ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ПРОФЕССИОНАЛЬНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
«НИЖЕГОРОДСКОЕ ОБЛАСТНОЕ УЧИЛИЩЕ
ОЛИМПИЙСКОГО РЕЗЕРВА ИМЕНИ В.С.ТИШИНА
(ГБПОУ «НОУОР ИМЕНИ В.С.ТИШИНА»)

КУРС ЛЕКЦИЙ

ЕН.01 МАТЕМАТИКА

Код и наименование специальности 49.02.01. Физическая культура

Нижний Новгород
2015 г.

Организация-разработчик: ГБПОУ «Нижегородское областное училище олимпийского резерва имени В.С.Тишина»

Разработчик: к.п.н., Чернякова И.Л.

Одобрена методическим объединением «20» ноября 2015 г. (протокол №3)

Лекция №1

Тема: Основы дискретной математики.

МНОЖЕСТВА

Величиной называется все что может быть измерено и выражено числом.
Множеством называется совокупность некоторых элементов, объединенных каким-либо общим признаком. Элементами множества могут быть числа, фигуры, предметы, понятия и т.п.

Множества обозначаются прописными буквами, а элементы множество строчными буквами. Элементы множеств заключаются в фигурные скобки.
Если элемент **x** принадлежит множеству **X**, то записывают **x** ∈ **X** (**∈** — принадлежит).
Если множество A является частью множества B, то записывают **A ⊂ B** (**⊂** — содержится).

Множество может быть задано одним из двух способов: перечислением и с помощью определяющего свойства.

Например, перечислением заданы следующие множества:

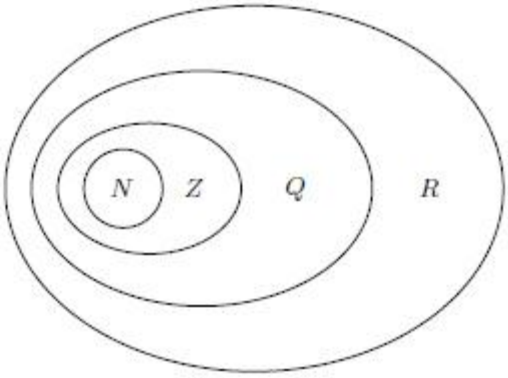
- A={1,2,3,5,7} — множество чисел
- X={x_1,x_2,...,x_n} — множество некоторых элементов x_1,x_2,...,x_n
- N={1,2,...,n} — множество натуральных чисел
- Z={0,±1,±2,...,±n} — множество целых чисел

Множество (-∞;+∞) называется **числовой прямой**, а любое число — точкой этой прямой. Пусть a — произвольная точка числовой прямой иδ — положительное число. Интервал (a-δ; a+δ) называется **δ- окрестностью точки a**.
Множество X ограничено сверху (снизу), если существует такое число c, что для любого x ∈ X выполняется неравенство x≤c (x≥c). Число c в этом случае называется **верхней (нижней) гранью** множества X. Множество, ограниченное и сверху и снизу, называется **ограниченным**. Наименьшая (наибольшая) из верхних (нижних) граней множества называется **точной верхней (нижней) гранью** этого множества.

Основные числовые множества

N	{1,2,3,...,n} Множество всех натуральных чисел
Z	{0, ±1, ±2, ±3,...} Множество **целых чисел.** Множество целых чисел включает в себя множество натуральных.
Q	Множество **рациональных чисел**. Кроме целых чисел имеются ещё и дроби. Дробь — это выражение вида p/q, где **p** — целое число, **q** — натуральное. Десятичные дроби также можно записать в виде p/q. Например: 0,25 = 25/100 = 1/4. Целые числа также можно записать в виде p/q. Например, в виде дроби со знаменателем "один": 2 = 2/1. Таким образом любое рациональное число можно записать десятичной

	дробью — конечно или бесконечной периодической.
R	Множество всех **вещественных чисел**. Иррациональные числа — это бесконечные непериодические дроби. К ним относятся: ▪ число π — отношение длины окружности к её диаметру; ▪ число e — названное в честь Эйлера и др.; Вместе два множества (рациональных и иррациональных чисел) — образуют множество действительных (или вещественных) чисел.



Если множество не содержит ни одного элемента, то оно называется **пустым множеством** и записывается **Ø**.

Элементы логической символики

→	"следует", "выполняется"
↔	равносильность утверждения
:	"такой, что"

Запись $\forall x$: $|x|<2 \rightarrow x^2 < 4$ означает: для каждого x такого, что $|x|<2$, выполняется неравенство $x^2 < 4$.

Квантор

При записи математических выражений часто используются кванторы.

Квантором называется логический символ, который характеризует следующие за ним элементы в количественном отношении.

- ∀- **квантор общности**, используется вместо слов "для всех", "для любого".
- ∃- **квантор существования**, используется вместо слов "существует", "имеется". Используется также сочетание символов ∃!, которое читается как существует единственный.

Операции над множествами

Два **множества А и В равны** (А=В), если они состоят из одних и тех же элементов.

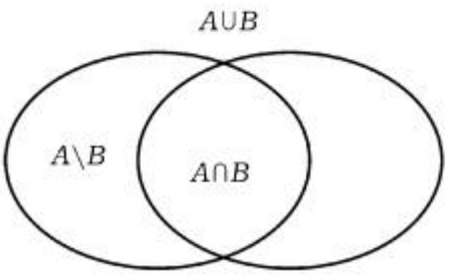
Например, если А={1,2,3,4}, В={3,1,4,2} то А=В.

Объединением (суммой) множеств А и В называется множество А ∪ В, элементы которого принадлежат хотя бы одному из этих множеств. Например, если А={1,2,4}, В={3,4,5,6}, то А ∪ В = {1,2,3,4,5,6}

Пересечением (произведением) множеств А и В называется множество А ∩ В, элементы которого принадлежат как множеству А, так и множеству В. Например, если А={1,2,4}, В={3,4,5,2}, то А ∩ В = {2,4}

Разностью множеств А и В называется множество АВ, элементы которого принадлежат множеству А, но не принадлежат множеству В. Например, если А={1,2,3,4}, В={3,4,5}, то АВ = {1,2}

Симметричной разностью множеств А и В называется множество А Δ В, являющееся объединением разностей множеств АВ и ВА, то есть А Δ В = (АВ) ∪ (ВА). Например, если А={1,2,3,4}, В={3,4,5,6}, то А Δ В = {1,2} ∪ {5,6} = {1,2,5,6}



Свойства операций над множествами

Свойства перестановочности

А ∪ В = В ∪ А

А ∩ В = В ∩ А

Сочетательное свойство

(А ∪ В) ∪ С = А ∪ (В ∪ С)

(А ∩ В) ∩ С = А ∩ (В ∩ С)

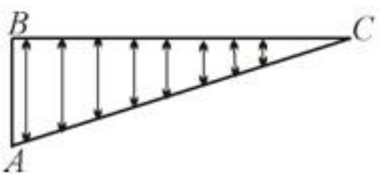
Счетные и несчетные множества

Для того, чтобы сравнить два каких-либо множества А и В, между их элементами устанавливают соответствие.

Если это соответствие взаимооднозначное, то множества называются эквивалентными или равномощными, А□ В или В □А.

Пример 1

Множество точек катета ВС и гипотенузы АС треугольника АВС являются равномощными.



Лекция №2

Тема 1.2.

Элементы математической логики

Понятие о науке логике

Logos (греч.) - слово, понятие, рассуждение, разум. Слово “логика” обозначает совокупность правил, которым подчиняется процесс мышления или обозначает науку о правилах рассуждения и тех формах, в которых оно осуществляется. Логика изучает абстрактное мышление как средство познания объективного мира, исследует формы и законы, в которых происходит отражение мира в процессе мышления.

Основными формами абстрактного мышления являются: ПОНЯТИЯ, СУЖДЕНИЯ, УМОЗАКЛЮЧЕНИЯ.
ПОНЯТИЕ - форма мышления, в которой отражаются существенные признаки отдельного предмета или класса однородных предметов: *портфель трапеция ураганный ветер.*
СУЖДЕНИЕ - мысль, в которой что-либо утверждается или отрицается о предметах. Суждения являются повествовательными предложениями, истинными или ложными. Они могут быть простыми и сложными: *Весна наступила, и грачи прилетели.*
УМОЗАКЛЮЧЕНИЕ - прием мышления, посредством которого из исходного знания получается новое знание; из одного или нескольких истинных суждений, называемых посылками, мы по определенным правилам вывода получаем заключение. *Есть несколько видов* умозаключений.
Все металлы - *простые вещества.*
Литий - *металл.*
Литий - *простое вещество.*

Чтобы достичь истины при помощи умозаключений, надо соблюдать законы логики.

ФОРМАЛЬНАЯ ЛОГИКА - наука о законах и формах правильного мышления.
МАТЕМАТИЧЕСКАЯ ЛОГИКА изучает логические связи и отношения, лежащие в основе дедуктивного (логического) вывода.

Формальная логика связана с анализом наших обычных содержательных умозаключений, выражаемых разговорным языком. *Математическая логика* изучает только умозаключения со строго определенными объектами и суждениями, для которых можно однозначно решить, истинны они или ложны.

Умение правильно рассуждать необходимо в любой области человеческой деятельности: науке и технике, юстиции и дипломатии, планировании, военном деле и т.д.

Но хотя умение это восходит к древнейшим временам, логика, т.е. наука о том, какие формы рассуждений правильны, возникла лишь немногим более 2 тысяч лет назад.

Операции над высказываниями

1. *Операция отрицания.*

Отрицанием высказывания A называется высказывание, обозначаемое $\overline{A}$ (читается «не A», «неверно, что A»), которое истинно, когда A ложно и ложно, когда A – истинно.

Отрицающие друг друга высказывания A и $\overline{A}$ называются **противоположными.**

Построим отрицание высказывания «число π $=3,14$». Это истинное высказывание. Тогда его отрицание будет следующим: «$\pi \neq 3,14$» – ложное высказывание.

2. *Операция конъюнкции.*

Конъюнкцией высказываний A и B называется высказывание, обозначаемое $A \wedge B$ (читается «A и B»), истинные значения которого определяются в том и только том случае, когда оба высказывания A и B истинны.

Конъюнкцию высказываний называют логическим произведением и часто обозначают AB.

Пусть дано высказывание A – «в марте температура воздуха от $0^0 C$ до $+7^0 C$» и высказывание B – «в Витебске идет дождь». Тогда $A \wedge B$ будет следующей: «в марте температура воздуха от $0^0 C$ до $+7^0 C$ и в Витебске идет дождь». Данная конъюнкция будет истинной, если будут высказывания A и B истинными. Если же окажется, что температура была меньше $0^0 C$ или в Витебске не было дождя, то $A \wedge B$ будет ложной.

3. *Операция дизъюнкции*.

Дизъюнкцией высказываний A и B называется высказывание $A \vee B$ (A или B), которое истинно тогда и только тогда, когда хотя бы одно из высказываний истинно и ложно – когда оба высказывания ложны.

