных игр (орлянка, кости, рулетка). Первоначально её основные понятия не имели строго математического вида, к ним можно было относиться как к некоторым эмпирическим фактам, как к свойствам реальных событий и они формулировались в наглядных представлениях. Самые ранние работы учёных в области теории вероятностей относятся к XVII веку. Исследуя прогнозирование выигрыша в азартных играх, Блез Паскаль и Пьер Ферма открыли первые вероятностные закономерности, возникающие при бросании костей. Важный вклад в теорию вероятностей внёс Якоб Бернулли - он дал доказательство закона больших чисел в простейшем случае независимых испытаний. В первой половине XIX века теория вероятностей начинает применяться к анализу ошибок наблюдений; Лаплас и Пуассон доказали первые предельные теоремы. Во второй половине XIX века основной вклад внесли русские учёные П. Л. Чебышев, А. А. Марков и А. М. Ляпунов. В это время были доказаны закон больших чисел, центральная предельная теорема, а также разработана теория цепей Маркова. Современный вид теория вероятностей получила благодаря аксиоматизации, предложенной Андреем Николаевичем Колмогоровым. В результате теория вероятностей приобрела строгий математический вид и окончательно стала восприниматься как один из разделов математики.

Тема 1.2 События и их виды

Всякое действие, явление, наблюдение с несколькими различными исходами, реализуемое при данном комплексе условий, будем называть ***испытанием.*** Результат этого действия или наблюдения называется ***событием***. Если, например, испытание состоит в бросании монеты, то выпадение герба является событием; если испытание — изготовление подшипника данного типа, то соответствие подшипника стандарту — событие; если испытание — бросание игральной кости, т. е. кубика, на гранях которого проставлены цифры (очки) от 1 до 6, то выпадение пятерки — событие.

События принято обозначать заглавными буквами латинского алфавита: $A, B, C, \ldots$. Если событие при заданных условиях может произойти или не произойти, то оно называется ***случайным***. В том случае, когда событие должно непременно произойти, его называют ***достоверным***, а в том случае, когда оно заведомо не может произойти,- ***невозможным***.

События называются ***несовместными***, если каждый раз возможно появление только одного из них.

События называются ***совместными***, если в данных условиях появление одного из этих событий не исключает появление другого при том же испытании.

События называются ***противоположными,*** если в условиях испытания они, являясь единственными его исходами, несовместны.

Полной системой событий A_1, A_2, A_3, … , A_n называется совокупность несовместных событий, наступление хотя бы одного из которых обязательно при данном испытании.

Если полная система состоит из двух несовместных событий, то такие события называются противоположными и обозначаются A и $\overline{A}$.

Задача1. В коробке находится 30 пронумерованных шаров. Установить, какие из следующих событий являются невозможными, достоверными, противоположными:

достали пронумерованный шар *(A);*

достали шар с четным номером *(B);*

достали шар с нечетным номером *(C);*

достали шар без номера *(Д).*

Какие из них образуют полную группу?

Решение. *A* – достоверное событие; *Д* – невозможное событие;

B и *C* – противоположные события.

Полную группу событий составляют *A* и *Д, B* и *C*.

Тема 1.3 Алгебра событий

Теорема сложения вероятностей несовместных событий

Суммой конечного числа событий называется событие, состоящее в наступлении хотя бы одного из них.

Сумму двух событий обозначают символом A+B, а сумму *n* событий символом

$$A_1+A_2+ \dots +A_n.$$

Теорема сложения вероятностей

Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий.

$$P(A+B)=P(A)+P(B) \text{ или}$$

$$P(A_1+A_2+\ldots+A_n)=P(A_1)+P(A_2)+\ldots+P(A_n)$$

Следствие 1. Если событие A_1, A_2, … ,A_n образует полную систему, то сумма вероятностей этих событий равна единице.

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

Следствие 2. Сумма вероятностей противоположных событий A и $\overline{A}$ равна единице.

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

Задача 2. Имеется 100 лотерейных билетов. Известно, что на 5 билетов попадает выигрыш по 20000 руб., на 10 - по 15000 руб, на 15 - по 10000 руб., на 25 - по 2000 руб. и на остальные ничего. Найти вероятность того, что на купленный билет будет получен выигрыш не менее 10000 руб.

Решение. Пусть А, В, и С- события, состоящие в том, что на купленный билет падает выигрыш, равный соответственно 20000, 15000 и 10000 руб. Так как события А, В и С несовместны, то

$$P(A+B+C)=P(A)+P(B)+P(C)=\frac{5}{100}+\frac{10}{100}+\frac{15}{100}=0{,}3\cdot$$

Задача 3. На заочное отделение техникума поступают контрольные работы по математике из городов *А, В* и *С*. Вероятность поступления контрольной работы из города *А* равна 0,6, из города *В* - 0,1. Найти вероятность того, что очередная контрольная работа поступит из города *С*.

Решение. События «контрольная работа поступила из города *А*», «контрольная работа поступила из города В» и «контрольная работа поступила из города С» образуют полную систему, поэтому сумма их вероятностей равна единице:

$$0{,}6+0{,}1+p=1,\text{ т. е. } p=1-0{,}7=0{,}3.$$

Задача 4. Вероятность того, что день будет ясным, $p=0{,}85$. Найти вероятность g того, что день будет облачным.

Решение. События «день ясный» и «день облачный» противоположные, поэтому

$$p+g=1,\text{ т.е } g=1-p=1-0{,}85=0{,}15.$$

Разностью событий А и *В* называется событие *А* \ *В*, происходящее тогда и только тогда, когда происходит А, но не происходит В.

Теорема умножения вероятностей независимых событий

При совместном рассмотрении двух случайных событий *А* и *В* возникает вопрос: как связаны события *А* и *В* друг с другом, как наступление одного из них влияет на возможность наступления другого?

Простейшим примером связи между двумя событиями служит причинная связь, когда наступление одного из событий обязательно приводит к наступлению другого, или наоборот, когда наступление одного исключает возможность наступления другого.

Для характеристики зависимости одних событий от других вводится понятие ***условной вероятности.***

Определение. Пусть *А* и *В* - два случайных события одного и того же испытания. Тогда условной вероятностью события *А* или вероятностью события А при условии, что наступило событие В, называется число $\frac{P(AB)}{P(B)}$.

Обозначив условную вероятность $P(A/B)$, получим формулу

$$P(A/B)=\frac{P(AB)}{P(B)},\ (P(B)\neq 0). \qquad (1)$$

Задача 5. Вычислить вероятность того, что в семье, где есть один ребенок- мальчик, родится второй мальчик.

Решение. Пусть событие *А* состоит в том, что в семье два мальчика, а событие *В* - что один мальчик.

Рассмотрим все возможные исходы: мальчик и мальчик; мальчик и девочка; девочка и мальчик; девочка и девочка.

Тогда $P(AB)=\frac{1}{4}$, $P(B)=\frac{3}{4}$ и по формуле (1) находим

$$P(A/B)=\frac{1}{4}:\frac{3}{4}=\frac{1}{3}\approx 0{,}3.$$

Событие *A* называется ***независимым*** от события *B*, если наступление события *B* не оказывает никакого влияния на вероятность наступления события *A*.

Теорема умножения вероятностей

Вероятность одновременного появления двух независимых событий равна произведению вероятностей этих событий:

$$P(AB)=P(A)\cdot P(B). \qquad (2)$$

Вероятность появления нескольких событий, независимых в совокупности, вычисляется по формуле

$$P(A_1A_2...A_n)=P(A_1)\cdot P(A_2)\cdot...\cdot P(A_n). \qquad (3)$$

Задача 6. В первой урне находится 6 черных и 4 белых шара, во второй- 5 черных и 7 белых шаров. Из каждой урны извлекают по одному шару. Какова вероятность того, что оба шара окажутся белыми.

Решение. Пусть A_1 - из первой урны извлечен белый шар; A_2- из второй урны извлечен белый шар. Очевидно, что события A_1 и A_2 независимы.

Так как $P(A_1)=\frac{4}{10}=\frac{2}{5}$, $P(A_2)=\frac{7}{12}$, то по формуле (2) находим

$$P(A_1A_2)=\frac{2}{5}\cdot\frac{7}{12}=\frac{14}{60}=\frac{7}{30}.$$

Задача 7. Прибор состоит из двух элементов, работающих независимо. Вероятность выхода из строя первого элемента равна 0,2; вероятность выхода из строя второго элемента равна 0,3. Найти вероятность того, что: а) оба элемента выйдут из строя; б) оба элемента будут работать.

Решение. Пусть событие *A*- выход из строя первого элемента, событие *B*- выход их строя второго элемента. Эти события независимы (по условию).

а) одновременное появление *A* и *B* есть событие *AB*. Следовательно,

$$P(AB)=P(A)\cdot P(B)=0{,}2\cdot 0{,}3=0{,}06;$$

б) Если работает первый элемент, то имеет место событие $\overline{A}$ (противоположное событию *A*- выходу этого элемента из строя); если работает второй элемент- событие *B*. Найдем вероятности событий $\overline{A}$ и $\overline{B}$:

$$P(\overline{A})=1-P(A)=1-0{,}2=0{,}8;$$

$$P(\overline{B})=1-P(B)=1-0{,}3=0{,}7.$$

Тогда событие, состоящее в том, что будут работать оба элемента, есть $\overline{A}\overline{B}$ и значит

$$P(\overline{A}\overline{B})=P(\overline{A})\cdot P(\overline{B})=0{,}8\cdot 0{,}7=0{,}56.$$

Задача 8. Пусть опыт *G* заключается в проведении стрельбы наугад по квадрату "Ω", точки которого являются элементарными событиями *ω*. Пусть попадание в квадрат "Ω" есть достоверное событие Ω, а попадание в области "*A*" и "*B*" - события *A* и *B*. Тогда события *A* + *B*, *AB*, *A* \ *B*, *A* будут выглядеть следующим образом (рис.1):

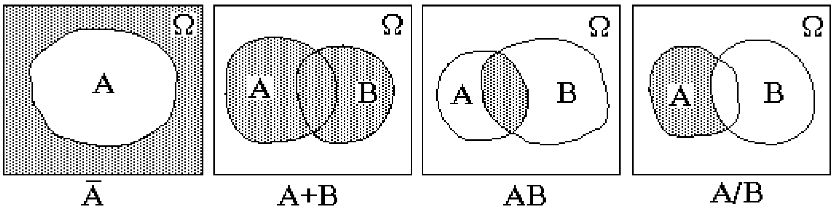


Рис.1. Диаграммы Венна

Графические изображения на плоскости соотношений между множествами называются ***диаграммами Венна*** (см. рис.1).

