Министерство образования и науки Российской Федерации
Государственное образовательное учреждение высшего профессионального образования
«Владимирский государственный университет имени Александра Григорьевича и
Николая Григорьевича Столетовых»
Кафедра электротехники и электроэнергетики

Максимов Ю.П.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ ПО ДИСЦИПЛИНЕ «Вероятностные и статистические задачи электроэнергетики»

Направление подготовки: электроэнергетика и электротехника

Квалификация (степень) выпускника: магистр

Владимир 2014

Занятие №1. *Элементы комбинаторики*

Размещениями без повторений из n элементов по m называются такие выборки, которые, имея по m элементов, выбранных из числа данных n элементов генеральной совокупности без повторений, отличаются одна от другой либо составом элементов, либо порядком их расположения.

Характерный пример размещений без повторений — вся совокупность четырехзначных номеров, в каждом из которых нет повторения цифр.

Число размещений из n элементов по m договоримся обозначать A_n^m.

Пример.

Сосчитать, сколько можно сделать перестановок в словах: замок, топор, ротор, колокол.

Решение:

замок: P' = $\frac{5!}{1!1!1!1!1!} = 120$; ротор: P' = $\frac{5!}{1!2!2!} = 30$

топор: P' = $\frac{5!}{1!1!2!} = 60$; колокол: P' = $\frac{7!}{2!2!2!} = 210$

Пример.

Я помню, что нужный мне телефонный номер начинается с цифры 9 и содержит три четверки и две пятерки. Однако расположение этих пяти цифр забыто. Сколько нужно сделать проб?

Решение:

P'= $\frac{5!}{3!2!} = 10$

Сочетаниями с повторениями из элементов n классов по m называются такие размещения с повторениями из n классов по m, которые отличаются одно от другого хотя бы одним элементом. Их число подсчитывается по формуле:

$$C_n^m = \frac{(m+n-1)!}{m!(n-1)!}$$

Пример.

В продажу поступили открытки 10 разных видов. Сколькими способами можно образовать набор из 12 открыток?

Решение:

$C_{10}^{12} = \frac{(10+12-1)!}{12!(10-1)!}$=293930.

Обобщим полученные сведения в таблице

Выборки с повторениями

Название	Характерный признак отличия	Пример	Формула подсчета вариантов
Размеще-ния	состав порядок	a, b,c из 3 по 2 ab, bc, ca, ba, cb, ac, aa, cc, bb	A_n^m=n^m
Переста-новки	порядок	a, b из 2 ab, ba, aa, bb	P'=$\frac{k!}{k_1!k_2!*...*k_n!}$
Сочетания	состав	a, b, c из 3 по 2 ab, bc, ca, aa, cc, bb	$C_n^m = \frac{(m+n-1)!}{m!(n-1)!}$

Занятие №2. *Случайные события. Операции над ними.*

Опыт (**эксперимент**, **испытание**) – это ситуация с более чем одним возможным исходом, из которых всегда имеет место точно одно так называемое **элементарное событие**. Исходом опыта может быть результат наблюдения или измерения.

Извлечение карты из колоды – эксперимент. Один из исходов эксперимента – извлечение дамы бубен. Бубновую даму можно извлечь из колоды, содержащей 36 карт и 52 карты. Число карт – условие испытания.

Единичный, отдельный исход эксперимента называется **элементарным событием**. Набор всех элементарных событий – **пространство событий (множество).**

Извлечение любой карты из колоды – элементарное событие. Полному набору событий соответствует полное множество X, относящееся к заданному эксперименту. **Полный набор событий** – набор всех возможных исходов эксперимента. Элементарному событию соответствует только одна точка пространства событий. Аналогом элементарного события является элемент множества.

Теория вероятностей изучает **случайные** события. **Случайным событием** называется событие, которое может произойти или не произойти в результате некоторого эксперимента (далее будем опускать термин «случайный»).

Событие – это любое подмножество пространства событий, набор элементарных исходов. В диаграммах Венна событию соответствует подмножество элементарных событий. Событие произошло, если в результате эксперимента произошло элементарное событие, принадлежащее этому поднабору. Например, элементарные события – «туз конкретной масти» – благоприятствуют случайному событию «туз».

События обычно обозначаются заглавными буквами латинского алфавита: A, B, C, D, E, F и т. д. События можно классифицировать.

Достоверное событие – это событие, которое обязательно произойдет в результате испытания (подброшенный камень обязательно упадет на Землю вследствие действия закона притяжения). Достоверные события условимся обозначать символом Ω.

Невозможное событие – это событие, которое не может произойти в результате данного опыта (извлечение черного шара из урны с белыми шарами есть событие невозможное). Невозможное событие обозначим $\varnothing$.

Достоверные и невозможные события не являются случайными.

Совместные события – несколько событий называют совместными, если в результате эксперимента наступление одного из них не исключает появления других. (В магазин вошел покупатель. События «в магазин вошел покупатель старше 60 лет» и «в магазин вошла женщина» – совместные, так как в магазин может войти женщина старше 60 лет.)

Несовместные события – несколько событий называют несовместными в данном опыте, если появление одного из них исключает появление других (выигрыш, ничейный исход и проигрыш при игре в шахматы как результат одной партии – три несовместных события).

События называют **единственно возможными**, если в результате испытания хотя бы одно из них обязательно произойдет. Некоторая фирма рекламирует свой товар по радио и в газете. Обязательно произойдет одно и только одно из следующих событий: «потребитель услышал о товаре по радио», «потребитель прочитал о товаре в газете», «потребитель получил информацию о товаре по радио и из газеты», «потребитель не слышал о товаре по радио и не читал газеты». Это четыре единственно возможных события.

Несколько событий называют **равновозможными**, если в результате испытания ни одно из них не имеет объективно большей возможности появления, чем другие (при бросании игральной кости выпадение каждой из ее граней – события равновозможные).

Два единственно возможных и несовместных события называются **противоположными** (купля и продажа определенного вида товара есть события противоположные).

Полная группа событий – совокупность всех единственно возможных и несовместных событий.

Полную группу можно определить так: если $\bigcup\limits_{i=1}^{n} A_i = \Omega$ и $A_i \cap A_j = \varnothing$ для любой пары $(i \neq j)$, тогда $\{A_1, A_2, ..., A_n\}$ – полная группа событий.

Занятие 3. *Классическое определение вероятности. Методика вычисления вероятностей событий.*

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.

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

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 свойства:

1. 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 $0/n = 0$.

Пример. 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.

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

Пример. 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 A. 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

$$C_{14}^{3}=\frac{14\cdot13\cdot12}{1\cdot2\cdot3}=14\cdot13\cdot2=364.$$

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

$$m=C_{4}^{2}\cdot C_{14}^{3}=6\cdot364=2184.$$

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

$$P(A)=\frac{m}{n}=\frac{2184}{8568}\approx0{,}255.$$

Занятие 4. *Произведение, сумма событий. Вероятности произведения и суммы событий. Теорема умножения вероятностей.*

Теорема сложения вероятностей несовместных событий

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

Сумму двух событий обозначают символом А+В, а сумму n событий символом $A_1+A_2+\ldots+A_n$.

Теорема сложения вероятностей.

Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий.

$P(A+B)=P(A)+P(B)$ 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 образуют 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.

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

Пример. 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

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

Пример. 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:

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

Пример. 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

$p+g=1$, т.е $g=1-p=1-0{,}85=0{,}15$.

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 $\dfrac{P(AB)}{P(B)}$.

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

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

Пример. 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 $P(AB)=\dfrac{1}{4}$, $P(B)=\dfrac{3}{4}$ и по формуле находим

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

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

Теорема умножения вероятностей

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

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

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

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

Пример. 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 $P(A_1)=\dfrac{4}{10}=\dfrac{2}{5}$, $P(A_2)=\dfrac{7}{12}$, то $P(A_1A_2)=P(A_1)\cdot P(A_2)=\dfrac{2}{5}\cdot\dfrac{7}{12}=\dfrac{14}{60}=\dfrac{7}{30}$

Пример. 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 А- выход из строя первого элемента, событие В- выход их строя второго элемента. Эти события независимы (по условию).

а) Одновременное появление А и В есть событие АВ. 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 А- выходу этого элемента из строя); если работает второй элемент- событие В. 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,

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

Занятие5. *Независимые события. Вероятность произведения независимых событий.*

Рассмотрим пример с двумя событиями. Пусть событие А – «извлечение короля», В – «извлечение карты с картинкой». Тогда вероятность появления короля равна 4/52, а вероятность появления короля, если извлеченная карта – картинка, равна 4/16.

Другой пример. В урне два белых и три черных шара. Чему равна вероятность появления белого шара при первом извлечении из урны? При втором извлечении из урны? Здесь возможны два случая.

Первый случай. Схема возвращенного шара, т. е. шар после первого испытания возвращается в урну.

Пусть событие А – «появление белого шара при первом испытании». Так как N = 5, а М = 2, то Р(А) = 2/5.

Пусть событие В – «появление белого шара при втором извлечении». Так как шар после первого испытания возвратился в урну, то N = 5, а M = 2 и P(B) = 2/5.

Таким образом, вероятность каждого из событий не зависит от того, произошло или не произошло другое событие. События А и В в этом случае называются независимыми. Итак, **события А и В называются независимыми, если вероятность каждого из них не зависит от того, произошло или нет другое событие.** Вероятности независимых событий называются безусловными.

Второй случай. Схема невозвращенного шара, т. е. шар после первого испытания в урну не возвращается.

Вероятность появления белого шара при первом испытании Р(А) = 2/5. Белый шар в урну не возвращается, следовательно, в урне остались один белый и три черных шара. Чему равна вероятность события В при условии, что событие А произошло? N = 4, M = 1.

Искомую вероятность обозначают Р(В/А) или $P(B)_A$ или $P_A(B)$. Итак, Р(В/А) = 1/4 называют **условной вероятностью**, а события А и В – зависимыми. В предыдущем примере с картами Р(А) = 4/52; Р(А/В) = 4/16.

Итак, **события А и В называются зависимыми, если вероятность каждого из них зависит от того, произошло или нет другое событие.** Вероятность события В, вычисленная в предположении, что другое событие А уже осуществилось, называется условной вероятностью. Очевидно, что если два события А и В независимые, то справедливы равенства:

Р(В) = Р(В/А), Р(А) = Р(А/В), или Р(В/А) – Р(В) = 0.

Теорема умножения вероятностей. Вероятность произведения двух зависимых событий А и В равна произведению вероятности одного из них на условную вероятность другого:

$$P(A \cdot B) = P(A \cap B) = P(B) \cdot P(A/B) = P(A) \cdot P(B/A). \qquad (1)$$

Произведением событий A и B называют событие, состоящее в одновременном появлении и события A, и события B.

Доказательство. Проиллюстрируем понятие условной вероятности для случая равновозможных элементарных исходов, где применимо классическое определение вероятности. Пусть даны два события A, B, такие, что $P(A) \neq 0$ и $P(B) \neq 0$, и пусть из всех возможных N исходов событию A благоприятствуют M исходов, событию B благоприятствуют K исходов, событию A и B благоприятствуют L исходов. Вероятности событий A, B, A·B соответственно равны $P(A) = M/N$, $P(B) = K/N$, $P(A \cdot B) = L/N$.

Подсчитаем условную вероятность события B/A. Событию B/A будут благоприятствовать L исходов из M исходов. Тогда $P(B/A) = L/M$. Разделим числитель и знаменатель дроби на N и получим

$$P(B/A) = \frac{L/N}{M/N} = \frac{P(A \cdot B)}{P(A)} = \frac{P(A \cap B)}{P(A)}, \qquad (2)$$

где $P(A) \neq 0$.

Вероятность наступления события B, вычисленная при условии, что событие A уже произошло, равна вероятности пересечения событий A и B, деленной на вероятность события A. Из формулы (3) следует (2).

Пример. Проиллюстрируем формулу (2). Предположим, мы подбросили игральную кость. Пусть событие A – «появилось число 6». Мы знаем, что $P(A) = 1/6$. Предположим, мы не знаем, какое именно число выпало при подбрасывании, но знаем, что оно четное (событие E). Информация о событии E уменьшает наше пространство событий, изменяет вероятность появления события A.

Пространство событий (полная группа событий) для первоначального события A выглядит как набор точек от 1 до 6. Пространство событий, корреспондирующее с событием B, уменьшилось сразу в два раза. Новое пространство имеет три равновозможные точки, отсюда вероятность выпадения «6» при условии, что выпавшее число – четное, возрастает от 1/6 до 1/3. Этот пример хорошо показывает обоснованность принятого определения вероятности. Из уравнения (2) имеем $P(A/E) = P(A \cap E) / P(E) = (1/6) / (1/2) = 1/3$.

Полученный результат согласуется с тем, что мы поняли из рассмотренного примера, когда уменьшали пространство событий до трех точек.

Пример. Консультационная фирма претендует на два заказа от двух крупных корпораций. Эксперты фирмы считают, что вероятность получения консультационной работы в корпорации A (событие A) равна 0,45. По предположению экспертов, если фирма получит заказ у корпорации A, то вероятность того, что и корпорация B обратится к ним, равна 0,9. Какова вероятность получения консультационной фирмой обоих заказов?

Решение: Согласно условиям $P(A) = 0{,}45$, $P(B/A) = 0{,}9$.

Необходимо найти $P(A \cdot B)$, которая является вероятностью того, что оба события (и событие A, и событие B) произойдут. Из формулы (1.8) имеем:

$P(A \cdot B) = P(A) \cdot P(B/A) = 0{,}45 \cdot 0{,}9 = 0{,}405$.

Пример. В большой рекламной фирме 21 % работников получают высокую заработную плату. Известно также, что 40 % работников фирмы – женщины, а 6.4 % работников – женщины, получающие высокую заработную плату. Можем ли мы утверждать, что на фирме существует дискриминация женщин в оплате труда?

Решение: Сформулируем решение этой задачи в терминах теории вероятностей и спросим: «Чему равна вероятность того, что случайно выбранный работник будет женщиной, имеющей высокую заработную плату?». Определим событие A – «случайно выбранный работник имеет высокую зарплату», событие B – «случайно выбранный работник – женщина», тогда:

$P(A/B) = P(A \cdot B)/P(B) = 0{,}064/0{,}40 = 0{,}16$.

Поскольку 0,16 меньше, чем 0,21, то можно заключить, что женщины, работающие в рекламной фирме, имеют меньше шансов получить высокую заработную плату по сравнению с мужчинами. Если события А и В – независимы, то имеет место следующая теорема. **Вероятность произведения двух независимых событий А и В равна произведению их вероятностей**:

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

Занятие 6. *Формула полной вероятности. Формула Байеса.*

Рассмотрим два события А и Н. Каковы бы ни были взаимоотношения между событиями А и Н, всегда можно сказать, что вероятность события А равна вероятности пересечения событий А и Н плюс вероятность пересечения А и дополнения Н (событие $\overline{H}$). Поясним сказанное на диаграмме Венна (рис.1). Разложение А на части зависит от H и $\overline{H}$.

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

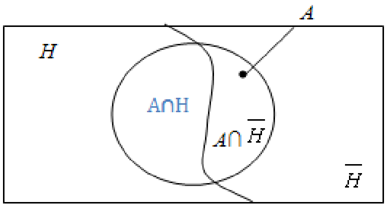


Рис. 1. Диаграмма Венна к формуле (1)

Наборы H и $\overline{H}$ – форма расчленения набора А на два подмножества взаимно несовместных событий. События Н и $\overline{H}$ взаимно противоположны. Событие А может произойти либо с Н, либо с $\overline{H}$, но не с двумя вместе (см. рис. 1).

