ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Б1.Б.1.10 Теория вероятностей и математическая статистика

Направление подготовки (специальность): 10.05.03 Информационная безопасность автоматизированных систем

Профиль образовательной программы Информационная безопасность автоматизированных систем критически важных объектов
Форма обучения: очная

СОДЕРЖАНИЕ

Конспект лекций

1.1 Лекция 1 (Л-1) (2 ч.)

Тема: Случайные события, классификация и вероятности

1.1.1 Вопросы лекции:

1.Случайные события, их классификация.
2.Вероятность случайных событий, ее интерпретации.
3. Основные теоремы теории вероятностей.

1.1.2. Краткое содержание вопросов:

1.Случайные события, их классификация.

Определение. *Два события называются* ***равными****, если одно из них наступает тогда и только тогда, когда наступает другое.*

Пример. Будут произведены 3 выстрела в мишень. A – число попаданий в мишень равно 0, B – число попаданий в мишень меньше, чем 0,5. Очевидно, что $A = B$.

Определение. *Два события называются* ***равновозможными****, если вероятности их наступления равны (в смысле статистического определения вероятности).*

На практике равновозможность событий обычно усматривается из симметрии ситуации.

Пример. Пусть испытание – бросание монеты. Тогда события A – выпадение “орла” и B – выпадение “решки” являются равновозможными.

Определение. *Событие называется* ***достоверным****, если оно наступает в каждом из испытаний.*

Достоверное событие будем обозначать через E. Такое событие определено однозначно для каждого вида испытания.

Пример. Пусть испытание – бросание игральной кости. Тогда $E = \{1,2,3,4,5,6\} = (m < 10) = (m > 0) = ...,$ где m – число выпавших очков.

Т.к. $N_E = N$, то $P(E) = \lim\limits_{N \to \infty} \dfrac{N_E}{N} = \lim\limits_{N \to \infty} 1 = 1,$ т.е.

$$P(E) = 1.$$

Определение. *Событие называется* ***невозможным****, если оно не наступает ни в одном из испытаний.*

Невозможное событие будем обозначать символом $\varnothing$. Это событие определено однозначно для каждого вида испытания.

Пример. Пусть измеряется рост наудачу взятого человека. Тогда $\varnothing$ = (значение роста – отрицательное число) = (рост – более 100 км) =….

Т.к. $N_{\varnothing} = 0$, то $P(\varnothing) = \lim\limits_{N \to \infty} \dfrac{N_{\varnothing}}{N} = \lim\limits_{N \to \infty} 0 = 0,$ т.е.

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

Определение. *Два события называются* ***несовместными (несовместимыми)****, если они не могут наступить одновременно.*

Пример. Испытание – извлечение карты из колоды. Если событие A – извлечена карта красной масти, событие B – извлечена карта черной масти, то A и B – несовместны.

Пример. Пусть по мишени производится 3 выстрела и m – число попаданий в мишень. Тогда события, например, $(m=3)$ и $(m\le 1)$ – несовместны.

Определение. *События* $A_1, A_2, \dots, A_k$ *называются* ***единственно возможными*** *для некоторого испытания, если в результате испытания хотя бы оно из них обязательно наступает.*

Пример. Пусть испытание – бросание игральной кости. $A=\{1,2,3\}$, $B=\{3,4,5,6\}$, $C=\{3,4,5\}$. Тогда события A и B – единственно возможны (т.к. не существует такого исхода бросания игральной кости, при котором ни A, ни B не наступило). Напротив, A и C не являются единственно возможными (т.к. при выпадении “6” ни A, ни C не наступают).

Определение. *Говорят, что события* $A_1, A_2, \dots, A_k$ *образуют* ***полную систему (группу),*** *если эти события попарно несовместимы и единственно возможны.*

Пример. Пусть испытание – бросание игральной кости. Тогда события $A_1=\{1\}, A_2=\{2\}, \dots, A_6=\{6\}$ образуют полную систему.

Пример. Пусть по мишени производится 3 выстрела и m – число попаданий в мишень. Тогда события, например, $(m=0), (1\le m\le 2), (m=3)$ образуют полную систему.

Заметим, что при заданном типе испытания полная система событий определена, вообще говоря, неоднозначно.

Определение. *Если два события образуют полную систему, то они называются парой* ***взаимно противоположных*** *событий.*

Если одно из событий такой пары обозначено, скажем, через A, другое будет обозначено $\overline{A}$.

Пример. Пусть испытание – бросание монеты. Тогда события A – выпадение “орла” и B – выпадение “решки” являются взаимно противоположными ($B=\overline{A}$).

Пример. Пусть по мишени производится 3 выстрела, и m – число попаданий в мишень. Тогда события, например, $(m<2)=(m=0 \text{ или } m=1)$ и $(m\ge 2)=(m=2 \text{ или } m=3)$ – взаимно противоположны.

Операции над событиями

Определение. ***Суммой*** *событий A и B называется такое событие* $C=A+B$, *которое считается наступившим тогда и только тогда, когда наступило или событие A, или событие B, или оба эти события вместе.*

Пример. Пусть испытание – извлечение карты из колоды, а следующие события состоят в извлечении: A – карты красной масти, B – картинки, D – числовой карты. Если в результате конкретного испытания из колоды достали, например, “семерку” крестей то событие $A+B$ не наступило, а события $A+D$ и $B+D$ наступили.

Пример. Пусть по мишени производится 3 выстрела, m – число попаданий в мишень $A=(m<2)$, $B=(m>0)$, $C=A+B$. Тогда $C=\{m=1\}$.

Замечание 1. Условие единственной возможности событий $A_1, A_2, \dots, A_k$ равносильно тому, что $A_1+A_2+\dots+A_k=E$. В частности, если события $A_1, A_2, \dots, A_k$ образуют полную систему, то $A_1+A_2+\dots+A_k=E$, и при $k=2$ имеем

$$A+\overline{A}=E.$$

Определение. ***Произведением*** *событий A и B называется такое событие* $C=AB$, *которое считается наступившим тогда и только тогда, когда события A и B наступили одновременно.*

Пример. Пусть испытание состоит в бросании игральной кости.
$A=\{1,2,3\}$, $B=\{3,4\}$. Тогда $AB=\{3\}$ и $A+B=\{1,2,3,4\}$.

Замечание 2. Произвольные события A и B являются несовместимыми тогда и только тогда, когда $AB=\varnothing$.

2.Вероятность случайных событий, ее интерпретации.

Определение. *Пусть некоторое испытание имеет* n *исходов, причем эти исходы*

а) попарно несовместимы;

б) единственно возможны;

в) равновозможны

и наступлению события A *благоприятствует* m *исходов из* n*. Тогда вероятность* $P(A)$ *наступления события A (в одном испытании) определяется по формуле*

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

Пример. В коробке имеется 10 хороших деталей и 5 бракованных. Наудачу из коробки извлекается одна деталь. Найти вероятность наступления события A – извлеченная деталь – хорошая.

Решение. Общее число исходов $n=15$ равно полному числу деталей в коробке. Извлечению хорошей детали благоприятствует $m=10$ исходов из общего числа (число хороших деталей). Тогда

$$P(A)=\frac{10}{15}=\frac{2}{3}.$$

Пример. Одновременно бросаются три монеты. Найти вероятность того, что на двух из них выпадет «орел».

Таблица 1.

Номер исхода	Номер монеты		
	1	2	3
1	О	О	О
2	О	Р	О
3	О	О	Р
4	О	Р	Р
5	Р	О	О
6	Р	Р	О
7	Р	О	Р
8	Р	Р	Р

Решение. Для удобства будем предполагать, что монеты некоторым образом занумерованы. Единичным исходом здесь является совокупный результат по трем монетам (другими словами, для того, чтобы задать единичный исход, надо сказать, *что* выпало на первой монете, на второй и на третьей). Перечислим возможные исходы (см. Таблицу 1, в которой выпадение «орла» на соответствующей монете обозначено буквой «О», «решки» – «Р»). Видно, что общее число n исходов равно 8. Число m благоприятствующих исходов равно 3 – это исходы с номерами 2, 3, 5 Таблицы 1. Тогда

$$P(A)=\frac{m}{n}=\frac{3}{8}.$$

Пример. В коробке 6 белых шаров и 8 красных. Наудачу одновременно извлекаются 3 шара. Найти вероятность, того, что среди них будут:

а) два белых шара;

б) не менее одного белого.

Решение. а) Для удобства будем предполагать, что имеющиеся шары некоторым образом перенумерованы. Пусть, например, белые шары имеют номера 1, 2, … ,6 красные – 7, 8 , … ,14. Тогда единичным исходом является произвольная тройка номеров: $\{1,2,3\}$, $\{1,2,4\}$, …, $\{12,13,14\}$.

Тогда общее число n исходов равно числу способов, которыми можно выбрать 3 номера из имеющихся 14-ти номеров. Напомним, что такое число равно соответствующему числу сочетаний:

$$n=C_{14}^{3}.$$

(В общем случае, $C_k^s=\dfrac{k!}{s!(k-s)!}$ *равно числу способов, которыми можно выбрать* s *объектов из* k *имеющихся объектов*.) Таким образом,

$$n=C_{14}^{3}=\frac{14!}{3!11!}=\frac{1\cdot2\cdot\ldots\cdot10\cdot11\cdot12\cdot13\cdot14}{1\cdot2\cdot3\cdot1\cdot2\cdot\ldots\cdot10\cdot11}=2\cdot13\cdot14=364.$$

Найдем теперь число m исходов, благоприятствующих появлению двух белых шаров среди трех извлеченных. Число способов, которыми можно выбрать 2 шара из имеющихся 6-ти белых шаров, равно C_6^2. Но число благоприятствующих исходов с фиксированной парой белых шаров равно числу способов, которыми можно выбрать оставшийся красный шар в тройку, т.е. равно C_8^1. Поэтому

$$m = C_6^2 \cdot C_8^1 = \frac{6!}{2! 4!} \cdot \frac{8!}{1! 7!} = 5 \cdot 3 \cdot 8 = 120.$$

Окончательно имеем

$$P(A) = \frac{m}{n} = \frac{120}{364} = \frac{30}{91},$$

где A – событие состоящее в том, что среди трех отобранных шаров ровно 2 белых шара.

б) Полное число n исходов найдено в п. а). Число троек, в которых не менее 2-х белых шаров, равно сумме троек с двумя белыми шарами и троек с тремя белыми шарами:

$$m = C_6^2 \cdot C_8^1 + C_8^3 = 120 + 56 = 176.$$

Окончательно имеем

$$P(B) = \frac{m}{n} = \frac{176}{364} = \frac{44}{91},$$

где B – событие состоящее в том, что среди трех отобранных шаров не менее 2-х белых шаров.

Теорема. Вероятность события, противоположного событию равна

Доказательство. Пусть полная система равновозможных элементарных исходов содержит n событий, из которых m ($m \le n$), благоприятны событию A. Тогда $n - m$ исходов неблагоприятны событию A, т.е. благоприятствуют событию $\overline{A}$. Таким образом,

$$P(\overline{A}) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(A).$$

Классическое определение вероятности предполагает, что
- число элементарных исходов конечно;
- эти исходы равновозможны.

Однако на практике встречаются испытания с бесконечным числом возможных исходов. Кроме того, нет общих методов, позволяющих результат испытания, даже с конечным числом исходов, представить в виде суммы равновозможных элементарных исходов. Поэтому применение классического определения вероятности весьма ограничено. ***Пример***: кубик со смещенным центром тяжести.

Классическое определение вероятности имеет ограниченную применимость. Так, оно неприемлемо, если результаты испытания не равновозможны.

Во многих случаях более удобным оказывается **статистическое определение вероятности**, которое связано с понятием *относительной частоты* появления события A в опытах. **Относительная частота** появления события A – это отношение числа m появлений события A в серии из n опытов к числу испытаний: $P^*(A) = \frac{m}{n}$.

Опыт показывает, что при проведении сравнительно малого числа испытаний относительная частота $P^*(A)$ принимает значения, которые могут сильно отличаться друг от друга. При однотипных **массовых** испытаниях во многих случаях наблюдается устойчивость относительной частоты события, т.е. с увеличением числа испытаний относительная частота колеблется около некоторого постоянного числа $P(A)$, причем эти отклонения тем меньше, чем больше произведено испытаний.

Вероятностью события в статистическом смысле называется число, относительно которого стабилизируется (устанавливается) относительная частота при неограниченном увеличении числа опытов.

Поэтому, на практике за вероятность события A принимается относительная частота $P^*(A)$ при достаточно большом числе испытаний.

Свойства вероятности, вытекающие из классического определения вероятности, сохраняются и при статистическом определении вероятности.

Если вероятность некоторого события близка к нулю, то, в соответствии со сказанным следует, что при единичном испытании в подавляющем большинстве случаев такое событие не наступит. Возникает вопрос: насколько малой должна быть вероятность, чтобы можно было пренебречь вероятностью наступления некоторого события в единичном испытании (например, землетрясение в Минске)? Достаточно малую вероятность, при которой наступление события можно считать практически невозможным, называют **уровнем значимости**. На практике уровень значимости обычно принимают равным 0,05 (пятипроцентный уровень) или 0,01 (однопроцентный уровень).

Чтобы преодолеть недостаток классического определения вероятности, связанный с его неприменимостью к испытаниям с бесконечным числом исходов, вводят понятие геометрической вероятности – вероятности попадания точки в некоторую область (отрезок, часть плоскости и т.д.).

В подобных случаях пространство элементарных исходов может быть представлено областью G, а под событием A можно понимать исходы, входящие в некоторую область g, принадлежащую области G.

Пусть на область G наугад бросается «точка». Какова вероятность того, что эта точка попадет в область g, являющуюся частью области G?

1. Пусть отрезок g длины l_g, составляет часть отрезка G длина которого l_G. На отрезок G наудачу поставлена точка. Предполагается, что

-поставленная точка может оказаться в любой точке отрезка G;

- вероятность попадания точки на отрезок g пропорциональна длине этого отрезка и не зависит от его расположения относительно отрезка G.

Тогда вероятность попадания точки на отрезок g определяется равенством $P = \frac{l_g}{l_G}$.

2. Пусть плоская фигура g с площадью S_g составляет часть плоской фигуры G, площадь которой S_G. На фигуру G наудачу брошена точка. Предполагается, что:

- брошенная точка может оказаться в любой точке фигуры G;

- вероятность попадания брошенной точки на фигуру g пропорциональна площади этой фигуры и не зависит ни от ее расположения относительно фигуры G, ни от формы g.

В этих предположениях вероятность попадания точки на фигуру g определяется равенством $P = \frac{S_g}{S_G}$.

3. Аналогично вводится понятие геометрической вероятности при бросании точки в пространственную область G объема V_G, содержащую область g объема

Множество попарно несовместных событий называют **полной группой событий**, если при любом исходе случайного эксперимента непременно наступает одно из событий, входящих в это множество. Другими словами, для полной группы событий $A_1, A_2, \dots, A_n$ выполнены следующие условия:

- появление одного из событий данного множества в результате испытания является достоверным событием, т.е. событие $A_1 + A_2 + \dots + A_n = \Omega$;

-события A_i и A_j ($i \neq j$) попарно несовместны и $A_i \cdot A_j$ – событие невозможное при любых $i \neq j$, т.е. $A_i \cdot A_j = \varnothing$.

Простейшим примером полной группы событий является пара противоположных событий A и $\overline{A}$.

Основные понятия комбинаторики.

При решении ряда теоретических и практических задач требуется из конечного множества элементов по заданным правилам составлять различные комбинации и производить подсчет числа всех возможных таких комбинаций. Такие задачи принято называть *комбинаторными*.

При решении задач комбинаторики используют правила суммы и произведения.

Правило суммы – если элемент *a* может быть выбран n способами, а элемент *b* – *m* способами, то один из этих элементов можно выбрать *n+m* способами.

Правило произведения – если элемент *a* может быть выбран n способами и после каждого такого выбора элемент *b* можно выбрать *m* способами, то пару (*ab*) из этих элементов в указанном порядке можно выбрать *nm* способами.

Упорядоченные наборы, состоящие из k различных элементов, выбранные из n данных элементов, называются **размещениями** из n элементов по k. Размещения могут отличаться как элементами, так и порядком.

Теорема. Число всех размещений из n элементов по k вычисляется по формуле:

$$A_n^k = n(n-1)(n-2)\dots(n-(k-1)) = \frac{n!}{(n-k)!}$$

Действительно, первый элемент размещения может быть выбран n способами. Для каждого из этих вариантов есть n-1 способов расположения одного из оставшихся элементов на втором месте. Следовательно, по правилу произведения, имеется n*(n-1) различных способов выбора элементов на первых двух местах. Продолжая это рассуждение по индукции, получаем доказательство.

Пример: Различными размещениями множества из трех элементов {1,2,3} по два будут наборы (1,2), (2,1), (1,3), (3,1), (2,3), (3,2)

В частном случае k=n размещения называются **перестановками** P_n.

Так как каждая перестановка содержит все ***n*** элементов множества, то различные перестановки отличаются друг от друга только порядком элементов и

$$P_n = A_n^n = n!$$

Пример: Различными перестановками множества элементов {1,2,3} будут (1,2,3), (1,3,2), (2,3,1), (2,1,3), (3,2,1), (3,1,2)

Неупорядоченные наборы из k элементов, взятых из данных n элементов, называются **сочетаниями** из n элементов по k.

Теорема. Число сочетаний из n элементов по k вычисляется по формуле

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

3. Основные теоремы теории вероятностей.

Теорема сложения вероятностей.

$$\boxed{P(A+B)=P(A)+P(B)-P(AB).}$$

Важным частным случаем этой теоремы является

Теорема сложения вероятностей для несовместных событий. *Вероятность суммы двух несовместных событий равна сумме их вероятностей, т.е.*

$$\boxed{P(A+B)=P(A)+P(B).}$$

Доказательство. Так как события A и B несовместны, то их произведение равно невозможному событию, т.е. $AB = \varnothing$. Поскольку вероятность невозможного события равна нулю**,** то из теоремы сложения вероятностей следует требуемое утверждение.

Отметим, что аналогичное утверждение справедливо для любого числа попарно несовместных событий: *вероятность суммы попарно несовместных событий равна сумме их вероятностей.*

Следствие. *Пусть события* $A_1, A_2, ..., A_k$ *образуют полную систему, тогда сумма их вероятностей равна 1 т.е.*

$$P(A_1)+P(A_2)+...+P(A_k)=1.$$

Доказательство. Из определения полной системы следует, что события $A_1, A_2, ..., A_k$, в частности, являются единственно возможными, поэтому $A_1 + A_2 + ... + A_k = E$. Тогда

$$P(A_1 + A_2 + ... + A_k) = P(E).$$

Вероятность достоверного события равна 1. События $A_1, A_2, ..., A_k$, в частности, являются попарно несовместными. Тогда из теоремы сложения вероятностей для несовместных событий следует требуемое утверждение.

Данное следствие при $k = 2$ представляет важное свойство противоположных событий: *сумма вероятностей взаимно противоположных событий равна 1, т.е.*

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

Определение. ***Условной вероятностью*** $P_B(A)$ *называется вероятность наступления события A в предположении наступления события B.*

Определение. *Два события называются* ***независимыми,*** *если вероятность наступления одного из них не зависит от того, считается ли другое событие наступившим или нет.*

Данное определение равносильно следующему:

$$\text{события } A \text{ и } B \text{ независимы} \Leftrightarrow \begin{cases} P_B(A)=P_{\overline{B}}(A), \\ P_A(B)=P_{\overline{A}}(B). \end{cases}$$

Пример. Пусть испытание состоит в извлечении карты из колоды. Событие A – извлечена «картинка», событие B – извлечена «7». Выяснить, являются ли события A и B независимыми.

Решение. Так как среди «картинок» нет «семерок», то $P_A(B)=\dfrac{0}{16}=0$**.** Так как среди «не картинок» – 4 «семерки», то $P_{\overline{A}}(B)=\dfrac{4}{36-16}=\dfrac{4}{20}=\dfrac{1}{5}$**.** Таким образом,

$P_A(B)\neq P_{\overline{A}}(B)$**,** поэтому события A и B зависимы. Аналогично, в общем случае *произвольные (неравные) несовместные события – зависимы.*

Теорема (*необходимое и достаточное условие независимости событий*). *События A и B независимы тогда и только тогда, когда*

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

Пример. Пусть испытание состоит в бросании игральной кости, $A = \{3, 4, 5\}$, $B = \{2, 3, 4, 6\}$. Выяснить, являются ли события A и B независимыми.

Решение. Очевидно, что $P(A) = \dfrac{3}{6} = \dfrac{1}{2}$. В предположении обязательного наступления события *B,* полное число возможных исходов равно 4, из которых 2 исхода благоприятствуют наступлению события *A,* поэтому $P_B(A) = \dfrac{2}{4} = \dfrac{1}{2}$. Так как $P(A) = P_B(A)$, то события A и B – независимы.

Теорема умножения вероятностей.

$$\boxed{\begin{gathered} P(AB) = P(A)P_A(B), \\ P(ABC) = P(A)P_A(B)P_{AB}(C), \\ P(ABCD) = P(A)P_A(B)P_{AB}(C)P_{ABC}(D), \end{gathered}}$$

Теорема умножения вероятностей для независимых событий. *Вероятность произведения двух независимых событий равна произведению их вероятностей, т.е.*

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

Аналогичное утверждение справедливо для любого числа независимых событий.

Пример. Два стрелка одновременно выстреливают в мишень. Вероятность попадания для первого стрелка равна 0,6, для второго – 0,8. Найти вероятность того, что в мишени будет:

а) одна пробоина;

б) хотя бы одна пробоина.

Решение. а) Прежде всего, укажем, когда может наступать интересующее нас событие, перебирая все возможные варианты.

В мишени будет одна пробоина
тогда и только тогда, когда
первый стрелок попал и второй стрелок промахнулся
или
первый стрелок промахнулся и второй стрелок попал.

Пусть событие A – в мишени будет одна пробоина, событие B_1 – первый стрелок попал, событие B_2 – второй стрелок попал. Тогда $\overline{B}_1$ – первый стрелок промахнулся, $\overline{B}_2$ – второй стрелок промахнулся. “Тогда и только тогда, когда” соответствует отношению равенства событий. **Соединительный союз “или” соответствует операции сложения событий. Соединительный союз “и” соответствует умножению событий.** Тогда фраза русского языка, в которой мы перечислили все возможности для наступления события *A,* равносильна следующему символическому равенству

$$A = B_1\overline{B}_2 + \overline{B}_1 B_2.$$

Откуда следует равенство вероятностей

$$P(A) = P(B_1\overline{B}_2 + \overline{B}_1 B_2).$$

Так как события $B_1\overline{B}_2$ и $\overline{B}_1 B_2$ несовместны, то, применяя теорему сложения вероятностей для несовместных событий, приходим к равенству

$$P(A)=P(B_1\overline{B}_2)+P(\overline{B}_1B_2).$$

События $B_1,\overline{B}_2$ и $\overline{B}_1,B_2$ попарно независимы, поэтому, применяя теорему умножения вероятностей для независимых событий, получаем

$$P(A)=P(B_1)P(\overline{B}_2)+P(\overline{B}_1)P(B_2).$$

По условию, $P(B_1)=0{,}6$ и $P(B_2)=0{,}8.$ Тогда, по свойству взаимно противоположных событий (см. следствие из теоремы сложения вероятностей для несовместных событий, $k=2$), $P(\overline{B}_1)=1-P(B_1)=1-0{,}6=0{,}4$ и $P(\overline{B}_2)=1-P(B_2)=1-0{,}8=0{,}2.$ Окончательно имеем

$$P(A)=0{,}6\cdot 0{,}2+0{,}4\cdot 0{,}8=0{,}44.$$

б) Пусть m – число попаданий в мишень, тогда искомой является вероятность $P(m\geq 1)$ (заметим, что слова «хотя бы один», «не менее чем один», «по-крайней мере один» являются синонимами). Событие $(m\geq 1)$ равносильно тому, что число попаданий в мишень будет равно 1 или 2, т.е.

$$(m\geq 1)=(m=1)+(m=2).$$

Тогда, учитывая несовместность событий $(m=1)$ и $(m=2)$, получаем

$$P(m\geq 1)=P(m=1)+P(m=2).$$

$P(m=1)=P(A)=0{,}44$ (см. п. а) данного примера). Событие $(m=2)$ (два попадания в мишень) наступает тогда и только тогда, когда первый стрелок попадет в мишень и второй стрелок попадет, т.е.

$$(m=2)=B_1B_2.$$

Поэтому

$$P(m=2)=P(B_1)P(B_2)=0{,}6\cdot 0{,}8=0{,}48$$

(см. теорему умножения вероятностей для независимых событий). Окончательно имеем

$$P(m\geq 1)=P(m=1)+P(m=2)=0{,}44+0{,}48=0{,}92.$$

Отметим, что эта задача допускает и другое решение. Так как события $(m\geq 1)$ и $(m=0)$ взаимно противоположны, то

$$P(m\geq 1)=1-P(m=0).$$

Но $P(m=0)=P(\overline{B}_1)P(\overline{B}_2)=0{,}4\cdot 0{,}2=0{,}08.$ Следовательно

$$P(m\geq 1)=1-P(m=0)=1-0{,}08=0{,}92.$$

Пример. В коробке лежат 4 белых шара и 6 красных. Наудачу, один за другим из коробки извлекается 2 шара. Найти вероятность того, что среди них будет:

а) один красный шар;

б) менее 2-х красных шаров.