Рассмотрим пространство элементарных событий Ω, соответствующее некоторому стохастическому эксперименту и пусть F – некоторая система случайных событий. Систему событий F называют ***алгеброй событий***, если выполняются условия: Ω Є F; из того, что A Є F, следует, что $\overline{A}$ Є F; из того, что A и B Є F, следует, что AB и A+ B Є F. Отсюда следует, что, применяя любые из введенных выше операций к произвольной системе событий из F, получим событие также принадлежащее F.

Замечание. ***Алгебру событий*** иногда называют также *алгеброй Буля.*

Тема 1.4 Основные аксиомы теории вероятностей

Вероятность события, рассматривается как мера объективной возможности появления случайного события.

Классическое определение вероятности

Число, являющееся выражением меры объективной возможности наступления события, называется ***вероятностью*** этого события и обозначается символом *P(A).*

Определение. Вероятностью события *A* называется отношение числа исходов m, благоприятствующих наступлению данного события *A*, к числу *n* всех исходов (несовместных, единственно возможных и равновозможных), т.е.

$$P(A) = \frac{m}{n}. \qquad (4)$$

Следовательно, для нахождения вероятности события необходимо, рассмотрев различные исходы испытания, подсчитать все возможные несовместные исходы *n,* выбрать число интересующих нас исходов m и вычислить отношение *m* к *n*.

Из этого определения вытекают следующие ***аксиомы***:

Вероятность любого испытания есть неотрицательное число, не превосходящее единицы.

Действительно, число m искомых событий заключено в пределах $0 \le m \le n$. Разделив обе части на *n*, получим

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

2. Вероятность достоверного события равна единице, т. к. $n/n = 1$.

3. Вероятность невозможного события равна нулю, поскольку $0/n = 0$.

Простая система аксиом сформулированная А. Н. Колмогоровым, позволила описать уже существовавшие к тому времени классические разделы теории вероятностей, дать толчок развитию её новых разделов, например, теории случайных процессов, и стала общепринятой в современной теории вероятностей.

Задача 9. В лотерее из 1000 билетов имеются 200 выигрышных. Вынимают наугад один билет. Чему равна вероятность того, что этот билет выигрышный?

Решение. Общее число различных исходов есть n=1000. Число исходов, благоприятствующих получению выигрыша, составляет m=200. Согласно формуле (4), получим

$$P(A) = \frac{200}{1000} = \frac{1}{5} = 0,2.$$

Контрольные вопросы и задания

1. Назовите основную задачу теории вероятности?
2. Что называется событием? Перечислите основные виды событий. Дайте им определения. Приведите примеры событий.
3. В чем заключается алгебра событий?
4. Перечислите основные аксиомы теории вероятностей.

Лекция 2 Основные понятия комбинаторики

Тема 2.1 Основные понятия комбинаторики

Основатели теории вероятности французские ученые XYII века Пьер Ферма и Блез Паскаль первыми нашли ключ к составлению количественной оценки вероятности события. Они использовали метод, который позже был назван комбинаторным анализом, или, проще, комбинаторикой.

Например, если взять 10 различных цифр 0, 1, 2, 3, …, 9 и составлять из них комбинации, то будем получать различные числа, например 143, 431, 5671, 1207, 43 и т.п.

Мы видим, что некоторые из таких комбинаций отличаются только порядком цифр (например, 143 и 431), другие - входящими в них цифрами (например, 5671 и 1207), третьи различаются и числом цифр (например, 143 и 43).

Таким образом, полученные комбинации удовлетворяют различным условиям.

В зависимости от правил составления можно выделить три типа комбинаций: ***перестановки, размещения, сочетания***.

Предварительно познакомимся с понятием ***факториала***.

Произведение всех натуральных чисел от 1 до n включительно называют

n- факториалом и пишут

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n.$$

Задача 10. Вычислить: а) $3!$; б) $7! - 5!$; в) $\frac{7!+5!}{6!}$.

Решение. а) $3! = 1 \cdot 2 \cdot 3 = 6$.

б) Так как $7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7$ и $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$, то можно вынести за скобки $5!$

Тогда получим

$5!(6 \cdot 7 - 1) = 5! \cdot 41 = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 41 = 120 \cdot 41 = 4920$.

в) $\frac{7!+5!}{6!} = \frac{5!(6 \cdot 7 + 1)}{5! \cdot 6} = \frac{6 \cdot 7 + 1}{6} = \frac{43}{6}$.

Комбинация из n элементов, которые отличаются друг от друга только порядком элементов, называются ***перестановками***.

Перестановки обозначаются символом $\boldsymbol{P_n}$, где n- число элементов, входящих в каждую перестановку. (P - первая буква французского слова *permutation*- перестановка).

Число перестановок можно вычислить по формуле

$$P_n = n\cdot(n-1)(n-2)\ldots 3\cdot 2\cdot 1$$

или с помощью факториала:

$$P_n = n!$$

Запомним, что ***0! = 1 и 1! = 1.***

Задача 11. Сколькими способами можно расставлять на одной полке шесть различных книг?

Решение. Искомое число способов равно числу перестановок из 6 элементов, т. е.

$$P_6 = 6! = 1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6 = 720\,.$$

Размещениями из m элементов в n в каждом называются такие соединения, которые отличаются друг от друга либо самими элементами (хотя бы одним), либо порядком их расположения.

Размещения обозначаются символом A_m^n, где m- число всех имеющихся элементов, n- число элементов в каждой комбинации. (A-первая буква французского слова *arrangement*, что означает «размещение, приведение в порядок»).

При этом полагают, что $n \le m$.

Число размещений можно вычислить по формуле

$$A_m^n = \underbrace{m\cdot(m-1)(m-2)\cdot\ldots}_{n.\text{множителей}}, \qquad (6)$$

т. е. число всех возможных размещений из m элементов по n равно произведению n последовательных целых чисел, из которых большее есть m.

Запишем эту формулу в факториальной форме:

$$A_m^n = \frac{m!}{(m-n)!}\,. \qquad (7)$$

Задача 12. Сколько вариантов распределения трех путевок в санатории различного профиля можно составить для пяти претендентов?

Решение. Искомое число вариантов равно числу размещений из 5 элементов по 3 элемента, т. е.

$$A_5^3 = 5\cdot 4\cdot 3 = 60\,.$$

Сочетаниями называются все возможные комбинации из m элементов по n, которые отличаются друг от друга по крайней мере хотя бы одним элементом (здесь m и n- натуральные числа, причем $n \le m$).

Число сочетаний из m элементов по n обозначаются C_m^n (C-первая буква французского слова *combination*- сочетание).

В общем случае число из m элементов по n равно числу размещений из m элементов по n, деленному на число перестановок из n элементов:

$$C_m^n = \frac{A_m^n}{P_n}\,. \qquad (8)$$

Используя для чисел размещений и перестановок факториальные формулы, получим:

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

Задача 13. В бригаде из 25 человек нужно выделить четырех для работы на определенном участке. Сколькими способами это можно сделать?

Решение. Так как порядок выбранных четырех человек не имеет значения, то это можно сделать C_{25}^4 способами.

Находим по формуле (8):

$$C_{25}^4 = \frac{25 \cdot 24 \cdot 23 \cdot 22}{1 \cdot 2 \cdot 3 \cdot 4} = 12650 . \qquad (9)$$

Кроме того, при решении задач используются следующие формулы, выражающие основные свойства сочетаний:

$$C_m^n = C_m^{m-n} \quad (0 \le n \le m) \qquad (10)$$

(по определению полагают $C_n^n = 1$ и $C_n^0 = 1$);

$$C_m^n + C_m^{n+1} = C_{m+1}^{n+1} . \qquad (11)$$

Тема 2.2. Решение комбинаторных задач

Задача 14. На втором курсе техникума изучается 16 предметов. На понедельник нужно в расписание поставить 3 предмета. Сколькими способами можно это сделать?

Решение. Способов постановки в расписание трех предметов из 16 столько, сколько можно составить размещений из 16 элементов по 3.

$$A_{16}^3 = \frac{16!}{(16-3)!} = \frac{16!}{13!} = \frac{13! \cdot 14 \cdot 15 \cdot 16}{13!} = 14 \cdot 15 \cdot 16 = 3360 .$$

Задача 15. Из 15 объектов нужно отобрать 10 объектов. Сколькими способами это можно сделать?

Решение.

$$C_{15}^{10} = \frac{15!}{(15-10)! \cdot 10!} = \frac{15!}{5! 10!} = \frac{10! \cdot 11 \cdot 12 \cdot 13 \cdot 14 \cdot 15}{5! \cdot 10!} = \frac{11 \cdot 12 \cdot 13 \cdot 14 \cdot 15}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = \frac{11 \cdot 3 \cdot 13 \cdot 3 \cdot 14}{2 \cdot 3 \cdot 1 \cdot 1} =$$

$$= \frac{11 \cdot 3 \cdot 13 \cdot 14}{2} = 11 \cdot 3 \cdot 13 \cdot 7 = 3003.$$

Задача16. В соревнованиях участвовало четыре команды. Сколько вариантов распределения мест между ними возможно?

Решение.

$$P_4 = 1 \cdot 2 \cdot 3 \cdot 4 = 24 .$$

Задача 17. Сколькими способами можно составить дозор из трех солдат и одного офицера, если имеется 80 солдат и 3 офицера?

Решение. Солдат в дозор можно выбрать

$$C_{80}^3 = \frac{80!}{77! 3!} = \frac{77! \cdot 78 \cdot 79 \cdot 80}{77! \cdot 1 \cdot 2 \cdot 3} = \frac{78 \cdot 79 \cdot 80}{2 \cdot 3} = 13 \cdot 79 \cdot 80 = 82160$$

способами, а офицеров $C_3^1 = 3$ способами. Так как с каждой командой из солдат может пойти любой офицер, то всего имеется $C_{80}^3 \cdot C_3^1 = 82160 \cdot 3 = 246480$ способов.

Задача 18. В партии из 18 деталей находятся 4 бракованных. Наугад выбирают 5 деталей. Найти вероятность того, что из этих 5 деталей две окажутся бракованными.

Решение. Число всех равновозможных независимых исходов *n* равно числу сочетаний из 18 по 5 т. е.

$$n = C_{18}^{5} = \frac{18 \cdot 17 \cdot 16 \cdot 15 \cdot 14}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 18 \cdot 17 \cdot 28 = 8568$$

Подсчитаем число m, благоприятствующих событию А. Среди 5 взятых наугад деталей должно быть 3 качественных и 2 бракованных. Число способов выборки двух бракованных деталей из 4 имеющихся бракованных равно числу сочетаний из 4 по 2:

$$C_4^2 = \frac{4 \cdot 3}{1 \cdot 2} = 6.$$

Число способов выборки трех качественных деталей из 14 имеющихся качественных равно

$$C_{14}^3 = \frac{14 \cdot 13 \cdot 12}{1 \cdot 2 \cdot 3} = 14 \cdot 13 \cdot 2 = 364.$$

Любая группа качественных деталей может комбинироваться с любой группой бракованных деталей, поэтому общее число комбинаций m составляет

$$m = C_4^2 \cdot C_{14}^3 = 6 \cdot 364 = 2184.$$

Искомая вероятность события А равна отношению числа исходов m, благоприятствующих этому событию, к числу n всех равновозможных независимых исходов:

$$P(A) = \frac{m}{n} = \frac{2184}{8568} \approx 0{,}255.$$

Задача 19. Найти x, если известно, что $C_{x-2}^2 = 21$.