Дизъюнкцию высказываний называют также логической суммой $A+B$.

Высказывание «$4<5$ или $4=5$» является истинным. Так как высказывание «$4<5$» – истинное, а высказывание «$4=5$» – ложное, то $A \vee B$ представляет собой истинное высказывание «$4 \leq 5$».

4. *Операция импликации*.

Импликацией высказываний A и B называется высказывание $A \Rightarrow B$ («если A, то B», «из A следует B»), значение которого ложно тогда и только тогда, когда A истинно, а B ложно.

В импликации $A \Rightarrow B$ высказывание A называют *основанием,* или посылкой, а высказывание B – *следствием,* или *заключением.*

С помощью таблиц истинности это можно определить так:

Дано высказывание «Если число 12 делится на 2 и на 3, то оно делится на 6». Так как высказывание A – «число 12делится на 2» истинно, высказывание B – «число 12 делится на 3» также истинно, то и импликация $A \Rightarrow B$ истинна.

5. *Эквиваленция высказываний*.

Эквиваленцией высказываний A и B называется высказывание, обозначаемое $A \Leftrightarrow B$ ($A;B$) (читается «A эквивалентно B», «A тогда и только тогда, когда B», «A необходимо и достаточно для B»), которое истинно тогда, когда A и B одновременно истинны или оба ложны.

Дано высказывание A – «число $5n$ делится на 2» и высказывание B – «число n является четным». Сформулируем эквиваленцию $A \Leftrightarrow B$:

a) число 5 делится на 2 тогда и только тогда, когда n – четное число;

b) условия: число $5n$ делится на 2 и что число n – четное, эквивалентны;

c) для того чтобы число $5n$ делилось на $2,$ необходимо и достаточно, чтобы n было четным;

d) для того чтобы n было четным, необходимо и достаточно, чтобы число $5n$ делилось на 2;

e) из того, что $5n$ делится на 2, следует, что n число четное и обратно.

Два высказывания, составленные из высказываний $A, B, C, \ldots$ с помощью знаков $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ и отрицания называют **равносильными,** если они имеют одну и ту же истинность при любых предположениях об истинности и ложности $A, B, C, \ldots$

Например, $A \wedge B = B \wedge A$, $A \Rightarrow B = \overline{A} \vee B$.

Составные высказывания, принимающие значения истинности при всех наборах значений входящих в них элементарных высказываний, называют **тавтологиями.** Их называют и **тождественно-истинными высказываниями** или **законами логики.**

$A \vee \overline{A}$ принимает значение истинности при любом наборе истинности высказывания A. Значит, формула $A \vee \overline{A}$ является тавтологией.

Пусть символы $X, Y, Z, \ldots$ обозначают произвольные высказывания. Каждое из них представляет собой переменное, которое может принимать два значения *И* и *Л,* и называется **переменным элементарным высказыванием.** В отличие от переменного, элементарное высказывание, имеющее определенное значение *И* или *Л,* называется **постоянным.**

Переменные элементарные высказывания $X, Y, Z, \ldots$ есть **формулы.**

Действия над числами в числовых выражениях выполняются в определенном порядке: умножение и деление, затем сложение и вычитание. Аналогично и в логике высказываний ***логические операции выполняют по следующему правилу:*** операцию конъюнкции раньше дизъюнкции, и обе эти операции выполняют раньше операций импликации и эквиваленции.

Указанное правило помогает сократить число скобок в формулах.

Если в формулу алгебры высказываний вместо переменных $X, Y, Z, \ldots$ подставить высказывания определенной истинности, то получим составное высказывание также определенной истинности. Заменив в конкретном составном высказывании элементарные высказывания соответствующими переменными, получим формулу, выражающую **логическую структуру** данного высказывания.

Любое высказывание можно формализовать, то есть, заменить его формулой.

Например, высказыванию «если число *60* делится на *3* и на *5*, то *60* делится на *15*» соответствует формула $(A \wedge B) \Rightarrow C$, где A – «число *60* делится на *3*», B – «число *60* делится на *5*», C – «число *60* делится на *15*».

Для ***формализации высказываний*** поступают следующим образом:

1) выделяют все элементарные высказывания и обозначают их соответствующими буквами;

2) выделяют все логические связки и заменяют их логическими символами;

3) расставляют скобки в соответствии со смыслом исходного высказывания, учитывая при этом правило расстановки скобок.

Если известно значение каждого высказывания, входящего в формулу, то с помощью таблиц истинности можно найти значение этой формулы.

Для ***составления таблицы*** (см. табл. 7) выписываются сначала элементарные переменные высказывания X и Y, затем более сложные высказывания $\overline{X}$ и $\overline{Y}$, входящие в эту формулу, затем более сложные высказывания

Нетрудно установить, что таблица имеет 2^n строк, где n – число элементарных переменных в формуле (в случае двух переменных – *4* строки, трех – *8* строк и т.д.).

Для удобства пользования таблицами значения истинности переменных записываются «*И*», значения лжи – «*Л*».

Значения формулы при любой комбинации значений переменных высказываний можно описать посредством таблицы.

Формула, принимающая значение истины хотя бы при одном значении входящих высказываний, называется **выполнимой.**

Например, $A \Rightarrow B$, $A \Leftrightarrow B$ – выполнимые.

Формулы, принимающие значение истинности при всех наборах значений входящих в них высказывательных переменных, называются **тождественно истинными формулами** или **тавтологиями.**

Формулы, принимающие значение лжи при всех наборах значений входящих в них высказывательных переменных, называются **тождественно ложными** или **противоречиями.**

Если формулы при всех наборах истинности и лжи входящих высказывательных переменных принимают одинаковые значения, то их называют **равносильными.** Запись $A \equiv B$ читается так: A равносильно B.

ЛЕКЦИЯ № 3
Тема:
ПРИБЛИЖЕННЫЕ ЧИСЛА И ДЕЙСТВИЯ НАД НИМИ.

2.1. Величины и их измерения

1. ***Приближенное значение величины. Абсолютная и относительная погрешности***

Решение практических задач, как правило, связано с числовыми значениями величин. Эти значения получаются либо в результате измерения, либо в результате вычислений. В большинстве случаев значения величин, которыми приходится оперировать, являются приближенными.

Пусть X - точное значение некоторой величины, а x - наилучшее из известных ее приближенных значений. В этом случае погрешность (или ошибка) приближения x определяется разностью X-x. Обычно знак этой ошибки не имеет решающего значения, поэтому рассматривают ее абсолютную величину.

Величина e_x, называемая *абсолютной погрешностью* приближенного значения x, в большинстве случаев остается неизвестной, так как для ее вычисления нужно точное значение X. Вместе с тем, на практике обычно удается установить верхнюю границу абсолютной погрешности. Число в этом случае называется *предельной абсолютной погрешностью,* или *границей абсолютной погрешности приближения x.*

Таким образом, предельная абсолютная погрешность приближенного числа x - это всякое число, не меньшее абсолютной погрешности e_x этого числа.

Пример: Возьмем число. Если же вызвать на индикатор 8-разрядного МК, получим приближение этого числа: Попытаемся выразить абсолютную погрешность значения. Получили бесконечную дробь, не пригодную для практических расчетов. Очевидно, однако, что следовательно, число $0,00000006 = 0,6 * 10^{-7}$ можно считать предельной абсолютной погрешностью приближения

Во многих случаях значения границы абсолютной ошибки так же как и наилучшие значения приближения *x*, получаются на практике в результате измерений. Пусть, например, в результате повторных измерений одной и той же величины *x* получены значения: 5,2; 5,3; 5,4; 5,3. В этом случае естественно принять за наилучшее приближение измеряемой величины среднее значение *x*=5,3. Очевидно также, что граничными значениями величины *x* в данном случае будут $НГ_X$= 5,2, $ВГ_X$ = 5,4, а граница абсолютной погрешности *x* может быть определена как половина длины интервала, образуемого граничными значениями $НГ_X$ и $ВГ_X$.

По абсолютной погрешности нельзя в полной мере судить о точности измерений или вычислений. Качество приближения характеризуется величиной *относительной погрешности,* которая определяется как отношение ошибки e_x к модулю значения *X*(когда оно неизвестно, то к модулю приближения *x*).

Предельной относительной погрешностью (или *границей относительной погрешности)* приближенного числа называется отношение предельной абсолютной погрешности к абсолютному значению приближения *x*:

Относительную погрешность выражают обычно в процентах.

Пример Определим предельные погрешности числа x=3,14 как приближенного значения π. Так как π=3,1415926…., то |π-3,14|<0,0015927<0,0016=по формуле связи получаем таким образом

2. ***Верные и значащие цифры. Запись приближенных значений***

Цифра числа называется *верной* (в широком смысле), если ее абсолютная погрешность не превосходит единицы разряда, в котором стоит эта цифра.

Пример. X=6,328 X=0,0007 X<0,001 следовательно цифра 8-верная

Пример: А). Пусть 0 = 2,91385, В числе *а* верны в широком смысле цифры 2, 9, 1.

Б) Возьмем в качестве приближения к числу = 3,141592... число = 3,142. Тогда следует, что в приближенном значении = 3,142 все цифры являются верными.

В) Вычислим на 8-разрядном МК частное точных чисел 3,2 и 2,3, получим ответ: 1,3913043. Ответ содержит ошибку, поскольку
разрядная сетка МК не вместила всех цифр результата и все разряды начиная с восьмого были опущены. (В том, что ответ неточен, легко убедиться, проверив деление умножением: 1,3913043 2,3 = 3,9999998.) Не зная истинного значения допущенной ошибки, вычислитель в подобной ситуации всегда может быть уверен, что ее величина не превышает единицы самого младшего из изображенных на индикаторе разряда результата. Следовательно, в полученном результате все цифры верны.

Первая отброшенная (неверная) цифра часто называется *сомнительной.*

Говорят, что приближенное данное записано *правильно,* если в его записи все цифры верные. Если число записано правильно, то по одной только его записи в виде десятичной дроби можно судить о точности этого числа. Пусть, например, записано приближенное число a = 16,784, в котором все цифры верны. Из того, что верна последняя цифра 4, которая стоит в разряде тысячных, следует, что абсолютная погрешность значения a не превышает 0,001. Это значит, что можно принять т.е. a = 16,784±0,001.

Очевидно, что правильная запись приближенных данных не только допускает, но и обязывает выписывать нули в последних разрядах, если эти нули являются выражением верных цифр. Например, в записи = 109,070 нуль в конце означает, что цифра в разряде тысячных верна и она равна нулю. Предельной абсолютной погрешностью значения , как следует из записи, можно считать Для сравнения можно заметить, что значение c = 109,07 является менее точным, так как из его записи приходится принять, что

Значащими цифрами в записи числа называются все цифры в его десятичном изображении, отличные от нуля, и нули, если они расположены между значащими цифрами или стоят в конце для выражения верных знаков.

Пример а) 0,2409 - четыре значащие цифры; б) 24,09 - четыре значащие цифры; в) 100,700 - шесть значащих цифр.