Решение. а) Пусть событие A – среди двух извлеченных шаров – ровно один красный. Это событие наступает тогда и только тогда, когда первый из извлеченных шаров – красный, а второй – белый или первый шар – белый, а второй – красный. Напомним, что соединительный союз “или” соответствует сложению событий, союзы “и”, “а” соответствуют умножению событий. Тогда описание всех возможностей наступления события A равносильно следующему формальному равенству

$$A=К_1Б_2+Б_1К_2,$$

где $К_1$ ($К_2$) – первый (второй) шар – красный, $Б_1$ ($Б_2$) – первый (второй) шар – белый. События $К_1Б_2$ и $Б_1К_2$ – несовместны, поэтому, используя теорему сложения вероятностей для несовместных событий, получаем

$$P(A)=P(К_1Б_2)+P(Б_1К_2).$$

Применяя теперь теорему умножения вероятностей, приходим к равенству

$$P(A)=P(K_1)P_{K_1}(Б_2)+P(Б_1)P_{Б_1}(K_2).$$

Для вычисления вероятностей из правой части последнего равенства используем классическое определение вероятности. Тогда

$$P(A)=\frac{6}{10}\cdot\frac{4}{9}+\frac{4}{10}\cdot\frac{6}{9}=\frac{8}{15}.$$

б) Пусть m – число красных шаров среди двух извлеченных. Тогда искомой является вероятность $P(m<2)$. Очевидно, что $(m<2)=(m=0)+(m=1)$, и $P(m=1)=P(A)$ (см. п. а) данного примера). Вместе с тем, событие $(m=0)$ – среди извлеченных шаров нет красных – равносильно тому, что первый шар окажется белым и второй – также белым, т.е. $(m=0)=Б_1Б_2$, поэтому

$$P(m=0)=P(Б_1Б_2)=P(Б_1)P_{Б_1}(Б_2)=\frac{4}{10}\cdot\frac{3}{9}=\frac{2}{15}.$$

Окончательно имеем

$$P(m<2)=P(m=0)+P(m=1)=\frac{2}{15}+\frac{8}{15}=\frac{10}{15}=\frac{2}{3}.$$

Заметим, что вероятность $P(m<2)$ может быть также найдена по-другому. События $(m<2)$ и $(m=2)$ взаимно противоположны, поэтому

$$P(m<2)=1-P(m=2).$$

Но $P(m=2)=P(K_1K_2)=P(K_1)P_{K_1}(K_2)=\frac{6}{10}\cdot\frac{5}{9}=\frac{1}{3}.$

Тогда $P(m<2)=1-\frac{1}{3}=\frac{2}{3}.$

1.2 Лекция 2 (Л-2) (2 ч.)

Тема: Следствия основных теорем теории вероятностей, схема повторных испытаний

1.2.1 Вопросы лекции:

1. Условная вероятность события. Формула полной вероятности, формула Байеса.
2. Схема повторных испытаний. Формулы Бернулли, Пуассона, Лапласа.
3. Простейший поток событий, его свойства.

1.2.2. Краткое содержание вопросов:

1. Условная вероятность события. Формула полной вероятности, формула Байеса

Определение. ***Условной вероятностью*** $P_B(A)$ *называется вероятность наступления события А в предположении наступления события В.*

Определение. *Два события называются* ***независимыми,*** *если вероятность наступления одного из них не зависит от того, считается ли другое событие наступившим или нет.*

Если при наступлении события A вероятность события B не меняется, то события A и B называются **независимыми**.

События $A_1, A_2, \ldots, A_n$ называются **попарно независимыми**, если независимы любые два из них.

События $A_1, A_2, \ldots, A_n$ называются **независимыми в совокупности (или просто независимыми)**, если независимы каждые два из них и независимы каждое событие и все возможные произведения остальных.

Пусть событие A может произойти только с одним из несовместных событий $H_1, H_2, \ldots, H_n$, образующих полную группу. Например, в магазин поступает одна и та же продукция от трех предприятий и в разном количестве. Вероятность выпуска некачественной продукции на этих предприятиях различна. Случайным образом отбирается одно из изделий. Требуется определить вероятность того, что это изделие некачественное (событие A). Здесь события H_1, H_2, H_3 – это выбор изделия из продукции соответствующего предприятия.

В этом случае вероятность события A можно рассматривать как сумму произведений событий $A = \sum_{i=1}^{n} AH_i$.

По теореме сложения вероятностей несовместных событий получаем $P(A) = \sum_{i=1}^{N} P(AH_i)$. Используя теорему умножения вероятностей, находим

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

Полученная формула называется **формулой полной вероятности**

Пусть событие A происходит одновременно с одним из n несовместных событий $H_1, H_2, \ldots, H_n$, вероятности которых $P(H_i)$ ($i = 1,2,\ldots,n$) известны до опыта (*вероятности априори*). Производится опыт, в результате которого зарегистрировано появление события A, причем известно, что это событие имело определенные условные вероятности $P(A / H_i)$ ($i = 1,2,\ldots,n$). Требуется найти вероятности событий H_i если известно, что событие A произошло (*вероятности апостериори*).

Задача состоит в том, что, имея новую информацию (событие А произошло), нужно переоценить вероятности событий $H_1, H_2, \ldots, H_n$.

На основании теоремы о вероятности произведения двух событий

$$P(H_i A) = P(A)P(H_i / A) = P(H_i)P(A / H_i),$$

откуда

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

Полученная формула носит название **формулы Байеса**.

2.Схема повторных испытаний. Формулы Бернулли, Пуассона, Лапласа.

Серия повторных независимых испытаний, в каждом из которых данное событие A имеет одну и ту же вероятность $P(A) = p$, не зависящую от номера испытания, называется **схемой Бернулли**. Таким образом, в схеме Бернулли для каждого испытания имеются только два исхода: событие A (успех), вероятность которого $P(A) = p$ и событие $\overline{A}$ (неудача), вероятность которого $P(\overline{A}) = q = 1 - p$.

Для того чтобы найти вероятность появления события ровно m раз в серии k опытов, достаточно произвести перемножение сомножителей в производящей функции. Коэффициент при члене и даст искомую вероятность.

Мы предполагали, что вероятность наступления события в каждом из опытов постоянна. На практике часто приходится встречаться с более сложным случаем, когда опыты производятся в неодинаковых условиях, и вероятность события от опыта к опыту меняется. Например, производится серия выстрелов при изменяющейся дальности.

Случай непостоянной вероятности появления события в опытах

Способ вычисления вероятности заданного числа появлений событий в таких условиях дает ***общая теорема о повторении опытов***.

Пусть проводится n независимых опытов, в каждом из которых может появиться или не появиться некоторое событие A, причем вероятность появления этого события в i-м опыте равна p_i, а вероятность его не появления соответственно $q_i = 1 - p_i \quad (i = 1, \ldots, n)$. Требуется найти вероятность $P_{m,n}$ того, что в результате n опытов событие A появится ровно m раз.

Решение данной задачи проводится с помощью так называемой ***производящей функции***, имеющей вид: $\varphi_n(z) = \prod_{i=1}^{n}(q_i + p_i z)$.

Пример. Производится 4 независимых выстрела по одной и той же цели с различных расстояний. Вероятности попадания при этих выстрелах равны соответственно

$p_1 = 0{,}1 \quad p_2 = 0{,}2 \quad p_3 = 0{,}3 \quad p_4 = 0{,}4$.

Найти вероятность трех попаданий.

Решение: Составим производящую функцию

$$\varphi_n(z) = \prod_{i=1}^{n}(q_i + p_i z) = (0{,}9 + 0{,}1z)\cdot(0{,}8 + 0{,}2z)\cdot(0{,}7 + 0{,}3z)\cdot(0{,}6 + 0{,}4z) =$$

$$0{,}302 + 0{,}440z + 0{,}215z^2 + 0{,}040z^3 + 0{,}002z^4$$

Отсюда вероятность трех попаданий равна 0,040. Легко найти и вероятности ни одного, одного, двух и четырех попаданий, выписывая коэффициенты при z^0, z^1, z^2 и z^4.

Число наступлений события A называется **наивероятнейшим**, если оно имеет наибольшую вероятность по сравнению с вероятностями наступления A любое другое количество раз.

Теорема. Наивероятнейшее число наступлений события в независимых испытаниях заключено между числами $m_0' = np - q$ и $m_0'' = np + p$.

Рассмотрим задачу – частный случай задач предыдущей темы. Наблюдение над решением позволит нам получить формулу, существенно упрощающую вычисления в аналогичных случаях.

Пример. Предполагается произвести 4 выстрела по мишени. Вероятность попадания при каждом выстреле считается известной и равной 0,7. Найти вероятность того, что число попаданий в мишень будет:

а) равно 2;

б) не менее 2-х;

в) менее 4-х.

. Число выстрелов по мишени обозначим через n (здесь $n = 4$), $p = 0{,}7 -$ вероятность попадания в мишень при каждом выстреле, $q = 1 - p = 1 - 0{,}7 = 0{,}3 -$ вероятность промаха при каждом выстреле, m – число попаданий. Требуется найти $P(m = 2)$, эту же вероятность обозначим через $P_{2,4}$. Перебирая все случаи, в которых число попаданий в мишень будет равно 2, получаем

$$P_{2,4} = ppqq + pqpq + pqqp + qppq + qpqp + qqpp =$$

$$= 6p^2q^2 = 6\cdot 0{,}7^2\cdot 0{,}3^2 = 0{,}2646.$$

Теорема. *Пусть произведено n повторных независимых испытаний, в каждом из которых некоторое событие A наступает с вероятностью p. Тогда вероятность $P_{m,n}$ того, что в этих n испытаниях событие A наступит m раз, вычисляется по формуле*

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

где C_n^m – число сочетаний из n по m , $q = 1 - p$.

Полученная формула носит название *формулы Бернулли.*

Завершим рассмотрение нашего примера.

б) Так как $(m \ge 2) = (m = 2) + (m = 3) + (m = 4),$ то, применяя теорему сложения вероятностей для несовместных событий, получаем

$$P(m \ge 2) = P(m = 2) + P(m = 3) + P(m = 4) = P_{2,4} + P_{3,4} + P_{4,4}.$$

Первое слагаемое последней суммы найдено в п. а) данного примера. Аналогично для остальных:

$$P_{3,4} = C_4^3 p^3 q^1 = \frac{4!}{3!\cdot 1!}\cdot 0{,}7^3\cdot 0{,}3 = 4\cdot 0{,}7^3\cdot 0{,}3 = 0{,}4116,$$

$$P_{4,4} = C_4^4 p^4 q^0 = \frac{4!}{4!\cdot 0!}\cdot 0{,}7^4\cdot 0{,}3^0 = 1\cdot 0{,}7^4\cdot 1 = 0{,}2401.$$

Окончательно имеем

$$P(m \ge 2) = 0{,}2646 + 0{,}4116 + 0{,}2401 = 0{,}9163.$$

в) По аналогии с предыдущим пунктом задания,

$$P(m < 4) = P_{0,4} + P_{1,4} + P_{2,4} + P_{3,4},$$

т.е. решение требует, вообще говоря, четырех применений формулы Бернулли. Однако возможно и более короткое решение. Действительно, события $(m < 4)$ и $(m = 4)$ – взаимно противоположны, следовательно

$$P(m < 4) = 1 - P(m = 4).$$

Вероятность $P(m = 4) = P_{4,4}$ найдена в п. б) примера. Таким образом, получаем

$$P(m < 4) = 1 - P_{4,4} = 1 - 0{,}2401 = 0{,}7599.$$

Если число испытаний велико, формулу Бернулли применять неудобно. В этом случае можно применять приближенные формулы, точность которых увеличивается с возрастанием n.

Теорема Пуассона.

Теорема. *Пусть произведено n повторных независимых испытаний, в каждом из которых некоторое событие A наступает с вероятностью p , причем*

а) число испытаний достаточно велико ($n \ge 100$);

б) $\lambda = np \le 10.$

Тогда вероятность $P_{m,n}$ того, что в этих n испытаниях событие A наступит m раз, вычисляется по следующей приближенной формуле

$$P_{m,n} = \frac{\lambda^m}{m!}e^{-\lambda}.$$

Эта формула и называется *формулой Пуассона (редких событий).*

Пример. По каналу связи передано 1000 сигналов. Вероятность ошибки при передаче каждого из сигналов равна 0,005. Найти вероятность того, что неверно передано:

а) 7 сигналов;

б) не менее 4-х сигналов.

Решение. а) Воспользуемся формулой Пуассона, т.к. условия ее применимости в данном случае выполнены: число испытаний достаточно велико $(n = 1000 \geq 100)$ и $\lambda = np = 1000 \cdot 0{,}005 = 5 \leq 10.$ Искомое значение $P_{7,1000}$ найдем по таблице функции Пуассона при $m = 7$ и $\lambda = 5$ (см. учебник Н.Ш. Кремера, с.556): $P_{7,1000} = 0{,}1045.$

б) Требуется найти $P(m \geq 4)$, где m – число неверно принятых сигналов. Так как $(m \geq 4) = (m = 4) + (m = 5) + ... + (m = 1000),$ то $P(m \geq 4) = P_{4,1000} + P_{5,1000} + ... + P_{1000,1000}.$

Искать каждое из слагаемых этой суммы и затем выполнять суммирование – такое решение не представляется рациональным из-за большого числа слагаемых и потому, что таблица функции Пуассона не дает искомых значений с требуемой в данном случае точностью. Воспользуемся переходом к противоположному событию: $P(m \geq 4) = 1 - P(m < 4) = 1 - (P_{0,1000} + P_{1,1000} + P_{2,1000} + P_{3,1000}).$

Находя вероятности из правой части последнего равенства по таблице функции Пуассона, окончательно получаем

$$P(m \geq 4) = 1 - (0{,}0067 + 0{,}0337 + 0{,}0842 + 0{,}1404) = 0{,}735.$$

Формула Пуассона находит применение в теории массового обслуживания. Она может рассматриваться как математическая модель *простейшего потока событий* с интенсивностью $\lambda = np$. Параметр $\lambda = np$ представляет при этом *среднее число успехов*.

Теоремы Лапласа

Лапласом была получена важная приближенная формула для вероятности $P_n(m)$ появления события A точно m раз, при условии, что n достаточно велико. В отличие от формулы Пуассона здесь нет ограничения на малость величины p в отдельном испытании, т.е. область применимости формулы Лапласа шире.

Теорема. *Пусть произведено n повторных независимых испытаний, в каждом из которых некоторое событие A наступает с вероятностью p , причем число испытаний достаточно велико (* $n \geq 100$ *). Тогда вероятность* $P_{m,n}$ *того, что в этих n испытаниях событие A наступит* m *раз, вычисляется по следующей приближенной формуле*

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

где $f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$ *– функция Гаусса,* $x = \frac{m - np}{\sqrt{npq}}$, $q = 1 - p$.

Пример. Имеется партия деталей, состоящая из 1000 штук. В среднем среди деталей такого вида стандартные детали составляют 90%. Найти вероятность того, что число стандартных деталей в данной партии окажется равным 890.

Решение. Число испытаний в данном случае достаточно велико $(n = 1000 \geq 10)$, поэтому локальная теорема Муавра-Лапласа применима. Из условия следует, что вероятность быть стандартной для произвольной детали данной партии равна

$p = \frac{90}{100} = 0{,}9$, $q = 1 - p = 1 - 0{,}9 = 0{,}1$, $m = 890$. Тогда

$$x = \frac{m - np}{\sqrt{npq}} = \frac{890 - 1000 \cdot 0{,}9}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}} = -1{,}05.$$

По локальной теореме Муавра-Лапласа,

$$P_{890,1000} = \frac{f(-1{,}05)}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}}.$$

Учитывая, что функция Гаусса четная, используя таблицу этой функции, находим $f(-1{,}05) = f(1{,}05) = 0{,}2299.$ Окончательно, получаем

$$P_{890,1000} = \frac{0,2299}{\sqrt{1000 \cdot 0,9 \cdot 0,1}} = 0,0242.$$

Интегральная теорема Муавра-Лапласа содержит приближенную формулу для вероятности $P_n(m_1, m_2)$ того, что событие A появится не менее m_1 раз и не более m_2 раз.

Теорема. *Пусть произведено n повторных независимых испытаний, в каждом из которых некоторое событие А наступает с вероятностью p , причем число испытаний достаточно велико* ($n \geq 100$). *Тогда вероятность того, что число m наступлений события А в этих n испытаниях будет заключено в границах от* m_1 *до* m_2, *вычисляется по следующей приближенной формуле*

$$P(m_1 \leq m \leq m_2) = \frac{1}{2}\left(\Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)\right),$$

где $\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} e^{-x^2/2} dx$ – *функция Лапласа,* $q = 1 - p$.

Пример. Каждая из 1000 деталей партии стандартна с вероятностью 0,9. Найти вероятность того, что число стандартных деталей этой партии будет не меньше 880.

Решение. Число *n* повторных независимых испытаний в данном случае равно числу деталей в партии (каждая из деталей партии будет проверяться на предмет качества, а в этой проверке и состоит испытание). $n = 1000 \geq 100$, поэтому интегральная теорема Муавра-Лапласа применима; неравенство $(m \geq 880)$, где m – число стандартных деталей в партии, здесь равносильно $(880 \leq m \leq 1000)$, поэтому $m_1 = 880$, $m_2 = 1000$; $p = 0,9$, $q = 1 - p = 1 - 0,9 = 0,1$; $np = 1000 \cdot 0,9 = 900$; $npq = 1000 \cdot 0,9 \cdot 0,1 = 90$. Тогда

$$P(880 \leq m \leq 1000) = \frac{1}{2}\left(\Phi\left(\frac{1000 - 900}{\sqrt{90}}\right) - \Phi\left(\frac{880 - 900}{\sqrt{90}}\right)\right) =$$

$$= \frac{1}{2}\left(\Phi(10,5) - \Phi(-2,11)\right).$$

По свойствам функции Лапласа (см. ниже), $\Phi(10,5) = 1$, $\Phi(-2,11) = -\Phi(2,11)$. По таблице функции Лапласа (см. учебник Н.Ш. Кремера, с. 555) находим $\Phi(2,11) = 0,9651$. Тогда окончательно имеем

$$P(880 \leq m \leq 1000) = \frac{1}{2}(1 + \Phi(2,11)) = \frac{1}{2}(1 + 0,9651) = 0,9826.$$

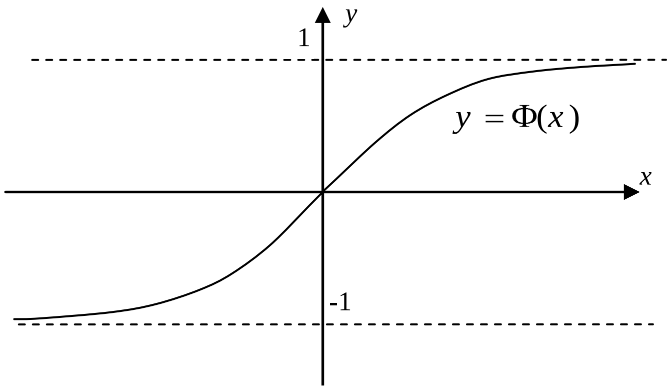


Рис. 2

Свойства функции Лапласа

1. Функция Лапласа нечетна: $\Phi(-x) = -\Phi(x)$.
2. Функция Лапласа – монотонно возрастающая;
3. $\lim\limits_{n \to +\infty} \Phi(x) = 1$, $\lim\limits_{n \to -\infty} \Phi(x) = -1$, т.е. прямые $y = 1$ и $y = -1$ являются горизонтальными асимптотами (правой и левой соответственно) графика $y = \Phi(x)$; на практике полагаем $\Phi(x) \approx 1$ при $x \geq 4$.

График функции Лапласа схематично изображен на рис. 2.

Пример. Имеется партия деталей, состоящая из 1000 штук. В среднем среди деталей такого вида стандартные детали составляют 90%. Найти вероятность того, что число стандартных деталей в данной партии окажется равным 890.

Решение. Число испытаний в данном случае достаточно велико $(n = 1000 \geq 10)$, поэтому локальная теорема Муавра-Лапласа применима. Из условия следует, что вероятность быть стандартной для произвольной детали данной партии равна

$p = \frac{90}{100} = 0{,}9$, $q = 1 - p = 1 - 0{,}9 = 0{,}1$, $m = 890$. Тогда

$$x = \frac{m - np}{\sqrt{npq}} = \frac{890 - 1000 \cdot 0{,}9}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}} = -1{,}05.$$

По локальной теореме Муавра-Лапласа,

$$P_{890,1000} = \frac{f(-1{,}05)}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}}.$$

Учитывая, что функция Гаусса четная, используя таблицу этой функции, находим $f(-1{,}05) = f(1{,}05) = 0{,}2299.$ Окончательно, получаем

$$P_{890,1000} = \frac{0{,}2299}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}} = 0{,}0242.$$

3. Простейший поток событий, его свойства.

Потоком событий называют последовательность событий, которые наступают в случайные моменты времени. **Интенсивностью потока** λ называют среднее число событий в единицу времени. Простейшим (пуассоновским) называют поток событий, который обладает свойствами *стационарности*, *отсутствия последействий* и *ординарности*.

Свойство стационарности характеризуется тем, что вероятность $P_t(m)$ появления m событий на любом промежутке времени зависит только от числа m и от длительности промежутка времени t и не зависит от начала его отсчёта.

Свойство отсутствия последействия характеризуется тем, что вероятность появления m событий на любом промежутке времени не зависит от того, появлялись или не появлялись события в моменты времени, предшествующие началу рассматриваемого промежутка, т.е. предыстория потока не сказывается на вероятности появления событий.

Свойство **ординарности** характеризуется тем, что появление двух и более событий за малый промежуток времени маловероятно по сравнению с вероятностью появления только одного события.

Если интенсивность простейшего потока λ известна, то вероятность появления m событий за время t определяется формулой

$$P_t(m) = \frac{(\lambda t)^m \cdot e^{-\lambda t}}{m!}$$

Пример простейшего потока событий. Среднее число заказов такси, поступающих на диспетчерский пункт в одну минуту, равно трем. Найти вероятность того, что за 2 минуты поступит 4 вызова.

Подставляя в вышеприведенную формулу $\lambda = 3,\ t = 2,\ m = 4$, получим $P_2(4) = 0.135$

1.3 Лекция 3-4 (Л-3-4) (4 ч.)

Тема: Случайные величины, их классификация, законы распределения, числовые характеристики.

1.3.1 Вопросы лекции:

1. Случайные величины, их классификация, закон распределения.
2. Функция распределения, плотность распределения, вероятность попадания в интервал.
3. Числовые характеристики ДСВ. Числовые характеристики НСВ.

1.3.2. Краткое содержание вопросов:

1. Случайные величины, их классификация, закон распределения.

Определение. ***Случайной величиной*** *называется переменная, которая в результате испытания принимает то или иное числовое значение.*

Пример. Число попаданий в мишень при n выстрелах – случайная величина.

Пример. Рост наудачу взятого человека – случайная величина.

Определение. *Случайная величина называется* ***дискретной****, если число ее возможных значений конечно или счетно., в противном случае – она является недескретной.* (Напомним, что множество называется *счетным*, если его элементы можно перенумеровать натуральными числами.)

В этом смысле, число попаданий в мишень – пример дискретной случайной величины. Рост человека – непрерывная случайная величина (такие случайные величины будут рассмотрены ниже).

Для обозначения случайных величин будем использовать заглавные буквы латинского алфавита (возможно с индексами), например, $X, Y, Z, \ldots,\ X_1, Y_2, Z_3, \ldots$ и т.п.

Соответствие между всеми возможными значениями дискретной случайной величины и их вероятностями, т.е. совокупность пар чисел $(x_i; p_i)$ называется **законом распределения** данной случайной величины.

Закон распределения можно задавать таблично, аналитически (в виде формулы) и графически.

Определение. ***Законом распределения*** *дискретной случайной величины называется такая таблица, в которой перечислены все возможные значения этой случайной величины (без повторений) с соответствующими им вероятностями.*

В общем виде закон распределения для случайной величины, например, X :

X :

x_i	x_1	x_2	$\ldots$	x_k
p_i	p_2	p_2	$\ldots$	p_k

где $p_i = P(X = x_i),\ i = 1, 2, \ldots, k$.