Решение. Так как $C_{x-2}^2 = \frac{(x-2)!}{(x-2-2)!2!} = \frac{(x-2)!}{(x-4)!\cdot 2} = \frac{(x-4)!\cdot(x-3)(x-2)}{(x-4)!\cdot 2} = \frac{(x-3)(x-2)}{2}$, то получим

$$\frac{(x-3)(x-2)}{2} = 21,$$

$$(x-3)(x-2) = 42,$$

$$x^2 - 5x + 6 - 42 = 0,$$

$$x^2 - 5x - 36 = 0$$

$$x_1 = -4,\ x_2 = 9.$$

По определению сочетания следует, что $x - 2 \le 2$, $x \le 4$. Таким образом $x = 9$.

Ответ: 9

Контрольные вопросы и задания

1. Перечислите основные задачи комбинаторики
2. Что называется n- факториалом? Чему равен n!.
3. Что называется перестановками, размещениями, сочетаниями.
4. Запишите формулы вычисления перестановок, размещений, сочетаний.

Лекция 3 Случайные величины

Тема 3.1 Случайные величины

Случайной называется величина, которая в результате испытания может принять то или иное числовое значение, причем заранее неизвестно, какое именно.

Если для какой- либо величины ее измерение повторять многократно в практически одинаковых условиях, то обнаружится, что всякий раз получаются несколько отличные друг от друга результаты. Это складывается влияние причин двух видов: 1) основных, определяющих главное значение результата; 2) второстепенных, обуславливающих их расхождение.

При совместном действии этих причин понятия необходимости и случайности оказываются тесно связанными между собой, но необходимое преобладает над случайным.

Таким образом, возможные значения случайных величин принадлежат некоторым числовым множествам.

Случайным является то, что на этих множествах величины могут принять любое значение, но какое именно, заранее сказать нельзя.

Случайная величина связана со случайным событием.

Если случайное событие – ***качественная характеристика*** испытаний, то случайная величина – его ***количественная характеристика***.

Случайные величины обозначают заглавными латинскими буквами $X, Y, Z,$ а их значение – прописными- x_i, y_i, z_i.

Вероятность того, что случайная величина X примет значение x_1 обозначают:

$$P(X = x_1) = p_1 \text{ и т.д.}$$

Случайные величины задают законами распределения.

Закон распределения случайной величины – это соответствие, установленное между возможными значениями случайной величины и их вероятностями.

Законы распределения могут быть заданы тремя способами: табличным, графическим, аналитическим. Способ задания зависит от типа случайной величины.

Различают два основных типа случайных величин: ***дискретные и непрерывно распределенные случайные величины.***

Тема 3.2. Дискретная и непрерывная случайные величины

Если значения, которые может принимать данная случайная величина X, образует дискретный (конечный или бесконечный) ряд чисел $x_1, x_2, ..., x_n, ...,$ то и сама случайная величина X называется ***дискретной.***

Если же значения, которые может принимать данная случайная величина X, заполняют конечный или бесконечный промежуток (а, в) числовой оси *Ох,* то случайная величина называется ***непрерывной.***

Каждому значению случайной величины дискретного типа x_n отвечает определенная вероятность p_n; каждому промежутку (а, в) из области значений случайной величины непрерывного типа также отвечает определенная вероятность $P(a < X < b)$ того, что значение, принятое случайной величиной, попадает в этот промежуток.

Тема 3.3. Закон распределения случайной величины

Соотношение, устанавливающее тем или иным способом связь между возможными значениями случайной величины и их вероятностями, называется ***законом распределения*** случайной величины.

Закон распределения дискретной случайной величины обычно задается ***рядом распределения:***

x_i	x_1	x_2	x_3	$\dots$	x_n
p_i	p_1	p_2	p_3	$\dots$	p_n

При этом $\sum_{i=1}^{n} p_i = 1$, где суммирование распространяется на все (конечное или бесконечное) множество возможных значений данной случайной величины X.

Закон распределения непрерывной случайной величины удобно задавать с помощью ***функции плотности вероятности*** $f(x)$.

Вероятность $P(a < X < b)$ того, что значение, принятое случайной величиной X, попадет в промежуток (а, в), определяется равенством

$$P(a < X < b) = \int_a^b f(x)dx.$$

График функции $f(x)$ называется ***кривой распределения***. Геометрически вероятность попадания случайной величины в промежуток (а, в) равна площади соответствующей криволинейной трапеции, ограниченной кривой распределения, осью *Ox* и прямыми *x=а, x=в* (рис.2).

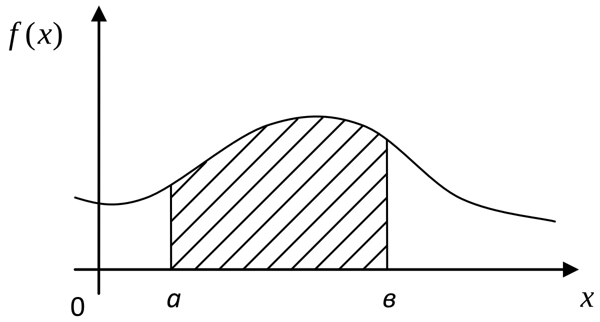


Рис. 2

Задача 20. Даны вероятности значений случайной величины X: значение 10 имеет вероятность 0,3; значение 2 – вероятность 0,4; значение 8 – вероятность 0,1; значение 4 – вероятность 0,2. Построить ряд распределения случайной величины X.

Решение. Расположив значения случайной величины в возрастающем порядке, получим ряд распределения:

x_i	2	4	8	10
p_i	0,4	0,2	0,1	0,3

Возьмем на плоскости xOp точки (2; 0,4), (4; 0,2), (8; 0,1) и (10; 0,3). Соединив последовательные точки прямолинейными отрезками, получим ***многоугольник*** (или ***полигон***) распределения случайной величины X (рис. 3).

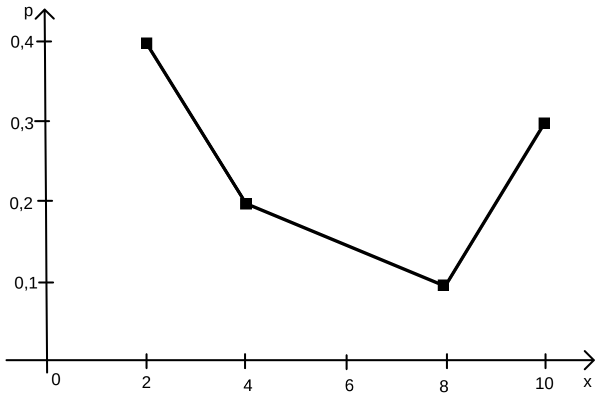


Рис.3

Задача 21. Разыгрываются две вещи стоимостью по 5000 руб и одна вещь стоимостью 30000 руб. Составить закон распределения выигрышей для человека, купившего один билет из 50.

Решение. Искомая случайная величина X представляет собой выигрыш и может принимать три значения: 0, 5000 и 30000 руб. Первому результату благоприятствует 47 случаев, второму результату - два случая и третьему – один случай. Найдем их вероятности:

$$P(x_1) = \frac{47}{50} = 0{,}94; \quad P(x_2) = \frac{2}{50} = 0{,}04; \quad P(x_3) = \frac{1}{50} = 0{,}02.$$

Закон распределения случайной величины имеет вид:

x_i	0	5000	30000
p_i	0,94	0,04	0,02

В качестве проверки найдем

$$P(x_1) + P(x_2) + P(x_3) = 0{,}94 + 0{,}04 + 0{,}02 = 1.$$

Задача 22. Случайная величина X подчинена закону распределения с плотностью $f(x)$, причем

$$f(x) = \begin{cases} 0 & при \quad x < 0, \\ a = (3x - x^2) & при \quad 0 \le x \le 3, \\ 0 & при \quad x > 0. \end{cases}$$

Требуется: 1) найти коэффициент а; 2) построить график распределения плотности $y = f(x)$; 3) найти вероятность попадания X в промежуток (1; 2).

Решение. 1. Так как все значения данной случайной величины заключены на отрезке [0;3] (см. рис.4), то

$$\int_0^3 a(3x - x^2)dx = 1, \text{ откуда}$$

$$a(\frac{3}{2}x^2 - \frac{1}{3}x^3)\Big|_0^3 = 1, \text{ или}$$

$$a(\frac{27}{2} - 9) = 1, \text{ т.е. } a = \frac{2}{9}.$$

2. Графиком функции $f(x)$ в интервале [0; 3] является парабола $y = \frac{2}{3}x - \frac{2}{9}x^2$, а вне этого интервала графиком служит сама ось абсцисс (см. рис. 4).

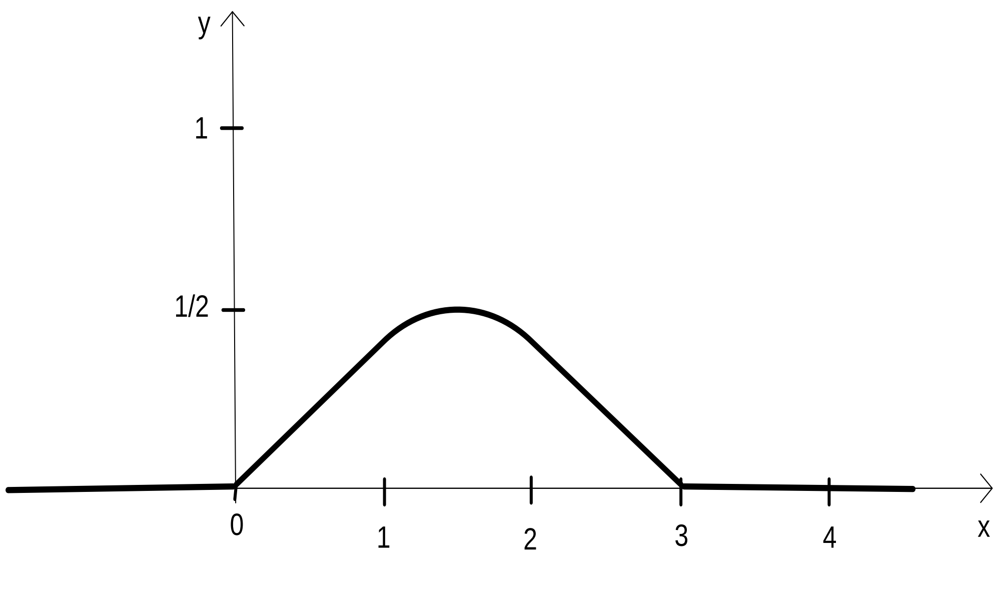


Рис. 4

3. Вероятность попадания случайной величины X в промежуток (1; 2) найдется из равенства

$$P(1 < X < 2) = \int_1^2 (\frac{2}{3}x - \frac{2}{9}x^2)dx = (\frac{x^2}{3} - \frac{2x^3}{27})\Big|_1^2 = \frac{4}{3} - \frac{16}{27} + \frac{2}{27} - \frac{1}{3} = \frac{13}{27}.$$

Тема 3.4 Вычисление вероятностей событий

Задача 23. В урне находится 10 шаров, из них 6 белых и 4 черных шара. Вынули из урны 2 шара. Какова вероятность того, что оба шара - белые?