Выдача числовых значений в ЭВМ, как правило, устроена таким образом, что нули в конце записи числа, даже если они верные, не сообщаются. Это означает, что если, например, ЭВМ показывает результат 247,064 и в то же время известно, что в этом результате верными должны быть восемь значащих цифр, то полученный ответ следует дополнить нулями: 247,06400.

В процессе вычислений часто происходит *округление чисел,* т.е. замена чисел их значениями с меньшим количеством значащих цифр. При округлении возникает погрешность, называемая погрешностью округления. Пусть x- данное число, а x_1 - результат округления. Погрешность округления определяется как модуль разности прежнего и нового значений числа.

В отдельных случаях вместо $\Delta_{окр}$ приходится использовать его верхнюю оценку.

Пример Выполним на 8-разрядном МК действие 1/6. На индикаторе высветится число 0,1666666. Произошло автоматическое округление бесконечной десятичной дроби 0,1(6) до числа разрядов, вмещающихся в регистре МК. При этом можно принять

Цифра числа называется *верной в строгом смысле,* если абсолютная погрешность этого числа не превосходит половины единицы разряда, в котором стоит эта цифра.

Правила записи приближенных чисел.

1. Приближенные числа записываются в форме x ± Δx. Запись X = x ± Δx означает, что неизвестная величина X удовлетворяет следующим неравенствам: $x-\Delta x <= X <= x+\Delta x$

При этом погрешность Δx рекомендуется подбирать так, чтобы а) в записи Δx было не более 1-2 значащих цифр;

б) младшие разряды в записи чисел x и Δx соответствовали друг другу.

Примеры: 23,4±0,2 ; 2,730±0,017 ; -6,97±0,10.

2. Приближенное число может быть записано без явного указания его предельной абсолютной погрешности. В этом случае в его записи (мантиссе) должны присутствовать только верные цифры (в широком смысле, если не сказано обратное). Тогда по самой записи числа можно судить о его точности.

Примеры. Если в числе A=5,83 все цифры верны в строгом смысле, то ΔA=0,005. Запись B=3,2 подразумевает, что ΔB=0,1. А по записи C=3,200 мы можем заключить, что ΔC=0,001. Таким образом, записи 3,2 и 3,200 в теории приближенных вычислений означают не одно и то же.

Цифры в записи приближенного числа, о которых нам неизвестно, верны они или нет, называются *сомнительными.* Сомнительные цифры (одну-две) оставляют в записи чисел промежуточных результатов для сохранения точности вычислений. В окончательном результате сомнительные цифры отбрасываются.

Округление чисел.

1. Правило округления. Если в старшем из отбрасываемых разрядов стоит цифра меньше пяти, то содержимое сохраняемых разрядов числа не изменяется. В противном случае в младший сохраняемый разряд добавляется единица с тем же знаком, что и у самого числа.
2. При округлении числа, записанного в форме x±Δx, его предельная абсолютная погрешность увеличивается с учетом погрешности округления.

Пример: Округлим до сотых число 4,5371±0,0482. Неправильно было бы записать 4,54±0,05 , так как погрешность округленного числа складывается из погрешности исходного числа и погрешности округления. В данном случае она равна 0,0482 + 0,0029 = 0,0511 . Округлять погрешности всегда следует с избытком, поэтому окончательный ответ: 4,54±0,06.

Пример Пусть в приближенном значении a = 16,395 все цифры верны в широком смысле. Округлим a до сотых: a_1 = 16,40. Погрешность округления Для нахождения полной погрешности, нужно сложить с погрешностью исходного значения a_1 которая в данном случае может быть найдена из условия, что все цифры в записи a верны: = 0,001. Отсюда следует, что в значении a_1 = 16,40 цифра 0 не верна в строгом смысле.

ЛЕКЦИЯ № 4.

Тема: Математическая статистика

Математи́ческая стати́стика — наука, разрабатывающая математические методы систематизации и использования статистических данных для научных и практических выводов.

Во многих своих разделах математическая статистика опирается на теорию вероятностей, позволяющую оценить надёжность и точность выводов, делаемых на основании ограниченного статистического материала (напр., оценить необходимый объём выборки для получения результатов требуемой точности при выборочном обследовании).

Математическая статистика — раздел математики, разрабатывающий методы регистрации, описания и анализа данных наблюдений и экспериментов с целью построения вероятностных моделей массовых случайных явлений[1]. В зависимости от математической природы конкретных результатов наблюдений статистика математическая делится на статистику чисел, многомерный статистический анализ, анализ функций (процессов) и временных рядов, статистику объектов нечисловой природы.

Выделяют описательную статистику, теорию оценивания и теорию проверки гипотез. Описательная статистика есть совокупность эмпирических методов, используемых для визуализации и интерпретации данных (расчет выборочных характеристик, таблицы, диаграммы, графики и т. д.), как правило, не требующих предположений о вероятностной природе данных. Некоторые методы описательной статистики предполагают использование возможностей современных компьютеров. К ним относятся, в частности, кластерный анализ, нацеленный на выделение групп объектов, похожих друг на друга, и многомерное шкалирование, позволяющее наглядно представить объекты на плоскости.

Методы оценивания и проверки гипотез опираются на вероятностные модели происхождения данных. Эти модели делятся на параметрические и непараметрические. В параметрических моделях предполагается, что характеристики изучаемых объектов описываются посредством распределений, зависящих от (одного или нескольких) числовых параметров. Непараметрические модели не связаны со спецификацией параметрического семейства для распределения изучаемых характеристик. В математической статистике оценивают параметры и функции от них, представляющие важные характеристики распределений (например, математическое ожидание, медиана, стандартное отклонение, квантили и др.), плотности и функции распределения и пр. Используют точечные и интервальные оценки.

Большой раздел современной математической статистики — статистический последовательный анализ, фундаментальный вклад в создание и развитие которого внес А. Вальд во время Второй мировой войны. В отличие от традиционных (непоследовательных) методов статистического анализа, основанных на случайной выборке фиксированного объема, в последовательном анализе допускается формирование массива наблюдений по одному (или, более общим образом, группами), при этом решение об проведении следующего наблюдения (группы наблюдений) принимается на основе уже накопленного массива наблюдений. Ввиду этого, теория последовательного статистического анализа тесно связана с теорией оптимальной остановки.

В математической статистике есть общая теория проверки гипотез и большое число методов, посвящённых проверке конкретных гипотез. Рассматривают гипотезы о значениях параметров и характеристик, о проверке однородности (то есть о совпадении характеристик или функций распределения в двух выборках), о согласии эмпирической функции распределения с заданной функцией распределения или с параметрическим семейством таких функций, о симметрии распределения и др.

Большое значение имеет раздел математической статистики, связанный с проведением выборочных обследований, со свойствами различных схем организации выборок и построением адекватных методов оценивания и проверки гипотез.

Задачи восстановления зависимостей активно изучаются более 200 лет, с момента разработки К. Гауссом в 1794 г. метода наименьших квадратов.

Разработка методов аппроксимации данных и сокращения размерности описания была начата более 100 лет назад, когда К. Пирсон создал метод главных компонент. Позднее были разработаны факторный анализ[2] и многочисленные нелинейные обобщения[3].

Различные методы построения (кластер-анализ), анализа и использования (дискриминантный анализ) классификаций (типологий) именуют также методами распознавания образов (с учителем и без), автоматической классификации и др.

В настоящее время компьютеры играют большую роль в математической статистике. Они используются как для расчётов, так и для имитационного моделирования (в частности, в методах размножения выборок и при изучении пригодности асимптотических результатов).

Математи́ческая **стати́стика** — наука о математических методах систематизации и использования статистических данных для научных и практических выводов. Во многих своих разделах математическая статистика опирается на теорию вероятностей, позволяющую оценить надежность и точность выводов, делаемых на основании ограниченного статистического материала (напр., оценить необходимый объем выборки для получения результатов требуемой точности при выборочном обследовании).

Математическая статистика - наука о математических методах анализа данных, полученных при проведении массовых наблюдений (измерений, опытов). В зависимости от математической природы конкретных результатов наблюдений статистика математическая делится на статистику чисел, многомерный статистический анализ, анализ функций (процессов) и временных рядов, статистику объектов нечисловой природы. Существенная часть статистики математической основана на вероятностных моделях. Выделяют общие задачи описания данных, оценивания и проверки гипотез. Рассматривают и более частные задачи, связанные с проведением выборочных обследований, восстановлением зависимостей, построением и использованием классификаций (типологий) и др.

Для описания данных строят таблицы, диаграммы, иные наглядные представления, например, корреляционные поля. Вероятностные модели обычно не применяются. Некоторые методы описания данных опираются на продвинутую теорию и возможности современных компьютеров. К ним относятся, в частности, кластер-анализ, нацеленный на выделение групп объектов, похожих друг на друга, и многомерное шкалирование, позволяющее наглядно представить объекты на плоскости, в наименьшей степени исказив расстояния между ними.

Методы оценивания и проверки гипотез опираются на вероятностные модели порождения данных. Эти модели делятся на параметрические и непараметрические. В параметрических моделях предполагается, что изучаемые объекты описываются функциями распределения, зависящими от небольшого числа (1-4) числовых параметров. В непараметрических моделях функции распределения предполагаются произвольными непрерывными. В статистике математической оценивают параметры и характеристики распределения (математическое ожидание, медиану, дисперсию, квантили и др.), плотности и функции распределения, зависимости между переменными (на основе линейных и непараметрических коэффициентов корреляции, а также параметрических или непараметрических оценок функций, выражающих зависимости) и др. Используют точечные и интервальные (дающие границы для истинных значений) оценки.

В математической статистике есть общая теория проверки гипотез и большое число методов, посвященных проверке конкретных гипотез. Рассматривают гипотезы о значениях параметров и характеристик, о проверке однородности (т.е. о совпадении характеристик или функций распределения в двух выборках), о согласии эмпирической функции распределения с заданной функцией распределения или с параметрическим семейством таких функций, о симметрии распределения и др.

Большое значение имеет раздел математической статистики, связанный с проведением выборочных обследований, со свойствами различных схем организации выборок и построением адекватных методов оценивания и проверки гипотез.

Задачи восстановления зависимостей активно изучаются более 200 лет, с момента разработки К. Гауссом в 1794 г. метода наименьших квадратов. В настоящее время наиболее актуальны методы поиска информативного подмножества переменных и непараметрические методы.

Различные методы построения (кластер-анализ), анализа и использования (дискриминантный анализ) классификаций (типологий) именуют также методами распознавания образов (с учителем и без), автоматической классификации и др.

Математические методы в статистике основаны либо на использовании сумм (на основе Центральной Предельной Теоремы теории вероятностей) или показателей различия (расстояний, метрик), как в статистике объектов нечисловой природы. Строго обоснованы обычно лишь асимптотические результаты. В настоящее время компьютеры играют большую роль в математической статистике. Они используются как для расчетов, так и для имитационного моделирования (в частности, в методах размножения выборок и при изучении пригодности асимптотических результатов).

ЛЕКЦИЯ №5

Задачи математической статистики

Установление закономерностей, которым подчинены массовые случайные явления, основано на изучении методами теории вероятностей ***статистических данных*** - результатов наблюдений.