Из определения закона распределения следует, что события $(X = x_1)$, $(X = x_2), \ldots, (X = x_k)$ образуют полную систему, поэтому (см. следствие из теоремы сложения вероятностей для несовместных событий в §1.6):

$$P(X = x_1) + P(X = x_2) + \ldots + P(X = x_k) = 1,$$

т.е.

$$\boxed{p_1 + p_2 + \ldots + p_k = 1.}$$

Данное равенство называется *основным свойством закона распределения*

Пример. Два стрелка одновременно выстреливают в мишень. Вероятность попадания для первого равна 0,6, для второго – 0,8. Составить закон распределения случайной величины Z – общего числа попаданий в мишень.

Решение. Возможные значения данной случайной величины: 0, 1, 2. Через B_1 и B_2 обозначим события, состоящие в попадании в мишень первого и второго стрелков (соответственно). Тогда аналогично упомянутому примеру получаем

$$P(Z=0)=P(\overline{B}_1)P(\overline{B}_2)=(1-0{,}6)\cdot(1-0{,}8)=0{,}4\cdot 0{,}2=0{,}08,$$

$$P(Z=1)=P(B_1)P(\overline{B}_2)+P(\overline{B}_1)P(B_2)=0{,}6\cdot 0{,}2+0{,}4\cdot 0{,}8=0{,}44,$$

$$P(Z=2)=P(B_1)P(B_2)=0{,}6\cdot 0{,}8=0{,}48.$$

Окончательно, закон распределения случайной величины Z имеет вид:

Z :

z_i	0	1	2	Σ
p_i	0,08	0,44	0,48	1

Упражнение. В коробке 3 белых шара и 2 красных. Составить закон распределения случайной величины X – числа белых шаров среди 2-х извлеченных шаров.

Ответ.

X :

x_i	0	1	2	Σ
p_i	0,1	0,6	0,3	1

Пример. В коробке – 3 белых шара и 2 красных. Шары извлекаются последовательно до появления белого шара. Составить закон распределения случайной величины X – числа извлеченных шаров.

Решение. Возможные значения данной случайной величины: 1, 2, 3. Событие $(X=1)$ (из коробки будет извлечен один единственный шар) наступает тогда и только тогда, когда первый из шаров оказывается белым, т.к. появление именно белого шара является сигналом к прекращению последующих извлечений (см. условие). Поэтому

$$P(X=1)=P(Б_1)=\frac{3}{5},$$

где событие $Б_1$ – первый из извлеченных шаров – белый. Событие $(X=2)$ (из коробки будет извлечено ровно 2 шара) наступает тогда и только тогда, когда первый из извлеченных шаров оказывается красным, а второй – белым. Поэтому

$$P(X=2)=P(К_1Б_2)=P(К_1)P_{К_1}(Б_2)=\frac{2}{5}\cdot\frac{3}{4}=\frac{3}{10},$$

где событие $К_1$ – первый из извлеченных шаров – красный, $Б_2$ – второй шар – белый. Наконец событие $(X=3)$ (из коробки будет извлечено 3 шара) наступает тогда и только тогда, когда первый шар – красный, второй – красный и третий – белый. Поэтому

$$P(X=3)=P(К_1К_2Б_3)=P(К_1)P_{К_1}(К_2)P_{К_1К_2}(Б_3)=\frac{2}{5}\cdot\frac{1}{4}\cdot\frac{3}{3}=\frac{1}{10}.$$

Окончательно искомый закон распределения имеет вид:

X:

x_i	1	2	3	Σ
p_i	0,6	0,3	0,1	1

Определение. *Случайная величина X имеет* ***биномиальный закон распределения*** *с параметрами n и p, если ее закон распределения имеет вид*:

X:

x_i	0	1	2	…	n
p_i	$P_{0,n}$	$P_{1,n}$	$P_{2,n}$	…	$P_{n,n}$

,

где вероятности $P_{m,n}$ *вычисляются по формуле Бернулли:*

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

n *– положительное целое число,* $m = 0,1,2,...,n,\quad 0 < p < 1.$

В пределе при $n \to \infty$ и $\lambda = np = const$ биномиальное распределение переходит в так называемое распределение Пуассона.

Определение. *Говорят, что случайная величина X имеет* ***распределение Пуассона*** *с параметром* λ*, если ее закон распределения имеет вид:*

X:

x_i	0	1	2	…
p_i	P_0	P_1	P_2	...

,

где

$$P_m = \frac{\lambda^m}{m!}e^{-\lambda},$$

$m = 0,1,2,\ldots,$ λ *– положительное число.*

Убедимся в том, что для распределения Пуассона выполняется основное свойство закона распределения: $\sum\limits_{m=0}^{\infty} P_m = 1$. Действительно, имеем

$$\sum_{m=0}^{\infty} P_m = \frac{\lambda^0}{0!}e^{-\lambda} + \frac{\lambda^1}{1!}e^{-\lambda} + \frac{\lambda^2}{2!}e^{-\lambda} + \frac{\lambda^3}{3!}e^{-\lambda} + ... = \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + ...\right)\cdot e^{-\lambda} = e^{\lambda}\cdot e^{-\lambda} = 1$$

(см. курс математического анализа, разложение функции $y = e^x$ в ряд Маклорена).

Арифметические операции над случайными величинами

Определение. *Случайные величины X и Y называются* ***равными****, если их законы распределения точно совпадают, и для произвольного числа* α *справедливо равенство:* $(X = \alpha) = (Y = \alpha)$.

Пример. Пусть законы распределения случайных величин *X и Y* имеют вид:

X:

x_i	0	1
p_i	0,5	0,5

Y:

y_i	0	1
p_i	0,5	0,5

.

Эти случайные величины равны, если дополнительно справедливы равенства $(X = 0) = (Y = 0)$ и $(X = 1) = (Y = 1)$, т.е. случайная величина *X* принимает значение 0 тогда и только тогда, когда случайная величина *Y* принимает значение 0, и аналогично со значением 1.

Произвольная случайная величина допускает умножение на число. Действительно, пусть закон распределения случайной величины *X* имеет вид:

X :

x_i	x_1	x_2	…	x_k
p_i	p_2	p_2	…	p_k

и α – некоторое число.

Определение. *Случайной величиной* $Y = \alpha \cdot X$ *называется такая случайная величина, закон распределения которой имеет вид:*

Y :

y_i	$\alpha \cdot x_1$	$\alpha \cdot x_2$	$\cdots$	$\alpha \cdot x_k$
p_i	p_2	p_2	$\cdots$	p_k

Пример. Пусть закон распределения случайной величины X имеет вид:

X:

x_i	0	1	2
p_i	0,16	0,48	0,36

и $\alpha = 5$, $Y = \alpha \cdot X$. Тогда закон распределения Y :

Y :

y_i	0	5	10
p_i	0,16	0,48	0,36

Можно придумать, например, следующую интерпретацию данному примеру. Заметим, что X – биномиально распределена с параметрами $n = 2$, $p = 0{,}6$. Пусть X – число попаданий в мишень при 2-х выстрелах, при каждом из которых попадание случается с вероятностью 0,6, и дополнительно известно, что за каждое попадание стрелку выплачивается вознаграждение в размере 5 ден. ед. Тогда Y – заработок стрелка.

Определение. *Случайные величины X и Y называются **независимыми**, если для любых i и j события $(X = x_i)$ и $(Y = y_j)$ – независимы.*

Пример. Пусть из коробки, в которой – 6 белых и 8 красных шаров, извлекается 1 шар. Рассмотрим случайные величины X – число белых шаров, Y – число красных шаров из извлеченных. События, например, $(X = 1)$ и $(Y = 1)$ – несовместны, а поэтому – зависимы. Следовательно, и случайные величины X и Y зависимы.

Определение. ***Суммой (разностью, произведением)*** *случайных величин X и Y называется такая случайная величина $Z = X + Y$ ($Z = X - Y$, $Z = X \cdot Y$), которая принимает значение z_k в некотором испытании, если значения x_i и y_j случайных величин X и Y в этом испытании таковы, что $z_k = x_i + y_j$ ($z_k = x_i - y_j$, $z_k = x_i \cdot y_j$).*

Пример. Пусть заданы законы распределения независимых случайных величин X и Y:

X:

x_i	0	1
p_i	0,4	0,6

Y:

y_j	0	1
p_j	0,2	0,8

Составить закон распределения случайной величины $U = X - Y$.

Решение. Удобно использовать вспомогательную таблицу вида:

y_j \ x_i	0	1
0	0	1
1	–1	0

в каждой из центральных клеток которой записаны соответствующие произведения случайных величин X и Y. Такая таблица показывает, какие значения принимает случайная величина U и когда она принимает эти значения. Так $U = 0$ тогда и только тогда, когда $X = 0$ и $Y = 0$ или $X = 1$ и $Y = 1$. Поэтому

$$P(U=0)=P((X=0)(Y=0)+(X=1)(Y=1)).$$

Применяя теорему сложения вероятностей для несовместных событий, теорему умножения вероятностей – для независимых событий (по условию, случайные величины X и Y – независимы), получаем

$$P(U=0)=P(X=0)P(Y=0)+P(X=1)P(Y=1)=0{,}4\cdot 0{,}2+0{,}6\cdot 0{,}8=0{,}56.$$

Для наступления каждого из двух оставшихся значений случайной величины U (-1 и 1) имеется по одной возможности. Например, $U=1$ тогда и только тогда, когда $X=1$ и $Y=0$. Тогда получаем:

$$P(U=1)=P((X=1)(Y=0))=P(X=1)P(Y=0)=0{,}6\cdot 0{,}2=0{,}12.$$

Аналогично,

$$P(U=-1)=P((X=0)(Y=1))=P(X=0)P(Y=1)=0{,}4\cdot 0{,}8=0{,}32.$$

Окончательно, закон распределения случайной величины U имеет вид:

U:

u_i	–1	0	1
p_i	0,32	0,56	0,12

2. Функция распределения, плотность распределения, вероятность попадания в интервал.

Функция распределения дискретной случайной величины

Определение. ***Функцией распределения*** *случайной величины X называется такая функция* $F(x)$, *значение которой в точке x численно равно вероятности того, что в произвольном испытании значение случайной величины X окажется меньше чем x, т.е.*

$$\boxed{F(x)=P(X<x).}$$

Данное определение задает функцию распределения не только для дискретных, но и для непрерывных случайных величин.

Пример. Пусть закон распределения случайной величины *X* имеет вид

X:

x_i	1	2
p_i	0,3	0,7

Найти функцию распределения этой случайной величины.

Решение. Найдем сначала *F(x)* для некоторых значений переменной *x*. Например,

$$F(0)=P(X<0)=P(\varnothing)=0,$$

так как данная случайная величина не имеет значений меньших нуля, а потому событие (X<0) для нее является невозможным. Аналогично, при любом значении переменной *x*, которое менее или равно 1, будем иметь $F(x)=0.$ Далее имеем:

$$F(1{,}5)=P(X<1{,}5)=P(X=1)=0{,}3.$$

Аналогично, при любом значении переменной *x* таком, что $1<x\le 2$, будем иметь $F(x)=0{,}3.$

$$F(2{,}5)=P(X<2{,}5)=P(X=1)+P(X=2)=0{,}3+0{,}7=1.$$

(Или, другими словами, так как все значения данной случайной величины менее 2,5, то событие (X < 2,5) является достоверным, а потому его вероятность равна 1.) Аналогично, при любом значении переменной *x*, которое более или равно 2, будем иметь $F(x)=1.$

Окончательно имеем:

$$F(x)=\begin{cases}0 & \text{при } x\le 1;\\ 0{,}3 & \text{при } 1<x\le 2;\\ 1 & \text{при } x>2.\end{cases}$$

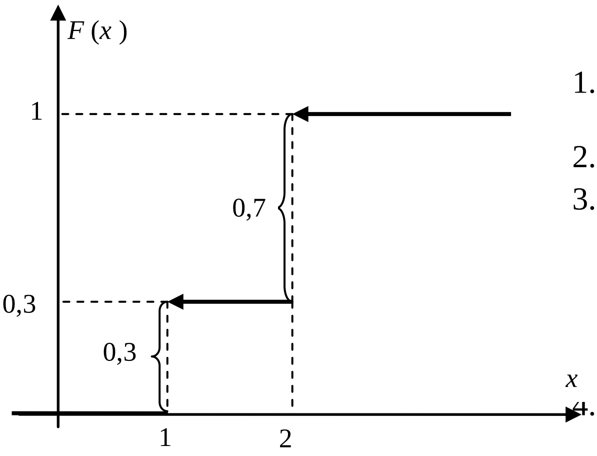


Рис. 3

График найденной функции распределения изображен на рис. 3.

Свойства функции распределения

1. Функция распределения является неубывающей функцией.
2. Область значений: $0 \le F(x) \le 1$.
3. Асимптотические свойства: $\lim\limits_{x\to-\infty} F(x) = 0, \quad \lim\limits_{x\to+\infty} F(x) = 1$ (другими словами, прямые $y=0$ и $y=1$ являются асимптотами (левой и правой соответственно) графика $y=F(x)$).
4. Вероятность того, что в произвольном испытании значение случайной величины X будет принадлежать полуинтервалу $[\alpha, \beta)$, где α и β – произвольные числа, вычисляется по формуле

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

Доказательство. Значение функции распределения равна вероятности соответствующего события, но область значений вероятности есть отрезок $[0,1]$ – тем самым доказано свойство 2.

Используя определение функции распределения, получаем $\lim\limits_{x\to-\infty} F(x) = \lim\limits_{x\to-\infty} P(X < x) = P(X < -\infty)$. Но произвольное значение случайной величины принадлежит числовой прямой, поэтому событие $(X < -\infty)$ является невозможным. Вероятность невозможного события равна нулю (см. § 1.3), поэтому $\lim\limits_{x\to-\infty} F(x) = 0$.

Аналогично, учитывая, что событие $(X < +\infty)$ является достоверным, а вероятность такого события равна 1, получаем $\lim\limits_{x\to+\infty} F(x) = 1$.

Нетрудно видеть, что

$$(X < \beta) = (X < \alpha) + (\alpha \le X < \beta),$$

причем события правой части этого равенства несовместны. Принимая во внимание определение функции распределения и теорему сложении вероятностей для несовместных событий, получаем

$$F(\beta) = P(X < \beta) = P(X < \alpha) + P(\alpha \le X < \beta) = F(\alpha) + P(\alpha \le X < \beta),$$

что равносильно свойству 4.

Доказательство свойства 1 мы оставляем читателю в качестве упражнения (указание: используйте рассуждении от противного и свойство 4).

Неформально говоря, случайная величина непрерывна, если ее значения полностью заполняют некоторый интервал. Более точно, справедливо

Определение. *Случайная величина называется* ***непрерывной****, если ее функция распределения непрерывна на всей числовой прямой и дифференцируема при всех x за исключением, быть может, отдельных значений.*

Определение. ***Плотностью распределения*** *непрерывной случайной величины X называется такая функция* $\varphi = \varphi(x)$, *что вероятность того, что в произвольном испытании значение случайной величины X окажется принадлежащим некоторому отрезку* $[\alpha, \beta]$, *вычисляется по формуле*

$$P(\alpha \le X \le \beta) = \int\limits_{\alpha}^{\beta} \varphi(x)\,dx.$$

Принимая во внимание геометрический смысл определенного интеграла, получаем

Геометрический смысл плотности распределения. *Вероятность того, что в произвольном испытании значение случайной величины X окажется принадлежащим некоторому отрезку* $[\alpha, \beta]$, *численно равна площади* $S(\alpha, \beta)$ *под кривой плотности распределения на данном отрезке* (см. рис. 4).

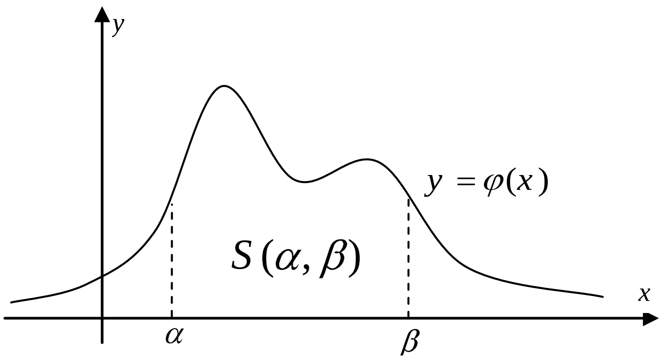


Рис. 4

Пример. Пусть плотность распределения случайной величины X имеет вид:

$$\varphi(x)=\begin{cases}1/2 & \text{при } x\in[-1,1];\\ 0 & \text{в остальных случаях.}\end{cases}$$

Найти вероятности:
а) $P(-2\le X\le -0{,}4)$; б) $P(X\le -3)$; в) $P(X\ge -2)$.

Решение. а) По определению плотности распределения,

$$P(-2\le X\le -0{,}4)=\int_{-2}^{-0{,}4}\varphi(x)dx.$$

Вместе с тем, данная плотность распределения задана аналитически по-разному на промежутках $[-2,-1]$ и $[-1;-0{,}4]$ отрезка интегрирования. Соответственно, используя свойства определенного интеграла, получаем

$$P(-2\le X\le -0{,}4)=\int_{-2}^{-0{,}4}\varphi(x)dx=\int_{-2}^{-1}0dx+\int_{-1}^{-0{,}4}1/2dx=\frac{1}{2}x\Big|_{-1}^{-0{,}4}=\frac{1}{2}(-0{,}4-(-1))=0{,}3.$$

По геометрическому смыслу плотности распределения, полученная вероятность численно равна площади под кривой плотности распределения (см. рис. 5) на отрезке $[-2;-0{,}4]$, т.е. равна площади фигуры, составленной из отрезка длины 1 и прямоугольника со сторонами $1/2$ и 0,6.

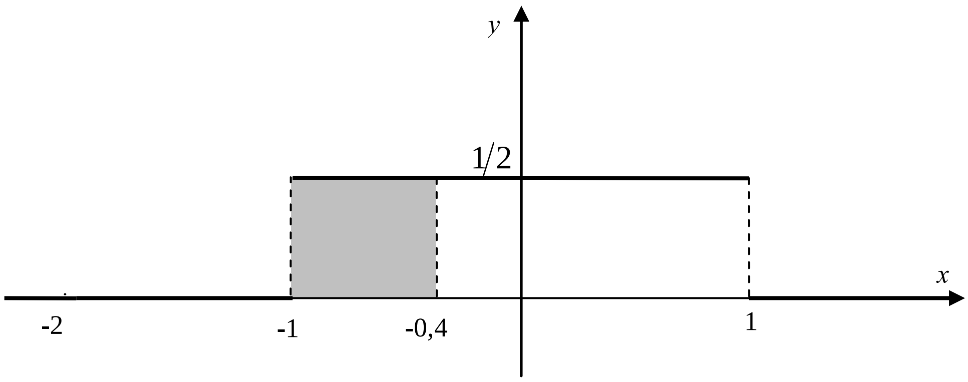


Рис. 5

б) Неравенство $(X\le -3)$ равносильно тому, что $(-\infty<X\le -3)$. Учитывая, что на промежутке $(-\infty;-3)$ данная плотность распределения равна 0, получаем

$$P(X\le -3)=P(-\infty<X\le -3)=\int_{-\infty}^{-3}0dx=0.$$

в) Аналогично предыдущим пунктам задачи, имеем

$$P(-2 \le X < +\infty) = \int_{-2}^{+\infty} \varphi(x)dx = \int_{-2}^{-1} 0dx + \int_{-1}^{1} 1/2dx + \int_{1}^{+\infty} 0dx = \frac{1}{2}x\Big|_{-1}^{1} = \frac{1}{2}(1-(-1)) = 1.$$

Рассмотрение геометрического смысла результатов последних двух пунктов данного примера мы оставляем читателю в качестве упражнения. ▸

Свойства плотности распределения

1. Плотность распределения неотрицательна, т.е. $\varphi(x) \ge 0$ при всех x.
2. Интеграл от плотности распределения на всей числовой прямой равен 1, т.е.

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

(Данное свойство называется *условием нормировки* плотности распределения.)

Доказательство. Предположим противное: пусть найдется такой отрезок $[\alpha, \beta]$, что плотность распределения $\varphi(x)$ отрицательна на этом отрезке. Тогда (см. свойства определенного интеграла) имеем

$$\int_{\alpha}^{\beta} \varphi(x)dx < 0.$$

Но, по определению плотности распределения, интеграл, стоящий в левой части последнего неравенства равен $P(\alpha \le X \le \beta)$. Так как вероятность события не может быть отрицательной, приходим к противоречию, что доказывает справедливость свойства 1.

По определению плотности распределения,

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

Но событие $(-\infty < X < +\infty)$ является достоверным, поэтому его вероятность равна 1. Тем самым доказано свойство 2.

Парадокс нулевой вероятности

Теорема. *Для непрерывной случайной величины вероятность принять произвольное числовое значение равно нулю.*

Доказательство. Пусть α – произвольное число. События $(X = \alpha)$ и $(\alpha \le X \le \alpha)$ – равны, поэтому, по определению плотности распределения, получаем

$$P(X = \alpha) = \int_{\alpha}^{\alpha} \varphi(x)dx = 0$$

(см. свойства определенного интеграла).

Из парадокса нулевой вероятности вытекает, что для любой непрерывной случайной величины вероятности попадания в произвольный отрезок числовой оси или в соответствующий полуинтервал (интервал) равны между собой, т.е. справедливо

Следствие. *Пусть X непрерывная случайная величина и α, β – произвольные числа. Тогда верно следующее равенство*

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

Доказательство. Очевидно, что

$$(\alpha \le X \le \beta) = (\alpha \le X < \beta) + (X = \beta),$$

причем события $(\alpha \le X < \beta)$ и $(X = \beta)$ – несовместны. Используя последнее равенство и теорему сложения вероятностей для несовместных событий, получаем

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

Но, согласно парадоксу нулевой вероятности, $P(X = \beta) = 0$. Тем самым доказано первое из трех равенств Следствия.

Доказательство оставшихся двух равенств мы оставляем читателю в качестве упражнения.

Функция распределения непрерывной случайной величины

Пусть X – непрерывная случайная величина и $\varphi=\varphi(x)$ – ее плотность распределения. Используя определения функции распределения (см. § 3.4) и плотности распределения, получаем

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

Обратно, если задана функция распределения непрерывной случайной величины, то (см. теорему об интеграле с переменным верхним пределом) плотность распределения этой случайной величины будет определяться равенством

$$\varphi(x)=F'(x).$$

Таким образом, имеется два равноправных способа задания непрерывной случайной величины: с помощью или плотности распределения, или функции распределения.

Пример. Пусть плотность распределения непрерывной случайной величины X имеет вид:

$$\varphi(x)=\begin{cases}\dfrac{1}{2} & \text{при } x\in[0;2],\\ 0 & \text{в остальных случаях.}\end{cases}$$

Найти функцию распределения.

Решение. Пусть $x<0$. Тогда

$$F(x)=\int_{-\infty}^{x}\varphi(x)dx=\int_{-\infty}^{x}0dx=0.$$

Если $x\in[0;2]$, то

$$F(x)=\int_{-\infty}^{x}\varphi(x)dx=\int_{-\infty}^{0}0dx+\int_{0}^{x}\frac{1}{2}dx=0+\frac{1}{2}x\Big|_{0}^{x}=\frac{1}{2}x.$$

Если $x>2$, то

$$F(x)=\int_{-\infty}^{x}\varphi(x)dx=\int_{-\infty}^{0}0dx+\int_{0}^{2}\frac{1}{2}dx+\int_{2}^{x}0dx=0+\frac{1}{2}x\Big|_{0}^{2}+0=\frac{1}{2}(2-0)=1.$$

Таким образом, окончательно, искомая функция распределения имеет вид

$$F(x)=\begin{cases}0 & \text{при } x<0,\\ \dfrac{1}{2}x & \text{при } x\in[0;2],\\ 1 & \text{при } x>2\end{cases}$$

(см. рис. 6).

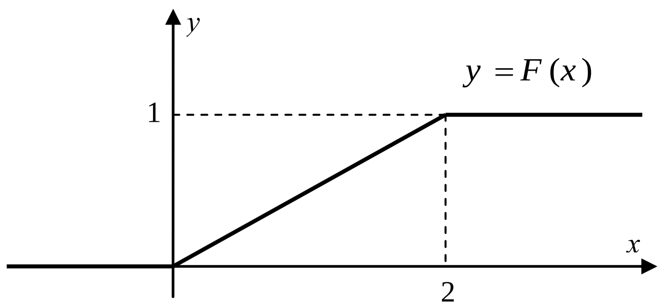


Рис. 6

3. Числовые характеристики ДСВ. Числовые характеристики НСВ.