Решение. Рассмотрим событие А – оба вынутых шара белого цвета.

Число всевозможных исходов равно количеству выборок 2 шаров из 10. Выборка без возвращения и без повторения, поэтому $n = C_{10}^2 = \frac{10!}{2! \cdot 8!} = 45$. Число исходов, благоприятствующих наступлению события А равно числу вариантов извлечения 2 белых шаров из 6, поэтому

$m = C_6^2 = \frac{6!}{2! \cdot 4!} = 15$. Тогда

$$p(A) = \frac{15}{45} = \frac{1}{3} = 0{,}33.$$

Ответ: $p(A) = \frac{1}{3} = 0{,}33$.

Задача 24. В секретном замке на общей оси 4 диска, каждый из которых разделен на 5 секторов, на которых написаны различные цифры. Замок открывается, если диски установлены так, что цифры на них составляют определенное четырехзначное число. Найти вероятность того, что при произвольной установке дисков замок будет открыт.

Решение. Рассмотрим событие А – замок будет открыт. Это событие равносильно тому, что цифры на дисках составляют определенное число.

Так как варианты набора цифр на дисках образуют выборку с возвращением (цифры могут повторяться) упорядоченную (при смене порядка цифр получается другое число), $n = \overline{A}_5^4 = 5^4 = 625.$ Благоприятный исход у этого события только один, поэтому

m = 1. Тогда $p(A) = \frac{1}{625}.$

Задача 25. Набирая номер телефона, абонент забыл последние 3 цифры и, помня лишь, что эти цифры различны, набрал их наудачу. Найти вероятность того, что набраны нужные цифры.

Решение. Пусть событие А – набран верный номер. Тогда число всевозможных исходов равно числу трехзначных чисел, составленных из различных цифр. Так как в этом случае мы имеем выборку без возвращения (цифры различны), но упорядоченную (меняя цифры местами, получаем новое число), то $n = A_{10}^3 = \frac{10!}{7!} = 720.$ Исход, благоприятствующий наступлению события А только 1. Поэтому $p(A) = \frac{1}{720}.$

Задача 26. В почтовом отделении имеются открытки 6 видов. Какова вероятность того, что среди четырех проданных открыток все открытки различны?

Решение. Пусть событие А - все проданные открытки различны.

Тогда число всевозможных исходов равно числу вариантов выбора четырех открыток. Эта выборка с возвращением (выбранные открытки могут быть одинаковые), неупорядоченная (так как важен лишь состав выборки, а не то, в каком порядке отобраны открытки). Значит $n = \overline{C}_6^4 = \frac{9!}{4!5!} = 126.$ Число исходов, благоприятствующих наступлению события А, есть число способов, которыми можно выбрать четыре различные открытки из 6 видов. Так как открытки теперь различны, то эта неупорядоченная выборка без повторения, значит $m = C_6^4 = 15.$ Тогда $p(A) = \frac{5}{42} \approx 0{,}12.$

Ответ: $p(A) = \frac{5}{42} \approx 0{,}12.$

Контрольные вопросы и задания

1. Какая величина называется случайной?
2. Какая величина называется дискретной?
3. Что называется случайной величиной. Типы случайных величин.
4. В чем заключается закон случайной величины?
5. Как вычисляется вероятность событий?

Лекция 4. Независимые испытания.

Тема 4.1. Независимые испытания. Формулы Байеса и Бернулли.

Определение 1. События A и B называются ***независимыми***, если $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$.

Определение 2. События $A_1, ..., A_n$ называются ***независимыми в совокупности*** или просто ***независимыми***, если каждое из них не зависит от произведения любой совокупности остальных. Если любые два события из $A_1, ..., A_n$ независимы, то $A_1, ..., A_n$ называются ***попарно независимыми***.

Замечание. Независимость событий не следует из их попарной независимости, но обратное утверждение верно.

Определение 3. Если события $A_1, ..., A_n$ независимы, то

$$P(A_1|A_2) = P(A_1),\ P(A_3|A_1A_2) = P(A_3),\ ...,\ P(A_n|A_1...A_{n-1}) = P(A_n).$$

Если событие А может наступить только при появлении одного из несовместных событий (гипотез) $H_1, H_2, ..., H_n$, то вероятность события А вычисляется по формуле *полной вероятности*:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + \ldots + P(H_n)P(A/H_n), \quad (12)$$

где $P(H_i)$ - вероятность гипотезы H_i, $\sum_{i=1}^{n} P(H_i) = 1$; $P(A/H_i)$- условная вероятность события А при выполнении гипотезы H_i ($i = 1, 2, \ldots, n$).

Проиллюстрируем формулу полной вероятности на графе с выделенной вершиной (см. рис. 5):

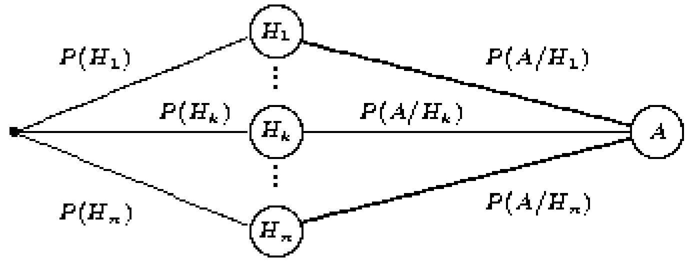


Рис. 5

Полная вероятность события А равна весу всего вероятностного графа с гипотезами.

С формулой полной вероятности тесно связана формула Байеса. Если до опыта вероятности гипотез были $P(H_1)$, $P(H_2)$ ..., $P(H_n)$, а в результате опыта появилось событие А, то с учетом этого события «новые», т. е. условные вероятности гипотез вычисляются по ***формуле Байеса***

$$P(H_k/A) = \frac{P(H_k) \cdot P(A/H_k)}{P(A)} \ (k = 1, 2, ...n), \text{ где } P(A) = \sum_{i=1}^{n} P(H_i) \cdot P(A/H_i). \tag{13}$$

Формула Байеса дает возможность «пересмотреть» вероятность гипотез с учетом наблюдавшегося результата опыта. Условная вероятность P(Hk/A) может находиться как отношение веса ветви, проходящей через вершину, соответствующую гипотезе Hk, к весу всего вероятностного графа.

Задача 27. На заводе, изготовляющем болты, первая машина производит 25%, вторая - 35%, третья - 40% всех изделий. В их продукции брак составляет соответственно 5, 4 и 2%.

1. Какова вероятность того, что случайно выбранный болт дефектный?
2. Случайно выбранный из продукции болт оказался дефектным. Какова вероятность того, что он был произведен первой, второй, третьей машиной?

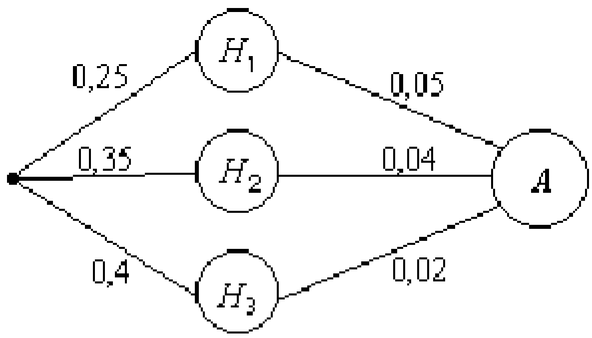
Решение. Пусть событие А = {выбрать дефектный болт}.

Выдвигаем три гипотезы:

H_1 ={болт изготовлен первой машиной}, $P(H_1)$ =0,25, $P(A/ H_1)$ =0,05;

H_2={болт изготовлен второй машиной}, $P(H_2)$=0,35, $P(A/ H_2)$=0,04;

H_3={болт изготовлен третьей машиной}, $P(H_3)$=0,4, $P(A/ H_3)$ =0,02.



P(A) = 0,25 • 0,05 + 0,35• 0,04 + 0,4• 0,02 = 0,0345

1.

$$P(H_1/A) = \frac{0,25 \cdot 0,05}{0,0345} = \frac{25}{69};$$

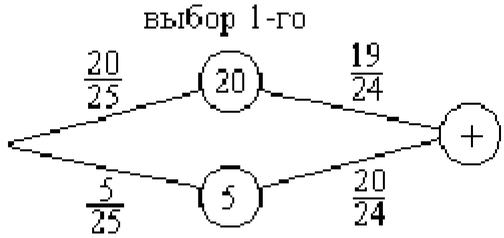
$$P(H_2/A) = \frac{0,35 \cdot 0,04}{0,0345} = \frac{28}{69};$$

$$P(H_3/A) = \frac{0,4 \cdot 0,02}{0,0345} = \frac{16}{69}.$$

Задача 28. Студент подготовил к экзамену 20 билетов из 25. В каком случае шансы взять известный билет больше - когда студент пришел на экзамен первым или вторым?

$$P_1 = \frac{20}{25} = \frac{4}{5}.$$

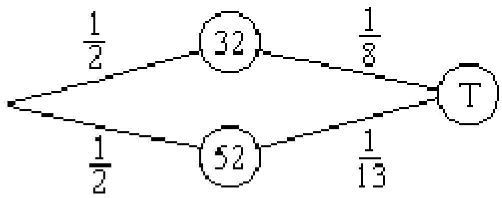
Решение. Найдем вероятность P_2 взять известный билет, придя на экзамен вторым, учитывая, что первый может взять как известный, так и неизвестный второму билет.

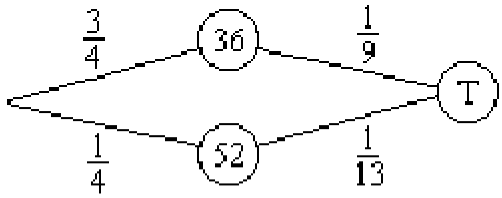


$$P_2 = \frac{20}{25} \cdot \frac{19}{24} + \frac{5}{25} \cdot \frac{20}{24} = \frac{4}{5} \quad .$$

Задача 29. Наудачу выбираем колоду, а из нее карту. В каком случае вероятность достать туз больше: если выбирать карту из двух колод, содержащих по 32 и 52 карты, или выбирать карту из трех колод в 36 карт и одной в 52?