Статистические данные представляют собой данные, полученные в результате обследования большого числа объектов или явлений; следовательно, математическая статистика имеет дело с массовыми явлениями.

Первая задача математической статистики - указать способы сбора и группировки статистических сведений, полученных в результате наблюдений или в результате специально поставленных экспериментов.

Вторая задача математической статистики - разработать методы анализа статистических данных в зависимости от целей исследования.

Современная математическая статистика разрабатывает способы определения числа необходимых испытаний до начала исследования, в ходе исследования и решает многие другие задачи. Современную математическую статистику определяют как науку о принятии решений в условиях неопределенности

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

Генеральная и выборочная совокупность статистических данных

Пусть требуется изучить совокупность однородных объектов относительно некоторого ***качественного*** или ***количественного*** признака, характеризующего эти объекты.

Качественными признаками объект обладает либо не обладает. Они не поддаются непосредственному измерению (например, спортивная специализация, квалификация, национальность, территориальная принадлежность и т. п.).

Количественные признаки представляют собой результаты подсчета или измерения. В соответствии с этим они делятся на дискретные и непрерывные.

Иногда проводиться ***сплошное*** обследование, т.е. обследуют каждый из объектов совокупности относительно признака, которым интересуются. На практике сплошное обследование применяют сравнительно редко. Например, если совокупность содержит очень большое число объектов, то провести сплошное обследование физически невозможно. В таких случаях случайно отбирают из всей совокупности ограниченное число объектов и подвергают их изучению. Различают ***генеральную*** и ***выборочную*** совокупности.

Выборочной совокупностью (выборкой) называют совокупность случайно отобранных объектов.

Генеральной (основной) совокупностью называют совокупность, объектов из которых производится выборка.

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.

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.

<u>На практике обычно пользуются бесповторным случайным отбором</u>.

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли (выборка должна правильно представлять пропорции генеральной совокупности) - выборка должна быть ***репрезентативной (представительной)***.

Выборка будет ***репрезентативной***, если:

- каждый объект выборки ***отобран случайно*** из генеральной совокупности;
- все объекты имеют ***одинаковую вероятность*** попасть в выборку.

ЛЕКЦИЯ № 6

Тема: Основы теории вероятностей и математической статистики

ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ.

Комбинаторика.

1. Введение

Обычно исследователь, получив данные эксперимента на одной или нескольких группах испытуемых и определив по ним некоторые обобщающие числовые характеристики (среднее, стандартное отклонение и др.). пытается найти ответ на следующие вопросы: насколько точно полученные результаты можно обобщить для более широкой совокупности (например, на всех жителей данного возраста)? Как хорошо его данные согласуются с данными других исследователей? Насколько достоверно различие экспериментальных данных, полученных в разных группах испытуемых или в одной и той же группе, но в разные промежутки времени? Существует ли связь между различными признаками, изучаемыми в проводимом исследовании, и если да, то насколько она сильна?

В ряде случаев исследователь пытается установить такую экспериментальную зависимость между изучаемыми признаками, чтобы по значениям одного из них, легко поддающегося измерению, установить значение другого, измерить который трудно или невозможно.

Конечно, в зависимости от целей конкретного исследования задачи могут быть различными и не ограничиваются приведенным перечнем.

Методы математической статистики, с помощью которых можно получить ответы на поставленные выше вопросы, рассматриваются в следующих главах. Чаще всего эти методы основаны на использовании тех или иных согласующихся с условиями проводимого эксперимента математических моделей, разработанных теорией вероятностей.

В данной главе рассматриваются некоторые ее элементарные, положения в том минимальном объеме, который необходим для дальнейшего изложения.

2. Определение вероятности

2.1. Испытание, событие, случайная величина

Под *испытанием* (*опытом*) в теории вероятностей принято понимать наблюдение какого-либо явления при соблюдении определенного комплекса условий, который должен каждый раз строго выполняться при повторении данного испытания. Если то же самое явление наблюдается при другом комплексе условий, то это уже другое испытание.

Когда речь идет о соблюдении комплекса условий данного испытания, имеется в виду постоянство значений всех факторов, контролируемых в данном испытании. Но при этом, как правило, имеет место большое число неконтролируемых факторов, которые трудно или невозможно учесть.

Результаты испытаний можно охарактеризовать качественно и количественно.

Качественная характеристика заключается в регистрации какого-либо явления, которое может наблюдаться или не наблюдаться при данном испытании. Любое из этих явлений называется в теории вероятностей *событием*.

События делятся на:

невозможные	достоверные	случайные
(в результате опыта никогда не произойдут),	(в результате опыта происходят всегда),	(в результате опыта событие может произойти или не произойти).

Теория вероятностей рассматривает именно случайные события. При этом предполагается, что испытание может быть повторено неограниченное (по крайней мере, теоретически) число раз. Например, выполнение штрафного броска в баскетболе есть испытание, а попадание в кольцо — событие.

Другим примером события, часто приводимым в учебниках по теории вероятностей, является выпадение определенного числа очков (от 1 до 6) при бросании игральной кости.

События в теории вероятностей принято обозначать начальными прописными латинскими буквами A, B, C, ...

Случайные события называются *несовместными* если появление одного исключает появление другого. В противном случае они называются *совместными*.

Если в результате опыта произойдет хоть одно из некой группы событий, то они образуют *полную группу*. Появление хотя бы одного события из полной группы – достоверное событие.

Если, по условиям испытания нет никаких оснований предполагать, что один из исходов появляется чаще других, то все исходы являются *равновозможными.*

Два события называются *независимыми*, если появление одного из них не изменяет вероятности другого.

Количественная характеристика испытания состоит в определении значений некоторых величин, которыми интересуются при данном испытании (например, число подтягиваний на перекладине или время на беговой дистанции). В силу действия большого числа неконтролируемых факторов эти величины могут принимать различные значения в результате испытания. Причем до испытания невозможно предсказать значение величины, поэтому она называется *случайной величиной*.

2.2. Вероятность событий

Вероятность какого либо события – численное выражение возможности его наступления.

В некоторых простейших случаях вероятности событий могут быть легко определены непосредственно исходя из условий испытаний.

Представим себе общую схему таких испытаний.

Пусть испытание имеет n возможных несовместных исходов, т. е. отдельных событий, могущих появиться в результате данного испытания; причем при каждом повторении испытания возможен один и только один из этих исходов. Кроме того, пусть по условиям испытания, нет никаких оснований предполагать, что один из исходов появляется чаще других, т. е. все исходы являются равновозможными.

Допустим теперь, что при n равновозможных несовместных исходах интерес представляет некоторое событие А, появляющееся при каждом из m исходов и не появляющееся при остальных $n–m$ исходах. Тогда принято говорить, что в данном испытании имеется n случаев, из которых m благоприятствуют появлению события А.

Вероятность события А равна отношению числа исходов, благоприятствующих событию А, к общему числу всех равновозможных несовместных исходов опыта:

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

(4.1)

Формула (4.1) представляет собой так называемое *классическое определение вероятности* по Лапласу, пришедшее из области азартных игр, где теория вероятностей применялась для определения перспективы выигрыша.

Статистическое определение вероятности.

Будем фиксировать число испытаний, в результате которых появилось некоторое событие А. Пусть было проведено N испытаний, в результате которых событие А появилось ровно n_N раз. Тогда число n_N называется частотой события, а отношение $\frac{n_N}{N}$ — частостью (относительной частотой) события.

Замечательным экспериментальным фактом является то, что частость события при большом числе повторений испытания начинает мало изменяться и стабилизируется около некоторого определенного значения, в то время как при малом числе повторений она принимает различные, совершенно случайные значения. Поэтому интуитивно ясно, что если при неограниченном повторении испытания частость события будет стремиться к вполне определенному числовому значению, то это значение можно принять и качестве объективной характеристики события А. Такое число $P(A)$, связанное с событием А, называется вероятностью события А.

Математически неограниченное число повторений испытания записывается в виде предела (*lim*) при *N*, стремящемся к бесконечности (∞):

$$P(A) = \lim_{N\to\infty} \frac{n_N}{N}.$$

Поскольку n_N никогда не может превзойти *N*, то вероятность оказывается заключенной в интервале $0 \le P(A) \le 1$.

Следует отметить, что приведенное определение вероятности является абстрактным, оно не может быть экспериментально проверено, так как на практике нельзя реализовать бесконечно большое число повторений испытания.

Пусть проводятся независимые испытания, при каждом из которых вероятность события А неизменна. Справедливо утверждение, называемое *законом больших чисел* или *теоремой Бернулли*: если *N* достаточно велико, то с вероятностью сколь угодно близкой к единице, отличие $\frac{n_N}{N}$ от *P(A)* меньше любого наперед заданного положительного числа или, в символьной записи, $P\left(\left|\frac{n_N}{N} - P(A)\right| > \varepsilon\right) = 1$. Т.е. много раз бросая монету, мы “почти наверняка” будем получать примерно равные частоты выпадения герба и цифры.

3. Действия над событиями

В этом разделе приводятся основные правила операций над событиями с использованием для наглядности их графического изображения в виде диаграмм.

Вначале введем понятие “поле событий” как совокупности всех случайных событий данного испытания, для которых определены вероятности. На рис. 4.1 поле событий изображено в виде заштрихованного прямоугольника.

1. Сумма (объединение) событий (рис. 4.2) представляет собой сложное событие, состоящее в появлении хотя бы одного из событий А и В. Объединение событий обозначается как $A \cup B$, или $A + B$.

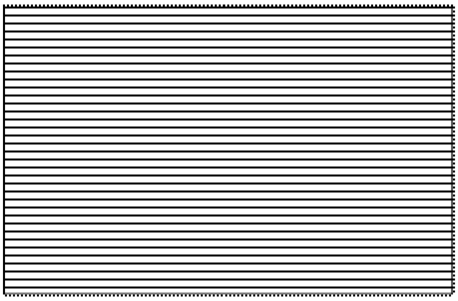
2. Произведением (пересечением) событий А и В называется их совместное появление (рис. 4.3). Обозначается произведение событий как $A \cap B$, или $A \bullet B$.

3. Достоверным событием называется событие, которое обязательно происходит в результате данного испытания (рис. 4.4). Оно обозначается обычно как Е.

4. Невозможное событие – событие, которое не может произойти в результате данного испытания. Принятое обозначение – $\varnothing$.

5. Несовместными называются события, которые в результате данного испытания не могут произойти вместе (рис. 4.5). Примеры несовместных событий: попадание и промах при выстреле, выпадение двух и трех очков при бросании игральной кости. Рис. 4.5 наглядно показывает, что для несовместных событий $A \bullet B = \varnothing$.

6. Противоположным к А событием называется событие, состоящее в не появлении события А (рис. 4.6). Обозначается противоположное событие символом $\overline{A}$. Примеры противоположных событий: промах и попадание при выстреле, выпадение герба или цифры при одном подбрасывании монеты.