Во многих практических случаях информация о случайной величине, которую дают закон распределения, функция распределения или плотность вероятностей, является избыточной. Часто проще и удобнее пользоваться числами, которые описывают случайную величину суммарно. К числу наиболее важных из таких числовых характеристик случайных величин относятся математическое ожидание, дисперсия и среднее квадратическое отклонение.

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.

Пусть закон распределения дискретной случайной величины X имеет вид

X :

x_i	x_1	x_2	$\cdots$	x_k
p_i	p_2	p_2	$\cdots$	p_k

Определение. ***Математическим ожиданием*** *дискретной случайной величины X называется число M(X), вычисляемое по формуле*

$$\boxed{M(X)=\sum_{i=1}^{k} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k.}$$

Математическое ожидание случайной величины есть число около которого группируются значения этой случайной величины.

Механическим аналогом математического ожидания дискретной случайной величины является центр масс (центр тяжести) системы точечных масс: если в точках числовой оси с абсциссами $x_1, x_2, ..., x_k$ расположены точечные массы $p_1, p_2, ..., p_k$, то абсцисса их центра масс находится точно по формуле для $M(X)$, приведенной выше.

Пример. Пусть случайная величина X биномиально распределена с параметрами $n=3$ и $p=0{,}8$ (см. пример из § 3.1):

X:

x_i	0	1	2	3
p_i	0,008	0,096	0,384	0,512

Тогда

$$M(X)=0\cdot 0{,}008+1\cdot 0{,}096+2\cdot 0{,}384+3\cdot 0{,}512=2{,}4.$$

Свойства математического ожидания

1. *Математическое ожидание постоянной случайной величины равно самой постоянной, т.е.*

$$M(C)=C,$$

где C – некоторое число.

(*Постоянной случайной величиной C* называется такая случайная величина, которая принимает единственное значение равное *C* с вероятностью 1.)

2. *Постоянный множитель можно выносить за знак математического ожидания, т.е.*

$$M(\alpha X)=\alpha M(X),$$

где α – произвольное число.

3. *Математическое ожидание суммы (разности) случайных величин равно сумме (разности) математических ожиданий этих случайных величин, т.е.*

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

4. *Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий, т.е.*

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

5. *Пусть* $X_1, X_2, ..., X_n$ *– такие случайные величины, математические ожидания которых равны между собой, т.е.* $M(X_i) = a$, *где* $i = 1, 2, ..., n$, *и* a *– некоторое число. Тогда среднее арифметическое этих случайных величин равно их общему математическому ожиданию, т.е.*

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

Заметим, что свойства 2 – 5 математического ожидания остаются справедливыми также для непрерывных случайных величин.

Пусть закон распределения случайной величины X тот же, что и выше (см. начало параграфа).

Определение. ***Дисперсией*** *дискретной случайной величины X называется число* $D(X)$, *определяемое равенством*

$$\boxed{D(X) = \sum_{i=1}^{k}(x_i - M(X))^2 p_i = (x_1 - M(X))^2 p_1 + (x_2 - M(X))^2 p_2 + ... + +(x_k - M(X))^2 p_k.}$$

Число $D(X)$ является мерой разброса значений случайной величины X около ее математического ожидания.

Пример. Пусть случайная величина X биномиально распределена с параметрами $n = 3$ и $p = 0{,}8$. Найдем дисперсию этой случайной величины.

В предыдущем примере найдено, что $M(X) = 2{,}4$. Тогда

$D(X) = (0 - 2{,}4)^2 \cdot 0{,}008 + (1 - 2{,}4)^2 \cdot 0{,}096 + (2 - 2{,}4)^2 \cdot 0{,}384 + (3 - 2{,}4)^2 \cdot 0{,}512 = 0{,}48.$

Свойства дисперсии

1. *Дисперсия постоянной случайной величины равна нулю, т.е.*

$$D(C) = 0.$$

2. *Постоянный множитель можно выносить за знак дисперсии, возводя его при этом в квадрат, т.е.*

$$D(\alpha X) = \alpha^2 D(X),$$

где α *– произвольное число.*

3. *Справедливо равенство:*

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

4. *Дисперсия суммы (разности) двух независимых случайных величин равна сумме дисперсий этих случайных величин, т.е.*

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

где случайные величины X и Y – независимы.

5. *Пусть случайные величины* $X_1, X_2, ..., X_n$ *– независимы и* $D(X_i) = \sigma^2$, *где* $i = 1, 2, ..., n$. *Тогда*

$$D\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{\sigma^2}{n}.$$

Замечание. $\sqrt{D(X)}$ называется *средним квадратическим отклонением* случайной величины X и обычно обозначается через σ.

Отметим также, что свойство 3 дисперсии более удобно для ее вычисления по сравнению с исходным определением дисперсии.

Пример. Пусть закон распределения случайной величины X имеет вид

X:

x_i	1	2
p_i	0,6	0,4

Найти $D(X)$, используя свойство 3 дисперсии.

Решение.

$$M(X) = 1 \cdot 0{,}6 + 2 \cdot 0{,}4 = 1{,}4, \qquad M(X^2) = 1^2 \cdot 0{,}6 + 2^2 \cdot 0{,}4 = 2{,}2,$$

$$D(X) = M(X^2) - M^2(X) = 2{,}2 - 1{,}4^2 = 0{,}24.$$

Математическое ожидание и дисперсия случайной величины называются *параметрами распределения* этой случайной величины.

Теорема. *Пусть случайная величина* $X \equiv m$ *– биномиально распределена с параметрами* n *и* p, *тогда параметры ее распределения могут быть найдены по формулам:*

$$M(m) = np, \qquad D(m) = npq.$$

Также справедливы равенства

$$M\left(\frac{m}{n}\right) = p, \qquad D\left(\frac{m}{n}\right) = \frac{pq}{n}.$$

Пример. Пусть случайная величина X биномиально распределена спараметрам $n = 3$и $p = 0{,}8$. Тогда

$$M(X) = np = 3 \cdot 0{,}8 = 2{,}4, \qquad D(m) = npq = 3 \cdot 0{,}8 \cdot 0{,}2 = 0{,}48.$$

Очевидно, что использование формул последней теоремы упрощает и ускоряет вычисление математического ожидания и дисперсии биномиально распределенной случайной величины по сравнению с применением исходных определений для *M(X)* и $D(X)$.

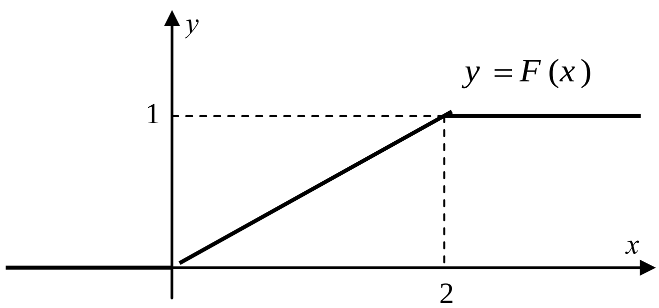


Рис. 7

Формулы для вычисления математического ожидания и дисперсии непрерывной случайной величины аналогичны соответствующим формулам для дискретной случайной величины. Действительно, рассмотрим следующую таблицу.

	Дискретная случайная величина	Непрерывная случайная величина
Способ описания	Закон распределения	Плотность распределения
$M(X)$	$M(X) = \sum_{i=1}^{k} x_i p_i$	$M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx$
$D(X)$	$D(X) = \sum_{i=1}^{k} (x_i - M(X))^2 p_i$	$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 \varphi(x)dx$

Пример. Функция распределения непрерывной случайной величины X имеет вид:

$$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ x^2/9 & \text{при } x \in [0,3], \\ 1 & \text{при } x > 3. \end{cases}$$

Найти математическое ожидание и дисперсию этой случайной величины.

Решение. Для нахождения $M(X)$ и $D(X)$ нам потребуется плотность распределения данной случайной величины (см. приведенные выше формулы). Получаем:

$$\varphi(x)=F'(x)=\begin{cases}0'=0 & \text{при } x<0,\\ \left(x^2/9\right)'=\dfrac{2}{9}x & \text{при } x\in[0,3],\\ 1'=0 & \text{при } x>3,\end{cases} \quad \text{или} \quad \varphi(x)=\begin{cases}\dfrac{2}{9}x & \text{при } x\in[0,3],\\ 0 & \text{в остальных случаях.}\end{cases}$$

Тогда имеем

$$M(X)=\int_{-\infty}^{+\infty}x\varphi(x)dx=\int_{-\infty}^{0}x\cdot 0dx+\int_{0}^{3}x\cdot\frac{2}{9}x\,dx+\int_{3}^{+\infty}x\cdot 0dx=\frac{2}{9}\cdot\frac{x^3}{3}\bigg|_0^3=2.$$

Геометрически, полученное значение математического ожидания есть абсцисса центра тяжести фигуры под графиком плотности распределения, т.е. абсцисса прямоугольного треугольника *OAB* (см. рис. 8; напомним, что центр тяжести треугольника есть точка пересечения медиан этого треугольника, а медианы в точке пересечения делятся в отношении 2:1, считая от вершины).

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).

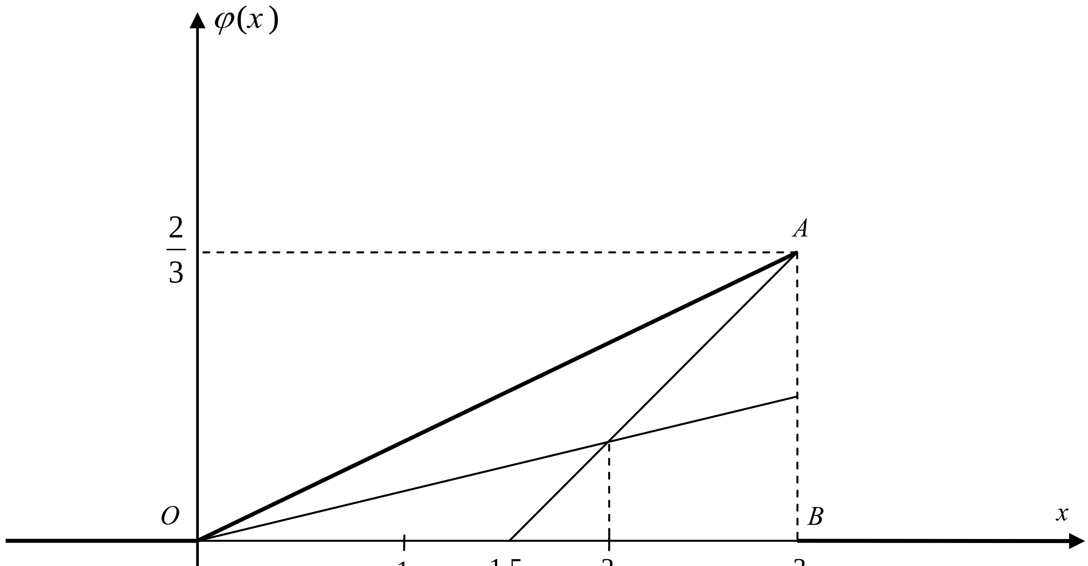


Рис. 8

1.4 Лекция 5-6 (Л-5-6) (4 ч.)

Тема: Основные законы распределения случайных величин.

1.4.1 Вопросы лекции:

1. Основные законы распределения ДСВ биномиальный, Пуассона.
2.Основные законы распределения НСВ: равномерный, показательный.
3. Нормальное распределение и его свойства.

1.4.2. Краткое содержание вопросов:

1. Основные законы распределения ДСВ биномиальный, Пуассона.

Случайную величину полностью задает закон ее распределения (в дискретном случае), а также функция распределения или плотность вероятностей (для непрерывной случайной величины).

Наиболее важными законами распределения дискретной случайной величины являются биномиальный закон, закон распределения Пуассона, геометрическое и гипергеометрическое распределение, а непрерывной – нормальное, равномерное и показательное распределения.

Закон распределения случайной величины X числа появлений события A в схеме Бернулли имеет вид $P_m(n) = P(X = m) = C_n^m p^m q^{n-m}$, где $m = 0,1,2,...,n$, $q = P(\overline{A}) = 1 - p$.

Эта формула еще называется **биномиальной**, так как её правая часть представляет собой $(m+1)$-й член бинома Ньютона: $(p+q)^n = \sum_{m=0}^{n} C_n^m p^m q^{n-m}$.

Очевидно, что для закона биномиального распределения вероятностей выполняется условие нормировки, т.е. сумма всех вероятностей равна единице: $\sum_{m=0}^{n} P_m(n) = (p+q)^n = 1$.

Биномиальное распределение для $n=10$ и некоторых значений p приведено ниже

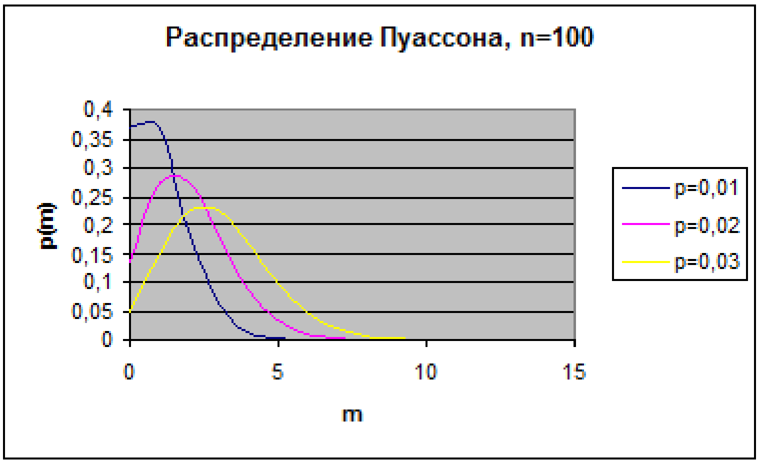
Математическое ожидание числа появлений события A в n независимых испытаниях для биномиального распределения равно произведению числа испытаний на вероятность появления события A в каждом испытании (т.е. среднему числу появления события в данной серии испытаний). $M(X) = np$

***Дисперсия** и среднее квадратическое отклонения равны соответственно:* $D(X) = npq$ $\sigma(X) = \sqrt{D(X)} = \sqrt{npq}$

Ранее отмечалось, что если при увеличении числа испытаний произведение $np = \lambda$ остается постоянным, то биномиальное распределение при больших значениях n сходится к распределению Пуассона.

Случайная величина X называется **распределенной по закону Пуассона,** если она может принимать значения $0,1,2,...,n$, соответствующая вероятность которых определяется по формуле Пуассона: $P_\lambda(m) = \frac{\lambda^m}{m!} e^{-\lambda}$, $(n \to \infty)$

Распределение Пуассона для $n=100,\ \lambda=1,2,3$ приведено ниже



Для распределения Пуассона математическое ожидание и дисперсия равны соответственно:

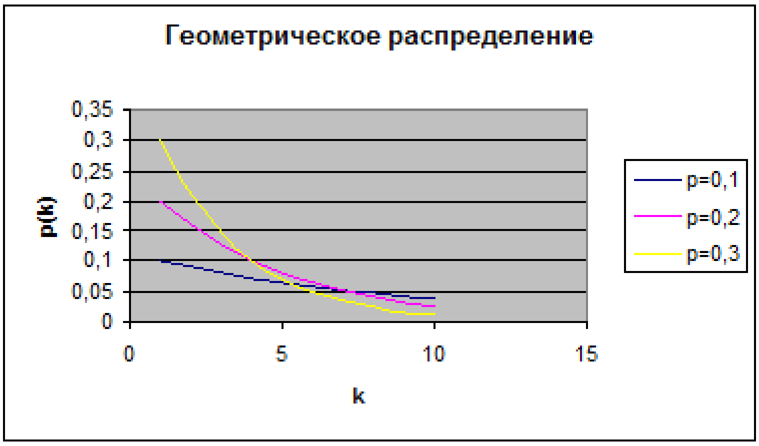
$$M(X)=\lambda \qquad D(X)=\lambda.$$

Равенство значений математического ожидания и дисперсии является уникальным свойством распределения Пуассона. Это свойство часто применяется на практике для решения вопроса, правдоподобна ли гипотеза о том, что случайная величина X распределена по закону Пуассона. Для этого определяют из опыта статистические характеристики случайной величины – математическое ожидание и дисперсию. Если их значения близки, то это может служить доводом в пользу гипотезы о пуассоновском распределении.

Дискретная случайная величина X имеет геометрическое распределение, если она принимает значения $k=1,2,3,\ldots$ (счетное множество значений) с вероятностями

$$p_k=P(X=k)=q^{k-1}\cdot p,\ 0<p<1,\ q=1-p.$$

Случайная величина, имеющая геометрическое распределение, представляет собой число испытаний в схеме Бернулли ***до первого успеха***. Геометрическое распределение для некоторых конкретных значений p приведено ниже



Можно показать, что математическое ожидание и дисперсия для геометрического распределения равны соответственно: $M(X)=\dfrac{1}{p} \qquad D(X)=\dfrac{q}{p^2}$

Пример. В большой партии изделий вероятность брака равна p. Контроль качества проводится до первого появления бракованного изделия. В результате серии проверок обнаружилось, что бракованное изделие впервые появлялось в среднем при десятом испытании. Оценить численное значение p .

Решение. Пусть X - число испытаний до первого появления бракованного изделия. Эта случайная величина имеет геометрическое распределение. По условию ее среднее значение равно $M(X)=10$. Таким образом $p=1/M(X)=0.1$

Гипергеометрическое распределение (урновая схема)

Дискретная случайная величина X имеет гипергеометрическое распределение, если она принимает значения m с вероятностями

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

p_m *представляет вероятность выбора m объектов, обладающих заданным свойством, из множества n объектов, случайно извлеченных (без возврата) из совокупности N объектов, среди которых M объектов обладают заданным свойством. Ниже приведен пример графика гипергеометрического распределения.*

Математическое ожидание и дисперсия случайной величины, имеющей гипергеометрическое распределение с параметрами n, M, N равны:

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

Пример. *Имеется 5 фирм, у трех из которых отчетность оформлена неправильно. 2 ревизора проверяют 2 произвольно выбранные фирмы. Какова вероятность того, что при проверке будет обнаружена неправильная отчетность а) ни в одной, б) в одной, в) в двух фирмах?*

Решение. Данная задача может быть решена с помощью гипергеометрического распределения. По условию задачи общее число объектов (фирм) равно N = 10, число фирм с неправильной отчетностью M=3. Проверяется всего две фирмы (n =2). Число фирм с неправильной отчетностью среди двух выбранных – величина переменная (m=0, 1, 2). Таким образом, имеем

а) $p_0 = P(X=0) = \dfrac{C_3^0\cdot C_2^2}{C_5^2}$ (ни одной неправильной отчетности)

б) $p_1 = P(X=1) = \dfrac{C_3^1\cdot C_2^1}{C_5^2}$ (одна неправильная отчетность)

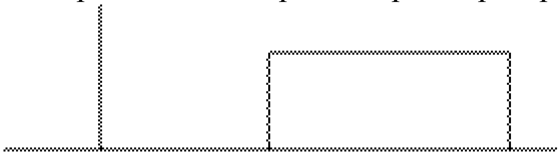
в) $p_2 = P(X=2) = \dfrac{C_3^2\cdot C_2^0}{C_5^2}$ (две неправильные отчетности).

2.Основные законы распределения НСВ: равномерный, показательный.

Непрерывная случайная величина считается **равномерно распределенной** на отрезке (a,b), если ее плотность вероятности имеет вид:

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

График плотности вероятности для равномерного распределения



Математическое ожидание и дисперсия непрерывной случайной величины, имеющей равномерное распределение, равны соответственно:

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

Пример. Интервал движения автобуса равен 15 мин. Какова вероятность того, что пассажир на остановке будет ждать автобус не более 5 минут?

Решение. Пусть случайная величина X - время ожидания автобуса. Она имеет равномерное распределение на отрезке [0,15]. Имеем

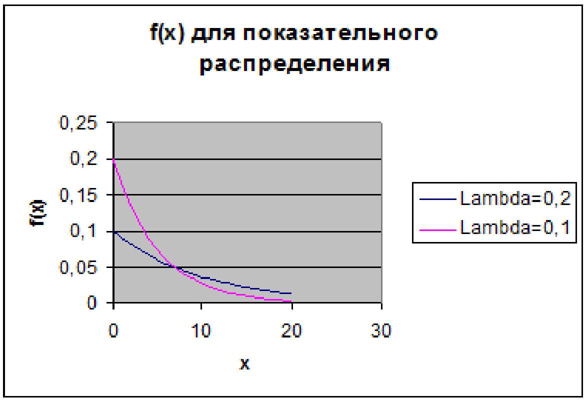
$$P(\alpha\le X\le\beta)=\int_\alpha^\beta f(x)\,dx$$

В рассматриваемом случае
$$P(0\le X\le 5)=\int_0^5\frac{1}{b-a}dx=\int_0^5\frac{1}{15-0}dx=\frac{5-0}{15-0}=\frac{1}{3}$$

Показательным (экспоненциальным) распределением непрерывной случайной величины X называется распределение, имеющее плотность вероятности вида:

$$f(x)=\begin{cases}0,(x<0)\\ \lambda e^{-\lambda x},(x\ge 0)\end{cases}$$

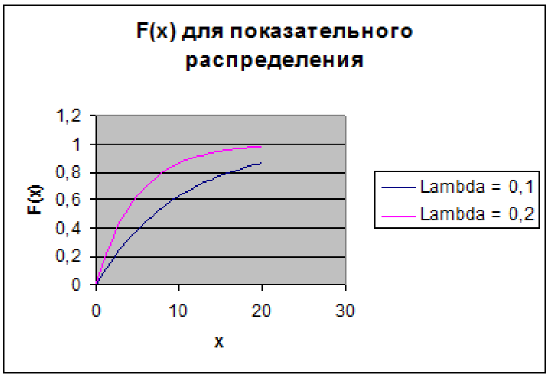
где λ – постоянная положительная величина. Плотность вероятностей для показательного распределения для $\lambda=0.1;\ 0.2$ приведена ниже



Функция распределения вероятности для показательного распределения имеет вид:

$$F(x)=\begin{cases}0,(x<0)\\1-e^{-\lambda x},(x\ge 0)\end{cases}$$

Функция распределения для $\lambda = 0.1;\ 0.2$ приведена ниже



Можно показать, что математическое ожидание и дисперсия случайной величины, имеющей экспоненциальное распределение, равны:

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

Пример. Установлено, что время горения электрической лампочки (Т) является случайной величиной, распределенной по показательному закону. Считая, что среднее значение этой величины равно 6 месяцам, найти вероятность того, что лампочка будет исправна более года.

Решение. Так как $M(T)=1/\lambda=6,\ \text{то}\ \lambda=1/6$ и функция распределения случайной величины Т имеет вид

$$F(x)=\begin{cases}1-\exp(-x/6),\ x\ge 0\\0,\ x<0\end{cases}$$

Поэтому

$$P(T>12)=P(12<T<+\infty)=F(\infty)-F(12)=1-(1-\exp(-12/6))=\exp(-2)\approx 0.135$$

3. Нормальное распределение и его свойства.

Непрерывная случайная величина X имеет нормальный закон распределения с параметрами a и σ, если ее плотность вероятности имеет вид функции Гаусса

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

где $\sigma>0$. С помощью непосредственного вычисления математического ожидания и дисперсии нормального распределения легко выяснить вероятностный смысл его параметров: a – есть математическое ожидание, а σ - среднее квадратическое отклонение нормального распределения. При $a=0,\ \sigma=1$ распределение называется **стандартным нормальным распределением**.

Графики $f(x)$ для ряда конкретных значений математического ожидания и среднего квадратического отклонения приведены ниже.