Решение. Пусть событие m = {достать туз}.



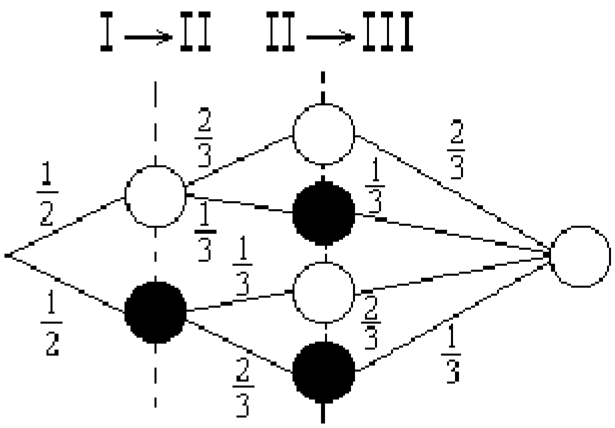


$$P_1(T) = \frac{1}{2}\cdot\frac{1}{8} + \frac{1}{2}\cdot\frac{1}{13} = \frac{21}{208}; \quad P_2(T) = \frac{3}{4}\cdot\frac{1}{9} + \frac{1}{4}\cdot\frac{1}{13} = \frac{4}{39};$$

$P_1(T) : P_2(T) = 63 : 64 < 1$, следовательно, в первом случае вероятность достать туз меньше, чем во втором.

Задача 30. В каждой из трех урн содержится по одному белому и одному черному шару. Из первой урны во вторую переложили один шар, из второй пополненной урны в третью тоже переложили один шар, а затем из третьей урны наудачу извлекли один шар. Какова вероятность извлечь белый шар из третьей пополненной урны?

Решение.

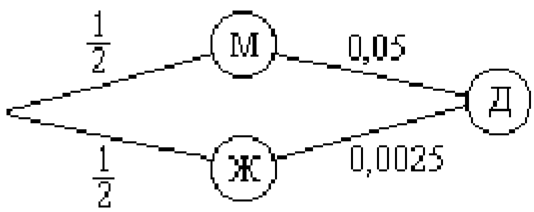


$$P() = \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{3} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{3} = \frac{1}{2}.$$

Какие гипотезы использовались в решении этой задачи?

Задача 31. Предположим, что 5 мужчин из 100 и 25 женщин из 10000 являются дальтониками. Наугад выбранное лицо страдает дальтонизмом. Какова вероятность, что это мужчина?

Решение.



$$P(\text{М} \,/\, \text{Д}) = \frac{\frac{1}{2}\cdot 0{,}05}{\frac{1}{2}\cdot 0{,}05 + \frac{1}{2}\cdot 0{,}0025} = \frac{20}{21}.$$

Тема 4.2. Биномиальное распределение.

Пусть производится определенное число n независимых опытов, причем в каждом из них с одной и той же вероятностью может наступить некоторое событие P. Рассмотрим случайную величину X, представляющую собой число наступлений событий A в n опытах. Закон ее распределения имеет вид

Значения x_i	0	1	2	…	n
Вероятности p_i	$P(A_{n,0})$	$P(A_{n,1})$	$P(A_{n,2})$		$P(A_{n,n})$

где $P(A_{n,k})$, вычисляется по ***формуле Бернулли***.

$$P(A_{n,k}) = C_n^k \cdot p^k \cdot (1-p)^{n-k} = C_n^k p^k g^{n-k}, \qquad (13)$$

где p – вероятность «успеха», g = 1 - p – вероятность «неудачи» в отдельном испытании.

Закон распределения, который характеризуется такой таблицей, называется ***биноминальным***.

Задача 32. Монету подбрасывают 5 раз. Составить закон распределения случайной величины X - числа выпадения герба.

Решение. Возможны следующие значения случайной величины X : 0, 1, 2, 3, 4, 5. Зная, что вероятность выпадения герба в одном испытании равна $\frac{1}{2}$, найдем вероятности значений случайной величины X по формуле Бернулли:

$$P(A_{5,0}) = C_5^0 p^0 g^5 = 1 \cdot (\frac{1}{2})^0 \cdot (\frac{1}{2})^5 = \frac{1}{32};$$

$$P(A_{5,1}) = C_5^1 p^1 g^4 = 5 \cdot (\frac{1}{2})^1 \cdot (\frac{1}{2})^4 = \frac{5}{32};$$

$$P(A_{5,2}) = C_5^2 p^2 g^3 = 10 \cdot (\frac{1}{2})^2 \cdot (\frac{1}{2})^3 = \frac{10}{32};$$

$$P(A_{5,3}) = C_5^3 p^3 g^2 = 10 \cdot (\frac{1}{2})^3 \cdot (\frac{1}{2})^2 = \frac{10}{32};$$

$$P(A_{5,4}) = C_5^4 p^4 g^1 = 5 \cdot (\frac{1}{2})^4 \cdot (\frac{1}{2})^1 = \frac{5}{32};$$

$$P(A_{5,5}) = C_5^5 p^5 g^0 = 1 \cdot (\frac{1}{2})^5 \cdot (\frac{1}{2})^0 = \frac{1}{32}.$$

Закон распределения имеет вид

Значения x_i	0	1	2	3	4	5
Вероятности p_i	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Сделаем проверку:

$$\frac{1}{32} + \frac{5}{32} + \frac{10}{32} + \frac{10}{32} + \frac{5}{32} + \frac{1}{32} = \frac{32}{32} = 1.$$

Контрольные вопросы и задания

1. Дайте определение полной вероятности. Запишите формулу Байеса.
2. Опишите схему Бернулли. Какие элементарные события повторяются в этих опытах?
3. Запишите формулу Бернулли.
4. Что называется законом распределения случайной величины?
5. Какой закон распределения называется биномиальным?

Лекция 5. Область применения и задачи математической статистики.

Тема 5.1. Задачи математической статистики.

«Статистика знает все» – такими словами начинается вторая часть романа И.Ильфа и Е.Петрова «Двенадцать стульев». «Известно, сколько какой пищи съедает в год средний гражданин республики… Известно, сколько в стране охотников, балерин, станков, собак всех пород, велосипедов, памятников, девушек, маяков и швейных машинок. Как много жизни, полной пыла, страстей и мысли, глядит на нас со статистических таблиц!» Зачем нужны эти таблицы, как их составлять и обрабатывать, какие выводы на их основании можно делать – на эти вопросы отвечает статистика (от итальянского stato – государство, латинского status – состояние).

Под ***математической статистикой*** понимают «раздел математики, посвященный математическим методам сбора, систематизации, обработки и интерпретации статистических данных, а также использование их для научных или практических выводов. Правила и процедуры математической статистики опираются на теорию вероятностей, позволяющую оценить точность и надежность выводов, получаемых в каждой задаче на основании имеющегося статистического материала». При этом ***статистическими данными*** называются сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающих теми или иными признаками. Современную математическую статистику определяют как *науку о принятии решений в условиях неопределенности.* Можно выделить две основные задачи математической статистики:

1. Указать способы сбора и группировки статистических сведений, полученных в результате наблюдений или в результате поставленных экспериментов.
2. Разработать методы анализа статистических данных в зависимости от целей исследования.

В связи с этим проводится:

а) оценка: неизвестной вероятности события, неизвестной функции распределения, параметров распределения, зависимости случайной величины от одной или нескольких случайных величин.

б) проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения.

Итак, задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Тема 5.2 Примеры применения теории вероятностей и математической статистики

Рассмотрим несколько примеров, когда вероятностно-статистические модели являются хорошим инструментом для решения управленческих, производственных, экономических, народнохозяйственных задач. Так, например, в романе А.Н. Толстого "Хождение по мукам" (т.1) говорится: "мастерская дает двадцать три процента брака, этой цифры вы и держитесь, - сказал Струков Ивану Ильичу". Встает вопрос, как понимать эти слова в разговоре заводских менеджеров, поскольку одна единица продукции не может быть дефектна на 23%. Она может быть либо годной, либо дефектной. Наверное, Струков имел в виду, что в партии большого объема содержится примерно 23% дефектных единиц продукции. Тогда возникает вопрос, а что значит "примерно"? Пусть из 100 проверенных единиц продукции 30 окажутся дефектными или из 1000-300, или из 100000-30000 и т.д., надо ли обвинять Струкова во лжи?

Или другой пример. Монетка, которую используют как жребий, должна быть «симметричной», т. е. при ее бросании в среднем в половине случаев должен выпадать герб, а в половине случаев - решетка (решка, цифра). Но что означает «в среднем»? Если провести много серий по 10 бросаний в каждой серии, то часто будут встречаться серии, в которых монетка 4 раза выпадает гербом. Для симметричной монеты это будет происходить в 20,5% серий. А если на 100000 бросаний окажется 40000 гербов, то можно ли считать монету симметричной? Процедура принятия решений строится на основе теории вероятностей и математической статистики. Рассматриваемый пример может показаться недостаточно серьезным. Однако это не так. Жеребьевка широко используется при организации промышленных технико-экономических экспериментов, например, при обработке результатов измерения показателя качества (момента трения) подшипников в зависимости от различных технологических факторов (влияния консервационной среды, методов подготовки подшипников перед измерением, влияния нагрузки подшипников в процессе измерения и т. п.). Допустим, необходимо сравнить качество подшипников в зависимости от результатов хранения их в разных консервационных маслах, т. е. в маслах состава A и B. При планировании такого эксперимента возникает вопрос, какие подшипники следует поместить в масло состава A, а какие - в масло состава B, но так, чтобы избежать субъективизма и обеспечить объективность принимаемого решения. Ответ на этот вопрос может быть получен с помощью жребия. Аналогичный пример можно привести и с контролем качества любой продукции. Чтобы решить, соответствует или не соответствует контролируемая партия продукции установленным требованиям, делается выборка. По результатам контроля выборки делается заключение обо всей партии. В этом случае очень важно избежать субъективизма при формировании выборки, т. е. необходимо, чтобы каждая единица продукции в контролируемой партии имела одинаковую вероятность быть отобранной в выборку. В производственных условиях отбор единиц продукции в выборку обычно осуществляют не с помощью жребия, а по специальным таблицам случайных чисел или с помощью компьютерных датчиков случайных чисел.