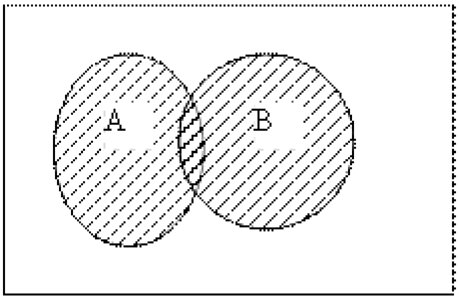
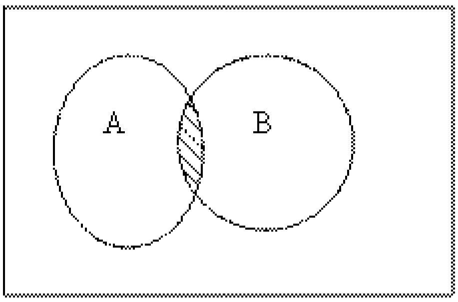


Рис. 4.1. Поле событий

Рис. 4.2. Сумма событий



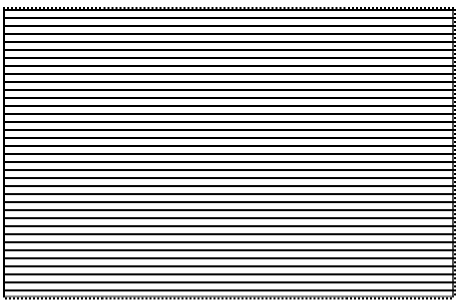
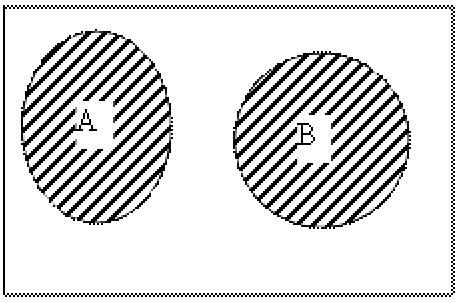


Рис. 4.3. Произведение событий

Рис. 4.4. Достоверное событие



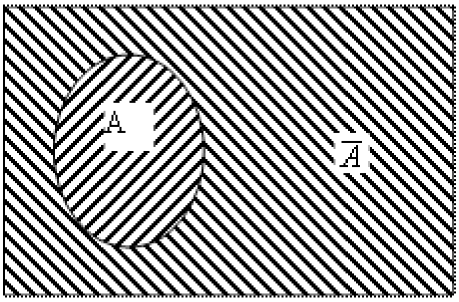


Рис. 4.5. Несовместные события

Рис. 4.6. Противоположные события

ЛЕКЦИЯ №7

4. Исчисление вероятностей

4.1.Примеры непосредственного определения вероятностей

Рассмотрим несколько примеров на вычисление вероятностей по формуле (4.1).

Пример 4.1

Испытание состоит в подбрасывании игральной кости, на каждой из граней которой проставлено число очков (от 1 до 6). Какова вероятность того, что: 1) выпадает 2 очка? 2) выпадает нечетное число очков?

Решение 1: В данном испытании имеется 6 равновозможных случаев (выпадение 1, 2, 3, 4, 5, 6 очков), так как нет оснований предполагать, что появление какого-то определенного числа очков более вероятно (если, конечно, кость симметрична). Поэтому вероятность выпадения любого числа очков, в том числе и 2, при одном подбрасывании равна $\frac{1}{6}$.

Событию А, заключающемуся в появлении нечетного числа очков, благоприятствуют три случая (выпадение 1, 3 и 5), поэтому по формуле (4.1) получаем

$$P(A) = \frac{3}{6} = 0{,}5.$$

Решение 2: В данном испытании имеется 2 равновозможных исхода (выпадение четного числа очков (т.е. 2, 4, 6) и нечетного), так как кость симметрична, то очевидно, что эти исходы равновозможные.

Событию А, заключающемуся в появлении нечетного числа очков, благоприятствуют 1 случай из двух, поэтому по формуле (4.1) получаем $P(A) = \frac{1}{2} = 0{,}5.$

Отметим, что построенную таким образом пространство элементарных событий непригодно для расчета вероятности того, что выпадает 2 очка, так как этому событию не благоприятствует не один из введенных нами элементарных исходов.

Пример 4.2

В урне 5 белых и 10 черных шаров, не отличающихся по размеру. Шары тщательно перемешивают и затем наугад вынимают 1 шар. Какова вероятность того, что вынутый шар окажется белым?

Решение. В этом примере имеется 15 равновозможных (шары не отличаются по размеру) исходов опыта, причем ожидаемому событию (появлению белого шара) благоприятствуют 5 из них, поэтому искомая вероятность составит $\frac{5}{15}$.

4.2. Основные правила вычисления вероятностей сложных событий

Ниже приведены основные правила, позволяющие определить вероятность появления сложного события на основании известных вероятностей составляющих его более простых событий.

1. *Вероятность достоверного события* равна единице:

$$P(E) = 1 \qquad (4.2)$$

.

2. *Вероятность объединения (суммы) несовместных событий* равна сумме их вероятностей:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) +, \ldots, + P(A_n) \qquad (4.3)$$

Эти два равенства являются аксиомами теории вероятностей, т. е. принимаются в качестве исходных, но требующих доказательства свойств вероятностей. На их основе строится вся теория вероятностей.

Все остальные, приведенные ниже без доказательств формулы могут быть выведены из принятых аксиом.

3. *Вероятность невозможного события* равна нулю:

$$P(\varnothing) = 0 \qquad (4.4)$$

.

4. *Вероятность события, противоположного* событию А, равна

$$P(\overline{A}) = 1 - P(A). \qquad (4.5)$$

Формула (4.5) оказывается полезной на практике в тех случаях, когда вычисление вероятности непосредственно события *A* затруднительно, в то время как вероятность противоположного события находится просто (см. ниже п.**9**).

5. *Теорема сложения вероятностей*. Вероятность объединения произвольных событий равна сумме их вероятностей за вычетом вероятности произведения событий:

$$P(A \cup B) = P(A) + P(B) - P(AB) \qquad (4.6)$$

.

Для несовместных событий $P(AB) = 0$ и формула (4.6) переходит в (4.3).

6. *Условная вероятность.* Если требуется найти вероятность события *B* при условии, что произошло некоторое другое событие *A*, то такую ситуацию характеризуют с помощью условной вероятности $P(B \mid A)$. Условная вероятность равна отношению вероятности произведения событий *A* и *B* к вероятности события *A*:

$$P(B|A) = \frac{P(AB)}{P(A)} \qquad (4.7)$$

В тех случаях, когда события *A* и *B* несовместны, $P(AB) = 0$ и соответственно $P(B|A) = 0$.

7. Определение условной вероятности в виде (4.7) дает возможность записать следующую формулу для вычисления вероятности произведения событий (*теорема умножения вероятностей*)

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

(4.8)

8. Поскольку вероятность события *A* (или *B*) для независимых событий по определению не изменяется при появлении другого события, то условная вероятность $P(A|B)$ совпадает с вероятностью события *A*, а условная вероятность $P(B|A)$ — с *P(B)*. Вероятности *P(A)* и *P(B*) в отличие от условных вероятностей называются безусловными.

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

(4.9)

Теорема умножения вероятностей для независимых событий записывается следующим образом:

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

, (4.10)

т. е. вероятность произведения независимых событий равна произведению их вероятностей.

9. Вычислим *вероятность появления хотя бы одного события в n испытаниях*

A – появление в *n* испытаниях **хотя бы** один раз интересующего нас события.

$\overline{A}$ – интересующее нас событие не появилось в *n* испытаниях **ни разу**.

A_1 – интересующее нас событие появилось в первом испытании.

A_2 – интересующее нас событие появилось во втором испытании.

….

A_n – интересующее нас событие появилось в *n*-ом испытании.

$$P(A) = 1 - P(\overline{A}) = 1 - P(\overline{A}_1\cdot\overline{A}_2\cdot\ldots\cdot\overline{A}_n) = 1 - P(\overline{A}_1)\cdot P(\overline{A}_2)\cdot\ldots\cdot P(\overline{A}_n)$$

(4.11)

10. *Формула полной вероятности*.

Если событие *A* может произойти только при появлении одного из несовместных событий $H_1, H_2, \ldots, H_n$, то

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \ldots + P(H_n)P(A|H_n)$$

. (4.12)

Пример 4.3

В урне 5 белых, 20 красных и 10 черных шаров, не отличающихся по размеру. Шары тщательно перемешивают и затем наугад вынимают 1 шар. Какова вероятность того, что вынутый шар окажется белым или черным?

Решение. Пусть событие A – появление белого или черного шара. Разобьем это событие на более простые. Пусть B_1 – появление белого шара, а B_2 – черного. Тогда, $A=B_1+B_2$по определению суммы событий. Следовательно $P(A)=P(B_1+B_2)$. Так как B_1 и B_2 – несовместные события, то по теореме о вероятности суммы несовместных событий (формула 4.3)$P(B_1+B_2) = P(B_1)+P(B_2)$.

Вычислим вероятности событий B_1 и B_2. В этом примере имеется 35 равновозможных (шары не отличаются по размеру) исходов опыта, событию B_1 (появлению белого шара) благоприятствуют 5 из них, поэтому $P(B_1)=\frac{5}{35}$. Аналогично, $P(B_2)=\frac{10}{35}$. Следовательно, $P(A)=\frac{5}{35}+\frac{10}{35}=\frac{15}{30}$.

Пример 4.4

Ведутся поиски двух преступников. Каждый из них независимо от другого может быть обнаружен в течение суток с вероятностью 0,5. Какова вероятность того, что в течение суток будет обнаружен хотя бы один преступник?

Решение. Пусть событие A – “обнаружен хотя бы один преступник”. Разобьем это событие на более простые. Пусть B_1 – обнаружен первый преступник, а B_2 – обнаружен второй преступник. Тогда, $A=B_1+B_2$ по определению суммы событий. Следовательно $P(A)=P(B_1+B_2)$. Так как B_1и B_2 – совместные события, то по теореме о вероятности суммы событий (формула 4.6)

$$P(B_1+B_2) = P(B_1)+P(B_2)-P(B_1\ B_2) = 0{,}5+0{,}5 - 0{,}25=0{,}75.$$

Можно решать и через обратное событие: $P(A)=1-P(\overline{A})=1-P(\overline{B}_1\cdot\overline{B}_2)=1-\frac{1}{2}\cdot\frac{1}{2}=1-\frac{1}{4}=\frac{3}{4}$.

Пример 4.5 а)

Преступник имеет 3 ключа. В темноте он открывает дверь выбирая ключ случайным образом. На открытие каждой из дверей он тратит 5 сек. Найти вероятность того, что он откроет все двери за 15 сек.

Решение. Пусть событие A – “открыты все двери”. Разобьем это событие на более простые. Пусть B – “открыта 1-я“, C – “ открыта 2-я“, а D – “ открыта 3-я“. Тогда, $A=BCD$ по определению произведения событий. Следовательно $P(A)=P(BCD)$. По теореме о вероятности произведения независимых событий (формула 4.10) $P(BCD) = P(B)P(C)\ P(D)$.