Рассмотрим более сложный случай. Пусть событие А может осуществляться лишь вместе с одним из событий $H_1, H_2, H_3, ..., H_n$, образующих полную группу, т. е. эти события являются единственно возможными и несовместными (рис. 2). Так как заранее неизвестно, какое из событий $H_1, H_2, H_3, ..., H_n$ наступит, то их называют **гипотезами**. Пусть также известны вероятности гипотез $P(H_1), P(H_2), \ldots, P(H_n)$ и условные вероятности события А, а именно: $P(A/H_1), P(A/H_2), \ldots, P(A/H_n)$.

Так как гипотезы образуют полную группу, то $\sum_{i=1}^{n} P(H_i) = 1$.

Рассмотрим событие А – это или $H_1 \cdot A$, или … $H_n \cdot A$. События $H_1 \cdot A, H_2 \cdot A, \ldots, H_n \cdot A$ – несовместные попарно, так как события $H_1, H_2, H_3, ..., H_n$ попарно несовместны. К этим событиям применяем теорему сложения вероятностей для несовместных событий:

$$P(A)=P(H_1 \cdot A)+P(H_2 \cdot A) +\ldots+ P(H_n \cdot A) = \sum_{i=1}^{n} P(A \cap H_i).$$

События H_1 и А, H_2 и А,..., H_n и А – зависимые. Применив теорему умножения вероятностей для зависимых событий, получим (рис. 2):

$$P(A) = P(H_1) \cdot P(A/H_1)+ P(H_2) \cdot P(A/H_2) +\ldots+P(H_n) \cdot P(A/H_n) = \sum_{i=1}^{n} P(H_i) \cdot P(A/H_i).$$

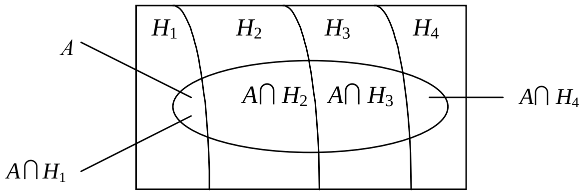


Рис. 2. Событие А может осуществляться лишь с одним из событий $H_1, H_2, ..., H_n$, образующих полную группу событий

Проиллюстрируем сказанное на примере с колодой карт (рис. 2.3). Определим А как событие, состоящее в извлечении карты с картинкой (т. е. карты с изображением или туза, или короля, или дамы, или валета). Пусть события В, С, D, Е означают извлечение карт различной масти («трефы», «бубны», «черви», «пики»). Мы можем сказать, что вероятность извлечь из колоды карту с изображением туза, короля, дамы или валета есть $P(A) = P(A \cap B) + P(A \cap C) + P(A \cap D) + P(A \cap E) = 4/52 + 4/52 + 4/52+4/52 = 16/52$. Это означает, как мы уже знаем, вероятность извлечения карты с картинкой из колоды в 52 карты. Событие А представляет собой набор, составленный из пересечений А с наборами В, С, D, Е (рис. 3).

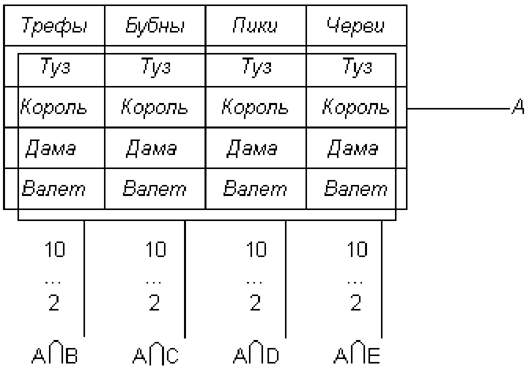


Рис. 3. Пример с колодой карт

Вывод. Если событие А может наступить только вместе с одним из событий H_1, H_2, H_3, ..., H_n, образующих полную группу несовместных событий и называемых гипотезами, то вероятность события А равна сумме произведений вероятностей каждого из событий H_1, H_2, H_3, ..., H_n на соответствующую условную вероятность события А.

Случай двух событий:

$$P(H) = P(H) \cdot P(A/H) + P(\bar{H}) \cdot P(A/\bar{H}).$$

Случай более чем двух событий:

$$P(H) = \sum_{i=1}^{n} P(H_i) \cdot P(A/H_i),$$

где i = 1, 2, ..., n.

Пример. Экономист полагает, что вероятность роста стоимости акций некоторой компании в следующем году будет равна 0,75, в случае успешного развития экономики страны, и эта же вероятность составит 0,30, если произойдет спад экономики. По его мнению, вероятность экономического подъема в будущем году равна 0,80. Используя предположения экономиста, оцените вероятность того, что акции компании поднимутся в цене в следующем году.

Решение: Событие А – «акции компании поднимутся в цене в будущем году». Составим рабочую таблицу:

H_i	Гипотезы H_i	P(	P(A	$P(H_i)P(A/H_i)$
1	H_1 – «подъем экономики»	0,	0,7	0,60
2	H_2 – «спад экономики»	0,	0,3	0,06
∑		1,	–	P(A) = 0,66

Пример. В каждой из двух урн содержится 6 черных и 4 белых шара. Из урны 1 в урну 2 наудачу переложен один шар. Найти вероятность того, что шар, извлеченный из урны 2 после перекладывания, окажется черным.

Решение: Событие А – «шар, извлеченный из урны 2, – черный». Составим рабочую таблицу:

H_i	Гипотезы H_i	$P(H_i)$	$P(A/H_i)$	$P(H_i)P(A/H_i)$

1	H_1 – «из урны 1 в урну 2 переложили черный шар»	6/10	7/11	42/110
2	H_2 – «из урны 1 в урну 2 переложили белый шар»	4/10	6/11	24/110
$\sum$		1,00	–	P(A) = 0,60

Вычисление вероятностей гипотез формула Бейеса

Представим, что существует несколько предположений (несовместных гипотез) для объяснения некоторого события. Эти предположения проверяются с помощью опыта. До проведения опыта бывает сложно точно определить вероятность этих предположений, поэтому им часто приписывают некоторые вероятности, которые называют **априорными** (доопытными). Затем проводят опыт и получают информацию, на основании которой корректируют априорные вероятности. После проведения эксперимента вероятность гипотез может измениться. Таким образом, доопытные вероятности заменяют послеопытными (**апостериорными**).

В тех случаях, когда стало известно, что событие А произошло, возникает потребность в определении условной вероятности $P(H_i/A)$. Пусть событие А может осуществляться лишь вместе с одной из гипотез H_i, (i = 1, 2,..., n). Известны вероятности гипотез $P(H_1)$, ..., $P(H_п)$ и условные вероятности А, т. е. $P(A/H_1)$, $P(A/H_2)$,…, $P(A/H_n)$. Так как $A \cdot H_i = H_i \cdot A$, то $P(A \cdot H_i) = P(H_i \cdot A)$ или $P(A)P(A/H_i)$, а отсюда по правилу пропорций:

$$P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{P(A)}.$$

Итак можно записать формулы Бейеса:

случай двух событий:

$$P(H/A) = \frac{P(H) \cdot P(A/H)}{P(H) \cdot P(A/H) + P(\bar{H}) \cdot P(A/\bar{H})};$$

случай более чем двух событий:

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

Формулы Бейеса позволяют переоценить вероятности гипотез после того, как становится известным результат испытания, в итоге которого появилось событие А.

Как видим из выражения (2.5), вероятность события Н, задаваемая при условии появления события А, получается из вероятностей событий H и $\bar{H}$ и из условной вероятности события А при заданном Н. Вероятности событий H и $\bar{H}$ называют **априорными** (предшествующими), вероятность Р(Н/А) называют **апостериорной** (последующей).

Пример. Экономист полагает, что в течение периода активного экономического роста американский доллар будет расти в цене с вероятностью 0,7, в период умеренного экономического роста доллар подорожает с вероятностью 0,4, и при низких темпах экономического роста доллар подорожает с вероятностью 0,2. В течение любого периода времени вероятность активного экономического роста равна 0,3, в периоды умеренного экономического роста – 0,5 и низкого роста – 0,2. Предположим, доллар дорожает в течение текущего периода, чему равна вероятность того, что анализируемый период совпал с периодом активного экономического роста?

Решение: Определим гипотезы: H_1 – «активный экономический рост»; H_2 – «умеренный экономический рост»; H_3 – «низкий экономический рост».

Определим событие А – «доллар дорожает». Имеем: $P(H_1) = 0,3$; $P(H_2) = 0,5$; $P(H_3) = 0,2$; $P(A/H_1) = 0,7$; $P(A/H_2) = 0,4$ и $P(A/H_3) = 0,2$. Найти: $P(H_1/A)$.

Используя формулу Бейеса (2.6) и подставляя заданные значения вероятностей, получаем:

$$P(H_1/A)=\frac{P(H_1)\cdot P(A/H_1)}{P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)}=\frac{0,3\cdot 0,7}{0,3\cdot 0,7+0,5\cdot 0,4+0,2\cdot 0,2}=0,467.$$

Пример. Партия деталей содержит 20% деталей, изготовленных заводом №1, 30% – заводом №2, 50% – заводом №3. Для завода №1 вероятность выпуска бракованной детали равна 0,05, для завода №2 – 0,01, для завода №3 – 0,06. Чему равна вероятность того, что наудачу взятая из партии деталь окажется бракованной?

Решение: Обозначим через B событие: наудачу взятая деталь – бракованная., через H_1, H_2, H_3 – деталь, изготовленная соответственно заводом №1, №2, №3. Из условия известны вероятности:

$P(H_1)=0.2$, $P(H_2)=0.3$, $P(H_1)=0.5$,

$P(B|H_1)=0.05$, $P(B|H_2)=0.01$, $P(B|H_3)=0.06$.

По формуле полной вероятности находим

$P(B)=0.2\cdot 0.05+0.3\cdot 0.01+0.5\cdot 0.06=0.043$.

Пример. Имеется пять урн. В 1-й, 2-й и 3-й урнах находится по 2 белых и 3 черных шара; в 4-й и 5-й урнах – по 1 белому и 1 черному шару. Случайно выбирается урна и из нее извлекается шар. Какова условная вероятность того, что выбрана 4-я или 5–я урна, если извлеченный шар оказался белым?

Решение: Обозначим через B событие – выбранный шар белый, H_1 –шар выбран из 1-й , 2-й или 3-й урны, через H_2 –шар выбран из 4-й или 5-й урны. Нужно определить $P(H_2|B)$. Определяем вероятности: $P(H_1)=3/5$, $P(H_2)=2/5$, $P(B|H_1)=2/5$, $P(B|H_2)=1/2$. По формулам Байеса находим

$$P(H_2\mid B)=\frac{2/5\cdot 1/2}{3/5\cdot 2/5+2/5\cdot 1/2}=\frac{5}{11}.$$

Занятие 7. *Понятие схемы Бернулли. Формула Бернулли. Локальная и интегральная формулы Муавра-Лапласа в схеме Бернулли*

Рассмотрим серию из n испытаний, в каждом из которых событие A появляется с одной и той же вероятностью p, причем результат каждого испытания не зависит от результатов остальных. Подобная постановка задачи называется **схемой повторения испытаний (схемой Бернулли)**. Найдем вероятность того, что в такой серии событие A произойдет ровно к раз (неважно, в какой последовательности). Интересующее нас событие представляет собой сумму равновероятных несовместных событий, заключающихся в том, что A произошло в некоторых к испытаниях и не произошло в остальных n – к испытаниях. Число таких событий равно числу сочетаний из n по к, то есть $C_n^к$, а вероятность каждого из них: $p^k q^{n-k}$, где q = 1 – p – вероятность того, что в данном опыте A не произошло. Применяя теорему сложения для несовместных событий, получим **формулу Бернулли**:

$$p_n(k)=C_n^k\cdot p^k\cdot q^{n-k}.$$

Пример. Для получения приза нужно собрать 5 изделий с особым знаком на этикетке. Найти вероятность того, что придется купить 10 изделий, если этикетки с этим знаком имеют 5% изделий.

Решение: Из постановки задачи следует, что последнее купленное изделие имеет особый знак. Следовательно, из предыдущих девяти эти знаки имели 4 изделия. Найдем вероятность этого по формуле Бернулли: $p_9(4)=C_9^4\cdot(0,05)^4\cdot(0,95)^5=0,0006092.$ Тогда

p = 0,0006092·0,05 = 0,0000304.

Локальная теорема Муавра – Лапласа

Теорема. Пусть в n независимых испытаниях, вероятность появления события A постоянна и равна p (0 < p < 1), тогда имеет место асимптотическая оценка

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{npq}}\varphi(x) \qquad x = \frac{k-np}{\sqrt{npq}} \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-x^2/2\right) \tag{14}$$

Доказательство теоремы сразу следует из центральной предельной теоремы, которая рассматривается в части 3 (п. 3.2).

Справедливость формулы (14) проиллюстрирована на рис. 1.

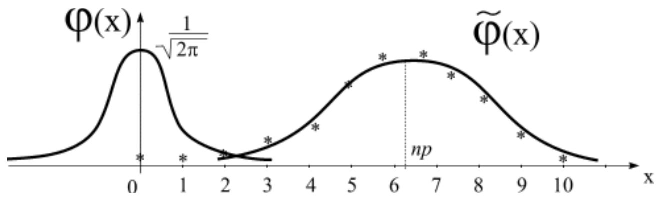


Рис. 1. Справедливость формулы.

Изобразим координаты (k, $P_n(k)$) звездочками. Функцию $P_n(k)$ аргумента k, можно приблизить, в соответствии с формулой (14):

$$\tilde{\varphi}(x) = \frac{1}{\sqrt{2\pi npq}}\exp\left(-(k-np)^2/(2npq)\right),$$

где np – координата центра тяжести (среднее значение), а $\sqrt{npq}$ характеризует меру «сжатости» около центра np.

Делая замену $x = (k-np)/\sqrt{npq}$, мы преобразуем произвольную функцию $\tilde{\varphi}(x)$ к стандартной φ(x), у которой координата центра тяжести np = 0, а $\sqrt{npq} = 1$. Из рисунка видно, что при $n \to \infty$, $P_n(k) \to 0$ (при этом всегда $\sum_{k=0}^{n} P_n(k) = 1$) ошибка уменьшается. Для удобства вычислений, функция φ(x) табулирована (см. приложение, табл. 3). Сама функция называется **кривой Гаусса** [5]. Функция φ(x) – четная, $\varphi(x) \to 0$ при $|x| \to \infty$, $\varphi(x) < 10^{-4}$, при $|x| > 5$, $\max\limits_x\{\varphi(x)\} = 1/\sqrt{2\pi} \approx 0{,}3989$.

Для практических приложений (при n > 10, $p \to \frac{1}{2}$) используют формулу

$$C_n^k \cdot p^k \cdot q^{n-k} \approx \frac{1}{\sqrt{npq}} \cdot \varphi\left(\frac{k-np}{\sqrt{npq}}\right).$$

Пример. Решить пример п 1.5, а).

Решение: Имеем

$$P_{5000}(50) = C_{5000}^{50} \cdot (0{,}01)^{50} \cdot (0{,}99)^{4950} \approx \frac{1}{\sqrt{npq}} \cdot \varphi\left(\frac{k-np}{\sqrt{npq}}\right),$$

k = 50, np = 50, $\sqrt{npq} = \sqrt{50 \cdot 0{,}99} \approx 7{,}04$.