Аналогичные проблемы обеспечения объективности сравнения возникают при сопоставлении различных схем организации производства, оплаты труда, при проведении тендеров и конкурсов, подбора кандидатов на вакантные должности и т.п. Всюду нужна жеребьевка или подобные ей процедуры. Поясним на примере выявления наиболее сильной и второй по силе команд при организации турнира по олимпийской системе (проигравший выбывает). Пусть всегда более сильная команда побеждает более слабую. Ясно, что самая сильная команда однозначно станет чемпионом. Вторая по силе команда выйдет в финал тогда и только тогда, когда до финала у нее не будет игр с будущим чемпионом. Если такая игра будет запланирована, то вторая по силе команда в финал не попадет. Тот, кто планирует турнир, может либо досрочно «выбить» вторую по силе команду из турнира, сведя ее в первой же встрече с лидером, либо обеспечить ей второе место, обеспечив встречи с более слабыми командами вплоть до финала. Чтобы избежать субъективизма, проводят жеребьевку. Для турнира из 8 команд вероятность того, что в финале встретятся две самые сильные команды, равна 4/7. Соответственно с вероятностью 3/7 вторая по силе команда покинет турнир досрочно.

При любом измерении единиц продукции (с помощью штангенциркуля, микрометра, амперметра и т. п.) имеются погрешности. Чтобы выяснить, есть ли систематические погрешности, необходимо сделать многократные измерения единицы продукции, характеристики которой известны (например, стандартного образца). При этом следует помнить, что кроме систематической присутствует и случайная погрешность.

Поэтому встает вопрос, как по результатам измерений узнать, есть ли систематическая погрешность. Если отмечать только, является ли полученная при очередном измерении погрешность положительной или отрицательной, то эту задачу можно свести к предыдущей. Действительно, сопоставим измерение с бросанием монеты, положительную погрешность - с выпадением герба, отрицательную - решетки (нулевая погрешность при достаточном числе делений шкалы практически никогда не встречается). Тогда проверка отсутствия систематической погрешности эквивалентна проверке симметричности монеты.

Целью этих рассуждений является сведение задачи проверки отсутствия систематической погрешности к задаче проверки симметричности монеты. Проведенные рассуждения приводят к так называемому «критерию знаков» в математической статистике.

При статистическом регулировании технологических процессов на основе методов математической статистики разрабатываются правила и планы статистического контроля процессов, направленные на своевременное обнаружение разладки технологических процессов, принятия мер к их наладке и предотвращению выпуска продукции, не соответствующей установленным требованиям. Эти меры нацелены на сокращение издержек производства и потерь от поставки некачественных единиц продукции. При статистическом приемочном контроле на основе методов математической статистики разрабатываются планы контроля качества путем анализа выборок из партий продукции. Сложность заключается в том, чтобы уметь правильно строить вероятностно-статистические модели принятия решений, на основе которых можно ответить на поставленные выше вопросы. В математической статистике для этого разработаны вероятностные модели и методы проверки гипотез, в частности, гипотез о том, что доля дефектных единиц продукции равна определенному числу P^1, например, $P^1 = 0{,}23$ (вспомните слова Струкова из романа А.Н. Толстого).

Тема 5.3 Первичная обработка статистических данных

Описание данных – предварительный этап статистической обработки. Используемые при описании данных величины применяются при дальнейших этапах статистического анализа – оценивании и проверке гипотез, а также при решении иных задач, возникающих при применении вероятностно-статистических методов принятия решений, например, при статистическом контроле качества продукции и статистическом регулировании технологических процессов.

Статистические данные – это результаты наблюдений (измерений, испытаний, опытов, анализов). Функции результатов наблюдений, используемые, в частности, для оценки параметров распределений и (или) для проверки статистических гипотез, называют **статистиками**. Если в вероятностной модели результаты наблюдений рассматриваются как случайные величины (или случайные элементы), то статистики, как функции случайных величин (элементов), сами являются случайными величинами (элементами). Статистики, являющиеся выборочными аналогами характеристик случайных величин (математического ожидания, медианы, дисперсии, моментов и др.) и используемые для оценивания этих характеристик, называют **статистическими характеристиками**.

Для статистических методов обработки информации используются специальные термины (табл. 1).

Таблица 1

Новый термин	Простое описание	Более научный термин	Определение
Общий ряд данных	То, откуда выбирают	**Генеральная совокупность**	Множество всех в принципе возможных результатов измерения
Выборка	То, что выбрали	**Статистическая выборка, статистический ряд**	Множество результатов, реально полученных в данном измерении
Варианта	Значение одного из результатов измерения	**Варианта**	Одно из значений элементов выборки
Ряд данных	Значения всех результатов измерения, перечисленные по порядку	**Вариационный ряд**	Упорядоченное множество всех вариант

Основополагающее понятие в вероятностно-статистических методах принятия решений – **выборка**. Как уже говорилось, выборка – набор наблюдаемых значений или множество объектов, отобранные из изучаемой совокупности. Например, единицы продукции, отобранные из контролируемой партии или потока продукции для контроля и принятия решений. Наблюдаемые значения обозначим $x_1, x_2, \ldots, x_n$, где n – объем выборки, т.е. число наблюдаемых значений, составляющих выборку.

При повторных наблюдениях будут получены иные наблюдаемые значения, соответствующие другому элементарному событию $\omega = \omega_1$. Цель обработки статистических данных состоит в том, чтобы по результатам наблюдений, соответствующим элементарному событию $\omega = \omega_0$, сделать выводы о вероятностной мере P и результатах наблюдений при различных возможных $\omega = \omega_1$.

Задача 33. Измерили рост 50 студентов техникума. Получили следующие результаты:

162, 168, 157, 176, 185, 160, 162, 158, 181, 179,

164, 176, 177, 180, 181, 179, 175, 180, 176, 165,

168, 164, 179, 163, 160, 176, 162, 178, 164, 190,

181, 178, 168, 165, 176, 178, 185, 179, 180, 168,

160, 176, 175, 177, 176, 165, 164, 177, 175, 181.

С некоторым запасом мы можем считать, что рост студентов находится в пределах от 140см до 210см. Значит числа 140, 141, 142, ..., 208, 209, 210 образуют генеральную совокупность. Добавление числа к указанному множеству не перестанет быть генеральной совокупностью. Выборка – это данные реального измерения роста, варианта – это любое из чисел выборки, а **вариационный ряд** – все реальные результаты измерения, выписанные в определенном порядке без повторений, например по возрастанию: 157, 158, 160, 162, 163, 164, 165, 168, 175, 176, 177, 178, 179, 180, 181, 185, 190.

Применяют и другие, более сложные вероятностные модели выборок. Например, цензурированные выборки соответствуют испытаниям, проводящимся в течение определенного промежутка времени. При этом для части изделий удается замерить время наработки на отказ, а для остальных лишь констатируется, что наработки на отказ для них больше времени испытания. Для выборок второго вида отбор объектов может проводиться в несколько этапов. Например, для входного контроля сигарет могут сначала отбираться коробки, в отобранных коробках – блоки, в выбранных блоках – пачки, а в пачках – сигареты. Четыре ступени отбора. Ясно, что выборка будет обладать иными свойствами, чем простая случайная выборка из совокупности сигарет.

Контрольные вопросы и задания:

1. Перечислите основные задачи математической статистики.
2. Какова область применения математической статистики?
3. Что включают статистические характеристики?
4. В чем заключается первичная обработка статистических данных?

Лекция 6. Понятие о генеральной совокупности, выборке и способы ее отбора.

Тема 6.1. 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. При этом предметом изучения являются какие-то качественные или количественные параметры объектов, составляющих данную совокупность (скажем, пригодность объектов к использованию, их вес, сорт, размер и (т. д.), законы распределения этих параметров. 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. Как правило, объём n выборки много меньше объёма N генеральной совокупности ($n < N$).

Тема 6.2. Способы выборки.

При составлении выборки можно поступать двумя способами: после того как объект отобран и над ним произведено наблюдение, он может быть возвращен либо не возвращен в генеральную совокупность. В соответствии со сказанным выборки подразделяют на *повторные и бесповторные*.

Повторной называют выборку, при которой отобранный объект (перед отбором следующего) возвращается в генеральную совокупность.

Бесповторной называют выборку, при которой отобранный объект в генеральную совокупность не возвращается.

На практике обычно пользуются бесповторным случайным отбором.

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли. Другими словами, выборка должна правильно представлять пропорции генеральной совокупности. Это требование коротко формулируют так: выборка должна быть ***репрезентативной (представительной)*** .

В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если ее осуществить случайно: каждый объект выборки отобран случайно из генеральной совокупности, если все объекты имеют одинаковую вероятность попасть в выборку.

На практике применяются различные способы отбора. Эти способы можно подразделить на два вида.

1. Отбор, не требующий расчленения генеральной совокупности на части. Сюда относятся:

а) простой случайный бесповторный отбор;

б) простой случайный повторный отбор.

2. Отбор, при котором генеральная совокупность разбивается на части. Сюда относятся:

а) типический отбор;

б) механический отбор;

в) серийный отбор.

Простым случайным называют такой отбор, при котором объекты извлекают по одному из всей генеральной совокупности и после обследования не возвращают (бесповторный отбор) или возвращают (повторный отбор) в генеральную совокупность.

Типическим называют отбор, при котором объекты отбираются не из всей генеральной совокупности, а из каждой ее «типической» части. Например, если детали изготовляют на нескольких станках, то отбор производят не из всей совокупности деталей, произведенных всеми станками, а из продукции каждого станка в отдельности.

Типическим отбором пользуются тогда, когда обследуемый признак заметно колеблется в различных типических частях генеральной совокупности. Например, если продукция изготовляется на нескольких машинах, среди которых есть более и менее изношенные, то здесь типический отбор целесообразен.

Механическим называют отбор, при котором генеральную совокупность «механически» делят на столько групп, сколько объектов должно войти в выборку, а из каждой группы отбирают один объект. Например, если нужно отобрать 20% изготовленных станком деталей, то отбирают каждую пятую деталь; если требуется отобрать 5% деталей, то отбирают каждую двадцатую деталь и т. д. Следует указать, что иногда механический отбор может не обеспечить репрезентативности выборки.

Серийным называют отбор, при котором объекты отбирают из генеральной совокупности не по одному, а «сериями», которые подвергаются сплошному обследованию. Например, если изделия изготовляются большой группой станков-автоматов, то подвергают сплошному обследованию продукцию только нескольких станков. Серийным отбором пользуются тогда, когда обследуемый признак колеблется в различных сериях незначительно.

На практике часто применяется комбинированный отбор, при котором сочетаются указанные выше способы.

Объекты выборки подвергаются сплошному обследованию, а затем, по результатам этого обследования, делаются определенные выводы и обо всей генеральной совокупности.

Контрольные вопросы и задания:

1. Сформулируйте определение генеральной совокупности, выборки.

2. Назовите виды выборок и способы их отбора.

Лекция 7. Статистическое распределение выборки.

Тема 7.1. Статистическое распределение выборки.