Вычислим вероятности событий B, C и D. В этом примере имеется 3 равновозможных (каждый ключ выбираем из 3-х) исходов опыта. Каждому из событий B, C и D благоприятствует 1 из них, поэтому $P(B)=P(C)=P(D)=\frac{1}{3}$

. $P(A)=\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{9}$.

Пример 4.5 б)

Изменим задачу: считаем, что преступник – забывчивый человек. Пусть преступник открыв дверь, оставляет ключ в ней. Какова тогда вероятность, что он откроет все двери за 15 сек?

Решение. Событие A – "открыты все двери". Опять, $A=BCD$ по определению произведения событий. Следовательно $P(A)=P(BCD)$. Но, теперь события B, C и D – зависимы. По теореме о вероятности произведения зависимых событий $P(BCD) = P(B)P(C|B)\,P(D|BC)$.

Вычислим вероятности : $P(B)=\frac{1}{3}$, $P(C\,|\,B)=\frac{1}{2}$ (ключа осталось только два и один из них подходит!), $P(D\,|\,BC)=\frac{1}{1}$ и, значит, $P(A)=\frac{1}{3}\cdot\frac{1}{2}\cdot\frac{1}{1}=\frac{1}{6}$.

Пример 4.6

Ведутся поиски двух преступников. Каждый из них независимо от другого может быть обнаружен в течение суток с вероятностью 0,5. После поимки одно из них, в связи с увеличением количества сотрудников, занятых в поисках, вероятность найти второго возрастает до 0,7. Какова вероятность того, что в течение суток будет обнаружены оба преступника.

Решение. Пусть событие A – "обнаружены два преступника". Разобьем это событие на более простые. Пусть B_1 – обнаружен первый преступник, а B_2 – обнаружен второй преступник, после того, как пойман первый. Тогда, $A=B_1B_2$ по определению произведения событий. Следовательно $P(A)=P(B_1B_2)$. Так как B_1 и B_2 – зависимые события, то по теореме о вероятности произведения зависимых событий (формула 4.8) $P(B_1B_2) = P(B_1)P(B_2/B_1) = 0{,}5\ 0{,}7=0{,}35$.

Пример 4.7

Найти вероятность того, что при подбрасывании монеты 10 раз герб выпадет хотя бы 1 раз.

Решение. Пусть событие A – "герб выпадет **хотя бы** 1 раз". Рассмотрим обратное событие: $\overline{A}$ – "герб не выпадет **ни разу**". Очевидно, что обратное событие легче чем исходное разбить на более простые. Пусть A_1 – герб не выпал при первом броске, A_2 – герб не выпал при втором броске, … A_{10} – герб не выпал при 10-м броске. Все события $A_1 \ldots A_{10}$ независимы, следовательно, (формула 4.11)

$$P(A)=1-P(\overline{A})=1-P(\overline{A}_1\cdot\overline{A}_2\cdot\ldots\cdot\overline{A}_{10})=1-P(\overline{A}_1)\cdot P(\overline{A}_2)\cdot\ldots\cdot P(\overline{A}_{10})=1-\frac{1}{2}\cdot\frac{1}{2}\cdot\ldots\cdot\frac{1}{2}=1-\frac{1}{1024}=\frac{1023}{1024}.$$

Пример 4.8

В проведении операции по освобождению заложников участвуют 2 группы снайперов: 10 человек с винтовкой ОП21 и 20 человек с АКМ47. Вероятность поражения из ОП21 – 0,85, а АКМ47 – 0,65. Найти вероятность того, что при одном выстреле произвольного снайпера преступник будет поражен.

Решение. Пусть событие A – “преступник поражен”. Разобьем это событие на более простые. Преступник может быть поражен либо из ОП21, либо из АКМ47. Вероятность того, что произвольный снайпер вооружен ОП21 (событие H_1) равна 10/30. Вероятность того, что произвольный снайпер вооружен АКМ47 (событие H_2) равна 20/30.

Вероятность того, что преступник поражен равна (формула 4.12)

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \frac{10}{30}\cdot 0{,}85 + \frac{20}{30}\cdot 0{,}65.$$

.

4.3. Комбинаторика

Правило сложения.

Если первое действие можно выполнить n различными способами, а второе — m способами, то выполнить первое ***ИЛИ*** второе действие можно $n+m$ способами.

Правило умножения.

Если первое действие можно выполнить n различными способами, а второе — m способами, то выполнить первое ***И*** второе действие (в таком порядке) можно $n\cdot m$ способами.

Эти правила можно обобщить на случай 3-х, 4-х и более действий.

Пример 4.9.

Рекламный плакат мебельной фабрики утверждает, что возможно составить 100000 различных вариантов расстановки производимых ей шкафов если купить хотя бы 5 шкафов. Верно ли это, если выпускается всего 10 различных типов шкафов?

Решение.

Вычислим, сколькими способами можно расставить 5 шкафов рядом друг с другом. Первую позицию можно заполнить 10-ю различными способами (принцип сложения), вторую — также 10-ю (никто не мешает купить и второй шкаф той же модели, что и первый), третью — опять 10-ю и т.д. Вообще имеем (принцип умножения) $10\cdot 10\cdot 10\cdot 10\cdot 10=100000$ различных вариантов расстановки пяти шкафов рядом друг с другом. Если же купить шкафов больше, чем 5, то, очевидно, вариантов расстановки будет еще больше. **Вывод:** реклама является добросовестной.

Пример 4.10.

Из тщательно перемешанной колоды в 52 карты вытягивают 3 карты. Сколько существует различных вариантов карт на руках у игрока?

Решение.

В данном опыте производится 3 действия: вытягивание 1-й карты, 2-й карты и 3-й карты.

Вычислим, сколькими способами можно вытянуть 1-ую карту. Так как всего в колоде 52 карты, то имеем 52 различных способа. (Здесь мы применили принцип сложения: карта может быть двойка пик ***ИЛИ*** тройка пик ***ИЛИ*** … ***ИЛИ*** туз червей. Значит, всего имеем 1+1+…+1=52 способа.)

Вычислим, сколькими способами можно вытянуть 2-ую карту. Так как в колоде осталось 51 карта, то, значит, второе действие можно выполнить 51-м способом.

Аналогично рассуждая, находим, что 3-е действие можно осуществить 50-ю способами.

Всего различных вариантов **расположения** карт на руках у игрока будет $52\cdot51\cdot50 = 132600$ способов. Для ответа осталось разделить это число на $3\cdot2\cdot1$ – это кол-во способов перетасовать эти 3 розданные карты.

Ответ:

22100.

Число сочетаний, размещений и перестановок

Если стоит задача вычислить сколькими способами можно расположить “в ряд” (т.е. важен порядок их следования) вытянутые m предметов из коробки содержащей различных n предметов, то имеем так называемую ситуацию” перестановок”.

Вычислим это количество: первую позицию можно заполнить n способами, вторую – n – 1 способом, третью – n – 2 способом, и т.д. Искомое количество способов заполнить все n позиций равно (по принципу умножения)

$n(n-1)(n-2)(n-3)\dots(n-m+1)$

и обозначается A_n^m.

Если стоит задача вычислить сколькими способами можно расположить “в ряд”(т.е. важен порядок их следования) вытянутые m предметов из коробки содержащей различных n предметов, то имеем так называемую ситуацию” перестановок”. Вычислим это количество: первую позицию можно заполнить n способами, вторую – n – 1 способом, третью – n – 2 способом, и т.д. Искомое количество способов равно (по принципу умножения) $A_n^m = n(n-1)(n-2)(n-3)\dots(n-m+1)$.

Но, поскольку нам не важно какой именно элемент стоит на каком месте, то необходимо A_n^m разделить на количество способов по разному переставлять уже выбранные элементы. А это количество равно A_n^n= n(n – 1)(n – 2) (n – 3)... 3 2 1 = n! (читается n факториал). Искомое количество способов заполнить все n позиций равно A_n^m / n! и обозначается C_n^m.

Пример 4.11.

В совбезе ООН 11 членов: 5 постоянных и 6 так называемые ”малые нации”. Для принятия решении, надо, чтобы было 7 голосов ”ЗА”, причем следующим образом: все постоянные + как минимум 2 временных. Сколько всего вариантов голосования? Сколько всего можно организовать выигрышных коалиций? (Выигрышной коалицией называется такая, когда как бы не голосовали противники решение все равно будет принято.)

Решение.

Так, как голосуют 11 делегаций и у них есть 2 выбора (“за”,”против”), то по принципу умножения имеем $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^{11} = 2048$– вариантов голосования. Так как все постоянные члены должны проголосовать ”за”, то выигрышная коалиция определяется только временными членами , а кол-во – количеством способов выбрать2 или 3 или 4 или 5 или 6 временных членов, голосующих ”за”.

Имеем $1 \cdot \left(C_6^2 + C_6^3 + C_6^4 + C_6^5 + C_6^6\right) = 15 + 20 + 15 + 6 + 1 = 57$ способов, причем 15 – число так называемых **минимальных** выигрышных коалиций.

4.4. Схема Бернулли

Пусть производится n одинаковых независимых опытов. В каждом испытании некоторое событие А может произойти с вероятностью p (а, значит, не произойти с вероятностью $q = 1 - p$).

Вычислим вероятность того, что событие произойдет *ровно* k раз в проведенных n опытах.

$P_n(0) = \underbrace{q \cdot q \cdot \ldots \cdot q}_{n} = q^n$ – вероятность того, что **во всех** опытах событие не произойдет (см. также пример 4.7)

$P_n(1) = \underbrace{p \cdot q \cdot \ldots \cdot q}_{n} + \underbrace{q \cdot p \cdot \ldots \cdot q}_{n} + \ldots + \underbrace{q \cdot q \cdot \ldots \cdot p}_{n} = n \cdot p \cdot q^{n-1}$ – вероятность того, что событие произойдет **ровно в одном** опыте.

$P_n(2) = \frac{n(n-1)}{2} p^2 q^{n-2}$ – вероятность того, что событие произойдет **ровно 2 раза** в *n* опытах.

…………………………

$P_n(k) = C_n^k p^k q^{n-k}$ – вероятность того, что событие произойдет **РОВНО k раз** в *n* попытках.

…………………………

$P_n(n) = \underbrace{p \cdot p \cdot \ldots \cdot p}_{n} = p^n$ – вероятность того, что событие произойдет **во всех** опытах.

4.5. Случайные величины

Случайные величины (с.в.) – численное значение появляющееся в результате опыта, и принимающее произвольное значение из заранее определенного множества.

Существует два типа случайных величин: дискретные и непрерывные.

Дискретные случайные величины принимают в результате испытания одно из изолированного дискретного множества значений. Они хорошо подходят для описания результатов измерений, связанных с подсчетом и выражаемых целыми числами.

Примеры дискретных случайных величин: оценка, полученная на экзамене, число попаданий в мишень в серии из 10 выстрелов и т. п.

Вероятность принятия дискретной случайной величиной каждого из возможных ее значений больше нуля. Эта вероятность может быть записана как

$P(\{X = x_i\}) = P_i$,

где i =... −1, 0, 1 ...