Итак,

$$P_{5000}(50) \approx \frac{1}{7{,}04}\varphi(0) = \left\{ \begin{array}{c} \text{по табл. 3} \\ \text{для } \varphi(x) \end{array} \right\} = \frac{0{,}3989}{7{,}04} \approx 0{,}057.$$

Интегральная теорема Муавра-Лапласа

Теорема. Пусть в n независимых испытаниях вероятность появления события А в каждом испытании постоянна и равна p, 0< p < 1, тогда, для любых $-\infty < a < b < \infty$, равномерно относительно a, b, при $n \to \infty$, имеет место асимптотическая оценка

$$P\{k_1 \le k \le k_2\} = \sum_{k=k_1}^{k_2} C_n^k \cdot p^k \cdot q^{n-k} \xrightarrow[n\to\infty]{} \int_{\grave{a}}^{b} \varphi(x)dx, \qquad (*)$$

где φ(x) - кривая Гаусса, $a = \frac{k_1 - np}{\sqrt{npq}}$, $b = \frac{k_2 - np}{\sqrt{npq}}$.

Функция $\varPhi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ называется функцией Лапласа.

Так как $P_n\{0 \le k \le n\} = 1$ для любого n , то из (*) должно следовать, что $\varPhi(+\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1$.

В самом деле, положим $\Im = \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt$, тогда

$$\Im^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right)^2 = 4\int_0^{\infty} e^{-\frac{t^2}{2}} dt \cdot \int_0^{\infty} e^{-\frac{x^2}{2}} dx = 4\int_0^{\infty}\int_0^{\infty} e^{-\frac{t^2+x^2}{2}} dtdx.$$

Введем полярные координаты:

$t = r\cos\varphi$, $x = r\sin\varphi$, $\varphi \in [0, \pi/2]$, $r \ge 0$, $dtdx = rdrd\varphi$.

Отсюда

$\Im^2 = 4\int_0^{\frac{\pi}{2}}\int_0^{\infty} re^{-\frac{r^2}{2}} drd\varphi = 4\int_0^{\frac{\pi}{2}} d\varphi \cdot \int_0^{\infty} e^{-\frac{r^2}{2}} rdr = 2\pi \cdot 1 = 2\pi$, $\Im = \sqrt{2\pi}$ - интеграл Пуассона.

Следовательно, $\varPhi(+\infty) = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1$.

Для практических приложений вместо (*) используют формулу:

$$P\{k_1 \le k \le k_2\} \approx \Phi(в) - \Phi(a),$$

где

$a = \frac{k_1 - np}{\sqrt{npq}}$, $b = \frac{k_2 - np}{\sqrt{npq}}$.

Учитывая, что $\Phi(+\infty) = 1$, легко получить $\Phi(x) + \Phi(-x) = 1$.

При $x > 0$, $\varPhi(x) = \int_{-\infty}^{x} \varphi(t)dt$, $\varPhi(-x) = \int_{-\infty}^{-x} \varphi(t)dt = \left\{ \begin{matrix} \text{замена} \\ y = -x, \quad dx = -dy \end{matrix} \right\} = -\int_{+\infty}^{x} = \int_{x}^{\infty}$.

Отсюда

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

Функция $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ - табулирована.

Таблица составлена для x < 0, а для x > 0, значения находятся по формуле
Ф (x) + Ф (-x) = 1.

Пример. Решить пример п 1.5, б).

Решение: Имеем

$$P_{5000}\{0 \le k \le 50\} = \sum_{k=0}^{50} C_{5000}^{k} (0,01)^k \cdot (0,99)^{5000-k} \approx \Phi(b) - \Phi(a),\ a = \frac{0-50}{7,07} \approx -7,\ b = \frac{50-50}{7,07} = 0.$$

По табл. 5 приложения находим

$\Phi(0) - \Phi(-7) = 0,5 - 0,0 = 0,5$.

Отсюда $P_{5000}\{0 \le k \le 50\} \approx 0,5$.

Сравнивая решение задачи п.1.5. а), б), можно предположить, что, так как $k = 50$ – наивероятнейшее число, с большой вероятностью реализуется событие { 40 ≤ k ≤ 60}, с центром в точке k_0:

$P_{5000}\{40 \le k \le 60\} \approx 0,037 \cdot 20 \approx 0,7$.

Заметим, что $\sqrt{npq}$ характеризует средние отклонения от среднего значения пр (чем меньше $\sqrt{npq}$, тем «круче» кривая Гаусса в точке симметрии)

Занятие 8. *Понятие случайной величины. Понятие дискретной случайной величины (ДСВ). Методика записи распределения функции от двух независимых ДСВ.*

В этом разделе теории вероятностей мы познакомимся с числовыми оценками, соответствующими исходам испытаний, например таким, как подбрасывание кости. Отсюда исходы испытаний, определяемые случаем, – случайные величины (СВ). Определим случайную величину следующим образом.

Случайная величина – это величина, которая в результате эксперимента (опыта, испытания) принимает одно из своих возможных значений, причем заранее неизвестно, какое именно. **Примеры случайных величин:**

- число дефектных деталей в партии при контроле качества;
- процент завершенного строительства жилого дома спустя 6 месяцев;
- число клиентов операционного отдела банка в течение рабочего дня;
- число продаж автомобилей в течение месяца.

Случайные величины обозначаются заглавными латинскими буквами: X, Y, Z и т. п. Строчные буквы используются для обозначения определенных значений случайной величины. Например, случайная величина X принимает значения $x_1, x_2, ..., x_n$. Различают случайные, дискретные и непрерывные величины.

Дискретной (прерывной) случайной величиной называют случайную величину, которая принимает конечное или бесконечное (но счетное) число отдельных, изолированных возможных значений с определенными вероятностями. Число студентов на лекции – дискретная случайная величина.

Совокупность значений может быть задана таблицей, функцией или графиком. Соотношение, устанавливающее связь между отдельными возможными значениями случайной величины и соответствующими им вероятностями, называется **законом распределения дискретной случайной величины**.

Простейшей формой закона распределения для дискретных случайных величин является ряд распределений.

Рядом распределения дискретной случайной величины X называется таблица, в которой перечислены возможные (различные) значения этой случайной величины x_1, x_2, ..., $x_п$ с соответствующими им вероятностями p_1, p_2, ..., p_n.

x_i	x_1	x_2	…	$x_п$
p_i	p_1	p_2	…	$p_п$

Таким образом, случайная величина X в результате испытания может принять одно из возможных значений x_1, x_2, ..., $x_п$ с вероятностями

$P\ (X = x_1) = p_1;\ P(X = x_2) = p_2;\ P(X = x_п) = p_n$.

Можно использовать более короткую запись: P(x) = P(5) = 0,2. Так как события $(X = x_1)$, $(X = x_2)$, …, $(X = x_п)$ составляют полную группу событий, то сумма вероятностей p_1, p_2, ..., p_n равна единице:

$$\sum_{i=1}^{n} p_i = 1 .$$

Ряд распределения случайной дискретной величины должен удовлетворять следующим условиям:

$$P(x) \geq 0 ,$$

$$\sum_{i=1}^{n} P(x_i) = 1 .$$

Пример. Каждый день местная газета получает заказы на новые рекламные объявления, которые будут напечатаны в завтрашнем номере. Число рекламных объявлений в газете зависит от многих факторов: дня недели, сезона, общего состояния экономики, активности местного бизнеса и т. д. Пусть X – число новых рекламных объявлений, напечатанных в местной газете в определенный день. X – случайная величина, которая может быть только целым числом. В нашем примере случайная величина X принимает значения 0; 1; 2; 3; 4; 5 с вероятностями 0,1; 0,2; 0,3; 0,2; 0,1; 0,1 соответственно (табл. 1).

Т а б л и ц а 1 . Ряд распределения случайной величины X

x_i	0	1	2	3	4	5
$P(x_i)$	0	0	0	0	0	0

Поскольку появления различных значений случайной величины X – несовместные события, то вероятность того, что в газету будут помещены или 2 или 3 рекламных объявления, равна сумме вероятностей P(2) + P(3) = 0,3 + 0,2 = 0,5. Вероятность же того, что их число будет находиться в пределах от 1 до 4 (включая 1 и 4), равна 0,8, т. е. $P(1 \leq X \leq 4) = 0,8$; а $P(X = 0) = 0,1$. Ряд распределения можно изобразить графически. Для этого по оси абсцисс откладывают возможные значения случайной величины, а по оси ординат – соответствующие им вероятности. Если точки $(x_i,\ p_i)$ соединить отрезками прямых, то полученная ломаная линия есть **многоугольник** (или **полигон**) **распределения**.

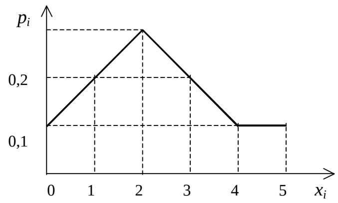


Рис. 1. Полигон распределения

Пример. В книжном магазине организована лотерея. Разыгрываются две книги стоимостью по 10 руб. и одна – стоимостью в 30 руб. Составить закон распределения случайной величины X – суммы чистого (возможного) выигрыша для того, кто приобрел один билет за 1 руб., если всего продано 50 билетов.

Решение: Случайная величина X может принимать три значения: 1 руб. (если владелец билета не выиграет, а фактически проиграет 1 руб., уплаченный им за билет); 9

руб.; 29 руб. (фактический выигрыш уменьшается на стоимость билета – 1 руб.). Первому результату благоприятствуют 47 исходов из 50, второму – два, а третьему – один. Поэтому их вероятности таковы: P(X = –1) = 47/50 = 0,94; P(X =9) = 2/50 = 0,04; P(X = 29) = 1/50 = 0,02;

Закон распределения случайной величины X имеет вид:

Сумма выигрыша,	–1	9	29
ВЕРОЯТНОСТЬ, Р	0,9	0,0	0,0

Контроль: $\sum_{i=1}^{n} p_i$ = 0,94 + 0,04 + 0,02 = 1.

Числовые характеристики ДСВ

Основной характеристикой случайной величины является математическое ожидание.

Пусть случайная величина X принимает значения x_k, k= 1,2,… с вероятностями p_k. Математическое ожидание (или среднее значение) дискретной случайной величины обозначается MX и равняется сумме числового ряда $\sum_k x_k p_k$, если ряд сходится абсолютно.

Пример. Средний выигрыш в примере 1 составляет:

MX= 100*(1/3)+200*(4/15)+300*(1/5)+400*(2/15)+500*(1/15)=233,(3).

Свойства математических ожиданий:

1) для любой постоянной величины C: MC=C;
2) для любой постоянной a: M(aX)=a*MX;
3) для любых случайных величин X и Y, имеющих математические ожидания MX и MY: M(X+Y)=MX+MY;
4) если случайные величины X(ω) и Y(ω) таковы, что X(ω) ≤ Y(ω) для всех $\omega\in\Omega$, то MX ≤ MY;
5) $|MX| \le M|X|$.

Все свойства математических ожиданий вытекают из свойств абсолютно-сходящихся числовых рядов.

Еще одна характеристика случайных величин – дисперсия. Дисперсия случайной величины X обозначается DX и равняется $M(X–MX)^2$.

Дисперсия - это средний квадрат отклонения значений случайной величины от ее математического ожидания.

Из определения дисперсии сразу следуют ее свойства:

1) для любой постоянной величины C: DC=0;
2) для любой постоянной a: $D(aX)=a^2*D(X)$.

Утверждение. Пусть X – случайная величина, MX – ее математическое ожидание, а MX^2 – математическое ожидание случайной величины X^2. Тогда

$$DX = MX^2 - (MX)^2 .$$

Доказательство.

$$M(X - MX)^2 = M\left(X^2 - 2X \cdot MX + (MX)^2\right) = MX^2 - 2(MX)^2 + (MX)^2 = MX^2 - (MX)^2$$

Наряду с дисперсией рассматривают среднее квадратическое отклонение

$$\sigma_X = \sqrt{DX} .$$

. **Пример.** Дисперсия выигрыша в рулетку $DX= MX^2-(MX)^2$.

$MX^2= 100^2*1/3+200^2*4/15+300^2*1/5+400^2*2/15+500^2*1/15=70000$; DX = 15555,(5).

$$\sigma_X = \sqrt{DX} \approx 126 .$$

Занятие 9. *Биноминальное распределение ДСВ. Понятие геометрического распределения.*

Рассмотрим наиболее часто встречающиеся распределения дискретных случайных величин. Одно из них – биномиальное.

Пусть проводится серия из n одинаковых и независимых между собой испытаний. В каждом из них событие A может наступить с положительной вероятностью p. Такие испытания называются испытаниями Бернулли.

Событие A будем называть «успехом», а событие $\overline{A}$ – «неудачей». $P(\overline{A}) = q = 1 - p$.

Рассмотрим случайную величину X – число успехов в n испытаниях. Она может принимать значения 0, 1, 2,…, n. Вероятность, что X примет значение k, т.е. в n испытаниях k раз наступит успех $P\{X = k\} = C_n^k p^k (1-p)^{n-k}$. Действительно, вероятность наступления k успехов в k фиксированных испытаниях и (n – k) неудач в остальных (n – k) испытаниях равна $p^k (1-p)^{n-k}$. Распределить k успехов среди n испытаний можно C_n^k способами.

Распределение случайной величины X называется распределением Бернулли или биномиальным распределением.

Пример. Монету подбрасывают 10 раз. Какова вероятность, что герб выпадет 4 раза?

Решение: При каждом подбрасывании «успех» – выпадение герба, n = 10, k = 4, p = 1/2. $P(X = 4) = C_{10}^4 \cdot (1/2)^4 \cdot (1/2)^6 = 210/2^{10} \approx 0{,}206$.

Биномиально распределенная случайная величина X – это целочисленная величина. Введем для нее производящую функцию.

$$\Pi(z) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} z^k = (pz+q)^n, \quad \text{т.к. } (a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k} \text{ (бином Ньютона)}$$

Математическое ожидание $MX = \Pi'(1) = n \cdot p$.

Дисперсия $DX = \Pi''(1) + \Pi'(1) - (\Pi'(1))^2 = n \cdot p \cdot q$.

Пример. Среднее количество выпадений герба при 10 подбрасываниях монеты равно MX = np = 10*(1/2) = 5, дисперсия равна DX = npq = 5*(1/2) = 5/2. Пусть теперь испытания Бернулли проводятся до наступления первой неудачи. Случайная величина X – число проведенных испытаний. Распределение X можно задать с помощью таблицы.

X	1	2	...	k	...
P	q	pq	...	$p^{k-1}q$	...

$P(X = k) = p^{k-1} * q$, k = 1, 2, 3,…

Такое распределение называется геометрическим.

Пример. Вероятность закатить хотя бы один шар в лузу при одном ударе бильярдиста постоянна и рана 0,7. Если при ударе закатить шар не удается, право удара переходит к другому игроку. Какова вероятность, что бильярдист сделает не менее 4 ударов?

Пусть X – число ударов, сделанных игроком.[Найдем вероятность дополнительного события. $P(X< 4) = 0{,}3+0{,}7*0{,}3+(0{,}7)^2*0{,}3 = 0{,}657$. Тогда $P(X \geq 4) = 1– 0{,}657 = 0{,}343$.

Производящая функция случайной величины с геометрическим распределением $\Pi(z) = \sum_{k=1}^{\infty} p^{k-1} q \cdot z^k = \frac{qz}{(1-pz)}$. Математическое ожидание $MX = \Pi'(1) = 1/q$. Дисперсия $DX = \Pi''(1) + \Pi'(1) - (\Pi'(1))^2 = p/q^2$.

Пример. Среднее число ударов бильярдиста MX=1/q=1/0,3=10/3=3,(3). Дисперсия числа ударов $DX= p/q^2 = 0{,}7/(0{,}3)^2 = 70/9 = 7{,}(7)$.

Занятие 10. *Понятие непрерывной случайной величины (НСВ). Формула вычисления вероятностей.*

Непрерывной случайной величиной называют случайную величину, которая может принимать любые значения на числовом интервале.