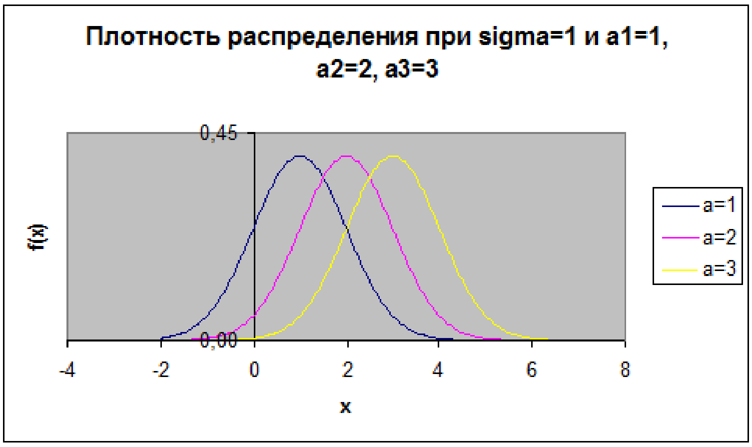


Рис. 1. Изменение вида функции $f(x)$ при изменении математического ожидания

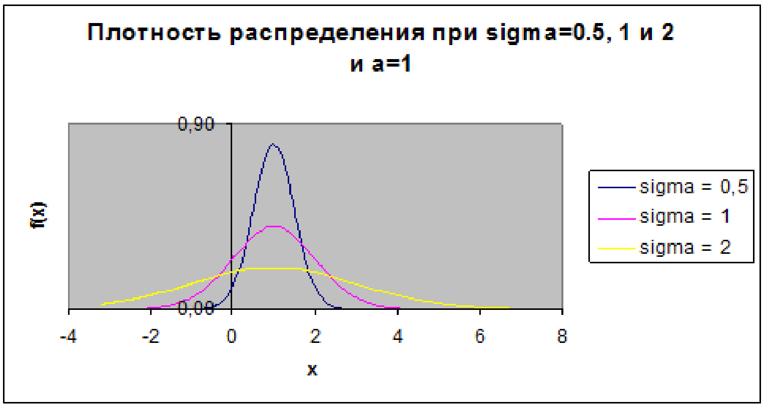


Рис. 2. Изменение $f(x)$ при изменении среднего квадратического отклонения

Функция распределения в случае нормального распределения, очевидно, равна

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt .$$

Графики функции $F(x)$ для ряда значений математического ожидания и среднего квадратического отклонения изображены на приводимых ниже рисунках

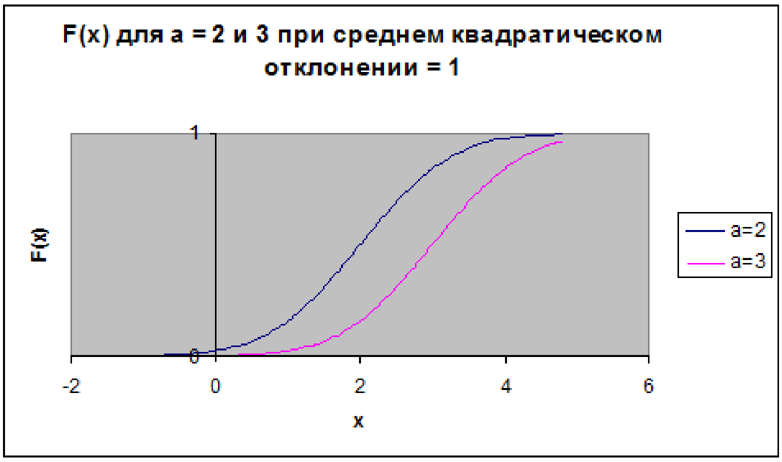


Рис. 3. Зависимость функции распределения от величины a

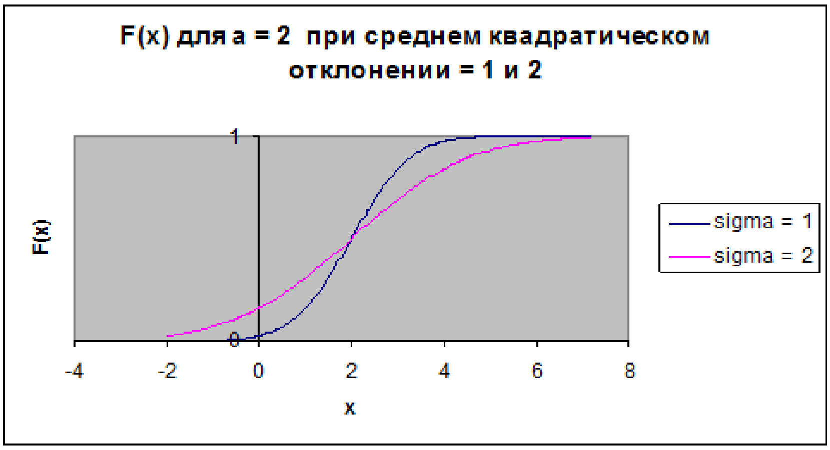


Рис. 4. Зависимость функции распределения от величины σ

Нормальное распределение имеет исключительно важное значение для практических применений, так как многие непрерывные случайные величины описываются именно этим распределением. Оказывается, что суммирование большого числа случайных величин с различными законами распределения приводит к нормальному распределению результирующей суммы. Это свойство подтверждается центральной предельной теоремой (теорема **Ляпунова**). Смысл этой теоремы состоит в следующем. Если случайная величина X представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то X имеет распределение, близкое к нормальному.

Следует иметь в виду, что при усилении влияния отдельных факторов могут появляться отклонения от нормального распределения результирующего параметра. Поэтому большое значение на практике уделяется экспериментальной проверке выдвинутых гипотез, в том числе и гипотезы о нормальном распределении.

График плотности нормального распределения называют **нормальной кривой Гаусса.**

Исследуем поведение функции плотности вероятности $f(x)$.

1. Очевидно, что функция определена на всей оси x.
2. Функция принимает лишь положительные значения, т.е. нормальная кривая расположена над осью Ox.
3. Ось Ox служит горизонтальной асимптотой графика. Других асимптот у графика нет.
4. При $x = a$ функция имеет максимум, равный $\varphi(x)_{max} = \dfrac{1}{\sigma\sqrt{2\pi}}$
5. Функция четная: ее график симметричен относительно прямой $x = a$
6. При $x = a \pm \sigma_x$ график функции имеет точки перегиба.

При любых значениях параметров a и σ, площадь, ограниченная нормальной кривой и осью x равна единице.

Часто требуется определить вероятность попадания случайной величины в заданный интервал. Эта вероятность может быть выражена в виде разности функции распределения вероятности в граничных точках этого интервала:

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx .$$

В случае нормального распределения $P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{(x-a)^2}{2\sigma^2}}dx$

Используя замену переменной: $t = \frac{x-a}{\sigma}$, $x = t\sigma + a$, $dx = \sigma dt$, получим

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{t^2}{2}}dt,$$

где $t_1 = \frac{x_1 - a}{\sigma}$, $t_2 = \frac{x_2 - a}{\sigma}$.

Разобьем полученный интеграл на два:

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}}\left(\int_{t_1}^{0} e^{-\frac{t^2}{2}}dt + \int_{0}^{t_2} e^{-\frac{t^2}{2}}dt\right) = \frac{1}{\sqrt{2\pi}}\left(\int_{0}^{t_2} e^{-\frac{t^2}{2}}dt - \int_{0}^{t_1} e^{-\frac{t^2}{2}}dt\right)$$

Тогда искомая вероятность может быть выражена в виде:

$$P(x_1 < X < x_2) = \Phi_0(t_2) - \Phi_0(t_1)$$

Функция Лапласа протабулирована, что существенно упрощает расчет попадания нормально распределенной случайной величины в любой заданный интервал.

Функция Лапласа не выражается через элементарные функции:

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}}\int_{0}^{x} e^{-\frac{t^2}{2}}dt.$$

Для ее вычисления используются специальные таблицы или методы приближенного вычисления.

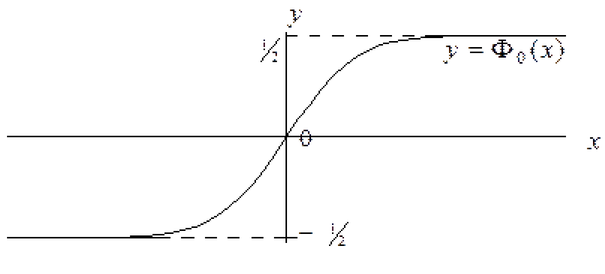
Функция $\Phi_0(x)$ *обладает следующими свойствами:*

1. $\Phi_0(0) = 0$;

2. $\Phi_0(+\infty) = \frac{1}{2}$;

3. функция $\Phi_0(x)$ *– нечетная, т.е.* $\Phi_0(-x) = -\Phi_0(x)$ *, поэтому в таблицах обычно приводятся значения* $\Phi_0(x)$ *только для положительных* x;

4.



функция $\Phi_0(x)$ *– монотонно возрастающая функция (это следует из того, что* $\Phi_0'(x) = \varphi_0(x) > 0$*).* Часто требуется вычислить вероятность того, что отклонение нормально распределенной случайной величины X по абсолютной величине от математического ожидания меньше заданного положительного числа δ, т.е. требуется найти вероятность того, что выполняется неравенство $|X - M_X| < \delta$.

Заменим это неравенство равносильным ему двойным неравенством $M_X - \delta < X < M_X + \delta$.

Воспользуемся формулой: $P(x_1 < X < x_2) = \Phi_0(t_2) - \Phi_0(t_1)$

Получим: $P(M_X - \delta < X < M_X + \delta) =$

$$= \Phi_0\left(\frac{(M_x + \delta) - M_x}{\sigma}\right) - \Phi_0\left(\frac{(M_x - \delta) - M_x}{\sigma}\right)$$

$$= \Phi_0\left(\frac{\delta}{\sigma}\right) - \Phi_0\left(-\frac{\delta}{\sigma}\right) = 2\Phi_0\left(\frac{\delta}{\sigma}\right)$$

Если в качестве δ взять утроенное значение среднего квадратического отклонения s, то получим:

$$P(|X - M_X| < 3\sigma) = 2\Phi_0(3) = 0.9973,$$

Таким образом, вероятность того, что абсолютная величина отклонения превысит (утроенное среднее квадратическое отклонение, очень мала 0,0027 или 0,27%). Такие события можно считать практически невозможными.

Другими словами, если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратичного отклонения. В этом и состоит сущность **правила «трех сигм»**.

На практике правило «трех сигм» применяют так: если распределение изучаемой случайной величины неизвестно, но правило «трех сигм» выполняется, то есть основания полагать, что изучаемая величина распределена нормально, и наоборот.

Пример 1. Текущая цена ценной бумаги представляет собой нормально распределенную случайную величину X со средним 100 у.е. и дисперсией 9. Найти вероятность того, что цена актива будет находиться в пределах от 91 до 109 у.е.

Решение. Так как $a = 100,\ \sigma = \sqrt{9} = 3$, то

$$P(91 < X < 109) = \Phi_0\left(\frac{109-100}{3}\right) - \Phi_0\left(\frac{91-100}{3}\right) = \Phi_0(3) - \Phi_0(-3) = 2\Phi_0(3) = 0.9973$$

1.5 Лекция 7-8 (Л-7-8) (4 ч.)

Тема: Многомерные случайные величины, их числовые характеристики

1.5.1 Вопросы лекции:

1. Понятие многомерной случайной величины.
2. Закон распределения МСВ
3. Условные характеристики МСВ

1.5.2. Краткое содержание вопросов:

1. Понятие многомерной случайной величины.

До сих пор мы рассматривали случайные величины, возможные значения которых определялись одним числом (одномерные случайные величины). Например, число очков, которое может выпасть при бросании игральной кости (дискретная одномерная случайная величина) или расстояние от орудия до места падения снаряда (непрерывная одномерная случайная величина).

Часто приходится иметь дело с величинами, возможные значения которых определяются двумя или более числами. Такие величины называются n – мерными случайными величинами; n – мерную случайную величину можно рассматривать как систему n случайных величин. В данном контексте используется также термин **многомерный случай-**

ный вектор $\vec{x} = (x_1, x_2, \ldots, x_n)$, где каждая из величин $x_1, x_2, \ldots, x_n$ называется составляющей (компонентой). Аналогично одномерным случайным величинам различают **дискретные** многомерные случайные величины (их составляющие дискретны) и **непрерывные** многомерные случайные величины, составляющие которых непрерывны.

Пример. Станок штампует стальные плитки. Если контролируемыми размерами являются длина X, ширина Y и высота Z плитки, то мы имеем трехмерную случайную величину (X,Y,Z).

Остановимся более подробно на двумерных случайных величинах.

Законом распределения дискретной двумерной случайной величины $\xi = (X, Y)$ называют перечень возможных значений этой величины, т.е. пар чисел (x_i, y_j), где x_i и y_j $(i = 1,2,\ldots,n;\ j = 1,2,\ldots,m)$ – возможные значения величин X и Y, соответственно, и вероятностей $p_{ij} = p(x_i, y_j)$ их совместного появления $P(X = x_i, Y = y_j)$.

Двумерная дискретная случайная величина $\xi = (X, Y)$ задается в виде **таблицы распределения** вида:

$X \diagdown Y$	y_1	y_2	...	y_m
x_1	p_{11}	p_{12}	...	p_{1m}
x_2	p_{21}	p_{22}	...	p_{2m}
...	...	...	...	...
x_n	p_{n1}	p_{n2}	...	p_{nm}

где первая строка таблицы указывает возможные значения составляющей Y, а первый столбец – все возможные значения составляющей X.

Так как события $(X = x_i, Y = y_j)$ ($i = 1,2,\ldots,n$; $j = 1,2,\ldots,m$) образуют полную группу, то $\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1$.

Зная закон распределения двумерной дискретной случайной величины, можно найти законы распределения каждой из ее составляющих. Так, например, вероятность того, что X примет значение x_k, равна $P(X = x_k) = \sum_{j=1}^{m} p_{kj}$.

2. Закон распределения МСВ

Функция $F(x, y)$, определяющая для каждой пары чисел x, y вероятность того, что X примет значение меньшее x, и при этом Y примет значение меньшее y, называется **совместной функцией распределения** двух случайных величин $F(x, y) = P(X < x, Y < y)$.

Геометрически это равенство можно истолковать так: $F(x, y)$ – это вероятность того, что случайная точка (X, Y) попадет в бесконечный квадрант с вершиной (x, y), расположенный левее и ниже этой вершины.

1. Значения совместной функции распределения удовлетворяют неравенству: $0 \le F(x,y) \le 1$.

2. $F(x,y)$ – неубывающая функция по каждому аргументу, т.е.

$F(x_2,y) \ge F(x_1,y)$, если $x_2 \ge x_1$;

$F(x,y_2) \ge F(x,y_1)$, если $y_2 \ge y_1$.

Совместная функция распределения имеет следующие предельные значения:

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

$F(-\infty,-\infty) = 0$; $F(\infty,\infty) = 1$.

3. При $x = \infty$ или $y = \infty$ совместная функция распределения системы становится функцией распределения одной из составляющих: $F(x,\infty) = F_1(x)$; $F(\infty,y) = F_2(y)$

Плотность совместного распределения вероятностей

Непрерывную двумерную случайную величину можно задать с помощью плотности распределения. **Плотность совместного распределения вероятностей** $f(x,y)$ двумерной непрерывной случайной величины (X, Y) – это вторая смешанная частная производная от функции распределения $F(x,y)$:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

Зная плотность совместного распределения $f(x,y)$, можно найти совместную функцию распределения $F(x,y)$ по формуле $F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dxdy$

следующей из определения плотности распределения двумерной непрерывной случайной величины (X, Y).

Смысл плотности совместного распределения вероятностей: вероятность попадания случайной точки в прямоугольник (с вершиной в точке (x,y) и сторонами Δx и Δy равна произведению $f(x,y)\Delta x\Delta y$, когда стороны этого прямоугольника стремятся к нулю.

В связи с этим, вероятность попадания случайной точки в ***произвольную область D*** равна двойному интегралу по области D от функции $f(x,y)$

$$P((X,Y) \subset D) = \iint_D f(x,y)dxdy$$

Свойства двумерной плотности вероятности

1. Двумерная плотность вероятности неотрицательна: $f(x,y) \ge 0$.

2. Двойной несобственный интеграл с бесконечными пределами от двумерной плотности вероятности равен единице: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$.

Две случайные величины называются ***независимыми***, если закон распределения одной из них не зависит от того, какие возможные значения приняла другая величина.

Теорема. Для того чтобы случайные величины X и Y были независимыми, необходимо и достаточно, чтобы функция распределения системы (X, Y) была равна произведению функций распределения составляющих: $F(X,Y) = F_1(X)F_2(Y)$.

Следствие. Для того чтобы случайные величины X и Y были независимыми, необходимо и достаточно, чтобы плотность совместного распределения системы (X, Y) была равна произведению плотностей распределения составляющих: $f(X,Y) = f_1(X)f_2(Y)$.

2. Условные законы распределения СВ

Пусть известна плотность распределения системы двух случайных величин. Используя свойства функций распределения, можно вывести формулы для нахождения плотности распределения одной величины, входящей в систему:

$$f_1(x) = F_1'(x) = \int_{-\infty}^{+\infty} f(x,y)dy; \qquad f_2(y) = F_2'(y) = \int_{-\infty}^{+\infty} f(x,y)dx \qquad (*)$$

Перейдём теперь к решению обратной задачи: по известным законам распределения отдельных случайных величин, входящих в систему, найти закон распределения системы. Легко увидеть, что в общем случае эта задача неразрешима. Действительно, с одной стороны, законы распределения отдельных случайных величин, входящих в систему, характеризуют каждую из случайных величин в отдельности, но ничего не говорят о том, как они взаимосвязаны. С другой стороны, искомый закон распределения системы должен содержать все сведения о случайных величинах системы, в том числе и о характере связей между ними.

Таким образом, если случайные величины X, Y взаимозависимы, то закон распределения системы не может быть выражен через законы распределения отдельных случайных величин, входящих в систему. Это приводит к необходимости введения условных законов распределения.

Распределение одной случайной величины, входящей в систему, найденное при условии, что другая случайная величина, входящая в систему, приняла определённое значение, называется условным законом распределения.

Для дискретных случайных величин условным распределением составляющей при условии, что $Y = y_j$ называется совокупность условных вероятностей $p(x_1 \mid y_j), p(x_2 \mid y_j), \ldots, p(x_n \mid y_j)$, вычисленных в предположен, что случайная величина уже приняла значение. Для нахождения $p(x_i \mid y_j)$ пользуются формулой

$$p(x_i \mid y_j) = \frac{p(x_i, y_j)}{p(y_j)}, \quad i = \overline{1,n}.$$

Заметим, что $\sum_{i=1}^{n} p(x_i \mid y_j) = 1$. Аналогично находим $p(y_j \mid x_i) = \frac{p(x_i, y_j)}{p(x_i)}, \quad j = \overline{1,m}$.

Условный закон распределения можно задавать как функцией распределения, так и плотностью распределения. Условная функция распределения обозначается $F(x \mid y)$; условная плотность распределения обозначается $f(x \mid y)_1$.

Плотностью распределения для случайной величины при условии, что случайная величина приняла определённое значение (*условной плотностью распределения*), назовём величину

$$f(x \mid y) = \frac{f(x,y)}{f_2(y)}.$$

Аналогично, плотностью распределения для случайной величины при условии, что случайная величина приняла определённое значение, назовём величину

$$f(y \mid x)=\frac{f(x,y)}{f_1(x)}.$$

Отсюда получаем: $f(x,y)=f_1(x)f(y \mid x)=f_2(y)f(x \mid y)$

или, с учётом формул (*) $f(x \mid y)=\dfrac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)dx}$; $f(y \mid x)=\dfrac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)dy}$.

Условная плотность распределения обладает всеми свойствами безусловной плотности распределения. В частности, $\int_{-\infty}^{+\infty} f(x \mid y)dx=1;$ $\int_{-\infty}^{+\infty} f(y \mid x)dy=1$.

3. Условные характеристики МСВ

Для описания условных законов распределения можно использовать различные характеристики подобно тому, как для одномерных распределений.

Наиболее важной характеристикой является условное математическое ожидание.

Условным математическим ожиданием дискретной случайной величины при $Y=y$ (y – определённое возможное значение случайной величины) называется сумма произведений возможных значений на их условные вероятности:

$$M(X \mid Y)=\sum_{i=1}^{n} x_i P(x_i \mid y).$$

Для непрерывных случайных величин:

$$M(X \mid Y)=\int_{-\infty}^{+\infty} x f(x \mid y)dx,$$

где – условная плотность распределения случайной величины при .

Аналогично, условным математическим ожиданием дискретной случайной величины при $X=x$ (X – определённое возможное значение случайной величины) называется сумма произведений возможных значений на их условные вероятности:

$$M(Y \mid X)=\sum_{j=1}^{m} y_j P(y_j \mid x).$$

Для непрерывных случайных величин: $M(Y \mid X)=\int_{-\infty}^{+\infty} y f(y \mid x)dy$,

где $f(y \mid x)$ – условная плотность распределения случайной величины при.

Аналогично вводятся условные дисперсии и условные моменты более высоких порядков.

Числовые характеристики системы двух случайных величин

Две случайные величины называются независимыми, если закон распределения одной из них не зависит от того, какие возможные значения приняла другая величина. Из этого определения следует, что условные распределения независимых случайных величин равны их безусловным распределениям. Укажем необходимые и достаточные условия независимости случайных величин.

ТЕОРЕМА 1: Для того чтобы случайные величины X и Y были независимыми, необходимо и достаточно, чтобы функция распределения системы (X,Y) была равна произведению функций распределения составляющих:

$$F(x,y)=F_1(x)\cdot F_2(y)$$

ТЕОРЕМА 2: Для того чтобы случайные величины и были независимыми, необходимо и достаточно, чтобы плотность вероятности системы была равна произведению плотностей вероятностей составляющих:

$f(x,y) = f_1(x)\cdot f_2(y)$.

Для описания системы двух случайных величин кроме математических ожиданий и дисперсий, составляющих используют и другие характеристики, к которым относятся корреляционный момент и коэффициент корреляции.

Корреляционным моментом μ_{xy} случайных величин и называют математическое ожидание произведения отклонений этих величин:

$$\mu_{xy} = M\big((X - M(X))\cdot(Y - M(Y)) \big).$$

Для вычисления корреляционного момента дискретных величин используют формулу:

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - M(X))(y_j - M(Y))P(x_i, y_j),$$

а для непрерывных величин:

$$\mu_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - M(X))(y - M(Y))\,f(x,y)\,dxdy.$$

Корреляционный момент служит для характеристики связи между величинами и .

ТЕОРЕМА 3: Корреляционный момент двух независимых случайных величин и равен нулю.

Замечание: из теоремы 3 следует, что если корреляционный момент двух случайных величин и не равен нулю, то и – зависимые случайные величины.

Коэффициентом корреляции r_{xy} случайных величин и называют отношение корреляционного момента к произведению средних квадратических отклонений этих величин:

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

Очевидно, коэффициент корреляции двух независимых случайных величин равен нулю (так как $\mu_{xy} = 0$).

Коррелированность и зависимость случайных величин

Две случайные величины и называются *коррелированными*, если их корреляционный момент (или коэффициент корреляции) отличен от нуля; и называют *некоррелированными* величинами, если их корреляционный момент равен нулю.

Две коррелированные величины также и зависимы. Обратное утверждение не всегда имеет место, то есть если две величины зависимы, то они могут быть как коррелированными, так и некоррелированными. Другими словами, корреляционный момент двух зависимых величин может быть не равным нулю, но может и равняться нулю.

Заметим, что для нормально распределённых составляющих двумерной случайной величины понятия независимости и некоррелированности равносильны.

Если $r_{xy} = \pm 1$, то и связаны линейной зависимостью $Y = aX + b$,

Если $r_{xy} > 0$, то говорят о положительной (или прямой) корреляции между и, то есть с возрастанием одной случайной величины другая случайная величина также возрастает.

Если $r_{xy} < 0$, то говорят об отрицательной корреляции между и, то есть с возрастанием одной случайной величины другая случайная величина убывает.

Задача 1. Закон распределения двумерной дискретной случайной величины задан таблицей

Y / X	– 4	– 2	0
0	0,1	0,1	0,2

1	0,1	0,2	0,1
4	0	0,1	0,1

Найти:

• собственные законы распределения случайных величин и;

• математические ожидания $M(X),\ M(Y)$;

• дисперсии $D(X),\ D(Y)$;

• корреляционный момент;

• коэффициент корреляции;

• закон распределения случайной величины при условии, что случайная величина принимает своё наименьшее значение.