Здесь X — обозначение случайной величины; x_i — конкретные числовые значения, принимаемые дискретной случайной величиной; p_i — вероятности этих значений.

Индекс i может в общем случае пробегать значения от $-\infty$ до ∞.

Функция $P(\{X = x_i\})$, связывающая значения дискретной случайной величины с их вероятностями, называется ее ***распределением*** (***законом распределения***). Обычно закон распределения записывается в виде таблицы вида

	1	2		n		
	1	2		n		

Пример 4.12. Пусть X – число очков выпавшее на игральной кости при одном броске. Тогда, эта с.в. распределена по закону

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

Непрерывные случайные величины в результате испытания могут принимать любые значения из некоторого интервала.

Примеры непрерывных случайных величин: спортивный результат в беге или прыжках, рост и масса тела человека, сила мышц и др.

Поскольку число возможных значений непрерывной случайной величины бесконечно велико и чаще всего нет оснований предположить, что одни значения появляются существенно чаще других, то вероятность принятия непрерывной случайной величиной каждого отдельного значения оказывается равной нулю. По этой причине нельзя описать распределение непрерывной случайной величины в виде вероятностей ее отдельных значений, как в случае дискретных случайных величин. Здесь необходимы другие подходы, которые рассмотрены в разделах 4.6 и 4.7.

ЛЕКЦИЯ №8

Тема: Доверительный интервал. Доверительная вероятность

В предыдущих $\overset{\circ}{n}\overset{\circ}{n}$ мы рассмотрели вопрос об оценке неизвестного параметра a одним числом. Такая оценка называется «точечной». В ряде задач требуется не только найти для параметра a подходящее численное значение, но и оценить его точность и надежность. Требуется знать - к каким ошибкам может привести замена параметра a его точечной оценкой $\tilde{a}$ и с какой степенью уверенности можно ожидать, что эти ошибки не выйдут за известные пределы?

Такого рода задачи особенно актуальны при малом числе наблюдений, когда точечная оценка $\tilde{a}$ в значительной мере случайна и приближенная замена a на $\tilde{a}$ может привести к серьезным ошибкам.

Чтобы дать представление о точности и надежности оценки $\tilde{a}$, в математической статистике пользуются так называемыми доверительными интервалами и доверительными вероятностями.

Пусть для параметра a получена из опыта несмещенная оценка $\tilde{a}$. Мы хотим оценить возможную при этом ошибку. Назначим некоторую достаточно большую вероятность β (например, $\beta = 0{,}9,\ 0{,}95$ или $0{,}99$) такую, что событие с вероятностью β можно считать практически достоверным, и найдем такое значение ε, для которого

$P\left(\left|\tilde{a}-a\right|<\varepsilon\right)=\beta$.   (14.3.1)

Тогда диапазон практически возможных значений ошибки, возникающей при замене a на $\tilde{a}$, будет $\pm\varepsilon$; большие по абсолютной величине ошибки будут появляться только с малой вероятностью $\alpha=1-\beta$.

Перепишем (14.3.1) в виде:

$P(\tilde{a}-\varepsilon<a<\tilde{a}+\varepsilon)=\beta$.   (14.3.2)

Равенство (14.3.2) означает, что с вероятностью β неизвестное значение параметра a попадает в интервал

$I_\beta=(\tilde{a}-\varepsilon;\tilde{a}+\varepsilon)$.   (14.3.3)

При этом необходимо отметить одно обстоятельство. Ранее мы неоднократно рассматривали вероятность попадания случайной величины в заданный неслучайный интервал. Здесь дело обстоит иначе: величина a не случайна, зато случаен интервал I_β. Случайно его положение на оси абсцисс, определяемое его центром $\tilde{a}$; случайна вообще и длина интервала 2ε, так как величина ε вычисляется, как правило, по опытным данным. Поэтому в данном случае лучше будет толковать величину β не как вероятность «попадания» точки a в интервал I_β, а как вероятность того, что случайный интервал I_β накроет точку a (рис. 14.3.1).

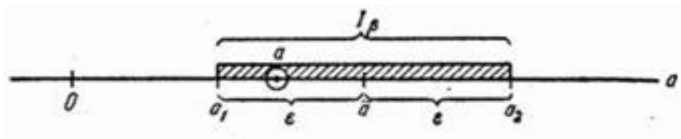


Рис. 14.3.1.

Вероятность β принято называть доверительной вероятностью, а интервал I_β - доверительным интервалом. Границы интервала I_β: $a_1=\tilde{a}-\varepsilon$ и $a_2=\tilde{a}+\varepsilon$ называются доверительными границами.

Дадим еще одно истолкование понятию доверительного интервала: его можно рассматривать как интервал значений параметра a, совместимых с опытными данными и не противоречащих им. Действительно, если условиться считать событие с вероятностью $\alpha=1-\beta$ практически невозможным, то те значения параметра a, для которых $\left|\tilde{a}-a\right|>\varepsilon$, нужно признать противоречащими опытным данным, а те, для которых $|\tilde{a}-a|<\varepsilon$, - совместимыми с ними.

Перейдем к вопросу о нахождении доверительных границ a_1 и a_2.

Пусть для параметра a имеется несмещенная оценка $\tilde{a}$. Если бы нам был известен закон распределения величины $\tilde{a}$, задача нахождения доверительного интервала была бы весьма проста: достаточно было бы найти такое значение ε, для которого

$$P\left(|\tilde{a}-a|<\varepsilon\right)=\beta.$$

Затруднение состоит в том, что закон распределения оценки $\tilde{a}$ зависит от закона распределения величины X и, следовательно, от его неизвестных параметров (в частности, и от самого параметра a).

Чтобы обойти это затруднение, можно применить следующий грубо приближенный прием: заменить в выражении для ε неизвестные параметры их точечными оценками. При сравнительно большом числе опытов n (порядка $20 \div 30$) этот прием обычно дает удовлетворительные по точности результаты.

В качестве примера рассмотрим задачу о доверительном интервале для математического ожидания.

Пусть произведено n независимых опытов над случайной величиной X, характеристики которой - математическое ожидание m и дисперсия D - неизвестны. Для этих параметров получены оценки:

$$\tilde{m}=\frac{\sum_{i=1}^{n} X_i}{n}; \quad \tilde{D}=\frac{\sum_{i=1}^{n}(X_i-\tilde{m})^2}{n-1}. \qquad (14.3.4)$$

Требуется построить доверительный интервал l_β, соответствующий доверительной вероятности β, для математического ожидания m величины X.

При решении этой задачи воспользуемся тем, что величина $\tilde{m}$ представляет собой сумму n независимых одинаково распределенных случайных величин X_i, и, согласно центральной предельной теореме, при достаточно большом n ее закон распределения близок к нормальному. На практике даже при относительно небольшом числе слагаемых (порядка $10 \div 20$) закон распределения суммы можно приближенно считать нормальным. Будем исходить из того, что величина $\tilde{m}$ распределена по нормальному закону. Характеристики этого закона - математическое ожидание и дисперсия - равны соответственно m и $\frac{D}{n}$ (см. гл. 13 n° 13.3). Предположим, что величина D нам известна, и найдём такую величину ε_β для которой

$$P\left(\left|\tilde{m}-m\right|<\varepsilon_\beta\right)=\beta. \qquad (14.3.5)$$

Применяя формулу (6.3.5) главы 6, выразим вероятность в левой части (14.3.5) через нормальную функцию распределения

$$P\left(\left|\tilde{m}-m\right|<\varepsilon_\beta\right)=2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_{\tilde{m}}}\right)-1. \qquad (14.3.6)$$

где $\sigma_{\tilde{m}}=\sqrt{\frac{D}{n}}$ - среднее квадратическое отклонение оценки $\tilde{m}$.

Из уравнения

$$2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_{\tilde{m}}}\right)-1=\beta$$

находим значение ε_β:

$$\varepsilon_\beta=\sigma_{\tilde{m}}\arg\Phi^*\left(\frac{1+\beta}{2}\right), \qquad (14.3.7)$$

где $\arg\Phi^*(x)$ - функция, обратная $\Phi^*(x)$, т. е. такое значение аргумента, при котором нормальная функция распределения равна x.

Дисперсия D, через которую выражена величина $\sigma_{\tilde{m}}$, нам в точности не известна; в качестве ее ориентировочного значения можно воспользоваться оценкой D (14.3.4) и положить приближенно:

$$\sigma_{\tilde{m}}=\sqrt{\frac{\tilde{D}}{n}}. \qquad (14.3.8)$$

Таким образом, приближенно решена задача построения доверительного интервала, который равен:

$$I_\beta = \left(\tilde{m} - \varepsilon_\beta; \tilde{m} + \varepsilon_\beta\right), \qquad (14.3.9)$$

где ε_β определяется формулой (14.3.7).

Чтобы избежать при вычислении ε_β обратного интерполирования в таблицах функции $\Phi^*(x)$, удобно составить специальную таблицу (см. табл. 14.3.1), где приводятся значения величины

$$t_\beta = \arg\Phi^*\left(\frac{1+\beta}{2}\right)$$

в зависимости от β. Величина t_β определяет для нормального закона число средних квадратических отклонений, которое нужно отложить вправо и влево от центра рассеивания для того, чтобы вероятность попадания в полученный участок была равна β.

Через величину t_β доверительный интервал выражается в виде:

$$I_\beta = \left(\tilde{m} - t_\beta \sigma_{\tilde{m}}; \tilde{m} + t_\beta \sigma_{\tilde{m}}\right).$$

Таблица 14.3.1

β	t_β	β	t_β	β	t_β	β	t_β
0,80	1,282	0,86	1,475	0,91	1,694	0,97	2,169
0,81	1,310	0,87	1,513	0,92	1,750	0,98	2,325
0,82	1,340	0,88	1,554	0,93	1,810	0,99	2,576
0,83	1,371	0,89	1,597	0,94	1,880	0,9973	3,000
0,84	1,404	0,90	1,643	0,95	1,960	0,999	3,290
0,85	1,439			0,96	2,053		

Пример 1. Произведено 20 опытов над величиной X; результаты приведены в таблице 14.3.2.

Таблица 14.3.2

i	x_i	i	x_i	i	x_i	i	x_i
1	10,5	6	10,6	11	10,6	16	10,9
2	10,8	7	10,9	12	11,3	17	10,8
3	11,2	8	11,0	13	10,5	18	10,7
4	10,9	9	10,3	14	10,7	19	10,9
5	10,4	10	10,8	15	10,8	20	11,0

Требуется найти оценку $\tilde{m}$ для математического ожидания m величины X и построить доверительный интервал, соответствующий доверительной вероятности $\beta = 0,8$.