Примеры непрерывных случайных величин: возраст студентов, длина ступни ноги человека, масса детали и т. д. Это положение относится ко всем случайным величинам, измеряемым на непрерывной шкале, таким как меры веса, длины, времени, температуры, расстояния. Измерение может быть проведено с точностью до какого-нибудь десятичного знака, но случайная величина – теоретически непрерывная величина. В экономическом анализе находят широкое применение относительные величины, различные индексы экономического состояния, которые также вычисляются с определенной точностью, скажем, до двух знаков после запятой, хотя теоретически их значения – непрерывные случайные величины.

У непрерывной случайной величины возможные значения заполняют некоторый интервал (или сегмент) с конечными или бесконечными границами.

Закон распределения непрерывной случайной величины можно задать в виде **интегральной функции распределения**, являющейся наиболее общей формой задания закона распределения случайной величины, а также в виде **дифференциальной функции** (плотности распределения вероятностей), которая используется для описания распределения вероятностей только непрерывной случайной величины.

Функция распределения (или интегральная функция) $F(x)$ – универсальная форма задания закона распределения случайной величины. Для непрерывной случайной величины функция распределения также определяет вероятность того, что случайная величина X примет значение, меньшее фиксированного действительного числа x, т. е.

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

При изменении x меняются вероятности $P(X < x) = F(x)$. Поэтому $F(x)$ и рассматривают как функцию переменной величины. Принято считать, что случайная величина X известна, если известна ее функция распределения $F(x)$.

Теперь можно дать более точное определение непрерывной случайной величины: **случайную величину называют непрерывной, если ее функция распределения есть непрерывная, кусочно-дифференцируемая функция с непрерывной производной**.

1. Функция распределения есть неотрицательная функция, заключенная между 0 и 1, т.е. $0 \le F(x) \le 1$.

2. Функция распределения есть неубывающая функция, т. е. $F(x_2) \ge F(x_1)$, если $x_2 > x_1$. Тогда $P(x_1 \le X < x_2) = P(X < x_2) – P(X < x_1) = F(x_2) – F(x_1)$.

Так как любая вероятность есть число неотрицательное, то $P(x_1 \le \le X < x_2) \ge 0$, а следовательно, $F(x_2) – F(x_1) \ge 0$ и $F(x_2) \ge F(x_1)$.

Следствие 1. Вероятность того, что случайная величина X примет значение, заключенное в интервале (α, β), равна приращению функции распределения на этом интервале, т. е.

$P(\alpha \le X < \beta) = F(\beta) – F(\alpha)$.

Следствие 2. Вероятность того, что непрерывная случайная величина X примет одно определенное значение, равна нулю.

$P(X = x_1) = 0$.

Согласно сказанному, равенство нулю вероятности $P(X = x_1)$ не всегда означает, что событие $X = x_1$ невозможно. Говоря о вероятности события $X = x_1$, априорно пытаются угадать, какое значение примет случайная величина в опыте.

Если x_1 лежит в области возможных значений непрерывной случайной величины X, то с некоторой уверенностью можно предсказать область, в которую случайная величина может попасть. В то же время невозможно хотя бы с малейшей степенью уверенности угадать, какое конкретное значение из бесконечного числа возможных примет непрерывная случайная величина.

Например, если метеослужба объявляет, что температура воздуха в полдень составила 5 °С, то это не означает, что температура будет точно равна этому значению. Вероятность такого события равна нулю.

3. Если все возможные значения случайной величины принадлежат интервалу (α, β), то

F(x) = 0 при x ≤ α; F(x) = 1 при x > β.

В самом деле, F(x) = 0 для всех значений x ≤ α и F(x) = 1 при x > β, поскольку события X < x для любого значения x ≤ α, являются в этом случае невозможными, а для любого значения x > β – достоверными.

Следствие. Если возможные значения непрерывной случайной личины расположены на всей оси OX, то справедливы следующие предельные соотношения:

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

или F(– ∞) = 0; F(+ ∞) = 1. Это следствие справедливо и для дискретных случайных величин.

Функция плотности НСВ и интегральные функции распределения НСВ. Методика расчёта вероятностей для НСВ.

Плотностью распределения вероятностей непрерывной случайной величины X называется функция W(x), равная первой производной от функции распределения F(x),

W(x) = F ′(x),

где W(x) – дифференциальная функция распределения. Дифференциальная функция применяется только для описания распределения вероятностей непрерывных случайных величин.

Вероятность попадания непрерывной случайной величины в заданный интервал.

Вероятность того, что непрерывная случайная величина примет значение, принадлежащее интервалу (α, β), равна определенному интегралу от дифференциальной функции, взятому в пределах от α до β,

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} W(x)\,dx.$$

Используя соотношения (5.2) и (5.1), получим P(α ≤ X < β) = P(α < < X < < β) = $\int_{\alpha}^{\beta} W(x)\,dx$.

Геометрически этот результат равен площади криволинейной трапеции, ограниченной осью OX, кривой распределения W(x) и прямыми x = α, x = β.

Зная плотность распределения W(x), можно найти функцию распределения F(x) по формуле

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

В самом деле, так как неравенство X < x можно записать в виде двойного неравенства – ∞ < X < x, то F(x) = P(– ∞ < X < x) = $\int_{-\infty}^{x} W(x)\,dx$ (рис. 1).

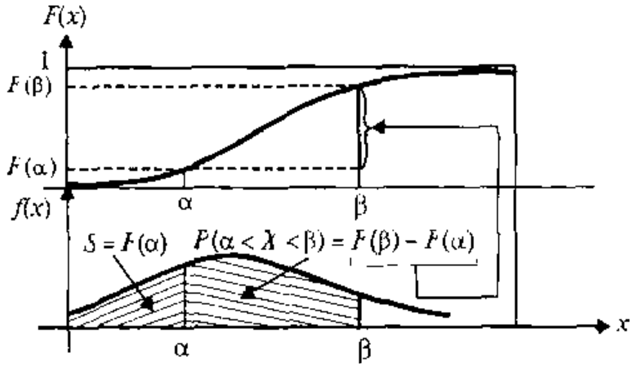


Рис. 1. Связь функции распределения с плотностью распределения вероятностей

Таким образом, для полной характеристики непрерывной случайной величины достаточно задать функцию распределения или плотность ее вероятности.

Свойства дифференциальной функции распределения

1. Дифференциальная функция – неотрицательная функция:

W(x) ≥ 0.

Это следует из того, что F(x) – неубывающая функция, а значит, ее производная неотрицательна.

2. Несобственный интеграл от дифференциальной функции в пределах от – ∞ до + ∞ равен 1

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

Очевидно, интеграл выражает вероятность достоверного события – ∞ < X + ∞.

Занятие 11. Характеристики НСВ. Методика вычисления математического ожидания, дисперсии, среднеквадратического отклонения НСВ по её функции плотности.

Математическое ожидание характеризует среднее ожидаемое значение случайной величины, т.е. приближенно равно ее среднему значению (вероятностный смысл математического ожидания). Иногда знания этой характеристики достаточно для решения задачи. Например, при оценке покупательной способности населения вполне может хватить знания среднего дохода, при анализе выгодности двух видов деятельности можно ограничиться сравнением их средних прибыльностей. Знание того, что выпускники данного университета зарабатывают в среднем больше выпускников другого, может послужить основанием для принятия решения о поступлении в данный ВУЗ и т.п.

Математическое ожидание **дискретной** случайной величины определяется соотношением:

$$M(X) = M_X = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i \text{, где } \sum_{i=1}^{n} p_i = 1.$$

Математическое ожидание **непрерывной** случайной величины равно

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

где $f(x)$ - плотность вероятности.

Свойства математического ожидания

Прежде чем формулировать свойства математического ожидания необходимо пояснить смысл арифметических операций $X+Y$, $X-Y$, $X \cdot Y$ и т.п., где X и Y – дискретные случайные величины.

Например, под суммой $X+Y$ понимается случайная величина Z, значениями которой являются все допустимые суммы $z_{ij}=x_i+y_j$, где x_i и y_j – все возможные значения соответственно случайных величин X и Y.

Свойства математического ожидания:

1. Математическое ожидание постоянной величины C равно этой величине.

$$M(C)=C\cdot 1=C.$$

2. Математическое ожидание суммы (разности) двух или нескольких случайных величин X и Y равно сумме (разности) их математических ожиданий:

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

Следствие. Если C – постоянная величина, то

$$M(X+C)=M(X)+C$$

3. Математическое ожидание произведения двух **независимых** случайных величин X и Y равно произведению их математических ожиданий:

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

Следствие. Математическое ожидание произведения нескольких **взаимно независимых** случайных величин равно произведению математических ожиданий этих величин.

Следствие. Постоянный множитель можно выносить за знак математического ожидания, т.е. $M(CX)=C\cdot M(X)$.

Дисперсия случайной величины и ее свойства.

На практике часто требуется оценить рассеяние случайной величины вокруг ее среднего значения. Например, акции двух компаний могут приносить в среднем одинаковые дивиденды, однако вложение денег в одну из них может быть гораздо более рискованной операцией, чем в другую. Поэтому возникает необходимость в числовой характеристике, оценивающей разброс возможных значений случайной величины относительно ее среднего значения (математического ожидания). Такой характеристикой является **дисперсия**.

Дисперсией (рассеянием) случайной величины X называют математическое ожидание квадрата отклонения этой величины от ее математического ожидания.

$$D(X)=M\left((X-M(X))^2\right).$$

Легко показать, что вышеприведенное выражение может быть записано в виде

$$D(X)=M(X^2)-\left(M(X)\right)^2$$

Действительно, используя основные теоремы о математическом ожидании, получим

$$D(X)=M\left((X-M(X))^2\right)=M\left(X^2-2XM(X)+M(X)^2\right)=$$
$$=M(X^2)-2M(X)M(X)+\left(M(X)\right)^2=M(X^2)-\left(M(X)\right)^2$$

В случае **дискретной** случайной величины, имеющей закон распределения $(x_i,p_i)\quad(i=1,2,...,n)$

$$D(X)=\sum_{i=1}^{n}(x_i-M(X))^2\,p_i\cdot$$

Для **непрерывной** случайной величины формула для расчета дисперсии имеет вид

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

Свойства дисперсии

1. Дисперсия постоянной величины равна нулю.

$$D(C)=0$$

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

$$D(CX)=C^2D(X).$$

3. Дисперсия суммы (разности) двух независимых случайных величин равна сумме дисперсий этих величин:

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

Следствие 1. Дисперсия суммы нескольких взаимно независимых случайных величин равна сумме дисперсий этих величин.

Следствие 2. Если C – постоянная величина, то $D(X+C)=D(X)$.

Математическое ожидание и дисперсия случайной величины являются ее основными числовыми характеристиками.

Пример. Пусть закон распределения дискретной случайной величины имеет вид

,07	,21	,55	,16	,01

Найти математическое ожидание и дисперсию случайной величины X.

Решение: Рассчитаем вначале математическое ожидание

$$M(X)=1\cdot 0.07+2\cdot 0.21+3\cdot 0.55+4\cdot 0.16+5\cdot 0.01=2.83$$

Дисперсия равна

$$D(X)=(1-2.83)^2\cdot 0.07+(2-2.83)^2\cdot 0.21+(3-2.83)^2\cdot 0.55+(4-2.83)^2\cdot 0.16+$$

$$(5-2.83)^2\cdot 0.01=0{,}661$$

Пример. Плотность вероятности непрерывной случайной величины равна

$$f(x)=\exp(-x),\ \text{где}\quad x>0$$

Найти ее математическое ожидание и дисперсию.

Решение: Найдем математическое ожидание:

$$M(X)=\int_{-\infty}^{\infty} x\,f(x)\,dx=\int_0^{\infty} xe^{-x}\,dx=\begin{bmatrix} U=x & dU=dx \\ dV=e^{-x}dx & V=-e^{-x}\end{bmatrix}=[-e^{-x}\cdot x]\Big|+\int_0^{\infty} e^{-x}dx=1$$

Далее,

$$M(X^2)=\int_0^{\infty} x^2\cdot e^{-x}\,dx=\begin{bmatrix} U=x^2 & dU=2x\,dx \\ dV=e^{-x} & V=-e^{-x}\end{bmatrix}=[-e^{-x}\cdot x^2]\Big|+2\int_0^{\infty} e^{-x}\,dx=2$$

Найдем дисперсию, используя формулу

$$D(X)=M(X^2)-(M(X))^2=2-1=1.$$

Среднее квадратическое отклонение.

Для оценки рассеяния возможных значений случайной величины вокруг ее среднего значения кроме дисперсии служат и некоторые другие характеристики. К их числу относится среднее квадратическое отклонение.

Средним квадратическим отклонением σ (или стандартом) случайной величины X называется корень квадратный из дисперсии $D(X)$ этой величины:

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

Легко показать, что дисперсия имеет размерность, равную квадрату размерности случайной величины. Поэтому размерность $\sigma(X)$ совпадает с размерностью X. В тех случаях, когда желательно, чтобы оценка рассеяния имела размерность случайной величины, вычисляют среднее квадратичное отклонение, а не дисперсию.

Понятие дисперсии и среднего квадратического отклонения широко используется практически во всех областях человеческой деятельности, связанных с процессами измерений. Так, например, в технике, они характеризуют точность измерительной аппаратуры (чем выше среднеквадратическое отклонение (разброс) при измерениях, тем хуже качество прибора).

Примерами использования данных параметров в экономике могут служить изучение риска различных действий со случайным исходом, в частности, при анализе риска инвестирования в ту или иную отрасль, при оценивании различных активов в портфеле ценных бумаг и т.д.

Пример.

Пусть имеется два варианта инвестирования со следующими характеристиками

Вероятности возможной чистой прибыли								
	Сравнение вариантов решений							
Чистая прибыль, млн.д.е.	-3	-	-1	0	1	2	3	4
Вероятности:								
Инвестиция 1	0	0	0,1	0,2	0,3	0,2	0,2	0
Инвестиция 2	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2

Ожидаемая чистая прибыль инвестирования определяется математическим ожиданием и составляет:

Инвестиция 1: $M(X)=\sum x_i \cdot p_i = 1.2$ *млн.*

Инвестиция 2: $M(X)=\sum x_i \cdot p_i = 1.1$ *млн.*

По ожидаемой прибыли предпочтительнее 1-й вариант. Однако мы не учли риск, связанный с инвестициями. Этот риск может быть определен с помощью дисперсии и (или) среднего квадратического отклонения. Используя результаты таблицы, получим

Инвестиция 1: $\sigma(X)=\sqrt{D(X)}=1.25$

Инвестиция 2: $\sigma(X)=\sqrt{D(X)}=2.385$

Т.е. риск по варианту для инвестиции 1 меньше. Выбор – за ЛПР.

Занятие 12. *Сущность выборочного метода. Генеральная совокупность и выборка.*

Предметом математической статистики является изучение случайных событий и случайных величин по результатам наблюдений. Совокупность предметов или явлений, объединенных каким-либо общим признаком, называется **статистической совокупностью**. Результатом наблюдений над статистической совокупностью являются **статистические данные –** сведения о том, какие значения принял в итоге наблюдений интересующий нас признак (случайная величина X).

Обработка статистических данных методами математической статистики приводит к установлению определенных закономерностей, присущих массовым явлениям. При этом **точность** статистических выводов повышается с ростом числа наблюдений.

Статистические данные, как правило, представляют собой ряд значений $\{x_1, x_2, ..., x_n\}$ некоторой случайной величины. Обработка этого ряда значений представляет собой первый этап исследования случайной величины.