Решение. Складывая вероятности по строкам, получим закон распределения случайной величины в виде ряда распределения

x_i	0	1	4	
$p(x_i)$	0,4	0,4	0,2	$\sum_{i=1}^{3} p(x_i) = 1$

Складывая вероятности по столбцам, получим закон распределения случайной величины в виде ряда распределения

y_j	-4	-2	0	
$p(y_j)$	0,2	0,4	0,4	$\sum_{j=1}^{3} p(y_j) = 1$

Найдём математические ожидания и дисперсии составляющих:

$$M(X) = \sum_{i=1}^{3} x_i p_i = 0 \cdot 0{,}4 + 1 \cdot 0{,}4 + 4 \cdot 0{,}2 = 1{,}2;$$

$$M(Y) = \sum_{j=1}^{3} y_j p_j = -4 \cdot 0{,}2 - 2 \cdot 0{,}4 + 0 \cdot 0{,}4 = -1{,}6;$$

$$D(X) = \sum_{i=1}^{3} (x_i - M(X))^2 \cdot p_i = (0 - 1{,}2)^2 \cdot 0{,}4 + (1 - 1{,}2)^2 \cdot 0{,}4 + (4 - 1{,}2)^2 \cdot 0{,}2 = 2{,}16;$$

$$\sigma_x = \sqrt{2{,}16} \approx 1{,}47;$$

$$D(Y) = \sum_{j=1}^{3} (y_j - M(Y))^2 \cdot p_j = (-4 + 1{,}6)^2 \cdot 0{,}2 + (-2 + 1{,}6)^2 \cdot 0{,}4 + 1{,}6^2 \cdot 0{,}4 = 2{,}24;$$

$$\sigma_y = \sqrt{2{,}24} \approx 1{,}5.$$

Найдём корреляционный момент и коэффициент корреляции:

$$\mu_{xy} \overset{def}{=} \sum_{i=1}^{3}\sum_{j=1}^{3} (x_i - M(X))(y_j - M(Y)) P(x_i, y_j) = (-1{,}2) \cdot (-2{,}4) \cdot 0{,}1 + (-1{,}2) \cdot (-0{,}4) \cdot 0{,}1 +$$
$$+ (-1{,}2) \cdot (1{,}6) \cdot 0{,}2 + (-0{,}2) \cdot (-2{,}4) \cdot 0{,}1 + (-0{,}2) \cdot (-0{,}4) \cdot 0{,}2 + (-0{,}2) \cdot 1{,}6 \cdot 0{,}1 +$$
$$+ 2{,}8 \cdot (-2{,}4) \cdot 0 + 2{,}8 \cdot (-0{,}4) \cdot 0{,}1 + 2{,}8 \cdot 1{,}6 \cdot 0{,}1 = 0{,}32;$$

$$r_{xy} \overset{def}{=} \frac{\mu_{xy}}{\sigma_x \sigma_y} = \frac{0{,}32}{1{,}47 \cdot 1{,}5} \approx 0{,}145.$$

Найдём закон распределения случайной величины при условии, что случайная величина принимает своё наименьшее значение, то есть при условии, что $Y = -4\ (y_1 = -4)$.

Искомый закон распределения, как ранее отмечалось, определяется совокупностью условных вероятностей $p(x_i \mid y_1),\ i=\overline{1,3}$, где

$$p(x_1 \mid y_1)=\frac{p(x_1,y_1)}{p(y_1)}=\frac{0{,}1}{0{,}2}=0{,}5;$$

$$p(x_2 \mid y_1)=\frac{p(x_2,y_1)}{p(y_1)}=\frac{0{,}1}{0{,}2}=0{,}5;$$

$$p(x_3 \mid y_1)=\frac{p(x_3,y_1)}{p(y_1)}=\frac{0}{0{,}2}=0.$$

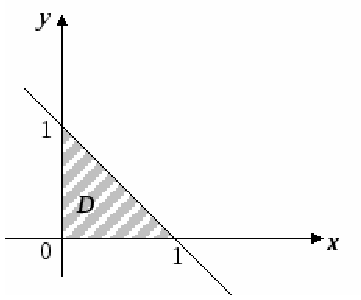
Следовательно, искомый закон распределения имеет вид:

	0	1	4
$p(x_i \mid y_1)$	0,5	0,5	0

Задача 2. Вне области $D=\{(x,y)\mid x+y\le 1,\ x\ge 0,\ y\ge 0\}$ плотность распределения двумерной случайной величины равна 0; в области плотность распределения $f(x,y)=Axy^2$.

Найти:

- коэффициент A;
- вероятность $P((X,Y)\in G)$, где $G=\left\{(x,y)\mid 0\le x\le\frac{1}{2},\ 0\le y\le\frac{1}{2}\right\}$;
- одномерные плотности распределения $f_1(x),\ f_2(y)$;
- математические ожидания;
- дисперсии;
- корреляционный момент;
- коэффициент корреляции.



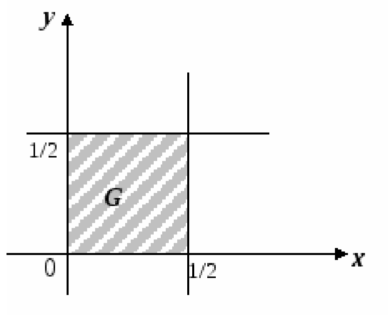
Решение. Для нахождения параметра A воспользуемся формулой

$$\iint_D f(x,y)\,dxdy=1.$$

Тогда

$$\iint_D Axy^2dxdy=A\int_0^1 xdx\int_0^{1-x}y^2dy=\frac{A}{3}\int_0^1 x\left(y^3\Big|_0^{1-x}\right)dx=$$

$$=\frac{A}{3}\int_0^1 x(1-x)^3dx=\frac{A}{60}\ \Rightarrow\ \frac{A}{60}=1\ \Rightarrow\ A=60.$$



Получим:

$$f(x,y)=\begin{cases}60xy^2, \text{ если } (x,y)\in D;\\ 0, \text{ если } (x,y)\notin D.\end{cases}$$

Найдём теперь вероятность попадания двумерной случайной величины (X,Y) в плоскую область G:

$$P((X,Y)\in G)\overset{def}{=}\iint_G f(x,y)\,dx\,dy=$$

$$=\iint_G 60xy^2dxdy=\int_0^{\frac{1}{2}}60xdx\int_0^{\frac{1}{2}}y^2dy=\frac{15}{48}\approx 0{,}3125.$$

Далее, найдём одномерные плотности распределения:

$$f_1(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\int_0^1 60xy^2 dy=20x, \quad x\in[\,0;1\,];$$

$$f_2(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\int_0^1 60xy^2 dx=30y^2, \quad y\in[\,0;1\,].$$

Итак:

$$f_1(x)=\begin{cases} 20x, & \text{если } x\in[0;1], \\ 0, & \text{если } x\notin[0;1]. \end{cases} \qquad f_2(y)=\begin{cases} 30y^2, & \text{если } y\in[0;1], \\ 0, & \text{если } y\notin[0;1]. \end{cases}$$

Найдём математические ожидания и дисперсии составляющих X и Y[1]:

$$M(X)=\iint_D xf(x,y)dxdy=\iint_D x60xy^2 dxdy=\frac{1}{3};$$

$$M(Y)=\iint_D yf(x,y)dxdy=\iint_D y60xy^2 dxdy=\frac{1}{2}.$$

Далее

$$M(X^2)=\iint_D x^2 f(x,y)dxdy=\iint_D 60x^3y^2 dxdy=\frac{1}{7};$$

$$M(Y^2)=\iint_D y^2 f(x,y)dxdy=\iint_D 60xy^4 dxdy=\frac{2}{7};$$

$$M(XY)=\iint_D xy\ f(x,y)dxdy=\iint_D 60x^2y^3 dxdy=\frac{1}{7}.$$

Тогда

$$D(X)=M(X^2)-(M(X))^2=\frac{1}{7}-\frac{1}{9}=\frac{2}{63},$$

$$D(Y)=M(Y^2)-(M(Y))^2=\frac{2}{7}-\frac{1}{4}=\frac{1}{28},$$

$$\mu_{xy}=M(XY)-M(X)\cdot M(Y)=\frac{1}{7}-\frac{1}{3}\cdot\frac{1}{2}=-\frac{1}{42}.$$

Так как $\sigma_x=\sqrt{D(X)}=\sqrt{\frac{2}{63}};\quad \sigma_y=\sqrt{D(Y)}=\sqrt{\frac{1}{28}}$, то нетрудно вычислить

$$r_{xy}\overset{def}{=}\frac{\mu_{xy}}{\sigma_x\sigma_y}=\frac{-\frac{1}{42}}{\sqrt{\frac{2}{63}\cdot\frac{1}{28}}}=-\frac{1}{\sqrt{2}}\approx-0{,}71$$

.

1.6 Лекция 9 -10 (Л-9-10) (4 ч.)

Тема: Генеральная и выборочная совокупность. Оценки статистических параметров распределения

1.6.1 Вопросы лекции:

1. Статистический материал и его первичная обработка. Эмпирические законы распределения. Полигон частот, гистограмма.
2. Числовые характеристики выборки. Точечные оценки выборочных характеристик.
3. Интервальные оценки, их свойства. Метод доверительных интервалов при заданных условиях.
4. Метод моментов

1.6.2. Краткое содержание вопросов:

1. Статистический материал и его первичная обработка. Эмпирические законы распределения. Полигон частот, гистограмма.

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 $\{x_1, x_2, \ldots, x_n\}$ 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.

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.

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). 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.

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.

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. Вариационный ряд*** можно представить таблицей:

X	x_1	x_2	…..	x_m
n	n_1	n_2	….	n_m

Статистическим распределением выборки называют перечень вариант и соответствующих им относительных частот. Статистическое распределение можно представить как

X	x_1	x_2	…..	x_m
w	w_1	w_2	….	w_m

где относительные частоты $w_k = \dfrac{n_k}{n}$.

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 $\Delta_k = [a_{k-1}, a_k),\ (k = 1,\dots,m)$ (Δ_m замкнут также и справа) с помощью некоторой системы точек $a = a_0 < a_1 < \dots < a_m = b$. Часто разбиение [a,b] производят на равные части, тогда $\Delta_k = [a + (k-1)h,\ a + kh),\ k = 1,\dots,m$, 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. Вариационный ряд имеет в таком случае вид

X	Δ_1	Δ_2	…..	Δ_m
n	n_1	n_2	….	n_m

а статистическое распределение –

X	Δ_1	Δ_2	…..	Δ_m
n	w_1	w_2	….	w_m

Число интервалов k часто выбирают на основании формулы Стерджерса $k = 1 + 1{,}4\ln n$.

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

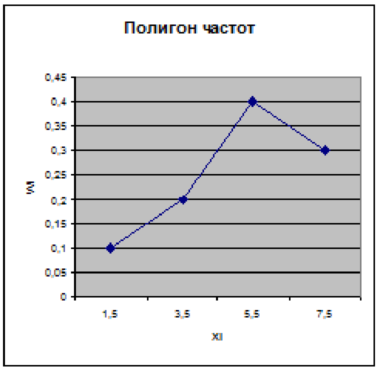
Полигоном частот называют ломаную линию, отрезки которой соединяют точки $(x_1;n_1), (x_2;n_2),\ldots,(x_k;n_k)$. Для построения полигона частот на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им частоты n_i и соединяют точки $(x_i;n_i)$ отрезками прямых.

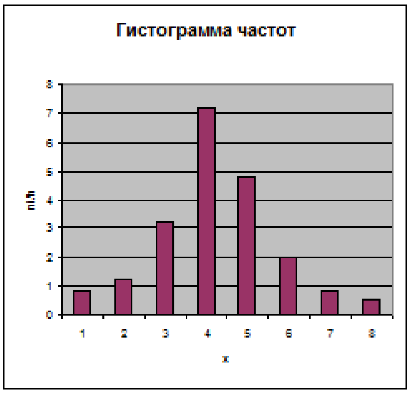
Полигон относительных частот строится аналогично, за исключением того, что на оси ординат откладываются относительные частоты w_i.

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 гистограммы частот равна сумме всех частот, т.е. ***объему выборки***.

В случае гистограммы ***относительных*** частот по оси ординат откладываются относительные частоты 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***.





2. Числовые характеристики выборки. Точечные оценки выборочных характеристик.

Пусть известно статистическое распределение частот количественного признака X. Обозначим через n_x число наблюдений, при которых наблюдалось значение признака, меньшее x и через n – общее число наблюдений. Очевидно, относительная частота события X < x равна n_x/n и является функцией x. Так как эта функция находится эмпирическим (опытным) путем, то ее называют эмпирической.

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события X < x. Таким образом, по определению $F^*(x) = n_x/n$, где n_x - число вариант, меньших x, n – объем выборки.

В отличие от эмпирической функции распределения выборки, функцию распределения $F(x)$ генеральной совокупности называют **теоретической функцией распределения.** Различие между этими функциями состоит в том, что теоретическая функция $F(x)$ определяет **вероятность** события X < x, тогда как эмпирическая – **относительную частоту** этого же события.

При росте n относительная частота события X < x, т.е. $F^*(x)$ стремится по вероятности к вероятности $F(x)$ этого события.

Иными словами: $$\lim_{n\to\infty} P[|F(x)-F^*(x)|<\varepsilon]=1 \qquad (\varepsilon>0)$$

Свойства эмпирической функции распределения:

1) Значения эмпирической функции принадлежат отрезку [0,1]

2) $F^*(x)$ - неубывающая функция

3) Если x_1- наименьшая варианта, то $F^*(x)$= 0 при $x \le x_1$, если x_k - наибольшая варианта, то $F^*(x)$=1 при $x > x_k$.

Эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Пример. Построим эмпирическую функцию по распределению выборки:

Варианты x_i	2	6	10
Частоты n_i	12	18	30

Найдем объем выборки: 12+18+30=60. Наименьшая варианта равна 2, поэтому $F^*(x)$=0 при x £ 2. Значение x<6, т.е. $x_1=2$, наблюдалось 12 раз, следовательно, $F^*(x)$=12/60=0,2 при 2< x £6. Аналогично, значения X < 10, т.е. $x_1=2$ и $x_2=6$ наблюдались 12+18=30 раз, поэтому $F^*(x)$=30/60 =0,5 при 6< x £10. Так как x=10 – наибольшая варианта, то $F^*(x)$=1 при x> 10. таким образом, искомая эмпирическая функция имеет вид:

$$F^*(x)=\begin{cases}0 & x\le 2\\ 0,2 & 2<x\le 6\\ 0,5 & 6<x\le 10\\ 1 & x>10\end{cases}$$

Важнейшие свойства статистических оценок

Пусть требуется изучить некоторый количественный признак генеральной совокупности. Допустим, что из теоретических соображений удалось установить, **какое именно** распределение имеет признак и необходимо оценить параметры, которыми оно определяется. Например, если изучаемый признак распределен в генеральной совокупности нормально, то нужно оценить математическое ожидание и среднее квадратическое отклонение; если признак имеет распределение Пуассона – то необходимо оценить параметр l.

Обычно имеются лишь данные выборки, например, значения количественного признака $x_1, x_2, \ldots, x_n$, полученные в результате n независимых наблюдений. Рассматривая $x_1, x_2, \ldots, x_n$ как независимые случайные величины $X_1, X_2, \ldots, X_n$ можно сказать, что ***найти статистическую оценку неизвестного параметра теоретического распре-***

деления – значит найти функцию от наблюдаемых случайных величин, которая дает приближенное значение оцениваемого параметра. Например, для оценки математического ожидания нормального распределения роль функции выполняет среднее арифметическое $\overline{X} = (X_1 + X_2 + ... + X_n)/n$

Для того чтобы статистические оценки давали корректные приближения оцениваемых параметров, они должны удовлетворять некоторым требованиям, среди которых важнейшими являются требования ***несмещенности*** и ***состоятельности*** оценки.

Пусть Θ^*- статистическая оценка неизвестного параметра Θ теоретического распределения. Пусть по выборке объема n найдена оценка Θ_1^*. Повторим опыт, т.е. извлечем из генеральной совокупности другую выборку того же объема и по ее данным получим другую оценку Θ_2^*. Повторяя опыт многократно, получим различные числа $\Theta_1^*, \Theta_2^*, ...\Theta_k^*$. Оценку Θ^* можно рассматривать как случайную величину, а числа $\Theta_1^*, \Theta_2^*, ...\Theta_k^*$- как ее возможные значения.

Если оценка Θ^* дает приближенное значение Θ **с избытком**, т.е. каждое число $\Theta_i^* \ (i = 1,...,k)$ больше истинного значения Θ то, как следствие, математическое ожидание (среднее значение) случайной величины Θ^* больше, чем Θ :$M(\Theta^*) > \Theta$. Аналогично, если Θ^* дает оценку **с недостатком**, то $M(\Theta^*) < \Theta$.

Таким образом, использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, привело бы к систематическим (одного знака) ошибкам. Если, напротив, $M(\Theta^*) = \Theta$, то это гарантирует от систематических ошибок.

Несмещенной называют статистическую оценку Θ^*, математическое ожидание которой равно оцениваемому параметру Θ при любом объеме выборки $M(\Theta^*) = \Theta$.

Смещенной называют оценку, не удовлетворяющую этому условию.

Несмещенность оценки еще не гарантирует получения хорошего приближения для оцениваемого параметра, так как возможные значения Θ_i^* могут быть ***сильно рассеяны*** вокруг своего среднего значения, т.е. дисперсия $D(\Theta_*)$ может быть значительной. В этом случае найденная по данным одной выборки оценка, например, Θ_1^*, может оказаться значительно удаленной от среднего значения $\overline{\Theta^*}$,а значит, и от самого оцениваемого параметра.

Эффективной называют статистическую оценку, которая, при заданном объеме выборки n, имеет ***наименьшую возможную дисперсию***.

При рассмотрении выборок большого объема к статистическим оценкам предъявляется требование ***состоятельности***.

Состоятельной называется статистическая оценка, которая при n®¥ стремится по вероятности к оцениваемому параметру. Например, если дисперсия несмещенной оценки при n®¥ стремится к нулю, то такая оценка оказывается и состоятельной.

Пусть для изучения генеральной совокупности относительно количественного признака X извлечена выборка объема n.

Выборочным средним $\overline{x_B}$ называют среднее арифметическое значение признака выборочной совокупности. Если все значения $x_1, x_2, ..., x_n$ признака выборки объема n различны, то $\overline{x_B} = (x_1 + x_2 + ... + x_n)/n$.

Если значения признака $x_1, x_2, \ldots, x_k$ имеют частоты $n_1, n_2, \ldots, n_k$ соответственно, причем $n_1+n_2+\ldots+n_k=n$, то $\overline{x_B}=(n_1x_1+n_2x_2+\ldots+n_kx_k)/n=(1/n)\sum_{i=1}^{k}n_ix_i$.

Выборочное среднее, найденное по данным одной выборки, равно определенному числу. При извлечении других выборок того же объема выборочное среднее будет меняться от выборки к выборке. То есть выборочное среднее можно рассматривать как случайную величину и говорить о его распределениях (теоретическом и эмпирическом) и о числовых характеристиках этого распределения (например, о математическом ожидании и дисперсии).

Для характеристики рассеяния наблюдаемых значений количественного признака выборки вокруг среднего значения $\overline{x_B}$ вводится ***выборочная дисперсия***. ***Выборочной дисперсией*** D_B называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\overline{x_B}$. Если все значения $x_1, x_2, \ldots, x_n$ признака выборки объема n различны, то $D_B=(1/n)\sum_{i=1}^{n}(x_i-\overline{x_B})^2$.

Если значения признака $x_1, x_2, \ldots, x_k$ имеют частоты $n_1, n_2, \ldots, n_k$ соответственно, причем $n_1+n_2+\ldots+n_k=n$, то $D_B=(1/n)\sum_{i=1}^{k}n_i(x_i-\overline{x_B})^2$.

Аналогично выборочным среднему и дисперсии определяются ***генеральные среднее и дисперсия***, характеризующие генеральную совокупность в целом. Для расчета этих характеристик достаточно в вышеприведенных соотношениях заменить объем выборки n на объем генеральной совокупности N.

Фундаментальное значение для практики имеет нахождение среднего и дисперсии признака ***генеральной совокупности*** по соответствующим известным ***выборочным*** параметрам. Можно показать, что ***выборочное среднее*** является несмещенной состоятельной оценкой генерального среднего. В то же время, несмещенной состоятельной оценкой генеральной дисперсии оказывается не выборочная дисперсия D_B, а так называемая «***исправленная» выборочная дисперсия***, равная $s^2=\frac{n}{n-1}D_B$.

Таким образом, в качестве оценок генерального среднего и дисперсии в математической статистике принимают выборочное среднее и исправленную выборочную дисперсию.

3. Интервальные оценки, их свойства. Метод доверительных интервалов при заданных условиях.

До сих пор мы рассматривали ***точечные*** оценки, т.е. такие оценки, которые определяются одним числом. При выборке ***малого объема*** точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. В связи с этим при небольшом объеме выборки пользуются интервальными оценками.

Интервальной называют оценку, определяющуюся двумя числами – концами интервала. Пусть найденная по данным выборки статистическая характеристика Θ^* служит оценкой неизвестного параметра Θ. Очевидно, Θ^* тем точнее определяет параметр Θ, чем меньше абсолютная величина разности $|\Theta-\Theta^*|$. Другими словами, ли $\delta > 0$ и $\left|\Theta - \Theta^*\right| < \delta$, то чем меньше d, тем точнее оценка. Таким образом, положительное число d характеризует ***точность оценки***.

Статистические методы ***не позволяют*** утверждать, что оценка Θ^* удовлетворяет неравенству $\left|\Theta - \Theta^*\right| < \delta$; можно говорить лишь о вероятности, с которой это неравенство осуществляется.

Надежностью (доверительной вероятностью) оценки Θ по Θ^* называют вероятность g, с которой осуществляется неравенство $\left|\Theta - \Theta^*\right| < \delta$. Обычно надежность оценки задается заранее, причем в качестве g берут число, близкое к единице – как правило 0,95; 0,99 или 0,999.

Пусть вероятность того, что $\left|\Theta - \Theta^*\right| < \delta$ равна g: $P\left[\left|\Theta - \Theta^*\right| < \delta\right] = \gamma$.

Заменим неравенство $\left|\Theta - \Theta^*\right| < \delta$ равносильным ему двойным неравенством

$P\left[\Theta^* - \delta < \Theta < \Theta^* + \delta\right] = \gamma$.

Это соотношение следует понимать так: вероятность того, что интервал $(\Theta^* - \delta,\ \Theta^* + \delta)$ заключает в себе (покрывает) неизвестный параметр Q, равна γ.

Таким образом, **доверительным** называют интервал $(\Theta^* - \delta,\ \Theta^* + \delta)$, который покрывает неизвестный параметр с заданной надежностью γ.

Величину 1 - g = a называют уровнем значимости или вероятностью ошибки.

Для построения интервальной оценки параметра необходимо знать закон его распределения как случайной величины

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

Пусть количественный признак X генеральной совокупности распределен нормально, причем среднее квадратическое отклонение s этого распределения ***известно***. Требуется оценить неизвестное математическое ожидание a по выборочному среднему $\bar{x}$. Найдем доверительные интервалы, покрывающие параметр a с надежностью γ.

Будем рассматривать выборочное среднее $\bar{x}$ как случайную величину $\bar{X}$ (т.к. $\bar{x}$ меняется от выборки к выборке) и выборочные значения $x_1; x_2; \ldots; x_N$ - как одинаково распределенные независимые случайные величины $X_1, X_2, \ldots, X_N$ (эти числа также меняются от выборки к выборке). Другими словами, математическое ожидание каждой из этих величин равно a и среднее квадратическое отклонение - s. Так как случайная величина X распределена нормально, то и выборочное среднее $\bar{X}$ также распределено нормально. Параметры распределения $\bar{X}$ равны $M(\bar{X}) = a$; $\sigma(\bar{X}) = \sigma / \sqrt{n}$.

Потребуем, чтобы выполнялось соотношение $P(\left|\bar{X} - a\right| < \delta) = \gamma$,

где γ - заданная надежность. Используем формулу $P(\left|X - a\right| < \delta) = 2\,\Phi(\delta / \sigma)$.

Заменим X на $\bar{X}$ и s на $\sigma(\bar{X}) = \sigma / \sqrt{n}$ и получим

$$P(|\overline{X}-a|<\delta)=2\,\Phi(\delta\sqrt{n}/\sigma)=2\Phi(t)$$

где $t=\delta\sqrt{n}/\sigma$. Выразив из последнего равенства δ, получим $P(|\overline{X}-a|<t\sigma/\sqrt{n})=2\Phi(t)$

Так как вероятность P задана и равна γ, окончательно имеем

$$P(\overline{x}-t\sigma/\sqrt{n}<a<\overline{x}+t\sigma/\sqrt{n})=2\Phi(t)=\gamma.$$

Таким образом, с надежностью γ можно утверждать, что доверительный интервал $(\overline{x}-t\sigma/\sqrt{n}<a<\overline{x}+t\sigma/\sqrt{n})$ покрывает неизвестный параметр а, причем точность оценки равна $t\sigma/\sqrt{n}$.

Число t определяется из равенства $\Phi(t)=\gamma/2$; по таблице функции Лапласа находят аргумент t, которому соответствует значение функции Лапласа, равное $\gamma/2$.

Отметим два момента:

1) при возрастании **объема** выборки n число δ убывает и, следовательно, точность оценки увеличивается,

2) увеличение **надежности** оценки $\gamma/2=\Phi(t)$ приводит к увеличению t (так как функция Лапласа возрастающая функция) и, следовательно, к возрастанию δ, то есть ***увеличение надежности*** оценки влечет за собой ***уменьшение ее точности***.

Если требуется оценить математическое ожидание с наперед заданной точностью δ и надежностью γ, то минимальный объем выборки, который обеспечит эту точность, находят по формуле $n=t^2\sigma^2/\delta^2$, следующей из равенства $\delta=t\sigma/\sqrt{n}$.

Доверительный интервал для математического ожидания нормального распределения при *неизвестной* дисперсии

Пусть количественный признак X генеральной совокупности распределен нормально, причем среднее квадратическое отклонение s этого распределения ***неизвестно***. Требуется оценить неизвестное математическое ожидание с помощью доверительных интервалов.