Решение. Имеем:

$$\tilde{m} = \frac{1}{20}\sum_{i=1}^{20} x_i = 10,78.$$

Выбрав за начало отсчета $x = 10$, по третьей формуле (14.2.14) находим несмещенную оценку $\tilde{D}$:

$$\tilde{D} = \left(\frac{13,38}{20} - 0,78^2\right)\frac{20}{19} = 0,064;$$

$$\sigma_{\tilde{m}} = \sqrt{\frac{\tilde{D}}{n}} = 0,0564.$$

По таблице 14.3.1 находим $t_\beta = 1,282$;

$$\varepsilon_\beta = t_\beta \sigma_{\tilde{m}} = 0,072.$$

Доверительные границы:

$$m_1 = \tilde{m} - 0,072 = 10,71;$$

$$m_2 = \tilde{m} + 0,072 = 10,85.$$

Доверительный интервал:

$$I_\beta = (10,71; 10,85).$$

Значения параметра m, лежащие в этом интервале, являются совместными с опытными данными, приведенными в таблице 14.3.2.

Аналогичным способом может быть построен доверительный интервал и для дисперсии.

Пусть произведено n независимых опытов над случайной величиной X с неизвестными параметрами m и D, и для дисперсии D получена несмещенная оценка:

$$\widetilde{D} = \frac{\sum_{i=1}^{n}(X_i - \widetilde{m})^2}{n-1}, \qquad (14.3.11)$$

где

$$\widetilde{m} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Требуется приближенно построить доверительней интервал для дисперсии.

Из формулы (14.3.11) видно, что величина D представляет собой сумму n случайных величин вида $\frac{(X_i - \widetilde{m})^2}{n-1}$. Эти величины не являются независимыми, так как в любую из них входит величина $\widetilde{m}$, зависящая от всех остальных. Однако можно показать, что при увеличении n закон распределения их суммы тоже приближается к нормальному. Практически при $n = 20 \div 30$ он уже может считаться нормальным.

Предположим, что это так, и найдем характеристики этого закона: математическое ожидание и дисперсию. Так как оценка D - несмещенная, то

$$M[\widetilde{D}] = D.$$

Вычисление дисперсии $D[\widetilde{D}]$ связано со сравнительно сложными выкладками, поэтому приведем ее выражение без вывода:

$$D[\widetilde{D}] = \frac{1}{n}\mu_4 - \frac{n-3}{n(n-1)}D^2, \qquad (14.3.12)$$

где μ_4 - четвертый центральный момент величины X.

Чтобы воспользоваться этим выражением, нужно подставить в него значения μ_4 и D (хотя бы приближенные). Вместо D можно воспользоваться его оценкой $\widetilde{D}$. В принципе четвертый центральный момент μ_4 тоже можно заменить его оценкой, например величиной вида:

$$\mu_4^* = \frac{\sum_{i=1}^{n}(X_i - \widetilde{m})^4}{n}, \qquad (14.3.13)$$

но такая замена даст крайне невысокую точность, так как вообще при ограниченном числе опытов моменты высокого порядка определятся с большими ошибками. Однако на практике часто бывает, что вид закона распределения величины X известен заранее: неизвестны лишь его параметры. Тогда можно попытаться выразить μ_4 через D.

Возьмем наиболее часто встречающийся случай, когда величина X распределена по нормальному закону. Тогда ее четвертый центральный момент выражается через дисперсию (см. гл. 6 n° 6.2):

$$\mu_4 = 3D^2,$$

и формула (14.3.12) дает

$$D[\tilde{D}] = \frac{3}{n}D^2 - \frac{n-3}{n(n-1)}D^2$$

или

$$D[\tilde{D}] = \frac{2}{n-1}D^2. \qquad (14.3.14)$$

Заменяя в (14.3.14) неизвестное D его оценкой $\tilde{D}$, получим:

$$D[\tilde{D}] = \frac{2}{n-1}\tilde{D}^2, \qquad (14.3.15)$$

откуда

$$\sigma_{\tilde{D}} = \sqrt{\frac{2}{n-1}}\tilde{D}. \qquad (14.3.16)$$

Момент μ_4 можно выразить через D также и в некоторых других случаях, когда распределение величины X не является нормальным, но вид его известен. Например, для закона равномерной плотности (см. главу 5) имеем:

$$\mu_4 = \frac{(\beta-\alpha)^4}{80}; \quad D = \frac{(\beta-\alpha)^2}{12},$$

где (α, β) - интервал, на котором задан закон. Следовательно,

$$\mu_4 = 1{,}8D^2.$$

По формуле (14.3.12) получим:

$$D[\widetilde{D}] = \frac{0,8n+1,2}{n(n-1)} D^2,$$

откуда находим приближенно

$$\sigma_{\widetilde{D}} \approx \sqrt{\frac{0,8n+1,2}{n(n-1)}}\widetilde{D}. \qquad (14.3.17)$$

В случаях, когда вид закона распределения величины X неизвестен, при ориентировочной оценке величины $\sigma_{\widetilde{D}}$ рекомендуется все же пользоваться формулой (14.3.16), если нет специальных оснований считать, что этот закон сильно отличается от нормального (обладает заметным положительным или отрицательным эксцессом).

Если ориентировочное значение $\sigma_{\widetilde{D}}$ тем или иным способом получено, то можно построить доверительный интервал для дисперсии, аналогично тому, как мы строили его для математического ожидания:

$$I_\beta = \left(\widetilde{D} - t_\beta \sigma_{\widetilde{D}}; \widetilde{D} + t_\beta \sigma_{\widetilde{D}}\right), \qquad (14.3.18)$$

где величина t_β в зависимости от заданной вероятности β находится по таблице 14.3.1.

Пример 2. Найти приближенно 80%-й доверительный интервал для дисперсии случайной величины X в условиях примера 1, если известно, что величина X распределена по закону, близкому к нормальному.

Решение. Величина остается той же, что в примере 1:

$$t_\beta = 1,282.$$

По формуле (14.3.16)

$$\sigma_{\widetilde{D}} = \sqrt{\frac{2}{19}} \cdot 0,064 = 0,0207.$$

По формуле (14.3.18) находим доверительный интервал:

$$I_\beta = (0,043; 0,085).$$

Соответствующий интервал значений среднего квадратического отклонения: $(0,21;0,29)$.

ЛЕКЦИЯ №9

Тема: Статистические гипотезы

Пусть в (статистическом) эксперименте доступна наблюдению случайная величина X, распределение которой $\mathbb{P}$ известно полностью или частично. Тогда любое утверждение, касающееся $\mathbb{P}$, называется **статистической гипотезой**. Гипотезы различают по виду предположений, содержащихся в них:

- Статистическая гипотеза, однозначно определяющая распределение $\mathbb{P}$, то есть $H\,:\,\{\mathbb{P}=\mathbb{P}_0\}$, где $\mathbb{P}_0$ какой-то конкретный закон, называется **простой**.

- Статистическая гипотеза, утверждающая принадлежность распределения $\mathbb{P}$ к некоторому семейству распределений, то есть вида $H\,:\,\{\mathbb{P}\in\mathcal{P}\}$, где $\mathcal{P}$ — семейство распределений, называется **сложной**.

На практике обычно требуется проверить какую-то конкретную и как правило простую гипотезу H_0. Такую гипотезу принято называть *нулевой*. При этом параллельно рассматривается противоречащая ей гипотеза H_1, называемая *конкурирующей* или *альтернативной*.

Выдвинутая гипотеза нуждается в проверке, которая осуществляется статистическими методами, поэтому гипотезу называют статистической. Для проверки гипотезы используют критерии, позволяющие принять или опровергнуть гипотезу.

В большинстве случаев статистические критерии основаны на случайной выборке $(X_1,X_2,\ldots,X_n)$ фиксированного объема $n\geq 1$ из распределения $\mathbb{P}$. В последовательном анализе выборка формируется в ходе самого эксперимента и потому её объем является случайной величиной .

Пример

Пусть дана независимая выборка $(X_1,\ldots,X_n)\sim\mathcal{N}(\mu,1)$ из нормального распределения, где μ — неизвестный параметр. Тогда $H_0\,:\,\{\mu=\mu_0\}$, где μ_0 — фиксированная константа, является простой гипотезой, а конкурирующая с ней $H_1: \{\mu > \mu_0\}$ — сложной.

Этапы проверки статистических гипотез.

Формулировка основной гипотезы H_0 и конкурирующей гипотезы H_1.

1. Задание уровня значимости α, на котором в дальнейшем и будет сделан вывод о справедливости гипотезы. Он равен вероятности допустить ошибку первого рода.
2. Расчёт статистики критерия такой, что:
   - её величина зависит от исходной выборки;
   - по её значению можно делать выводы об истинности гипотезы H_0;
   - сама статистика ϕ должна подчиняться какому-то известному закону распределения, так как сама ϕ является случайной в силу случайности $\mathbf{X}$.
3. Построение критической области. Из области значений ϕ выделяется подмножество $\mathbb{C}$ таких значений, по которым можно судить о существенных расхождениях с предположением. Его размер выбирается таким образом, чтобы выполнялось равенство $P(\phi \in \mathbb{C}) = \alpha$. Это множество $\mathbb{C}$ и называется **критической областью**.
4. Вывод об истинности гипотезы. Наблюдаемые значения выборки подставляются в статистику ϕ и по попаданию (или непопаданию) в критическую область $\mathbb{C}$ выносится решение об отвержении (или принятии) выдвинутой гипотезы H_0.

Виды критической области

Выделяют три вида критических областей:

- *Двусторонняя критическая область* определяется двумя интервалами $(-\infty,\ x_{\alpha/2}) \cup (x_{1-\alpha/2}\ +\infty)$, где $x_{\alpha/2},\ x_{1-\alpha/2}$ находят из условий $P(\phi < x_{\alpha/2}) = \frac{\alpha}{2},\quad P(\phi < x_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$.
- *Левосторонняя критическая область* определяется интервалом $(-\infty,\ x_\alpha)$, где x_α находят из условия $P(\phi < x_\alpha) = \alpha$.
- *Правосторонняя критическая область* определяется интервалом $(x_{1-\alpha},\ +\infty)$, где $x_{1-\alpha}$ находят из условия $P(\phi < x_{1-\alpha}) = 1 - \alpha$.

Список литературы

1. Спирина М.С. Дискретная математика: учеб. – М.: Академия, 2006

2. Гончарова Г.А., Мочалин А.А. Элементы дискретной математики: учеб. пособ.- М.: Форум: ИНФРА-М, 2003

3. Пехлецкий И. Д., Математика: Учебник для средних специальных учебных заведений. – М.: Академия, 2003.

4. Данко П. Е., Попов А. П., Кожевникова Т. Я. Высшая математика в упражнениях и задачах. Часть 1, 2. - М., Высшая школа, 2003.

5. Валуцэ И. И., Дилигул Г.Д. Математика для техникумов. - М., Наука, 19906

6. Лисичкин В.Т., Соловейчик И.Л. Математика. – М., Высшая школа, 1991

Теория вероятностей и математическая статистика:

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособ. – М.: Высш. шк., 1998

2. Гмурман В.Е. Теория вероятностей и математическая статистика. Москва «Высшая школа» 1998

Интернет-ресурсы:

http://www.moeobrazovanie.ru/srednee_professionalnoe_obrazovanie.html

https://www.google.ru

http://bashmakov.su/load/srednee_proftekhobrazovanie/npo_i_spo/21