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

Второй задачей математической статистики является разработка **методов анализа** статистических данных в зависимости от целей исследования. К этой задаче относятся:

- Оценка неизвестной **вероятности события**; оценка неизвестной **функции распределения**; оценка **параметров распределения**, вид которого известен; оценка зависимости случайной величины от одной или нескольких случайных величин и т.п.

- Проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен.

В современной математической статистике есть много общего с **наукой о принятии решений в условиях неопределенности**, так как она разрабатывает способы определения числа необходимых испытаний до начала исследования (планирование эксперимента), в процессе исследования (последовательный анализ) и решает многие другие аналогичные задачи.

Выборочный метод и его основные понятия. Случайная выборка, объем выборки.

Пусть требуется изучить совокупность **однородных** объектов относительно некоторого качественного или количественного признака, характеризующего эти объекты. Например, для партии деталей качественным признаком может служить стандартность детали, а количественным – контролируемый размер детали.

В принципе, возможно проведение сплошного обследования, т.е. обследование всех объектов. На практике такое обследование применяется редко, например,

- ✓ из–за большого числа объектов
- ✓ из–за дороговизны проведения операции контроля,
- ✓ из–за того, что контроль часто связан с разрушением объекта (проверка электролампы на долговечность ее работы), и т.д.

В таких случаях случайно отбирается и изучается **ограниченное** число объектов из совокупности.

Выборочной совокупностью или **случайной выборкой** называют совокупность случайно отобранных объектов.

Генеральной совокупностью называют совокупность объектов, из которых производится выборка.

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

Пример. Число единиц товара N, произведенного некоторым предприятием в течение года, есть генеральная совокупность. Для исследования качества продукции на практике рассматривается выборка, состоящая из n единиц товара. Признаком, или случайной величиной, может быть число единиц товара, удовлетворяющих сертификационным требованиям.

При составлении выборки можно поступать двумя способами: после того как объект отобран и исследован, его можно возвратить или не возвращать в генеральную совокупность. В связи с этим выборки подразделяются на **повторные** и **бесповторные**.

Повторной называют выборку, при которой отобранный объект (перед отбором следующего) возвращается в генеральную совокупность. При **бесповторной** выборке отобранный объект в генеральную совокупность не возвращается.

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, т.е. выборка должна быть репрезентативной (представительной). **Пример** – изучение общественного мнения.

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.

Если объем выборки достаточно велик, а выборка составляет лишь незначительную часть совокупности, то различие между повторной и бесповторной выборкой стирается.

Способы отбора

На практике применяются различные способы отбора, которые можно подразделить на два вида:

- Отбор, не требующий расчленения генеральной совокупности на части. Сюда относятся а) простой случайный бесповторный отбор и б) простой случайный повторный отбор.
- Отбор, при котором генеральная совокупность разбивается на части. Сюда относятся а) **типический отбор**, б) **механический отбор** и в) **серийный отбор**.

Простым случайным называют отбор, при котором объекты извлекаются по одному из генеральной совокупности. Осуществить такой отбор для генеральной совокупности из N объектов можно, например, посредством записи на карточках номеров от 1 до N, последующем перемешиванием карточек и выниманием их наугад. При этом обследованию подлежат объекты, имеющие номера, совпадающие с номерами карточек. Если карточки возвращаются в пачку, то имеем простую случайную повторную выборку, в противном случае – простую бесповторную. При большом объеме генеральной совокупности более рациональным является использование таблиц случайных чисел. Например, чтобы выбрать 50 объектов из пронумерованной генеральной совокупности, открывают любую страницу таблицы случайных чисел и выписывают 50 чисел подряд; в выборку попадают те объекты, номера которых совпадают с выписанными случайными числами. Если случайное число таблицы превосходит число N, такое число пропускают. При проведении бесповторной выборки пропускают также случайные числа, уже встречавшиеся раньше.

Типическим называют отбор, при котором объекты отбираются не из всей генеральной совокупности, а из каждой ее “типической” части. Например, если детали изготовлены на нескольких станках, то отбор производят из продукции каждого станка в отдельности.

Механическим называют отбор, при котором генеральная совокупность механически делится на столько групп, сколько объектов должно войти в выборку, а из каждой группы выбирается один объект. Например, если нужно отобрать 20% изготовленных станком деталей, то отбирают каждую пятую деталь.

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

На практике часто применяют **комбинированный отбор**, при котором сочетаются указанные выше способы. Например, разбивают генеральную совокупность на серии одинакового объема, затем простым случайным отбором выбирают несколько серий и, наконец, из каждой серии простым случайным отбором извлекают отдельные объекты.

Вариационный ряд для дискретных и непрерывных случайных величин.

Пусть из генеральной совокупности извлечена выборка, причем значение исследуемого параметра x_1 наблюдалось n_1 раз, x_2 - n_2 раз и т.д. При этом $\sum_i n_i = n$ объем выборки. Наблюдаемые значения x_i называют **вариантами**, а последовательность вариант, записанных в возрастающем порядке – **вариационным рядом**. Числа наблюдений называют **частотами,** а их отношения к объему выборки n_i / n - **относительными частотами. Статистическим распределением выборки** называют перечень вариант и соответствующих им относительных частот.

Заметим, что в теории вероятностей под распределением понимают соответствие между возможными значениями случайной величины и их **вероятностями,** а в математической статистике – соответствие между наблюдаемыми **вариантами** и их **частотами** или **относительными частотами**.

Приведенный способ представления статистических данных применяют в случае дискретных случайных величин. Для непрерывных случайных величин удобнее разбить отрезок [a,b] возможных значений случайной величины на частичные полуинтервалы $\Delta_k = [a_{k-1}, a_k)$, $(k = 1,...,m)$ (Δ_m замкнут также и справа) с помощью некоторой системы точек $a = a_0 < a_1 < ... < a_m = b$. Часто разбиение [a,b] производят на равные части, тогда $\Delta_k = [a + (k-1)h,\ a + kh),\ k = 1,...,m$, где $h = \dfrac{b-a}{m}$

В качестве частот n_k теперь надо брать количество наблюдаемых значений, попавших на каждый из частичных интервалов Δ_k. Число интервалов k часто выбирают на основании формулы Стерджерса $k = 1 + 1{,}4 \ln n$.

Полигон и гистограмма

Графически статистическое распределение представляется в частности, с помощью **полигона** и **гистограммы**.

Полигоном частот называют ломаную линию, отрезки которой соединяют точки $(x_1; n_1), (x_2; n_2), ...(x_k, n_k)$. Для построения полигона частот на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им частоты n_i и соединяют точки $(x_i; n_i)$ отрезками прямых.

Полигон относительных частот строится аналогично, за исключением того, что на оси ординат откладываются относительные частоты w_i.

В случае непрерывного признака строится **гистограмма**, для чего интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько частичных интервалов длиной h и находят для каждого частичного интервала n_i- сумму частот вариант, попавших в i – й интервал.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которой служат частичные интервалы длиною h, а высоты равны отношению n_i / h. Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии (высоте) n_i / h. Площадь i-го прямоугольника равна $hn_i / h = n_i$ - сумме частот вариант i –го интервала, поэтому площадь гистограммы частот равна сумме всех частот, т.е. **объему выборки**.

В случае гистограммы **относительных** частот по оси ординат откладываются относительные чатоты w_i, на оси абсцисс – частичные интервалы, над ними проводят отрезки, параллельные оси абсцисс на высоте W_i / h. Площадь i-го прямоугольника равна относительной частоте вариант W_i, попавших в i-й интервал. Поэтому площадь гистограммы относительных частот равна сумме всех относительных частот, то есть **единице**.

Занятие 13. *Точечные оценки для генеральной средней (математического ожидания), генеральной дисперсии и генерального среднеквадратического отклонения.*

Генеральным средним называется среднее арифметическое значений признака X в генеральной совокупности (обозначение $\overline{X}_Г$). Выборочным средним называется среднее арифметическое значение признака X в выборочной совокупности:

$$\overline{X}_B = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + ... + x_k \cdot n_k}{n}.$$

Выборочное среднее является оценкой для генерального среднего или является оценкой неизвестного математического ожидания с.в., если выборка получена в результате наблюдения над некоторой случайной величиной.

Генеральной дисперсией называется дисперсия признака X в генеральной совокупности. Выборочной дисперсией называется дисперсия признака X в выборочной совокупности:

$$D_B = \frac{1}{n}\sum_{j=1}^{k}\left(x_j - \overline{X}\right)^2 \cdot n_j.$$

Для вычисления дисперсии используют формулу:

$$D_B = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum_{j=1}^{k} x^2{}_j n_j - \left(\overline{X}\right)^2.$$

Дисперсия равняется среднему квадратов без квадрата среднего. Выборочная дисперсия является оценкой неизвестной генеральной дисперсии, или, ели наблюдается некоторая с.в., то выборочная дисперсия служит оценкой для неизвестной дисперсии. Пример на вычисление выборочного среднего и выборочной дисперсии см. ниже в разделе ”Выборочное уравнение регрессии”.

Во многих случаях мы располагаем информацией о виде закона распределения случайной величины (нормальный, бернуллиевский, равномерный и т. п.), но не знаем параметров этого распределения, таких как Mξ, Dξ. Для определения этих параметров применяется выборочный метод.

Пусть выборка объема n представлена в виде вариационного ряда. Назовем **выборочной средней** величину

$$\overline{x} = \frac{x_1 m_1 + x_2 m_2 + ... + x_k m_k}{n} = x_1\frac{m_1}{n} + x_2\frac{m_2}{n} + ... + \frac{m_k}{n}$$

Величина $\omega_i = \dfrac{m_i}{n}$ называется **относительной частотой** значения признака x_i. Если значения признака, полученные из выборки не группировать и не представлять в виде вариационного ряда, то для вычисления выборочной средней нужно пользоваться формулой

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

Естественно считать величину $\overline{x}$ выборочной оценкой параметра Mξ. Выборочная оценка параметра, представляющая собой число, называется **точечной оценкой**.

Выборочную дисперсию

$$\sigma^2 = \sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 \omega_i = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

можно считать точечной оценкой дисперсии Dξ генеральной совокупности.

Приведем еще один пример точечной оценки. Пусть каждый объект генеральной совокупности характеризуется двумя количественными признаками x и y. Например, деталь может иметь два размера – длину и ширину. Можно в различных районах измерять концентрацию вредных веществ в воздухе и фиксировать количество легочных заболеваний населения в месяц. Можно через равные промежутки времени сопоставлять доходность акций данной корпорации с каким-либо индексом, характеризующим среднюю доходность всего рынка акций. В этом случае генеральная совокупность представляет собой двумерную случайную величину ξ, η. Эта случайная величина принимает значения x, y на множестве объектов генеральной совокупности. Не зная закона совместного распределения случайных величин ξ и η, мы не можем говорить о наличии или глубине корреляционной связи между ними, однако некоторые выводы можно сделать, используя выборочный метод.

Выборку объема n в этом случае представим в виде таблицы, где i-тый отобранный объект (i= 1,2,...n) представлен парой чисел x_i, y_i :

x_1	x_2	...	x_n
y_1	y_2	...	y_n

Выборочный коэффициент корреляции рассчитывается по формуле

$$r_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y}$$

Здесь

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \,;\; \sigma_x = \sqrt{\sigma_x{}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \,;\; \sigma_y = \sqrt{\sigma_y{}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}\,.$$

Выборочный коэффициент корреляции можно рассматривать как точечную оценку коэффициента корреляции $\rho_{\xi\eta}$, характеризующего генеральную совокупность.

Выборочные параметры $\bar{x}, s_x, r_{xy}$ или любые другие зависят от того, какие объекты генеральной совокупности попали в выборку и различаются от выборки к выборке. Поэтому они сами являются случайными величинами.

Пусть выборочный параметр δ рассматривается как выборочная оценка параметра Δ генеральной совокупности и при этом выполняется равенство

$$M\delta = \Delta.$$

Такая выборочная оценка называется **несмещенной.**

Для доказательства несмещённости некоторых точечных оценок будем рассматривать выборку объема n как систему n независимых случайных величин ξ_1, ξ_2,... ξ_n , каждая из которых имеет тот же закон распределения с теми же параметрами, что и случайная величина ξ, представляющая генеральную совокупность. При таком подходе становятся очевидными равенства: $Mx_i = M\xi_i = M\xi$; $Dx_i = D\xi_i = D\xi$ для всех k = 1,2,...n.

Теперь можно показать, что выборочная средняя $\bar{x}$ есть несмещенная оценка средней генеральной совокупности или , что то же самое, математического ожидания интересующей нас случайной величины ξ :

$$M\bar{x} = M\,\frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\left(M\xi_1 + M\xi_2 + ... + M\xi_n\right) = \frac{1}{n} nM\xi = M\xi\,.$$

Выведем формулу для дисперсии выборочной средней:

$$D\bar{x} = D\,\frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n^2}\left(D\xi_1 + D\xi_2 + ... + D\xi_{n1}\right) = \frac{1}{n^2} nD\xi = \frac{D\xi}{n}\,.$$

Найдем теперь, чему равно математическое ожидание выборочной дисперсии σ^2. Сначала преобразуем σ^2 следующим образом:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - M\xi + M\xi - \bar{x}\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\left(x_i - M\xi\right)^2 - 2\left(x_i - M\xi\right)\left(\bar{x} - M\xi\right) + \left(\bar{x} - M\xi\right)^2\right) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - M\xi\right)^2 - \left(\bar{x} - M\xi\right)^2$$

Здесь использовано преобразование:

$$\sum_{i=1}^{n} 2\left(x_i - M\xi\right)\left(\bar{x} - M\xi\right) = 2\left(\bar{x} - M\xi\right)\sum_{i=1}^{n}\left(x_i - M\xi\right) =$$

$$= 2(\overline{x} - M\xi)\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} M\xi\right) = 2(\overline{x} - M\xi)(n\overline{x} - nM\xi) = 2n(\overline{x} - M\xi)^2$$

Теперь, используя полученное выше выражение для величины σ^2, найдем ее математическое ожидание.

$$M\sigma^2 = M\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - M\xi\right)^2 - (\overline{x} - M\xi)^2\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} M\left(x_i - M\xi\right)^2 - M\left(\overline{x} - M\xi\right)^2 = \frac{1}{n} nD\xi - D\overline{x} = D\xi - \frac{D\xi}{n} = \frac{n-1}{n} D\xi.$$

Так как $M\sigma^2 \neq D\xi$, **выборочная дисперсия не является несмещенной оценкой дисперсии генеральной совокупности**.

Чтобы получить несмещенную оценку дисперсии генеральной совокупности, нужно умножить выборочную дисперсию на $\frac{n}{n-1}$. Тогда получится величина $s^2 = \frac{n}{n-1}\sigma^2$, называемая **исправленной выборочной дисперсией**.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

Пусть имеется ряд несмещенных точечных оценок одного и того же параметра генеральной совокупности. Та оценка, которая имеет наименьшую дисперсию, называется **эффективной**.

Полученная из выборки объема n точечная оценка δ_n параметра Δ генеральной совокупности называется **состоятельной**, если она сходится по вероятности к Δ. Это означает, что для любых положительных чисел ε и γ найдется такое число $n_{\varepsilon\gamma}$, что для всех чисел n, удовлетворяющих неравенству $n > n_{\varepsilon\gamma}$ выполняется условие $P(|\delta_n - \Delta| < \varepsilon) > 1 - \gamma$. $\overline{x}$ и s^2 являются несмещёнными, состоятельными и эффективными оценками величин $M\xi$ и $D\xi$.