Оказывается, что по данным выборки можно построить случайную величину $T=\dfrac{\overline{X}-a}{S/\sqrt{n}}$, которая имеет распределение Стьюдента с $k=n-1$ степенями свободы. В последнем выражении - $\overline{X}$ - выборочное среднее, S - исправленное среднее квадратическое отклонение, n - объем выборки; возможные значения случайной величины T мы будем обозначать через t. Плотность распределения Стьюдента имеет вид

$S(t,n)=B_n\left[1+\dfrac{t^2}{n-1}\right]^{-n/2}$, где B_n некоторая постоянная, выражающаяся через гамма – функцию.

Несколько слов о распределении Стьюдента. Пусть $\xi_0, \xi_1, \ldots, \xi_n$ - независимые стандартные нормальные величины. Тогда случайная величина $t_n=\dfrac{\xi_0}{\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2}}$

имеет ***распределение Стьюдента*** (В. Госсет) с n степенями свободы. При росте числа степеней свободы распределение Стьюдента стремится к нормальному распределению и уже при $n \geq 30$ использование нормального распределения дает хорошие результаты.

Как видно, распределение Стьюдента определяется параметром n – объемом выборки (или, что то же самое – числом степеней свободы $k = n - 1$) и не зависит от неизвестных параметров a, σ. Поскольку $S(t,n)$ - четная функция от t, то вероятность выполнения неравенства $\left|\frac{\overline{X}-a}{S/\sqrt{n}}\right| < \gamma$ определяется следующим образом: $P\left(\left|\frac{\overline{X}-a}{S/\sqrt{n}}\right| < t_\gamma\right) = 2\int_0^{t_\gamma} S(t,n)dt = \gamma$.

Заменив неравенство в круглых скобках двойным неравенством, получим выражение для искомого доверительного интервала $P\left(\overline{X} - \frac{t_\gamma S}{\sqrt{n}} < a < \overline{X} + \frac{t_\gamma S}{\sqrt{n}}\right) = \gamma$

Итак, с помощью распределения Стьюдента найден доверительный интервал $\overline{x} - \frac{t_\gamma s}{\sqrt{n}} < a < \overline{x} + \frac{t_\gamma s}{\sqrt{n}}$, покрывающий неизвестный параметр а с надежностью γ. По таблице распределения Стьюдента и заданным n и γ можно найти t_γ и используя найденные по выборке $\overline{x}$ и s, можно определить доверительный интервал.

Пример. Количественный признак X генеральной совокупности распределен нормально. По выборке объема n = 16 найдены генеральное среднее $\overline{x} = 20{,}2$ и исправленное среднее квадратическое отклонение $s = 0{,}8$. Требуется оценить неизвестное математическое ожидание при помощи доверительного интервала с надежностью 0,95.

Решение. Найдем t_γ по таблице распределения Стьюдента, используя значения $\gamma = 0.95;\ n = 16$. Этот параметр оказывается равным 2,13. Найдем границы доверительного интервала:

$$\overline{x} - t_\gamma\ s/\sqrt{n} = 20{,}2 - 2{,}13 \cdot 0{,}8/\sqrt{16} = 19.77$$

$$\overline{x} - +t_\gamma\ s/\sqrt{n} = 20{,}2 + 2{,}13 \cdot 0{,}8/\sqrt{16} = 20.63$$

То есть с надежностью 0,95 неизвестный параметр а заключен в доверительном интервале $19{,}77 < a < 20{,}63$

Можно показать, что при возрастании объема выборки n распределение Стьюдента стремится к нормальному. Поэтому практически при n > 30 можно вместо него пользоваться нормальным распределением. При ***малых*** n это приводит к значительным ошибкам.

Доверительный интервал для оценки среднего квадратического отклонения s нормального распределения

Пусть количественный признак X генеральной совокупности распределен нормально и требуется оценить неизвестное генеральное среднее квадратическое отклонение s по исправленному выборочному среднему квадратическому отклонению s. Найдем доверительные интервалы, покрывающие параметр s с заданной надежностью γ.

Потребуем, чтобы выполнялось соотношение

$P(|\sigma - s| \quad \delta \quad) = \gamma$ или $P(s - \delta < \sigma < s + \delta) = \gamma$

Преобразуем двойное неравенство $s - \delta < \sigma < s + \delta$ в равносильное неравенство $s(1 - \delta / s) < \sigma < s(1 + \delta / s)$ и обозначим d / s = q. Имеем $s(1 - q) < \sigma < s(1 + q)$ (A)

и необходимо найти q. С этой целью введем в рассмотрение случайную величину $\chi = \frac{S}{\sigma}\sqrt{n-1}$

Оказывается, величина $\frac{S^2}{\sigma^2}(n-1)$ распределена по закону χ^2 с n – 1 степенями свободы.

Несколько слов о распределении хи-квадрат. Если $\xi_i,\ i = 1,2,\dots,n$ - независимые стандартные нормальные величины, то говорят, что случайная величина $\chi_n^2 = \sum_{i=1}^{n} \xi_i^2$ имеет распределение ***хи-квадрат*** с n степенями свободы.

Плотность распределения с имеет вид

$$R(\chi, n) = \frac{\chi^{n-1} e^{-\chi^2/2}}{2^{(n-3)/2}\,\Gamma\left(\frac{n-1}{2}\right)}$$

Это распределение не зависит от оцениваемого параметра s, а зависит только от объема выборки n.

Преобразуем неравенство (A) так, чтобы оно приняло вид $\chi_1 < \chi < \chi_2$. Вероятность этого неравенства равна заданной вероятности γ, т.е. $\int_{\chi_1}^{\chi_2} R(\chi, n)\,d\chi = \gamma$.

Предполагая, что q < 1, перепишем (A) в виде

$$\frac{1}{S(1+q)} < \frac{1}{\sigma} < \frac{1}{S(1-q)},$$

далее, умножим все члены неравенства на $S\sqrt{n-1}$:

$$\frac{\sqrt{n-1}}{(1+q)} < \frac{S\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{(1-q)} \quad \text{или} \quad \frac{\sqrt{n-1}}{(1+q)} < \chi < \frac{\sqrt{n-1}}{(1-q)}.$$

Вероятность того, что это неравенство, а также равносильное ему неравенство (A) будет справедливо, равна

$$\int_{\sqrt{n-1}/(1+q)}^{\sqrt{n-1}/(1-q)} R(\chi, n)\,d\chi = \gamma.$$

Из этого уравнения можно по заданным n и γ найти q, используя имеющиеся расчетные таблицы. Вычислив по выборке s и найдя по таблице q, получим искомый интервал (A1), покрывающий s с заданной надежностью γ.

Пример. Количественный признак X генеральной совокупности распределен нормально. По выборке объема n = 25 найдено исправленное среднее квадратическое отклонение s = 0.8. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение s с надежностью 0,95.

Решение. По заданным γ и σ по таблице находим значение q = 0.32. Искомый доверительный интервал есть

$$0{,}8(1-0{,}32) < \sigma < 0.8(1+0.32).$$

Мы предполагали, что q < 1. Если это не так, то мы придем к соотношениям

$$\sqrt{n-1}/(1+q) < \chi < \infty,$$

и значение q >1 может быть найдено из уравнения

$$\int_{\sqrt{n-1}/(1+q)}^{\infty} R(\chi, n)\,d\chi = \gamma$$

4. Метод моментов

Параметрическое оценивание закона распределения

Результаты предварительной обработки наблюдений случайной величины, дополненные сведениями о сущности изучаемого явления, зачастую оказываются достаточными для того, чтобы сформулировать гипотезу о модели закона распределения изучаемой случайной величины, нормальный ли этот закон, биномиальный или какой-либо другой. Используя наблюдения, можно найти оценки параметров предполагаемой модели, т.е. оценки входящих в модель числовых характеристик. Подставив в модель вместо параметров найденные оценки, получим оценку предполагаемой модели закона распределения, которая называется **параметрической**. Оценивание закона распределения, не требующее предварительного выбора его модели и оценивания входящих в неё параметров, называется **непараметрическим.** Примерами непараметрических оценок неизвестного закона распределения являются вариационный ряд, выборочная функция распределения и выборочная плотность распределения.

Пример (*). Дано случайное распределение успеваемости 100 студентов-заочников, сдававших четыре экзамена:

Число сданных экзаменов	0	1	2	3	4
Число студентов	1	1	3	35	60

Здесь случайной величиной является число сданных экзаменов среди четырёх. Обозначим её X. Установим закон распределения этой величины.

Построим сначала его непараметрическую оценку. Величина X – дискретная. Дискретный вариационный ряд, заданный столбцами 2 и 4 табл. 1, даёт непараметрическую оценку закона распределения числа сданных экзаменов среди четырёх сдаваемых.

Теперь сформулируем гипотезу о модели закона распределения случайной величины X – числе сданных экзаменов среди четырёх сдаваемых. Процесс сдачи четырёх экзаменов представим как четыре испытания, относительно которых сделаем следующие допущения:

Таблица 1.

	Число сданных экзаменов x_i	Число студентов m_i	Частость $p_i{*}=\frac{m_i}{n}$	$p_i^{\text{теор}}=$ $=\tilde{N}_4^{x_i}\cdot 0{,}88^{x_i}\cdot 0{,}12^{4-x_i}$	$m_i^{\text{теор}}=np_i^{\text{теор}}$	$(m_i - m_i^{\text{теор}})^2$	$(m_i -$ $- m_i^{\text{теор}})^2:$ $:m_i^{\text{теор}}$
	2	3	4	5	6	7	8

	0	1	0,01	0,00021	0,021		
	1	1 5	0,01	0,00608	0,608 7,32	5,382	0,735
	2	3	0,03	0,06691	6,691		
	3	35	0,35	0,32711	32,711	5,239	0,160
	4	60	0,60	0,59969	59,969	0,001	0,000
Итого		*n* = 100	1,00	1,00000			0,895

- эти испытания независимы, т.е. вероятность сдачи любым студентом любого экзамена не зависит от того, будет сдано или нет любое количество других экзаменов;

- вероятность сдачи студентом любого отдельно взятого экзамена одна и та же и равна p, а вероятность «несдачи» равна $(1 - p)$.

Конечно, эти допущения могут вызывать некоторые сомнения, но возможно, что они не будут противоречить результатам наблюдений. При этих допущениях мы имеем дело с испытаниями Бернулли и число сданных экзаменов среди четырёх сдаваемых будет иметь биномиальный закон распределения, т.е. вероятность того, что студент сдаст λ экзаменов, равна

$$P(X = x) = C_4^x p^x (1 - p)^{4-x}, \quad x = 0, 1, 2, 3, 4. \qquad (1)$$

Найдём оценку параметра p, входящего в модель (6). В условиях испытаний Бернулли состоятельной, несмещённой и эффективной оценкой вероятности является частость. В рассматриваемом примере p – вероятность того, что студент сдаст экзамен, поэтому частость p^* этого события, учитывая, что имеются сведения об успеваемости 100 студентов, вычисляем следующим образом:

$$p^* = \frac{\text{числоэкзаменов, сданных 100 студентами}}{\text{числоэкзаменов, сдаваемых 100 студентами}} = \frac{\sum_{i=1}^{5} x_i m_i}{4\times 100} = \frac{0\times1+1\times1+2\times3+3\times35+4\times60}{100\times4} = 0{,}88.$$

Так как $\sum_{i=1}^{5} x_i m_i / 100 = \overline{X}$ - это среднее число экзаменов, сданных одним студентом, то p^* можно было бы определить и так:

$$p^* = \frac{\text{среднеечислоэкзаменов, сданныходнимстудентом}}{\text{числоэкзаменов, сдаваемыходнимстудентом}} = \frac{\overline{\tilde{O}}}{4} = 0{,}88.$$

Заметим, что если находить оценку параметра p в модели (6) методом максимального правдоподобия и при этом учесть, что число x_i наблюдалось m_i раз, то мы получили бы для p^* такую же формулу, а именно

$$p^*_{мп} = \sum_{i=1}^{5} x_i m_i /(4n).$$

Подставив в модель (1) вместо параметра p его оценку p^*, получим параметрическую оценку неизвестного закона распределения числа сданных экзаменов, построенную в предположении, что допустима биномиальная модель

$$P(X = x) = C_4^x 0{,}88^x 0{,}12^{4-x}; \quad x = 0,\ 1,\ 2,\ 3,\ 4. \qquad (2)$$

Теоретические вероятности $p_i^{теор}$ и частоты $m_i^{теор}$, вычисленные в предположении, что имеет место модель (2), содержатся в столбцах 5 и 6 табл. 5. Поскольку различия между соответствующими числами столбцов 4 и 5 или между числами столбцов 3 и 6 небольшие, можно сделать предварительное заключение о приемлемости биномиальной модели. Графически это заключение подтверждается рисунком, на котором кривая вероятностей $p_i^{теор}$ близка к кривой частостей p_i^*.

Метод более глубокого обоснования приемлемости той или иной модели называется **критерием согласия**.

1.7 Лекция 11-12 (Л-11-12) (4 ч.)

Тема: Статистические критерии, их виды

1.7.1 Вопросы лекции:

1. Статистические гипотезы, ошибки первого и второго рода.
2. Статистические критерии, их виды, мощность критерия.
3. Критерий Пирсона.
4. Выравнивание статистических рядов.

1.7.2. Краткое содержание вопросов:

1. Статистические гипотезы, ошибки первого и второго рода.

На прошлой лекции мы рассматривали задачу построения доверительных интервалов для неизвестных параметров генеральной совокупности. Сегодня мы продолжим изучение основных задач математической статистики и перейдем к вопросу ***проверки статистических гипотез***.

Проверка статистических гипотез представляет собой важнейший этап процесса принятия решения в управленческой деятельности, позволяя проводить подготовительный этап предстоящих действий с учетом реальных характеристик процесса производства, контроля качества продукции, коммерческой деятельности, и т.п.

Как известно, ***закон распределения*** определяет количественные характеристики генеральной совокупности.

Если закон распределения **неизвестен**, но есть основания предположить, что он имеет определенный вид (например, А), то выдвигают гипотезу: генеральная совокупность распределена по закону А. В этой гипотезе речь идет **о виде** предполагаемого распределения.

Часто закон распределения известен, но неизвестны его **параметры**. Если есть основания предположить, что неизвестный параметр Θ равен определенному значению Θ_0, то может выдвигаться гипотеза $\Theta = \Theta_0$. В этой гипотезе речь идет о **предполагаемой величине параметра** известного распределения.

Возможны и другие гипотезы: о равенстве параметров двух или нескольких распределений, о независимости выборок и т. д.

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 $x_1, x_2, \ldots, x_n$ этой величины, второй - $y_1, y_2, \ldots, y_n$. 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 $H_0: \lambda = 3$ - 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).

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближенное распределение которой известно.

2. Статистические критерии, их виды, мощность критерия.

Статистическим критерием (или просто критерием) называют случайную величину (K), которая служит для проверки нулевой гипотезы. Например, если проверяют гипотезу о равенстве дисперсий двух нормальных генеральных совокупностей, то в качестве критерия K принимают отношение исправленных выборочных дисперсий $F = s_1^2 / s_2^2$.

Очевидно, что эта величина случайная, т.к. в различных опытах исправленные дисперсии принимают различные, заранее неизвестные значения.

Наблюдаемым значением критерия $K_{набл}$ называют значение критерия, вычисленное по выборкам. Например, если в вышеприведенном случае $s_1^2 = 20,\ s_2^2 = 5$, то $K_{набл} = 20/5 = 4$.

После выбора определенного критерия множество всех его возможных значений разбивают на два ***непересекающихся*** подмножества, одно из которых содержит значения критерия, при которых нулевая гипотеза ***отвергается***, а другое – при которых она ***принимается***.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу ***отвергают***.

Соответственно, о***бластью принятия гипотезы*** (областью допустимых значений) называют совокупность значений критерия, при которых гипотезу ***принимают***.

Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если области принятия гипотезы – гипотезу принимают.

Так как критерий K – одномерная случайная величина, то все ее возможные значения принадлежат некоторому интервалу и, соответственно, должны существовать точки, разделяющие критическую область и область принятия гипотезы. Такие точки называются ***критическими точками***.

Различают ***одностороннюю*** (***правостороннюю*** и ***левостороннюю***) и ***двустороннюю*** критические области.

Правосторонней называют критическую область, определяемую неравенством $K > k_{кр}$, где $k_{кр}$ - положительное число.

Левосторонней называют критическую область, определяемую неравенством $K < k_{кр}$, где $k_{кр}$ - отрицательное число.

Двусторонней называют критическую область, определяемую неравенствами $K < k_1$, $K > k_2$, где $k_2 > k_1$. В частности, если критические точки симметричны относительно нуля, двусторонняя критическая область определяется неравенствами $K < -k_{кр}$, $K > k_{кр}$ или равносильным неравенством $|K| > k_{кр}$. Различия между вариантами критических областей иллюстрирует следующий рисунок.

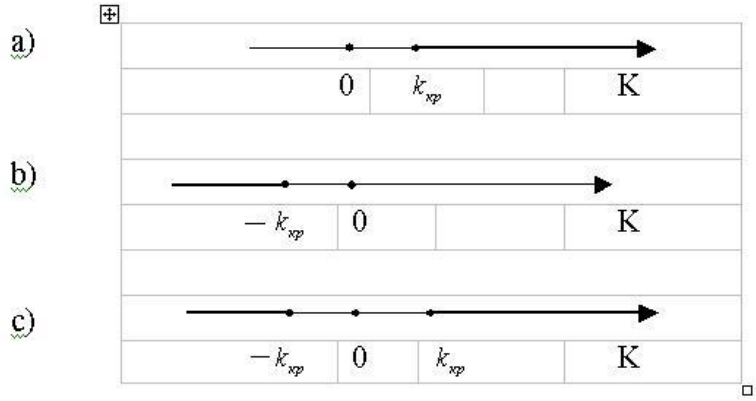


Рис. 1. Различные варианты критических областей а) правосторонняя, b) левосторонняя, c) двусторонняя

Мы подошли к вопросу об ***этапах проверки статистических гипотез***.

Выделим следующие ***этапы***:

- Формулируется нулевая гипотеза H_0
- Определяется критерий K, по значениям которого можно будет принять или отвергнуть H_0 и выбирается уровень значимости α
- По уровню значимости определяется критическая область
- По выборке вычисляется наблюдаемое значение критерия K, определяется, принадлежит ли оно критической области и на основании этого принимается гипотеза H_0 или альтернативная гипотеза H_1.

3. Критерий Пирсона.

Если закон распределения неизвестен, но есть основания предполагать, что он имеет определенный вид A, то проверяют нулевую гипотезу: генеральная совокупность распределена по закону A. Проверка этой гипотезы производится при помощи специально подобранной случайной величины – ***критерия согласия***.

Таким образом, к***ритерием согласия*** называют критерий проверки гипотезы о ***предполагаемом законе неизвестного распределения***.

Имеется несколько критериев согласия, причем наиболее часто используемым является критерий согласия К. Пирсона («хи квадрат»).

Пусть по выборке объема n получено эмпирическое распределение

Варианты…………………… x_i	x_1 x_2 …… x_s
Эмпирические частоты……. n_i	n_1 n_2 …… n_s

Для определенности рассмотрим сначала случай проверки статистической гипотезы о нормальном распределении генеральной совокупности.

Допустим, что в предположении нормального распределения генеральной совокупности вычислены теоретические частоты n_i^0. При уровне значимости α требуется проверить нулевую гипотезу: генеральная совокупность распределена нормально.

В качестве критерия проверки нулевой гипотезы примем случайную величину

$$\chi^2 = \sum_i (n_i - n_i^0)^2 / n_i^0 \qquad \text{(A)}$$

Естественно, чем меньше различаются эмпирические и теоретические частоты, тем меньше величина критерия, и, следовательно, он характеризует ***близость*** эмпирического и теоретического распределений.

Доказано, что при n больших распределения случайной величины (А) стремится к закону распределения χ^2 с k степенями свободы независимо от того, какому закону распределения подчинена генеральная совокупность. Поэтому сам критерий называют ***критерием согласия*** χ^2.

Число степеней свободы определяется из равенства $k = s - r - 1$, где s – число групп (частичных интервалов) выборки, r – число параметров предполагаемого распределения. В частности, если предполагаемое распределение – нормальное, то оценивают два параметра (математическое ожидание и среднее квадратическое отклонение), поэтому число степеней свободы $k = n - 3$.

Построим правостороннюю критическую область, исходя из требования, чтобы вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы была равна принятому уровню значимости α: $P[\chi^2 > \chi^2_{кр}(\alpha; k)] = \alpha$.

Таким образом, правосторонняя критическая область определяется неравенством $\chi^2 > \chi^2_{кр}(\alpha; k)$, а область принятия нулевой гипотезы – соответственно неравенством $\chi^2 < \chi^2_{кр}(\alpha; k)$. Обозначим значение критерия, вычисленного по данным наблюдений, через $\chi^2_{набл}$ и сформулируем правило проверки нулевой гипотезы:

Для того, чтобы при заданном уровне значимости проверить нулевую гипотезу H_0: генеральная совокупность распределена нормально, необходимо сначала вычислить теоретические частоты, а затем наблюдаемое значение критерия $\chi^2_{набл} = \sum_i (n_i - n_i^0)^2 / n_i^0$ и по таблице критических точек распределения χ^2, по заданному уровню значимости а и числу степеней свободы k = n – 3 найти критическую точку $\chi^2_{кр}(\alpha; k)$. Если $\chi^2_{набл} < \chi^2_{кр}$ -

нет оснований отвергать нулевую гипотезу. В противном случае нулевую гипотезу отвергают, считая, что генеральная совокупность не распределена по нормальному закону.

Отметим два обстоятельства.

- Объем выборки должен быть ***достаточно велик*** (не менее 50). Каждая группа должна содержать не менее 5-8 вариант, а малочисленные группы следует объединять в одну, суммируя частоты.
- Поскольку возможны ошибки первого и второго рода, следует проявлять осторожность. Например, можно повторить опыт, увеличить число наблюдений, построить предварительно график распределения и т.п.

Применение критерия согласия Пирсона не ограничивается случаем нормального распределения. Приведем примеры использования критерия Пирсона.

Пример1. При уровне значимости 0,05 проверить гипотезу о распределении по закону Пуассона генеральной совокупности, если по данным выборки объема $n=60$ получен следующий вариационный ряд:

Варианты k		1	2	3	4	5	6	7
Частоты n_k	8	17	16	10	6	2	0	1

Решение: Для расчета теоретических частот используем формулу Пуассона $P(k)=\frac{\lambda^k}{k!}e^{-\lambda}$

Для оценки параметра λ используем соотношение $\lambda=\frac{1}{n}\sum_{k=0}^{7}k\cdot n_k=2$

Вычислим $P(0)=0.1353$ $P(1)=0.2707$ $P(2)=0.2707$ $P(3)=0.1804$ $P(4)=0.0902$ $P(5)=0.0361$ $P(6)=0.0120$ $P(7)=0.0034$.

Значение χ^2 равно $\chi^2=\sum_{k=0}^{7}\frac{(n_k-n\cdot p_k)^2}{n\cdot p_k}=\frac{(8-60\cdot 0.1353)^2}{60\cdot 0.1353}+\frac{(17-60\cdot 0.2707)^2}{60\cdot 0.2757}+\ldots=0.2$.

Вычислим число степеней свободы $k=n-s-1=n-2=8-2=6$. По таблице критических точек распределения хи-квадрат при $n=6$ и $\alpha=0.05$ находим $\chi^2_{кр}=12{,}59$. Так как наблюдаемое значение меньше критического, то наблюдаемые значения согласуются с распределением Пуассона и нулевая гипотеза принимается.

4.Выравнивание статистических рядов.

Рассмотрим задачу «выравнивания» статистического распределения. Порядок решения этой задачи может быть следующим.

1. На основании статистических данных, оформленных в виде интервальной таблицы частот p^*, строят полигон или гистограмму и по внешнему виду этих графиков выдвигают гипотезу (делают предположение) о возможном теоретическом законе распределения случайной величины (кривой распределения).

Замечание. В некоторых случаях вид теоретической кривой распределения выбирается заранее из соображений, связанных с существом задачи.

2. Выясняют, от каких параметров зависит аналитическое выражение выбранной кривой распределения, и находят статистические оценки этих параметров. В этом случае задача выравнивания статистического распределения переходит в задачу рационального

выбора тех значений параметров, при которых соответствие между статистическим и теоретическим распределениями оказывается наилучшим.