Пусть из изучаемой генеральной совокупности сделана случайная выборка объёмом n. И пусть оказалось, что у n_1объектов, попавших в выборку, значение исследуемого признака X оказалось равным x_1, у n_2 объектов – значение x_2, …, у n_m объектов – значение x_m.

Тогда таблица (табл. 2) содержащая указанные данные, называется **статистическим распределением выборки.**

Таблица 2

Варианта	$\mathbf{x_1}$	$\mathbf{x_2}$**,**	**…**	$\mathbf{x_m}$	**Всего: Х вариант**
Кратность варианты	n_1	n_2	…	n_m	Сумма = п (объем выборки)
Частота варианты	$n_{1/}$ n	$n_{2/}$ n	…	$n_{m/}$ n	Сумма = 1

При этом числа ($x_1; x_2;. . . x_m$), представляющие собой все встретившиеся в выборке значения исследуемого признака X, называются *вариантами*, а количества ($n_1; n_2;. . . n_m$) объектов, имеющих соответствующие варианты, называются *частотами*. Частота варианты вычисляется по формуле:

Частота варианты = (кратность варианты : объем выборки).

Статистическое распределение выборки автоматически имеет вид, представленный в таблице, если исследуемый признак X является дискретной (прерывистой) величиной. Например, если исследуется экзаменационная оценка по какому-либо предмету большого количества студентов, то эта оценка X по своей природе является величиной дискретной (принимает лишь значения 2; 3; 4; 5). И если выборка составляет, например, 25 человек, то её статистическое распределение может быть, например, следующим (табл. 3):

Таблица 3

Варианта	**2**	**3**	**4**	**5**	**Всего:4 варианты**
Кратность варианты	2	8	10	5	Сумма = 25
Частота варианты	2/25	8/25	10/25	5/25	Сумма = 1

Статистическое распределение выборки для наглядности изображают графически – в виде так называемого *полигона частот,* представляющего собой ломаную линию с узлами в точках ($x_{1;}$ n_1), …, (x_m; n_m), (р; х) (рис. 6).

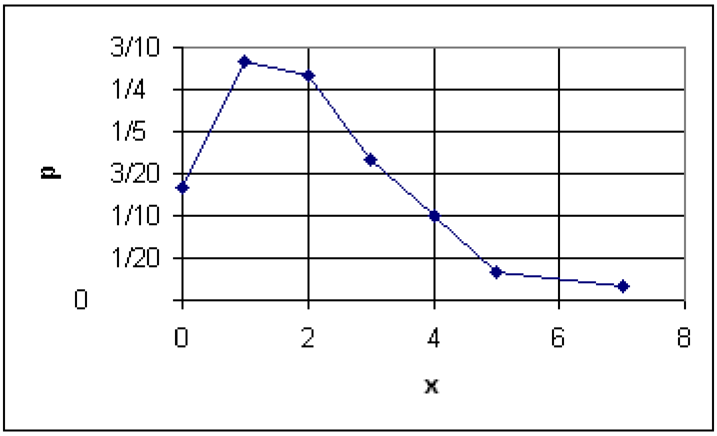


Рис. 6. Полигон относительных частот.

Если же исследуемый признак *X* является непрерывной величиной, то статистическое распределение выборки обычно оформляют в виде табл. 4:

Таблица 4

Варианта	**x_1-x_2**	**x_2-x_3**	**…**	**x_m-x_{m+1}**	**Всего: x вариант**
Кратность варианты	n_1	n_2	…	n_m	Сумма = п
Частота варианты	n_1/n	n_2/n	…	n_m/n	Сумма = 1

Здесь (x_1-x_2), (x_2-x_3), . . . (x_m-x_{m+1}) – интервалы (обычно одинаковые по длине), на которые разбивают весь интервал ($x_1; x_{m+1}$) значений признака *X* в выборке, а ($n_1; n_2; \ldots; n_m$) – частоты для соответствующих интервалов. Таким образом, получили *интервальный ряд данных*.

Например, если исследуется масса *X* (г) деталей, то статистическое распределение выборки для 100 деталей, случайно отобранных из общего количества, может быть таким (табл. 5):

Таблица 5

Варианта	0-40	40-80	80-120	120-160	160-200
Кратность варианты	1 2	2 0	2 8	25	15

По данным интервального ряда строят гистограмму частот или гистограмму относительных частот. Ступенчатая фигура, состоящая из прямоугольников, основания которых – частичные интервалы, высоты равны отношению частоты к длине частичного интервала (плотность частоты). Гистограмма частот может имеет вид (рис. 7):

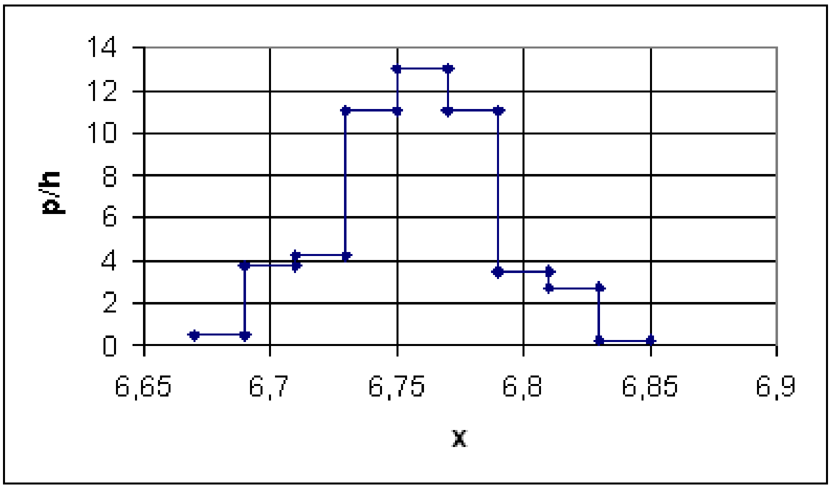


Рис. 7 Гистограмма распределения частот

Для гистограммы частот: площадь каждого прямоугольника равна частоте интервала, сумма площадей всех прямоугольников равна объему выборки.

Для гистограммы частот: площадь каждого прямоугольника равна частот интервала, сумма площадей всех прямоугольников равна 1.

Тема 7.2. Статистическая оценка параметров распределения.

Оценивание – это определение приближенного значения неизвестной характеристики или параметра распределения (генеральной совокупности), иной оцениваемой составляющей математической модели реального (экономического, технического и др.) явления или процесса по результатам наблюдений. Иногда формулируют более коротко: оценивание – это определение приближенного значения неизвестного параметра генеральной совокупности по результатам наблюдений. При этом параметром генеральной совокупности может быть либо число, либо набор чисел, либо функция, либо множество или иной объект нечисловой природы. Целью оценивания может быть нахождение упорядочения инвестиционных проектов по экономической эффективности или технических изделий (объектов) по качеству, формулировка правил технической или медицинской диагностики и т.д. Упорядочения в математической статистике называют также *ранжировками*. Это – один из видов объектов нечисловой природы. Оценивание проводят с помощью оценок – статистик, являющихся основой для оценивания неизвестного параметра распределения. Оценивание бывает двух видов – точечное оценивание и оценивание с помощью доверительной области. *Точечное оценивание* - способ оценивания, заключающийся в том, что значение оценки принимается как неизвестное значение параметра распределения. Перейдём теперь к основным числовым характеристикам статистического распределения выборки, с помощью которых происходит оценивание. Ими являются:

1. ***Размах выборки*** – это разница между наибольшей и наименьшей вариантой. На графике (рис. 6) - это длина области определения полигона частот.

2. ***Мода выборки*** – это наиболее часто встречающаяся ее варианта. На графике (рис. 6)– это точка, в которой достигается максимум полигона частот. Если эта точка

одна или если таких точек несколько, но подряд идущих, то выборку называют *унимодальной* (одна мода). Возможны и *бимодальные* (две моды) выборки и т.д.

3. ***Среднее значение признака X в выборке***, обозначаемое и называемое *выборочной средней*. Для нахождения среднего значения можно:

а) каждую варианту умножить на ее кратность;

б) сложить все полученные произведения;

в) поделить найденную сумму на сумму всех кратностей.

4. Величина, которая характеризует ***среднее значение квадратов отклонений вариант от выборочной средней*** . Она называется ***выборочной дисперсией***.

Определение. ***Отклонением*** называется разность между случайной величиной X и ее математическим ожиданием $M(X)$, т.е. $X - M(X)$.

Свойства математического ожидания.

А) Постоянный множитель можно выносить за знак математического ожидания:

$$M(C \cdot X) = C \cdot M(X)$$

Б) Математическое ожидание постоянной величины C равно самой этой величине:

$$M(C) = C$$

В) Математическое ожидание суммы двух случайных величин равно сумме их математических ожиданий:

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

Г) Математическое ожидание произведения независимых случайных величин равно произведению математических ожиданий этих величин:

$$M(X \cdot Y) = M(X) \cdot M(Y).$$

Отклонение $X - M(X)$ и его квадрат $(X - M(X))^2$ также являются случайными величинами.

Определение. ***Дисперсией дискретной*** случайной величины X называется математическое ожидание квадрата ее отклонения:

$$D(X) = M(X - M(X))^2. \qquad (14)$$

Свойства дисперсии.

А) Дисперсия постоянной величины C равна 0:

$$D(C) = 0.$$

Б) Если X - случайная величина, а C – постоянная, то

$$D(C \cdot X) = C^2 D(X)$$

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

В) Если X и Y - независимые случайные величины, то

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

Для вычисления дисперсий более удобной является формула

$$D(X) = M(X^2) - (M(X))^2. \qquad (15)$$

5. Величина, которая характеризует *среднее значение отклонения вариант от выборочной средней без учёта знака этого отклонения. Она называется* ***выборочным средним квадратическим отклонением.***

Средним квадратичным отклонением случайной величины называется корень квадратный из ее дисперсии:

$$\sigma(X) = \sqrt{D(X)}\ . \qquad (16)$$

6. Величина, называемая ***выборочным коэффициентом вариации***. Этот коэффициент характеризует *долю в процентах, которую составляет среднее отклонение от среднего по отношению к самому среднему.*

Формулы (14,15,16) можно использовать, если статистическое распределение выборки является дискретным. А если оно является непрерывным (интервальным), то его предварительно преобразуют в дискретное, в котором середины интервалов принимаются за его новые дискретные варианты.

Наличие нескольких методов оценивания одних и тех же параметров приводит к необходимости выбора между этими методами.