Занятие 14. *Понятие интервальной оценки. Надежность доверительного интервала.*

Точечные оценки параметров генеральной совокупности могут быть приняты в качестве ориентировочных, первоначальных результатов обработки выборочных данных. Их недостаток заключается в том, что неизвестно, с какой точностью оценивается параметр. Если для выборок большого объема точность обычно бывает достаточной (при условии несмещенности, эффективности и состоятельности оценок), то для выборок небольшого объема вопрос точности оценок становится очень важным.

Введем понятие интервальной оценки неизвестного параметра генеральной совокупности (или случайной величины ξ, определенной на множестве объектов этой генеральной совокупности). Обозначим этот параметр через Δ. По сделанной выборке по определенным правилам найдем числа Δ_1 и Δ_2, так чтобы выполнялось условие:

$$P(\Delta_1 < \Delta < \Delta_2) = P(\Delta \in (\Delta_1; \Delta_2)) = \gamma$$

Числа Δ_1 и Δ_2 называются **доверительными границами**, интервал (Δ_1, Δ_2) — **доверительным интервалом** для параметра Δ. Число γ называется **доверительной вероятностью** или **надежностью** сделанной оценки.

Сначала задается надежность. Обычно ее выбирают равной 0.95, 0.99 или 0.999. Тогда вероятность того, что интересующий нас параметр попал в интервал (Δ_1, Δ_2) достаточно высока. Число ($\Delta_1 + \Delta_2$) / 2 – середина доверительного интервала – будет давать значение параметра Δ с **точностью** ($\Delta_2 - \Delta_1$) / 2, которая представляет собой половину длины доверительного интервала.

Границы Δ_1 и Δ_2 определяются из выборочных данных и являются функциями от случайных величин x_1, x_2,..., x_n , а следовательно – сами случайные величины. Отсюда – доверительный интервал (Δ_1, Δ_2) тоже случаен. Он может покрывать параметр Δ или нет. Именно в таком смысле нужно понимать случайное событие, заключающееся в том, что доверительный интервал покрывает число Δ.

Доверительный интервал для математического ожидания нормального распределения при известной дисперсии.

Пусть случайная величина ξ (можно говорить о генеральной совокупности) распределена по нормальному закону, для которого известна дисперсия $D\xi = \sigma^2$ ($\sigma > 0$). Из генеральной совокупности (на множестве объектов которой определена случайная величина) делается выборка объема n. Выборка x_1, x_2,..., x_n рассматривается как совокупность n независимых случайных величин, распределенных так же как ξ (подход, которому дано объяснение выше по тексту).

Ранее также обсуждались и доказаны следующие равенства:

$Mx_1 = Mx_2 = ... = Mx_n = M\xi$;

$Dx_1 = Dx_2 = ... = Dx_n = D\xi$;

$M\,\bar{x} = M\xi$;

$D\,\bar{x} = D\xi/n$;

Достаточно просто доказать (мы доказательство опускаем), что случайная величина $\bar{x}$ в данном случае также распределена по нормальному закону.

Обозначим неизвестную величину $M\xi$ через a и подберем по заданной надежности γ число $d > 0$ так, чтобы выполнялось условие:

$$P(|\bar{x} - a| < d) = \gamma \qquad (1)$$

Так как случайная величина $\bar{x}$ распределена по нормальному закону с математическим ожиданием $M\bar{x} = M\xi = a$ и дисперсией $D\bar{x} = D\xi/n = \sigma^2/n$, получаем:

$$P(|\bar{x} - a| < d) = P(a - d < \bar{x} < a + d) = \Phi\left(\frac{a+d-a}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{a-d-a}{\frac{\sigma}{\sqrt{n}}}\right) = 2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right)$$

Осталось подобрать d таким, чтобы выполнялось равенство $2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right) = \gamma$ или

$\Phi\left(\frac{d\sqrt{n}}{\sigma}\right) = \frac{\gamma}{2}$.

Для любого $\gamma \in [0;1]$ можно по таблице найти такое число t, что $\Phi(t) = \gamma / 2$. Это число t иногда называют **квантилем**.

Теперь из равенства

$$\frac{d\sqrt{n}}{\sigma} = t$$

определим значение d: $d = \frac{\sigma t}{\sqrt{n}}$.

Окончательный результат получим, представив формулу (1) в виде:

$$P\left(\bar{x} - \frac{\sigma t}{\sqrt{n}} < a < \bar{x} + \frac{\sigma t}{\sqrt{n}}\right) = \gamma.$$

С надежностью γ доверительный интервал

$$\left(\bar{x}-\frac{\sigma t}{\sqrt{n}};\bar{x}+\frac{\sigma t}{\sqrt{n}}\right)$$

покрывает неизвестный параметр a = Mξ генеральной совокупности. Можно сказать иначе: точечная оценка $\bar{x}$ определяет значение параметра Mξ с точностью d=σ t / $\sqrt{n}$ и надежностью γ.

Задача. Пусть имеется генеральная совокупность с некоторой характеристикой, распределенной по нормальному закону с дисперсией, равной 6,25. Произведена выборка объема n = 27 и получено средневыборочное значение характеристики $\bar{x}$ = 12. Найти доверительный интервал, покрывающий неизвестное математическое ожидание исследуемой характеристики генеральной совокупности с надежностью γ =0,99.

Решение. Сначала по таблице для функции Лапласа найдем значение t из равенства Φ (t) = γ / 2 = 0,495. По полученному значению t = 2,58 определим точность оценки (или половину длины доверительного интервала) d: d = 2,5×2,58 / $\sqrt{27}$ ≈ 1,24. Отсюда получаем искомый доверительный интервал: (10,76; 13,24).

Доверительный интервал для математического ожидания нормального распределения при неизвестной дисперсии.

Пусть ξ – случайная величина, распределенная по нормальному закону с неизвестным математическим ожиданием Mξ, которое обозначим буквой a . Произведем выборку объема n. Определим среднюю выборочную $\bar{x}$ и исправленную выборочную дисперсию s^2 по известным формулам.

Случайная величина

$$t=\frac{(\bar{x}-a)\sqrt{n}}{s}$$

распределена по закону Стьюдента с n – 1 степенями свободы.

Задача заключается в том, чтобы по заданной надежности γ и по числу степеней свободы n – 1 **найти такое число t_γ, чтобы выполнялось равенство**

$$P\left(\left|\frac{(\bar{x}-a)\sqrt{n}}{s}\right|<t_\gamma\right)=\gamma \qquad (2)$$

или эквивалентное равенство

$$P\left(\bar{x}-t_\gamma\frac{s}{\sqrt{n}}<a<\bar{x}+t_\gamma\frac{s}{\sqrt{n}}\right)=\gamma. \qquad (3)$$

Здесь в скобках написано условие того, что значение неизвестного параметра a принадлежит некоторому промежутку, который и является доверительным интервалом. Его границы зависят от надежности γ, а также от параметров выборки $\bar{x}$ и s.

Чтобы определить значение t_γ по величине γ, равенство (2) преобразуем к виду:

$$P\left(\left|\frac{(\bar{x}-a)\sqrt{n}}{s}\right|\geq t_\gamma\right)=1-\gamma$$

Теперь по таблице для случайной величины t, распределенной по закону Стьюдента, по вероятности 1 – γ и числу степеней свободы n – 1 находим t_γ . Формула (3) дает ответ поставленной задачи.

Задача. На контрольных испытаниях 20-ти электроламп средняя продолжительность их работы оказалась равной 2000 часов при среднем квадратическом отклонении (рассчитанном как корень квадратный из исправленной выборочной дисперсии), равном 11-ти часам. Известно, что продолжительность работы лампы является нормально распределенной случайной величиной. Определить с надежностью 0,95 доверительный интервал для математического ожидания этой случайной величины.

Решение. Величина $1 - \gamma$ в данном случае равна 0,05. По таблице распределения Стьюдента, при числе степеней свободы, равном 19, находим: $t_\gamma = 2{,}093$. Вычислим теперь точность оценки: $2{,}093 \times 121/\sqrt{20} = 56{,}6$. Отсюда получаем искомый доверительный интервал:

(1943,4; 2056,6).

Доверительный интервал для дисперсии нормального распределения.

Пусть случайная величина ξ распределена по нормальному закону, для которого дисперсия $D\xi$ неизвестна. Делается выборка объема n . Из нее определяется исправленная выборочная дисперсия s^2. Случайная величина

$$\chi^2 = \frac{(n-1)s^2}{D\xi}$$

распределена по закону χ^2 с $n-1$ степенями свободы. По заданной надежности γ можно найти сколько угодно границ $\chi_1{}^2$ и $\chi_2{}^2$ интервалов, таких, что

$$\grave{o}\left(\chi_1{}^2 < \chi^2 < \chi_2{}^2\right) = \gamma \qquad (*)$$

Найдем $\chi_1{}^2$ и $\chi_2{}^2$ из следующих условий:

$$P(\chi^2 \le \chi_1{}^2) = (1 - \gamma)/2 \qquad (**)$$

$$P(\chi^2 \ge \chi_2{}^2) = (1 - \gamma)/2 \qquad (***)$$

Очевидно, что при выполнении двух последних условий справедливо равенство (*).

В таблицах для случайной величины χ^2 обычно дается решение уравнения $P(\chi^2 \ge \chi_q{}^2) = q$. Из такой таблицы по заданной величине q и по числу степеней свободы $n - 1$ можно определить значение $\chi_q{}^2$. Таким образом, сразу находится значение $\chi_2{}^2$ в формуле (***).

Для определения $\chi_1{}^2$ преобразуем (**):

$$P(\chi^2 \ge \chi_1{}^2) = 1 - (1 - \gamma)/2 = (1 + \gamma)/2$$

Полученное равенство позволяет определить по таблице значение $\chi_1{}^2$.

Теперь, когда найдены значения $\chi_1{}^2$ и $\chi_2{}^2$, представим равенство (*) в виде

$$P\left(\chi_1{}^2 < \frac{(n-1)s^2}{D\xi} < \chi_2{}^2\right) = \gamma .$$

Последнее равенство перепишем в такой форме, чтобы были определены границы доверительного интервала для неизвестной величины $D\xi$:

$$P\left(\frac{(n-1)s^2}{\chi_2{}^2} < D\xi < \frac{(n-1)s^2}{\chi_1{}^2}\right) = \gamma .$$

Отсюда легко получить формулу, по которой находится доверительный интервал для стандартного отклонения:

$$P\left(\frac{\sqrt{(n-1)}s}{\sqrt{\chi_2{}^2}} < \sqrt{D\xi} < \frac{\sqrt{(n-1)}s}{\sqrt{\chi_1{}^2}}\right) = \gamma \qquad (****)$$

Задача. Будем считать, что шум в кабинах вертолетов одного и того же типа при работающих в определенном режиме двигателях — случайная величина, распределенная по нормальному закону. Было случайным образом выбрано 20 вертолетов, и произведены замеры уровня шума (в децибелах) в каждом из них. Исправленная выборочная дисперсия измерений оказалась равной 22,5. Найти доверительный интервал, накрывающий неизвестное стандартное отклонение величины шума в кабинах вертолетов данного типа с надежностью 98%.

Решение. По числу степеней свободы, равному 19, и по вероятности $(1 - 0{,}98)/2 = 0{,}01$ находим из таблицы распределения χ^2 величину $\chi_2{}^2 = 36{,}2$. Аналогичным образом при вероятности $(1 + 0{,}98)/2 = 0{,}99$ получаем $\chi_1{}^2 = 7{,}63$. Используя формулу (****), получаем искомый доверительный интервал: (3,44; 7,49).

Занятие 15. *Проверка статистических гипотез. Сравнение выборочной средней с математическим ожиданием, двух дисперсий, двух математических ожиданий.*

Приведем несколько задач, которые могут быть решены с помощью проверки статистических гипотез.

1. Используется два метода измерения одной и той же величины. Первый метод дает оценки $x_1, x_2, ..., x_m$ этой величины, второй - $y_1, y_2, ..., y_n$. Требуется определить, обеспечивают ли оба метода **одинаковую точность измерений**.

2. Контроль точности работы некоторой производственной системы. Получаемые характеристики выпускаемой продукции характеризуются некоторым разбросом (дисперсией). Обычно величина этого разброса не должна превышать некоторого заранее заданного уровня. Требуется определить, обеспечивает ли система (например, линия сборки или отдельный станок) **заданную точность**.

Итак**, статистической** называют гипотезу о виде неизвестного распределения или о параметрах известных распределений. Примеры статистических гипотез: генеральная совокупность распределена по закону Пуассона; дисперсии двух нормальных распределений равны между собой.

Наряду с выдвинутой гипотезой всегда рассматривают и противоречащую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то принимается противоречащая гипотеза.

Нулевой (основной) называют выдвинутую гипотезу H_0.

Альтернативной (конкурирующей) называют гипотезу H_1, которая противоречит нулевой. Например, если нулевая гипотеза состоит в предположении, что математическое ожидание нормального распределения равно 5, то альтернативная гипотеза, например, может состоять в предположении, что $a \neq 5$. Кратко это записывают так: H_0:a=5; $H_1 : a \neq 5$.

Простой называют гипотезу, содержащую только одно предположение. Например, если λ - параметр показательного распределения, то гипотеза H_0:λ=3 - простая. **Сложной** называют гипотезу, состоящую из конечного или бесконечного числа простых гипотез. Например, сложная гипотеза $H : \lambda > 3$ состоит из бесконечного множества простых гипотез вида $H_i : \lambda = b_i$, где b_i - любое число, большее 3.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость ее проверки. Так как проверку производят статистическими методами, то ее называют **статистической**. В итоге **статистической проверки гипотезы** в двух случаях может быть принято неправильное решение, т.е. могут быть допущены ошибки двух родов.

Ошибка первого рода состоит в том, что будет **отвергнута правильная** гипотеза. **Ошибка второго рода** состоит в том, что будет **принята неправильная** гипотеза. Следует отметить, что последствия ошибок могут оказаться различными. Если отвергнуто правильное решение "продолжать строительство жилого дома", то эта ошибка первого рода повлечет материальный ущерб; если же принято неправильное решение "продолжать строительство" несмотря на опасность обвала дома, то эта ошибка второго рода может привести к многочисленным жертвам. Иногда, наоборот, ошибка первого рода влечет более тяжелые последствия.

Естественно, правильное решение может быть принято также в двух случаях, когда **принимается правильная** гипотеза или **отвергается неверная** гипотеза.

Вероятность совершения ошибки **первого рода** называют **уровнем значимости** и обозначают α. Чаще всего уровень значимости принимают равным 0,05 или 0,01. Если, например, принят уровень значимости 0,05, то это означает, что в пяти случаях из ста имеется риск допустить ошибку первого рода (отвергнуть правильную гипотезу).

Занятие 16. *Регрессионный анализ. Линейная регрессия.*

Понятие о регрессионном анализе

При рассмотрении взаимосвязей, как правило, рассматривают одну из величин (X) как независимую (объясняющую), а другую (Y) как зависимую (объясняемую). При этом изменение первой из них может служить причиной изменения другой. Например, рост дохода ведет к увеличению потребления; рост цены – к снижению спроса; снижение процентной ставки увеличивает инвестиции и т.д. Эта зависимость не является однозначной в том смысле, что каждому конкретному значению объясняющей переменой X может соответствовать не одно, а **множество** значений Y. Другими словами, каждому конкретному значению независимой переменной соответствует некоторое **вероятностное распределение** зависимой переменной. Поэтому анализируют, как объясняющая переменная (или переменные) влияет (или влияют) на зависимую переменную "в среднем". Зависимость такого типа, выражаемая соотношением $M(Y|x) = f(x)$ называется **функцией регрессии** Y на X. При рассмотрении зависимости двух случайных величин говорят о **парной регрессии**.

Зависимость **нескольких** переменных, выражаемую функцией $M(Y|x_1, x_2, ..., x_m) = f(x_1, x_2, ..., x_m)$ называют **множественной регрессией**.

Под **регрессией** понимается функциональная зависимость между объясняющими переменными и **условным математическим ожиданием** (средним значением) зависимой переменной Y, которая строится с целью предсказания (прогнозирования) среднего значения Y при некоторых значениях независимых переменных.

Установление формы зависимости и оценка параметров функции регрессии являются задачами **регрессионного анализа**.

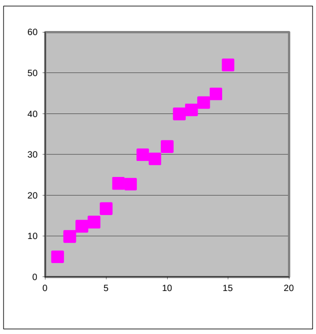
Так как реальные значения зависимой переменной могут быть различными при данном X (или $x_1, x_2, ..., x_m$), зависимость должна быть дополнена некоторым слагаемым ε, которое, по существу, является **случайной величиной**. Получающиеся в результате соотношения $Y = f(x) + \varepsilon$ или

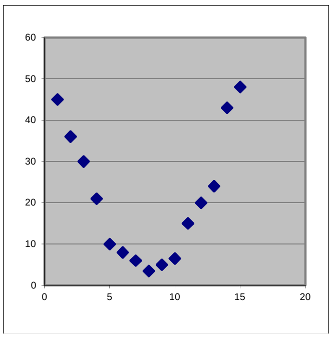
$Y = f(x_1, x_2, ..., x_m) + \varepsilon$ называются **регрессионными уравнениями** (или **моделями**).

Построение уравнения регрессии, описывающего эмпирические данные, включает три этапа:

- выбор **формулы** уравнения регрессии;
- определение **параметров** выбранного уравнения;
- анализ **качества уравнения** и проверка **адекватности** уравнения эмпирическим данным и, при необходимости, **совершенствование уравнения**.

В случае **парной** регрессии выбор уравнения обычно осуществляется по графическому изображению реальных статистических данных - **корреляционному полю.**





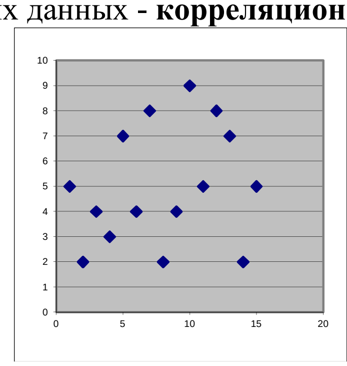


Рис.1 Корреляционные поля. А) – линейная регрессия; Б) – квадратичная регрессия; В) – отсутствие выраженной связи Y и X.

Выборочные уравнения регрессии.

Для определения значений теоретических коэффициентов, входящих в уравнения регрессии, необходимо знать и использовать **все** значения переменных генеральной совокупности, что практически невозможно. В связи с этим **по выборке ограниченного объема** строится так называемое **выборочное (эмпирическое) уравнение регрессии**. Из-за **ограниченности выборки** оценки коэффициентов, входящих в выборочное уравнение регрессии, отличаются от истинных (теоретических) значений, что приводит к несовпадению эмпирической и теоретической линий регрессии. Различные выборки из одной и той же генеральной совокупности обычно приводят к отличающимся друг от друга оценкам. Задача состоит в том, чтобы по конкретной выборке (x_i, y_i), $i = 1,...,n$ найти оценки неизвестных параметров так, чтобы построенная линия регрессии являлась наилучшей среди всех других линий.

Линейная регрессия

Если функция регрессии линейна, то говорят о **линейной регрессии**. Линейная регрессия (линейное уравнение) является распространенным (и простым) видом зависимости между экономическими переменными. Для простейшего случая **парной линейной регрессии**

$$M(Y|X = x_i) = \beta_0 + \beta_1 x_i \quad \text{или} \quad y_i = M(Y|X = x_i) + \varepsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ ,$$

где β_0 , β_1 - теоретические параметры регрессии; ε_i - случайное отклонение.

По выборке ограниченного объема строится **выборочное** уравнение регрессии $\hat{y}_i = b_0 + b_1 x_i$ (1)

где b_0 , b_1 - оценки неизвестных параметров β_0 , β_1, называемые **выборочными коэффициентами регрессии**, $\hat{y}_i$ - оценка условного математического ожидания $M(Y|X = x_i)$. Для величин y_i справедлива формула

$y_i = b_0 + b_1 x_i + e_i$ (2),где e_i - оценка теоретического отклонения ε_i.

Построенная прямая выборочной регрессии должна наилучшим образом описывать эмпирические данные, т.е. коэффициенты b_0 , b_1 должны быть такими, чтобы случайные отклонения e_i были минимальны. Наиболее распространенным методом нахождения коэффициентов уравнения регрессии является **метод наименьших квадратов (МНК)**.

Если по выборке (x_i, y_i), $i = 1,...,n$ требуется определить оценки b_0 , b_1 выборочного уравнения регрессии (2), то вводится в рассмотрение и минимизируется функция

$$Q(b_0, b_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \ .$$

Необходимым условием существования минимума данной функции двух переменных является равенство нулю ее частных производных по неизвестным параметрам b_0 , b_1:

$$\frac{\partial Q}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial Q}{\partial b_1} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) x_i = 0 \ .$$

Отсюда $$n b_0 + b_1 \sum x_i = \sum y_i$$

$b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i$,

и выразив из последних соотношений коэффициенты, получим

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \qquad b_0 = \bar{y} - b_1 \bar{x} \quad . \qquad (3)$$

где введены обозначения $\bar{x} = \frac{1}{n} \cdot \sum x_i$, $\bar{y} = \frac{1}{n} \cdot \sum y_i$, $\overline{x^2} = \frac{1}{n} \cdot \sum x_i^2$, $\overline{xy} = \frac{1}{n} \cdot \sum x_i y_i$.

Пример.

Для анализа зависимости объема потребления Y(у.е.) хозяйства от располагаемого дохода X(у.е.) отобрана следующая выборка объема $n=16$

	07	09	10	13	20	22	23	28	36	40	45	50
	02	05	08	10	15	17	19	25	32	30	41	44

Необходимо определить вид уравнения регрессии и по методу наименьших квадратов оценить параметры уравнения регрессии, а также спрогнозировать потребление при доходе X=160.

План решения. Строится корреляционное поле. По расположению точек на корреляционном поле предполагается, что зависимость Y от X – линейная. По МНК определяются коэффициенты $b_0 = 3.7$; $b_1 = 0.934$. Таким образом, уравнение парной регрессии имеет вид: $\hat{Y} = 0.934\, X + 3.7$

Множественная линейная регрессия

На экономический показатель чаще всего оказывает влияние не один, а несколько факторов. Например, спрос на некое благо определяется не только ценой данного блага, но и ценами на замещающие и дополняющие блага, доходом потребителей и многими другими факторами. В этом случае рассматривается множественная регрессия $M(Y|x_1, x_2, \ldots\ldots, x_m) = f(x_1, x_2, \ldots\ldots, x_m)$.

Теоретическое **линейное** уравнение регрессии имеет вид $Y = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_m x_m + \varepsilon$.

Для оценки параметров уравнения множественной линейной регрессии также, как правило, используется метод наименьших квадратов.

Занятие 17. *Дисперсионный анализ.*
Схема однофакторного дисперсионного анализа

При решении многих статистических задач приходится рассматривать зависимость результативного признака от многомерного фактора x=(x1, x2, xp), составляющие xк, к=1,…, p которого в дальнейшем будем называть его уровнями.

Примером таких задач могут служить задачи изучения зависимости качества воспроизведения учебного материала от скорости его подачи, урожайности от дозы веденных удобрений, прибыли от вложения в рекламу и т.п.

При решении подобных задач обычно используется дисперсионный метод, заключающийся в сравнении факторной и остаточной дисперсии, на которой разбивается общая дисперсия результативного признака y.

Факторная дисперсия S2факт. вызвана действием на случайную величину y фактора x, а остаточная (S2ост) целиком обусловлена случайными причинами.

Однофакторный дисперсионный анализ

Пусть на результаты признака у воздействует фактор х, имеющий р постоянных уровней, т.е. х=(х1, х2, …хр). На каждом из р уровней произведем одинаковое число испытаний, равное q, и результаты этих испытаний оформляем в виде таблицы 28. Х разбивает наблюдаемые значения на р групп и значения этих групп образуют столбцы таблицы. В последней строке таблицы приведены найденные значения групповых средних.

Общая выборочная средняя у= $\frac{\sum_{j=1}^{p} \overline{y_{ir}}}{p}$.

Используя данные таблицы, найдем:

1. Общую сумму квадратов отклонений наблюдаемых значений от общей средней $\overline{y}$: Rобщ= $\sum_{j=1}^{p} \sum_{i=1}^{q} (y_{ij} - \overline{y})^2$ - рассеяние всех наблюдаемых значений от общей средней.

2. $R_{факт.}$=q$\sum_{j=1}^{p} (\overline{y_{rj}} - \overline{y})^2$ - факторная сумма квадратов отклонений групповых средних от общей средней и она характеризует рассеяние групп около общей средней.

3. $R_{ост.}$=$\sum_{j=1}^{p} \sum_{i=1}^{q} (y_{ij} - \overline{y_{ri}})^2$ - остаточная сумма квадратов отношений наблюдаемых значений группы от своей групповой средней, которая уже характеризует рассеяние внутри группы. Общая сумма дисперсии δ^2 складывается из сумм межгрупповой и внутригрупповой дисперсий.

Аналогично: $R_{общ}$=$R_{факт.}$+$R_{ост.}$.

При практическом использовании дисперсионного анализа обычно вычисляют Rобщ и Rфакт ,а Rост. находят из разности : Rобщ.-Rфакт .

Таблица 1.

№ испытания	**Уровни фактора**			
	X_1	X_2	…	X_p
1	y_{11}	y_{12}	…	y_p
2				
. .				
Q	y_{q1}	y_{q2}	…	y_{qp}
Групповая Средняя	y_{r1}	y_{r2}	…	y_{r3}

Решение статистических задач методом дисперсионного анализа

Пусть на конечный результативный признак у, имеющий нормальное распределение, воздействует фактор х(х1, х2… хр) с р-уровнями и необходимо установить, оказывает ли этот фактор существенное влияние на признак у.

Если изучаемый фактор х(х1, х2… хр) оказывает существенное влияние на у, то групповые средние $\overline{y_{rj}}$ j=$\overline{1, p}$ существенно различаются между собой.

Поэтому одновременно значимо будут различаться Rфакт и Rост . А значит отношение $\frac{R_{факт.}}{R_{ост.}}$ будет заметно больше 1.

При решении практических задач обычно рассматривают отношение факторной и остаточной дисперсии, подсчитывается:

Fнабл= $\frac{S^2_{факт}}{S^2_{ост}}$, где Sфакт.= $\frac{R_{факт}}{p-1}$. $S^2_{ост} = \frac{R_{ост}}{p(q-1)}$

и задача сводится к сравнению факторной и остаточной дисперсий по критерию Фишера-Снедекора

Если Fнабл.>Fкр., то нулевая гипотеза о равенстве факторной и остаточной дисперсий отвергаются.

Пример. При уровне значимости α=0,05 исследовать степень влияния стажа рабочих предприятия на мотивацию труда.

Таблица 2

Стаж	1-5 л.	6-10 л.	11л и более
1 бр.	20	22	24
2 бр.	21	23	25
3 бр.	22	24	26
	21	23	25

Результат признака y – мотивация труда зависит от стажа, который имеет 3 уровня

P=q=3 $\bar{y}$ =23

rобщ.= $\sum\sum$ $3^2+2^2+1^2+1^2+1^2+1^2+2^2+3^2$ =4+8+18=12+18=30.

rфакт. =3(22+22)=24, rост. =30-24=6.

$S^2_{факт} = \frac{24}{2} = 12$, $S^2_{ост} = \frac{6}{3*2} = 1$.

F= $\frac{12}{1}$, Fkp(0,05;2;6)=5,14(из таблицы 10 Приложения 2).

Fнабл.>Fкр., поэтому нулевая гипотеза о равенстве факторной и остаточной дисперсий отвергаются. А это означает, что факторная дисперсия значимо отличается от остаточной дисперсии.

Занятие 18. *Сущность метода статистических испытаний. Моделирование сложных испытаний и их результатов*

Основным методом моделирования таких систем на ЭВМ является **метод статистического моделирования**, составляющий методологическую основу построения **имитационных моделей** систем на ЭВМ.

Сущность метода статистического моделирования сводится к построению для процесса функционирования исследуемой системы некоторого моделирующего алгоритма, имитирующего поведение и взаимодействие элементов системы с учетом случайных входных воздействий и воздействий внешней среды, и реализация этого алгоритма с использованием программно-технических средств ЭВМ.

В результате статистического моделирования системы получается серия частных значений искомых величин или функций, статистическая обработка которых позволяет получить сведения о поведении реального объекта или процесса в произвольные моменты времени. Если количество реализаций достаточно велико, то полученные результаты моделирования системы приобретают статистическую устойчивость и с достаточной точностью могут быть приняты в качестве оценок искомых характеристик процесса функционирования системы.

Теоретической основой метода статистического моделирования являются предельные теоремы теории вероятностей. В соответствии с данными теоремами множества случайных событий подчиняются определенным закономерностям, позволяющим не только прогнозировать их поведение, но и количественно оценить некоторые средние их характеристики. Характерные закономерности наблюдаются также в распределениях случайных величин, которые образуются при сложении множества воздействий. К основным предельным теоремам, используемым при статистическом моделировании, относятся центральная предельная теорема, теоремы Бернулли, Пуассона, Чебышева, Маркова, Лапласа. Принципиальное значение данных теорем состоит в том, что они гарантируют высокое качество статистических оценок при весьма большом числе испытаний (реализаций), которое с легкостью может быть получено при использовании ЭВМ.

Поясним сущность метода статистического моделирования на следующем примере: необходимо найти оценку математического ожидания выходной величины системы автоматического регулирования - $y(t)=F(x(t),f(t))$ если внешнее возмущающее воздействие $f(t)=F(\alpha)$ есть случайный процесс с известным законом распределения случайной величины α. Схема алгоритма, реализующего метод статистического моделирования для оценки $M[y]$, представлена на рис.1. Из него видно, что для учета стохастического возмущающего воздействия при статистическом моделировании на ЭВМ необходим механизм формирования значений случайных величин. При этом результаты статистического моделирования зависят как от количества реализаций, так и от качества исходных последовательностей случайных чисел.

Метод постановки статистического эксперимента для решения разнообразных практических задач известен давно – это **метод статистических испытаний (метод Монте-Карло)**.

Фактически современное имитационное моделирование является его развитием применительно к сложным системам.

Сущность метода Монте-Карло состоит в следующем: требуется найти значение α некоторой изучаемой величины. Для этого выбирают такую случайную величину X, математическое ожидание которой равно α - $M(X)=\alpha$. Практически поступают так: производят n испытаний, в результате которых получают n возможных значений X; вычисляют их среднее арифметическое - $\bar{x}=\sum_i x_i / n$ и принимают его в качестве оценки (приближенного значения) α^* искомого числа α - $\alpha \cong \alpha^* = \bar{x}$.

Другими словами метод Монте-Карло состоит в «разыгрывании случайных величин» и использовании их выборок для получения искомых оценок.

Таким образом статистическое моделирование систем и процессов на ЭВМ требует большого объема действий со случайными числами, а его результаты существенно зависят от качества исходных последовательностей случайных чисел.