Задача 32. Найти математическое ожидание случайной величины X, зная закон ее распределения

X	-1	0	1	2	3
p	0,2	0,1	0,25	0,15	0,3

Решение:

$$M(X) = -1\cdot 0{,}2 + 0\cdot 0{,}1 + 1\cdot 0{,}25 + 2\cdot 0{,}15 + 3\cdot 0{,}3 = -0{,}2 + 0 + 0{,}25 + 0{,}3 + 0{,}9 = 1{,}25\,.$$

Задача 34. Найдем математическое ожидание случайных величин X и Y, зная законы их распределения

1)

X	-8	-4	-1	1	3	7
p	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{4}$

2)

Y	-2	-1	0	1	2	3
p	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$	0	$\frac{1}{4}$

Решение: $M(X) = -8\cdot\frac{1}{12} - 4\cdot\frac{1}{6} - 1\cdot\frac{1}{4} + 1\cdot\frac{1}{6} + 3\cdot\frac{1}{12} + 7\cdot\frac{1}{4} = \frac{7}{12}$,

$$M(Y) = -2\cdot\frac{1}{6} - 1\cdot\frac{1}{6} - 0\cdot\frac{1}{12} + 1\cdot\frac{1}{3} + 2\cdot 0 + 3\cdot\frac{1}{4} = \frac{7}{12}\,.$$

Получили любопытный результат: законы распределения величин X и Y разные, а их математические ожидания одинаковы.

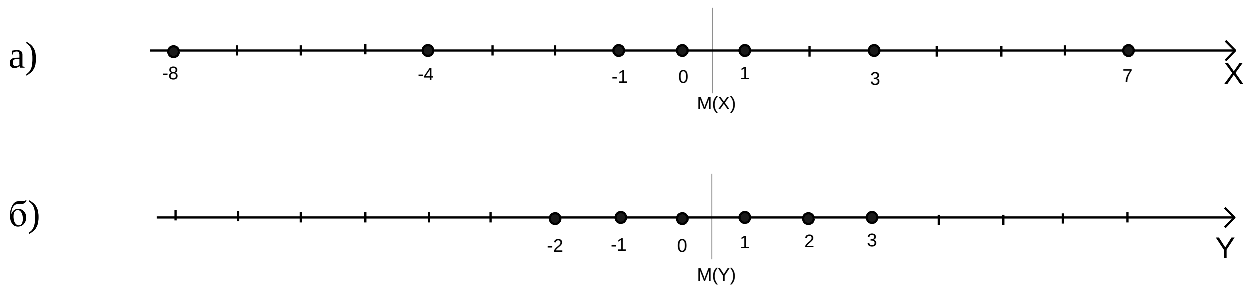


Рис.8

Из рис.8, *б* видно, что значения величины Y более сосредоточены около математического ожидания $M(Y)$, чем значения величины X, которые разбросаны (рассеяны) относительно ее математического ожидания $M(X)$ (рис.8, *а*).

Основной числовой характеристикой степени рассеяния значений случайной величины X относительно ее математического ожидания $M(X)$ является дисперсия, которая обозначается через $D(X)$.

Задача 35. Дискретная случайная величина распределена по закону:

X	-1	0	1	2
p	0,2	0,1	0,3	0,4

Найти $D(X)$.

Решение: Сначала находим $M(X)$.

$$M(X) = -1 \cdot 0{,}2 + 0 \cdot 0{,}1 + 1 \cdot 0{,}3 + 2 \cdot 0{,}4 = 0{,}9,$$

а затем $M(X^2)$.

$$M(X^2) = (-1)^2 \cdot 0{,}2 + 0^2 \cdot 0{,}1 + 1^2 \cdot 0{,}3 + 2^2 \cdot 0{,}4 = 2{,}1.$$

По формуле $D(X) = M(X^2) - (M(X))^2$ имеем

$$D(X) = 2{,}1 - (0{,}9)^2 = 2{,}1 - 0{,}81 = 1{,}29.$$

Контрольные вопросы и задания:

1. Перечислите характеристики статистического распределения выборки.
2. Что называется частотой и относительной частотой варианта?
3. Алгоритм составления дискретного вариационного ряда. Полигон частот.
4. Алгоритм составления интервального вариационного ряда. Гистограмма частот и ее геометрический смысл.
5. Построить статистическое распределение выборки оценок, полученных Вами по математике в течение этого семестра. Построить полигоны частот и относительных частот данной выборки.

Лекция 8. Понятие о статистической проверке гипотез.

Тема 8.1. Понятие о статистической проверке гипотез.

Статистическая проверка гипотез - система приемов в математической статистике, предназначенная для проверки соответствия опытных данных проверяемой гипотезе. К проблеме статистической проверки гипотез приводит большое число связанных с экспериментом вопросов, возникающих в приложениях, например, сравнение урожайности сортов каких-либо сельскохозяйственных культур, эффективности лекарственных препаратов и др. Правило, по которому принимается или отклоняется данная гипотеза, называют ***статистическим критерием***.

Статистическая гипотеза – любое предположение, касающееся неизвестного распределения случайных величин (элементов). Приведем формулировки нескольких статистических гипотез:

1. Результаты наблюдений имеют нормальное распределение с нулевым математическим ожиданием.

2. Результаты наблюдений имеют функцию распределения $N(0,1)$.

3. Результаты наблюдений имеют нормальное распределение.

4. Результаты наблюдений в двух независимых выборках имеют одно и то же нормальное распределение.

5. Результаты наблюдений в двух независимых выборках имеют одно и то же распределение. Различают нулевую и альтернативную гипотезы. *Нулевая гипотеза* – гипотеза, подлежащая проверке. *Альтернативная гипотеза* – каждая допустимая гипотеза, отличная от нулевой. Нулевую гипотезу обозначают H_0, альтернативную – H_1 (от Hypothesis – «гипотеза» (англ.)). Выбор тех или иных нулевых или альтернативных гипотез определяется стоящими перед менеджером, экономистом, инженером, исследователем прикладными задачами. Конкретная задача проверки статистической гипотезы полностью описана, если заданы нулевая и альтернативная гипотезы. Выбор метода проверки статистической гипотезы, свойства и характеристики методов определяются как нулевой, так и альтернативной гипотезами. Для проверки одной и той же нулевой гипотезы при различных альтернативных гипотезах следует использовать различные методы. При обработке реальных данных большое значение имеет правильный выбор гипотез H_0 и H_1. Принимаемые предположения, например, нормальность распределения, должны быть тщательно обоснованы, в частности, статистическими методами. Отметим, что в подавляющем большинстве конкретных прикладных постановок распределение результатов наблюдений отлично от нормального. Часто возникает ситуация, когда вид нулевой гипотезы вытекает из постановки прикладной задачи, а вид альтернативной гипотезы не ясен. В таких случаях следует рассматривать альтернативную гипотезу наиболее общего вида и использовать методы, решающие поставленную задачу при всех возможных H_1.

Статистические гипотезы бывают *параметрические и непараметрические*. Предположение, которое касается неизвестного значения параметра распределения, входящего в некоторое параметрическое семейство распределений, называется параметрической гипотезой.

Предположение, при котором вид распределения неизвестен (т.е. не предполагается, что оно входит в некоторое параметрическое семейство распределений), называется непараметрической гипотезой. Если H_0, H_1 – параметрические гипотезы, то задача проверки статистической гипотезы – параметрическая. Если хотя бы одна из гипотез H_0 и H_1 – непараметрическая, то задача проверки статистической гипотезы – непараметрическая. Другими словами, если вероятностная модель ситуации – параметрическая, т.е. полностью описывается в терминах того или иного параметрического семейства распределений вероятностей, то и задача проверки статистической гипотезы – параметрическая. Если же вероятностная модель ситуации – непараметрическая, т.е. ее нельзя полностью описать в терминах какого-либо параметрического семейства распределений вероятностей, то и задача проверки статистической гипотезы – непараметрическая. Статистическая гипотеза называется простой, если она однозначно задает распределение результатов наблюдений, вошедших в выборку. В противном случае статистическая гипотеза называется сложной. Однозначно определенный способ проверки статистических гипотез называется ***статистическим критерием***. Статистические критерии делятся на параметрические и непараметрические. Параметрические критерии используются в параметрических задачах проверки статистических гипотез, а непараметрические – в непараметрических задачах.

При проверке статистической гипотезы возможны ошибки. Есть два рода ошибок.

1. Нулевая гипотеза H_0 отвергается несмотря на то, что она верна. Это ошибка 1-го рода. Ее вероятность принято обозначать буквой α.

2. Нулевая гипотеза H_0 принимается несмотря на то, что она неверна. Эта ошибка 2-го рода. Ее вероятность принято обозначать буквой β.

Все возможные ситуации с принятием и непринятием гипотезы H_0 изображены ниже в табл. 6 (в скобках указаны вероятности соответствующих ситуаций). Таблица 6

Заключение о гипотезе H_0	**На самом деле гипотеза H_0**	
	Верна	**Неверна**
Отвергается	Ошибка 1-го рода (α)	Правильное решение (1-β)
Принимается	Правильное решение (1-α)	Ошибка 2-го рода (β)

Проиллюстрируем это утверждение на таком житейском примере. Пусть гипотеза H_0 состоит в том, что переход пешеходом перекрестка на красный свет окончится для него плохо. Тогда отвергнуть эту гипотезу – это значит переходить перекресток. Принять эту гипотезу – это значит стоять на месте (ждать зеленый сигнал).

1. Если гипотеза H_0 верна, то переходить перекресток – это отвергнуть правильную гипотезу, то есть совершить ошибку 1- го рода (смертельную).

2. Если гипотеза H_0 неверна, то оставаться на месте – это принять неправильную гипотезу, т.е. совершить ошибку 2-го рода (потерять время на ожидание зеленого сигнала светофора).

Очевидно, что тяжесть последствий каждой из этих двух ошибок совершенно разная. И еще очевидно, что уменьшая вероятность α ошибки 1-го рода, пешеход автоматически будет увеличивать вероятность ошибки 2-го рода. Действительно, уменьшая ошибку 1-го рода, пешеход не будет переходить на красный свет даже в тех случаях, когда машины от перекрестка далеко и переход ему практически ничем бы не грозил.

Контрольные вопросы и задания

1. Что называют статистической проверкой гипотез?
2. Виды проверок гипотез. В чем они заключаются. Приведите примеры.

Список литературы:

1. Тюрин Ю.Н. Теория вероятностей / Ю.Н. Тюрин, А.А. Макаров, Г.И. Симонов. – М.: МЦНМО, 2009
2. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике./ В. Е. Гмурман. - М.: Высш. шк., 2004.
3. Соловейчик И.Л. Сборник задач по математике для техникумов / И.Л.Соловейчик, В.Т. Лисичкин. – М.: Оникс 21 век, 2003.
4. Григулецкий В.Г. Математика для студентов экономических специальностей. Часть 2 / В.Г. Григулецкий, И.В. Лукьянова, И.А. Петунина. – Краснодар, 2002.
5. Богомолов Н.В. Практические занятия по математике/ Н.В.Богомолов. – М.: Высшая школа, 1990.
6. Валуцэ И.И. Математика для техникумов / И.И. Валуцэ, Г.Д. Дилигул. - М.: Наука, 1989.

Оглавление