Рассмотрим возможности и особенности получения последовательностей случайных чисел при статистическом моделировании систем на ЭВМ. На практике используются три основных способа генерации случайных чисел: аппаратный (физический), табличный (файловый) и алгоритмический (программный).

Реализация аппаратного способа генерации случайных чисел основана на использовании внешнего электронного устройства, подключаемого к ЭВМ. В качестве физического эффекта, лежащего в основе таких генераторов чисел, чаще всего используются шумы в электронных и полупроводниковых приборах. Напряжение с выхода источника шума, являющееся случайным процессом, стробируется напряжением полезного сигнала и квантуется относительно заданного порога. В результате получается серия импульсов, расстояния по времени между которыми являются случайными числами. Основным недостатком данного метода является невозможность получения при моделировании одинаковых последовательностей чисел. А его достоинства, связанные с не использованием ЭВМ, в настоящее время фактически нивелированы все возрастающими возможностями последних.

Табличный способ заключается в предварительном формировании таблицы случайных чисел для требуемого количества реализаций в виде файла. При достаточно большом количестве реализаций основным недостатком данного способа являются большие затраты машинных ресурсов на частое обращение к соответствующему файлу. Однако он позволяет легко осуществлять повторное воспроизведение последовательности чисел.

Алгоритмический способ получений последовательности случайных чисел основан на формировании случайных чисел в ЭВМ с помощью специальных алгоритмов и реализующих их программ. При этом вычисление случайных чисел может быть организовано как по мере необходимости, так и путем периодической генерации множества случайных чисел. Данный способ является наиболее предпочтительным ввиду большей гибкости реализации различных законов распределения.

При моделировании систем на ЭВМ программная имитация случайных воздействий любой сложности сводится к генерированию некоторых стандартных (базовых) процессов и к их последующему функциональному преобразованию. В качестве базового может быть принят любой удобный случайный процесс (нормальный, пуассоновский и т.п.). Однако при дискретном моделировании в качестве базового процесса используют последовательности чисел, представляющие собой реализации равномерно распределенных на интервале [0,1] случайных величин. Ввиду того, что ЭВМ оперирует с конечным множеством чисел, получаемые последовательности являются не идеальными случайными, а так называемыми пвсевдослучайными. При этом детерминированность получаемой последовательности определяется разрядностью чисел, с которыми оперирует ЭВМ.

Генератор случайных чисел должен удовлетворять следующим требованиям: получаемые последовательности должны состоять из квазиравномерно распределенных чисел, содержать статистически независимые числа, быть воспроизводимыми, иметь неповторяющиеся числа, использовать минимум машинного времени и памяти.

Наибольшее применение в практике моделирования на ЭВМ для генерации последовательностей псевдослучайных чисел находят алгоритмы вида -

$$x_{i+1} = \Phi(x_i), \qquad (1)$$

представляющие собой рекуррентные соотношения первого порядка, для которых начальное число x_0 и параметры соотношения заданы.

Прежде чем приступать к реализации моделирующих алгоритмов на ЭВМ, необходимо проверить **качество последовательности псевдослучайных чисел**. Данная процедура включает проверку равномерности, стохастичности и независимости.

Проверка равномерности может быть выполнена следующим образом. Интервал значений случайных чисел (0,1) разбивается на m частей. Тогда при генерации последовательности $\{x_i\}$ каждое из чисел x_i с вероятностью $p_j=1/m$ должно попадать в один из подынтервалов. Всего в каждый подынтервал попадет N_j чисел ($\sum_{j=1}^{m} N_j = N$). Относительная частота попадания случайных чисел в каждый из подынтревалов буде равна N_j/N. Вид получаемой гистограммы представлен на рис.2. Очевидно, что последовательность тем равномернее, чем ближе ломаная линия к теоретическому значению p_j. Оценка данной степени приближения может быть проведена с использованием критериев согласия (Пирсона, Стьюдента, Фишера). При этом на практике обычно принимается m=20-50, $N=(10^2\text{-}10^3)m$.

Существо применения критерия Пирсона заключается в следующем: находят $\chi^2 = \sum_{j=1}^{m} \left(N_j - Np_j\right)/(Np_j)$. По вычисленному значению χ^2 и числу степеней свободы $k = m - r - 1$ (r – число параметров теоретического закона распределения, для равномерного закона r=1) по таблице находят P. Если эта вероятность превышает некоторый уровень значимости γ, то гипотеза о равномерности принимается.

Проверка стохастичности осуществляется следующим образом. Последовательность разбивается на элементы первого и второго рода (a и b):

$$x_i = \begin{cases} a, \ \textit{если} \ \ x_1 < p \\ b, \ \textit{в противном случае} \end{cases} \qquad 0 < p < 1$$

Серией называется любой отрезок последовательности, состоящий из идущих друг за другом элементов одного и того же рода, а число элементов в этом отрезке называют длиной серии. В результате получаем последовательность –

aaabbbbbbbaabbaabaaaabbbbbbbb….

Для равномерно распределенной последовательности случайных чисел вероятность появления серии длиной j в последовательности длиной l определяется формулой Бернули:

$$P(j,l) = C_l^j \, p^j (1-p)^{l-j}; \ j = 0,...,l; l = 1,...,n$$

При экспериментальной проверке оцениваются частоты появлений серий длиной j. В результате получают теоретическую и экспериментальную зависимости $P(j,l)$, сходимость которых проверяется по известным критериям согласия при различных значениях p и l.

Проверка независимости последовательности псевдослучайных равномерно распределенных чисел проводится на основе вычисления корреляционного момента. Данная проверка осуществляется путем введения в рассмотрение последовательности $\{y_i\} = \{x_{i+\tau}\}$, где τ - величина сдвига последовательностей.

В общем случае корреляционный момент дискретных случайных величин ζ, η с возможными значениями x_i, y_i определяется по формуле:

$$K_{\zeta\eta} = \sum_i \sum_j (x_i - M[\zeta])(y_j - M[\eta]) p_{ij},$$

где p_{ij} – вероятность того, что (ζ, η) примет значение (x_i, y_j).

Для независимых случайных величин $K_{\zeta\eta} = 0$. При проведении оценок независимости используют понятие коэффициента корреляции –

$$\rho_{\zeta\eta} = K_{\zeta\eta} / (\sigma_x \sigma_y),$$

где σ_x, σ_y - средние квадратические отклонения величин ζ, η.

В качестве критерия корреляционной независимости используют соотношение:

$$\left|\rho_{\zeta\eta}\right| \le \beta\sqrt{1/N}.$$

Его выполнение означает, что с доверительной вероятностью β справедлива гипотеза корреляционной независимости.

Расчет коэффициента корреляции осуществляют следующим образом:

$$\rho_{\zeta\eta} = \frac{\frac{1}{N-\tau}\sum_{i=1}^{N-\tau} x_i x_{i+\tau} - \frac{1}{(N-\tau)^2}\sum_{i=1}^{N-\tau} x_i \sum_{i=1}^{N-\tau} x_{i+\tau}}{\sqrt{D[x_i] D[x_{i+\tau}]}},$$

$$D[x_i]=\frac{1}{N-\tau}\sum_{i=1}^{N-\tau}x_i^2-\frac{1}{(N-\tau)^2}\left(\sum_{i=1}^{N-\tau}x_i\right)^2,$$

$$D[x_{i+\tau}]=\frac{1}{N-\tau}\sum_{i=1}^{N-\tau}x_{i+\tau}^2-\frac{1}{(N-\tau)^2}\left(\sum_{i=1}^{N-\tau}x_{i+\tau}\right)^2.$$

При анализе качества программных генераторов псевдослучайных равномерных последовательностей важной характеристикой является длина отрезка апериодичности L, то есть длина отрезка генерируемой последовательности чисел, на котором ни одно число не повторяется. Очевидно, что использование при моделировании последовательности чисел большей чем L длины приведет к повторению опытов, что не позволит получить более лучших статистических оценок при увеличении затрат машинных ресурсов. С теоретическими и практическими способами оценки данного показателя вы можете ознакомиться в литературе (Советов Б.Я., Яковлев С.А. Моделирование систем).

Моделирование случайных воздействий на системы

В общем случае для моделирования случайных воздействий на системы используют случайные события, дискретные и непрерывные случайные величины, векторы и процессы. Формирование на ЭВМ случайных объектов любой природы из перечисленных сводится к генерации и преобразованию последовательностей случайных чисел. Рассмотрим вопросы из преобразования для генерации воздействий на моделируемую систему.

Простейшими случайными объектами являются случайные события. Процедура моделирования того или иного случайного события зависит от его формулировки. Например, необходимо смоделировать случайное событие A, наступающее с вероятностью p_A. В этом случае одним из вариантов моделирования является последовательный анализ значений x_i из сгенерированной последовательности случайных чисел и сравнения их с p_A. Если неравенство $x_i \le p_A$ выполняется, то исходом испытания является событие A.

Если искомый результат испытания является сложным событием, зависящим от двух и более простых событий, то процедура моделирования будет следующей:

а) для двух независимых простых событий A и B – осуществляют последовательную проверку условия совместного исхода испытания ($AB, \overline{A}B, A\overline{B}, \overline{AB}$) на основе проверки условий истинности каждого события ($x_i \le p_A, x_{i+1} \le p_B$);

б) для двух зависимых простых событий A и B – задается условная вероятность наступления события B при условии, что событие A уже произошло p(B/A); из последовательности $\{x_i\}$ извлекается очередное число и проверяется справедливость неравенства $x_i \le p_A$; далее процесс проверки организуется в зависимости от условия совместного исхода событий; например для события AB поступают следующим образом - если выполняется $x_i \le p_A$, то извлекается следующее число и проверяется неравенство $x_{i+1} \le p(B/A)$, если и оно выполняется то значит заданное событие наступило. Путем алгоритмической организации проверки приведенных неравенств может быть реализован любой совместный исход данных событий $AB, \overline{A}B, A\overline{B}, \overline{AB}$.

Для формирования значений случайных величин с заданным законом распределения в качестве базовой последовательности используют последовательности случайных чисел, имеющие равномерное распределение в интервале (0,1), которые преобразуются в значения случайной величины с заданным законом распределения.

Для моделирования дискретной случайной величины принимающей значения $y_1 \le y_2 \le ... \le y_j \le ...$ с вероятностями $p_1, p_2, ..., p_j, ...$ используют метод обратной функции: если ζ - равномерно распределенная на интервале (0,1) случайная величина, то искомая случайная величина η получается с помощью преобразования –

$\eta = F_{\eta}^{-1}(\zeta)$, где $F_{\eta}(\zeta)$ - функция распределения случайной величины ζ

($F_{\eta}(\zeta) = P(\eta \le \zeta)$).

Алгоритм формирования η сводится к следующему:

если $x_1 < p_1$, то $\eta = y_1$

если $x_1 < p_1 + p_2$, то $\eta = y_2$

………………………………

если $x_j < \sum_{j=1}^{m} p_j$, то $\eta = y_m$

Для моделирования непрерывных случайных величин с заданным законом распределения также пользуются методом обратной функции. При этом, чтобы получить число принадлежащее последовательности $\{y_j\}$ с плотностью распределения $f_{\eta}(y)$ необходимо разрешить относительно y_j уравнение вида –

$\int_{-\infty}^{y_j} f_{\eta}(y)dy = x_i$, где x_i – случайное число, имеющее равномерное распределение в интервале (0,1).

Основным недостатком данного способа является сложность интегрирования функции плотности распределения вероятностей.

Поэтому часто используют приближенные способы преобразования случайных чисел – универсальные и специализированные. Первые позволяют получать случайные числа с любым законом распределения, а вторые – только с конкретным законом.

Один из универсальных способов получения случайных чисел:

- генерируется случайное равномерно распределенное число x_i;

- с помощью этого числа случайным образом выбирается интервал (a_k,a_{k+1}) на основе предварительно сформированной таблицы коэффициентов масштабирования для каждого интервала, определенных из условия –

$\int_{a_k}^{a_{k+1}} f_{\eta}(y)dy = 1/m$, m- количество интервалов разбиения плотности $f_{\eta}(y)$ на участке (a,b);

- генерируется x_{i+1} и осуществляется ее масштабирование - $(a_{k+1} - a_k)x_{i+1}$;

- вычисляется случайное число с требуемым законом распределения - $y_i = a_k + (a_{k+1} - a_k)x_{i+1}$.

Для моделирования реализаций случайных векторов, обладающих заданными вероятностными характеристиками используют понятие совместного закона распределения случайных величин.

Рассмотрим дискретный двумерный случайный процесс. Составляющая ζ принимает возможные значения $\{x_i\}_{i=\overline{1,n}}$, а составляющая η принимает значения $\{y_j\}_{j=\overline{1,n}}$, причем каждой паре (x_i, y_j) соответствует вероятность p_{ij}. Тогда каждому x_i будет соответствовать $p_i = \sum_{j=1}^{n} p_{ij}$. На основании данного распределения можно определить конкретные значения x_i . Аналогично определяются значения $y_{j=i}$ и получаются пары реализаций случайного вектора.

Непрерывный двумерный случайный процесс описывается совместной функцией плотности вероятности - $f(x,y)$. Эта функция может быть использована для определения функции плотности случайной величины ζ как - $f_\zeta(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy$.

Имея функцию плотности $f_\zeta(x)$, можно найти случайное число x_i, а затем при условии, что $\zeta = x_i$, определить условное распределение случайной величины η - $f_\eta(y/\zeta = x_i) = f(x,y)/f_\zeta(x_i)$. В соответствии с этой функцией плотности можно определить случайное число y_i. Тогда пара чисел (x_i,y_i) будет являться реализацией вектора (ζ,η).

Общие принципы построения и эксплуатации имитационных моделей

Во-первых, необходимо выделить основные взаимодействия компонентов системы между собой и с внешней средой, которые являются существенными с точки зрения получения требуемых оценок ее функционирования, а также выбрать единицу времени, отражающую природу моделируемой системы.

Во-вторых, определить количество и законы распределения разыгрываемых случайных величин (векторов) и с учетом качества их разыгрывания выбрать продолжительность прогона модели и число прогонов (наблюдений).

Поскольку основная цель состоит в получении наблюдений с наименьшей ошибкой, то используют либо очень длительный прогон модели, либо повторения более коротких прогонов модели с различными последовательностями случайных чисел. Применение первого способа связано с большими затратами машинного времени. Применение второго способа ограничено необходимостью правильного выбора длительности прогона, соответствующей переходу модели в стационарный режим. В рамках второго способа могут быть использованы различные методы получения наблюдений – метод повторения, метод подынтервалов, метод циклов и др.

Метод повторения заключается в организации нескольких прогонов модели при одних и тех же начальных условиях, но с различными последовательностями случайных чисел. Его преимуществом является статистическая независимость получаемых наблюдений (необходимое условие для любого статистического теста). А недостаток состоит в том, что наблюдения могут оказаться сильно смещенными под влиянием начальных условий (переходное состояние).

Метод подынтервалов направлен на уменьшение влияние переходных условий, которому подвержен метод повторения. Он основан на разбиении каждого прогона модели на равные промежутки времени. Преимущество данного метода в том, что со временем влияние переходных условий уменьшается, а недостатком – не выполнение условия о независимости наблюдений от интервала к интервалу, так как между ними возникает автокорреляция. Ее влияние можно уменьшить путем увеличения длины прогона и длину интервалов.

Метод циклов позволяет уменьшить влияние указанной автокорреляции. Он подразумевает выбор интервалов таким образом, чтобы в их начальных точках условия были одинаковыми (с точки зрения рассматриваемой переменной). Однако его недостатком является уменьшение числа получаемых наблюдений. При этом ввиду нерегулярности циклов усложняется оценка значения каждого наблюдения.