Например, если выдвигается гипотеза о нормальном законе распределения $X \sim N(a; \sigma)$, то он зависит только от двух параметров: математического ожидания a и среднего квадратического отклонения σ. Их наилучшими статистическими оценками будут соответственно среднее выборочное $\bar{x}$ и выборочное среднее квадратическое отклонение $\tilde{\sigma}$, т. е.

$$a \approx \bar{x},\ \sigma \approx \tilde{\sigma}.$$

3. С учетом выдвинутой гипотезы о законе распределения случайной величины находят вероятности p_i попадания случайной величины в каждый из интервалов, указанных в статистической таблице распределения; записывают их в третьей строке таблицы и сравнивают полученные значения вероятностей p_i с соответствующими заданными частотами p_i* (для наглядности можно изобразить графически). Проводя такое сравнение, делается приблизительная оценка степени согласования статистического и теоретического распределений. На этом первый этап решения задачи по определению закона распределения случайной величины заканчивается.

Пример. Для разумного планирования и организации работы ремонтных мастерских специальной техники оказалось необходимым изучить длительность ремонтных операций, производимых мастерскими.
Результаты (сгруппированные по интервалам) соответствующего статистического обследования (фиксированы длительности операций в
100 случаях) представлены в таблице:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
n_i	36	24	16	10	7	4	3

Требуется выровнять это статистическое распределение с помощью показательного закона $f(t) = \lambda e^{-\lambda t}$ (при $t \geq 0$), где λ – длительность операции в единицу времени.

Решение

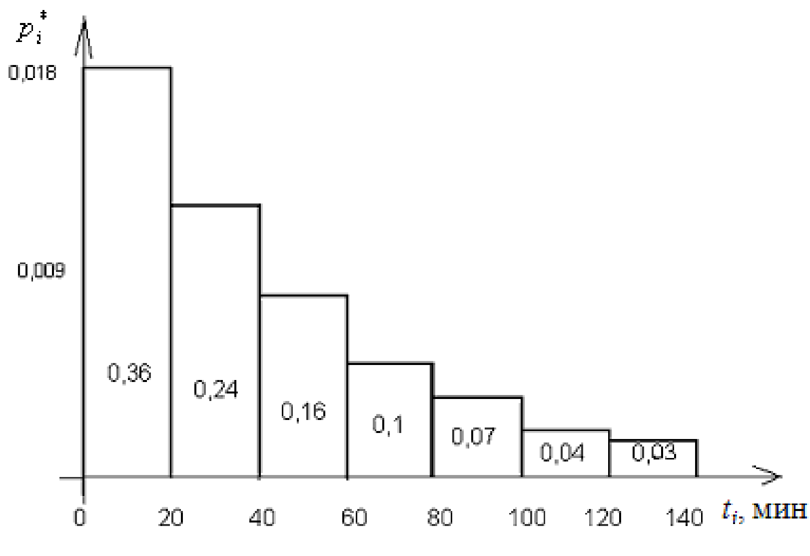
1. По данной таблице абсолютных частот построим таблицу относительных частот и соответствующую ей гистограмму:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
p_i*	0,36	0,24	0,16	0,10	0,07	0,04	0,03

.

$$\sum p_i^* = 0,36 + 0,24 + 0,16 + 0,1 + 0,07 + 0,04 + 0,03$$

Гистограмма относительных частот имеет вид:



Высоты прямоугольников гистограммы равны:

$$\Delta_1 = \frac{0,36}{20} = 0,018;$$

$$\Delta_2 = \frac{0,24}{20} = 0,012;$$

$$\Delta_3 = \frac{0,16}{20} = 0,008;$$

$\Delta_4 = \frac{0,10}{20} = 0,005$; $\Delta_5 = \frac{0,07}{20} = 0,0035$; $\Delta_6 = \frac{0,04}{20} = 0,002$; $\Delta_7 = \frac{0,03}{20} = 0,0015$.

2. По внешнему виду гистограммы выдвигаем гипотезу, что случайная величина T (время ремонта) подчиняется показательному закону

$$f(t) = \lambda e^{-\lambda t},$$

который зависит только от одного параметра λ (длительность операции в единицу времени).

Параметр $\lambda = \frac{1}{m_i}$, где m_i – математическое ожидание (среднее время ремонта) случайной величины T.

Следовательно, для выравнивания статистического распределения с помощью кривой показательного распределения найдем статистическую оценку параметра m_i:

$$m_i \approx \bar{t} = 10 \cdot 0,36 + 30 \cdot 0,24 + 50 \cdot 0,16 + 70 \cdot 0,1 + 90 \cdot 0,07 + 110 \cdot 0,04 + 130 \cdot 0,03 = 40$$

(числа 10, 30, 50, 70, 90, 110, 130 – это середины интервалов).

Тогда параметр $\lambda = \frac{1}{40}$.

3. Запишем теоретический закон распределения в виде функции плотности вероятности с учетом значения $\lambda = \frac{1}{40}$:

$$f(t) = \frac{1}{40} e^{-\frac{1}{40}t}.$$

По формуле вероятности попадания случайной величины (распределенной по показательному закону) на заданный интервал (α, β)

$$P(\alpha < X < \beta) = e^{-\lambda\alpha} - e^{-\lambda\beta}$$

найдем теоретические вероятности p_i, попадания случайной величины Т в каждый из семи интервалов и сравним их с соответствующими статистическими частотами p_i^*:

$$p_1 = P(0 < T < 20) = e^{-\frac{0}{40}} - e^{-\frac{20}{40}} = e^0 - e^{-0,5} \approx 1 - 0,6 = 0,4;$$

$$p_2 = P(20 < T < 40) = e^{-\frac{20}{40}} - e^{-\frac{40}{40}} = e^{-0,5} - e^{-1} \approx 0,6 - 0,37 = 0,23;$$

$$p_3 = P(40 < T < 60) = e^{-\frac{40}{40}} - e^{-\frac{60}{40}} = e^{-1} - e^{-1,5} \approx 0,37 - 0,22 = 0,15;$$

$$p_5 = P(60 < T < 80) = e^{-\frac{60}{40}} - e^{-\frac{80}{40}} = e^{-1,5} - e^{-2} \approx 0,22 - 0,14 = 0,08;$$

$$p_5 = P(80 < T < 100) = e^{-\frac{80}{40}} - e^{-\frac{100}{40}} = e^{-2} - e^{-2,5} \approx 0,14 - 0,08 = 0,06;$$

$$p_5 = P(100 < T < 120) = e^{-\frac{100}{40}} - e^{-\frac{120}{40}} = e^{-2,5} - e^{-3} \approx 0,08 - 0,05 = 0,03;$$

$$p_7 = P(120 < T < 140) = e^{-\frac{120}{40}} - e^{-\frac{140}{40}} = e^{-3} - e^{-3,5} \approx 0,05 - 0,03 = 0,02.$$

Для удобства сравнения теоретических вероятностей p_i с частотами p_i^* запишем полученные вероятности p_i в третью строку таблицы:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
p_i^*	0,36	0,24	0,16	0,10	0,07	0,04	0,03
p_i	0,40	0,23	0,15	0,08	0,06	0,03	0,02

Замечаем, что расхождение между опытными частотами p_i^* и теоретическими вероятностями p_i незначительны. Следовательно, вполне допустима гипотеза о показательном законе распределения изучаемой случайной величины Т.

4. Построим на одном графике с гистограммой выравнивающую ее кривую распределения f(t). Для этого вычислим значения

$$f(t)=\frac{1}{40}e^{-\frac{1}{40}t}$$

например, на правых концах интервалов:

$$f(20)=\frac{1}{40}e^{-\frac{1}{40}\cdot 20}=\frac{1}{40}e^{-0,5}\approx 0,015;\quad f(40)=\frac{1}{40}e^{-\frac{1}{40}\cdot 40}=\frac{1}{40}e^{-1}\approx 0,009;$$

$$f(60)=\frac{1}{40}e^{-\frac{1}{40}\cdot 60}=\frac{1}{40}e^{-1,5}\approx 0,006;\quad f(80)=\frac{1}{40}e^{-\frac{1}{40}\cdot 80}=\frac{1}{40}e^{-2}\approx 0,004;$$

$$f(100)=\frac{1}{40}e^{-\frac{1}{40}\cdot 100}=\frac{1}{40}e^{-2,5}\approx 0,002;\quad f(120)=\frac{1}{40}e^{-\frac{1}{40}\cdot 120}=\frac{1}{40}e^{-3}\approx 0,0013;$$

$$f(140)=\frac{1}{40}e^{-\frac{1}{40}\cdot 140}=\frac{1}{40}e^{-3,5}\approx 0,0008.$$

Построим график полученной кривой распределения f(t), в той же системе координат, что и гистограмма относительных частот.

Из рисунка видно, что теоретическая кривая f(t) сохраняет в основном существенные особенности статистического распределения.

1.9 Лекция 13-14 (Л-13-14) (4 ч.)

Тема: Стохастическая зависимость, функция регрессии.

1.9.1 Вопросы лекции:

1. Виды зависимостей между величинами. Функция регрессии.
2. Корреляционное отношение, коэффициент детерминации. Корреляционная зависимость.

1.9.2. Краткое содержание вопросов:

1. Виды зависимостей между величинами. Функция регрессии.

Две или несколько случайных величин могут быть связаны либо ***функциональной***, либо ***статистической*** (***стохастической***) зависимостью.

В экономике строгая функциональная зависимость реализуется редко, так как экономические показатели подвержены действию случайных, часто неконтролируемых факторов. Чаще имеет место так называемая ***статистическая*** зависимость, при которой изменение одной из величин влечет изменение ***распределения*** другой. В частности, при изменении одной из величин может изменяться ***среднее значение*** другой.

Пример статистической зависимости: урожай зерна Y зависит от количества внесенных удобрений X . С одинаковых по площади участков при равных количествах внесенных удобрений снимают разные урожаи. Это связано с влиянием случайных факторов (осадки, температура воздуха и др.). Вместе с тем, средний урожай зависит от количества удобрений, т.е. Y связано с X статистической зависимостью.

При изучении статистических зависимостей различают ***корреляцию*** и ***регрессию***. Основным методом исследования статистических зависимостей выступает ***корреляционно – регрессионный*** анализ.

Корреляционный анализ состоит в определении ***степени связи*** между случайными величинами.

Регрессионный анализ устанавливает ***формы зависимости*** между случайной величиной Y (зависимой переменной) и значениями одной или нескольких переменных величин X (независимыми переменными).

Одна из наиболее распространенных задач статистического исследования состоит в изучении связи между наблюдаемыми переменными. Знание взаимосвязей отдельных признаков дает возможность прогнозировать развитие ситуации при изменении конкретных характеристик объекта исследования. Основное содержание экономической политики, в конечном счете, может быть сведено к регулированию экономических переменных, осуществляемому на базе выявленной информации об их взаимовлиянии. Поэтому проблема изучения взаимосвязей показателей является одной из важнейших в статистическом анализе экономических систем.

Корреляция в широком смысле слова означает связь, соотношение между объективно существующими явлениями. Если случайные переменные причинно обусловлены, то имеется корреляция.

Корреляция может быть:
- положительной или отрицательной;
- в зависимости от числа переменных – простой или множественной;
- в зависимости от формы связи – линейной или нелинейной.

Важнейшими задачами *корреляционного анализа* являются:
- измерение силы связи двух или более факторов;
-отбор факторов, оказывающих существенное влияние на результативный признак (зависимую переменную) на основании измерения тесноты связи между факторами.

2.Корреляционное отношение, коэффициент детерминации. Корреляционная зависимость.

В случае лишь одной независимой переменой X в качестве меры связи между ней и зависимой переменной Y служит ***коэффициент корреляции***. Он оценивается по выборке объема n связанных пар наблюдений (x_i, y_i). В случае ***нескольких*** переменных необходимо последовательно вычислять коэффициенты корреляции по нескольким рядам числовых данных. Полученные коэффициенты сводят в таблицы, называемые ***корреляционными матрицами***.

Корреляционная матрица представляет собой квадратную матрицу, на пересечении строки и столбца которой находится коэффициент корреляции между соответствующими переменными.

Если в результате n испытаний система двух случайных величин (X,Y) приняла значения $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$, то коэффициент корреляции равен

$$\rho_{XY}=\frac{(1/n)\cdot\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sigma_x\cdot\sigma_y}$$

где $\bar{x},\bar{y}$ - средние значения, а σ_x,σ_y - средние квадратические отклонения случайных величин X,Y соответственно.

Для ***многомерной*** выборки (т. е. в случае более двух факторов) необходимо рассчитать ***корреляционную матрицу*** $\rho(X_i, X_j) = \rho(i,j) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{k1} & \rho_{k2} & \cdots & \rho_{kk} \end{pmatrix}$,которая является симметричной относительно главной диагонали.

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 $M(Y|x_1, x_2, \ldots, x_m) = f(x_1, x_2, \ldots, x_m)$ 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 $x_1, x_2, \ldots, x_m$), зависимость должна быть дополнена некоторым слагаемым е, которое, по существу, является ***случайной величиной***. Получающиеся в результате соотношения $Y = f(x) + \varepsilon$ или

$$Y = f(x_1, x_2, \ldots, x_m) + \varepsilon$$

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.***

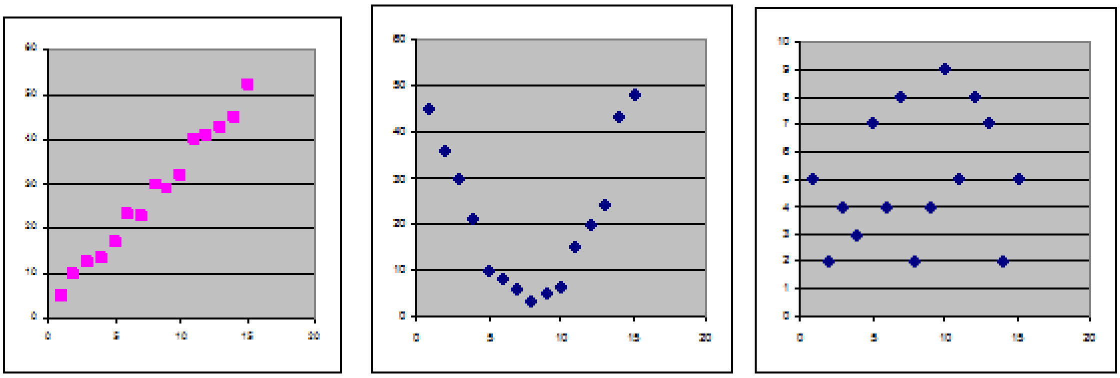


Рис.1 Корреляционные поля. А) – линейная регрессия; Б) – квадратичная регрессия; В) – отсутствие выраженной связи Y и X.

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 $(x_i, y_i),\ i = 1, \ldots, n$ найти оценки неизвестных параметров так, чтобы построенная линия регрессии являлась наилучшей среди всех других линий. 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 $(x_i, y_i),\ i = 1, \ldots, n$ 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 $nb_0 + b_1\sum x_i = \sum y_i$

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

и выразив из последних соотношений коэффициенты, получим

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

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

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 $M(Y|x_1, x_2, \ldots, x_m) = f(x_1, x_2, \ldots, x_m)$.

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

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

Нелинейная регрессия

Многие экономические зависимости не являются линейными. Например, при анализе эластичности спроса по цене применяется так называемая логарифмическая модель, при анализе издержек от объема выпуска – полиномиальная (кубическая) модель. Часто применяются и другие модели – например, обратная и экспоненциальная. Кратко рассмотрим некоторые из моделей нелинейной регрессии.

Пусть некоторая экономическая зависимость моделируется формулой

$$Y = AX^{\beta}$$

где A, b - параметры модели. Эта функция может отражать зависимость спроса Y на благо от его цены X (в этом случае b < 0) или от дохода X (b>0 – функция Энгеля). Прологарифмировав обе части последнего соотношения, получим $\ln Y = \ln A + \beta \ln X$; замена переменных $Y^* = \ln Y, \; X^* = \ln X$ позволяет свести уравнение к линейному виду

$$Y^* = \beta_0 + \beta X^* + \varepsilon.$$

По МНК можно рассчитать значения параметров аналогично случаю линейной модели (при этом вместо (x_i, y_i) рассматриваются $(\ln x_i, \ln y_i)$).

Обратная модель.

Обратная модель имеет вид $Y = \beta_0 + \beta_1\cdot\frac{1}{X} + \varepsilon$.

Заменой $X^* = \frac{1}{X}$ эта модель сводится к линейной. Обратная модель применяется, например, для характеристики связи удельных расходов сырья, материалов, топлива с объемом выпускаемой продукции.

Степенная функция вида $Y = \beta_0 + \beta_1 X + ... + \beta_m X^m + \varepsilon$

при m=3 (кубическая функция) в микроэкономике моделирует зависимость общих издержек от объема выпуска; квадратичная функция (m=2) отражает зависимость между объемом выпуска и средними или предельными издержками. Модель может быть сведена к ***линейной*** модели множественной регрессии с помощью замены $X \to X_1, X^2 \to X_2, ..., X^m \to X_m$. Параметры модели определяют с помощью МНК.

Показательная функция $Y = b_0 e^{\beta X} = b_0 m^X$

может использоваться при анализе изменения переменной Y с постоянным темпом прироста во времени. Примером может служить производственная функция Кобба – Дугласа с учетом научно – технического прогресса

$$Y = A K^{\alpha} L^{\beta} e^{\gamma t},$$

где K – затраты капитала, L – затраты труда, g характеризует темпы роста объема производства.

Прологарифмировав, получаем соотношение

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \gamma t,$$

которое сводится к линейному виду с помощью замен $y \to \ln Y,\ k \to \ln K,\ l \to \ln L,\ a \to \ln A$.

В заключение отметим, что построение и проверка качества уравнения регрессии требуют применения методов ***корреляционного*** анализа, позволяющих производить отбор существенных для описания регрессионной зависимости факторов.

1.10 Лекция 15 (Л-15) (2 ч.)

Тема: Понятие о случайной функции. Характеристики случайной функции

1.10.1 Вопросы лекции:

1. Два подхода к понятию о случайной функции, примеры случайных функций. Основные виды задач в теории случайных функций.
2. Закон распределения случайной функции.
4. Корреляционная теория случайной функции.
5. Математическое ожидание случайной функции, свойства.
6. Дисперсия случайной функции, свойства.

1.10.2. Краткое содержание вопросов:

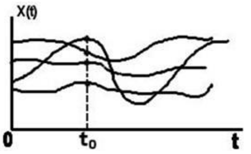
1. Два подхода к понятию о случайной функции, примеры случайных функций. Основные виды задач в теории случайных функций.

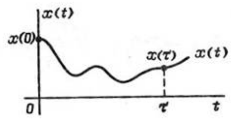
Случайным (стохастическим, вероятностным) процессом называется функция действительного переменного t, значениями которой являются соответствующие случайные величины X(t). В теории случайных процессов t трактуется как время, принимающее значения из некоторого подмножества T множества действительных чисел (tT, TR). В рамках классического математического анализа под функцией y=f(t) понимается такой тип зависимости переменных величин t и y, когда конкретному числовому значению аргумента t соответствует и притом единственное числовое значение функции y.

Для случайных процессов ситуация принципиально иная: задание конкретного аргумента t приводит к появлению случайной величины X(t) с известным законом распределения (если это дискретная случайная величина) или с заданной плотностью распределения (если это непрерывная случайная величина). Другими словами, исследуемая характеристика в каждый момент времени носит случайный характер с неслучайным распределением. Значения, которые принимает обычная функция y=f(t) в каждый момент времени, полностью определяет структуру и свойства этой функции.

Для случайных процессов дело обстоит иным образом: здесь совершенно не достаточно знать распределение случайной величины X(t) при каждом значении t, необходима информация об ожидаемых изменениях и их вероятностях, то есть информация о степени зависимости предстоящего значения случайного процесса от его предыстории.

Случайным процессом X(t) называется процесс, значение которого при любом фиксированном $t = t_0$ является случайной величиной $X(t_0)$. Случайная величина $X(t_0)$, в которую обращается с. п. при $t = t_0$, называется **сечением случайного процесса**, соответствующим данному значению аргумента t.





Реализацией случайного процесса X(t) называется неслучайная функция x(t), в которую превращается случайный процесс X(t) в результате опыта.

Основные виды задач в теории случайных функций.

В инженерных исследованиях решается два вида задач: задача анализа и синтеза. Задача анализа: известны характеристики входной функции и параметры системы, определить характеристики выходной функции. Задача синтеза: подобрать параметры системы так, чтобы функция на входе переходила в функцию на выходе с известными характеристиками.

2. Закон распределения случайной функции.

В рамках общего подхода к описанию случайных процессов характеристика сечений и любых их совокупностей осуществляется с помощью многомерных распределений. В частности, любое сечение характеризуется либо одномерной плотностью вероятности, либо одномерной функцией распределения $F(t; x)=P(X(t)\leq x)$.

Взаимосвязь любой пары сечений характеризуется двумерной плотностью вероятности или двумерной функцией распределения $F(t_1; t_2; x_1; x_2)=P(X(t_1)\leq x_1; X(t_2)\leq x_2)$, где $t_{1,2}$- два фиксированных момента времени; $x_{1,2}$- возможные значения случайных величин, соответствующих этим сечениям. Аналогично вводятся плотности и функции распределения трёх и более сечений, однако для большого числа случайных процессов оказывается достаточным ограничиться одномерными и двумерными распределениями.

Наиболее общий подход в описании случайных процессов состоит в задании всех его многомерных распределений, когда определена вероятность одновременного выполнения следующих событий:

$t_1, t_2,\ldots,t_n T, n N$: $X(t_i)\leq x_i$; $i=1,2,\ldots,n$;

$F(t_1;t_2;\ldots;t_n;x_1;x_2;\ldots;x_n)=P(X(t_1)\leq x_1; X(t_2)\leq x_2;\ldots; X(t_n)\leq x_n)$.

Такой способ описания случайных процессов универсален, но весьма громоздок.

Для получения существенных результатов выделяют наиболее важные частные случаи, допускающие применение более совершенного аналитического аппарата.

Если аргумент t принимает все действительные значения или все значения из некоторого интервала T действительной оси, то говорят о случайном процессе **с непрерывным временем**. Если t принимает только фиксированные значения, то говорят о случайном процессе **с дискретным временем**. Если сечение случайного процесса - дискретная случайная величина, то такой процесс называется **процессом с дискретными состояниями**. Если же любое сечение - непрерывная случайная величина, то случайный процесс называется **процессом с непрерывными состояниями**. В общем случае задать случайный процесс аналитически невозможно. Исключение составляют так называемые **элементарные случайные процессы**, вид которых известен.

3. Корреляционная теория случайной функции.

Аппарат числовых характеристик представляет собой весьма гибкий и мощный аппарат, позволяющий сравнительно просто решать многие практические задачи. Совершенно аналогичным аппаратом пользуются и в теории случайных функций. Для случайных функций также вводятся простейшие основные характеристики, аналогичные числовым характеристикам случайных величин, и устанавливаются правила действий с этими характеристиками. Такой аппарат оказывается достаточным для решения многих практических задач.

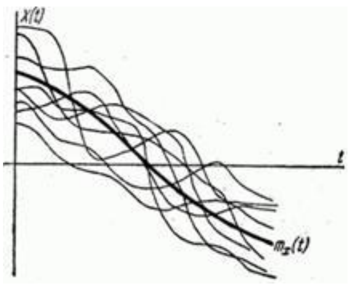
В отличие от числовых характеристик случайных величин, предоставляющих собой определённые числа, характеристики случайных функций представляют собой в общем случае не числа, а функции.

4. Математическое ожидание случайной функции, свойства.

Математическое ожидание случайной функции определяется следующим образом. Рассмотрим сечение случайной функции $X(t)$ при фиксированном t. В этом сечении мы имеем обычную случайную величину; определим ее математическое ожидание. Очевидно, в общем случае оно зависит от t, т. е. представляет собой некоторую функцию t: $m_x(t) = M[X(t)]$.

Таким образом, математическим ожиданием случайной функции $X(t)$ называется неслучайная функция $m_x(t)$, которая при каждом значении аргумента t равна математическому ожиданию соответствующего сечения случайной функции.

По смыслу математическое ожидание случайной функции есть некоторая средняя функция, около которой различным образом варьируются конкретные реализации случайной функции.



5. Дисперсия случайной функции, свойства.

Дисперсией случайной функции $X(t)$ называется неслучайная функция $D_x(t)$, значение которой для каждого t равно дисперсии соответствующего сечения случайной функции: $D_x(t) = D[X(t)]$.

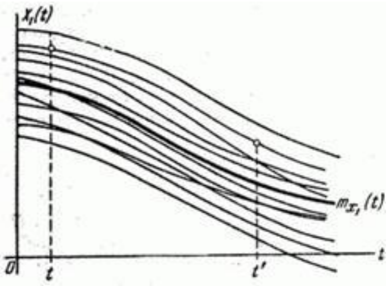
Дисперсия случайной функции при каждом t характеризует разброс возможных реализаций случайной функции относительно среднего, иными словами, «степень случайности» случайной функции.

Очевидно, $D_x(t)$ есть неотрицательная функция. Извлекая из нее квадратный корень, получим функцию $\sigma_x(t)$ - среднее квадратическое отклонение случайной функции:

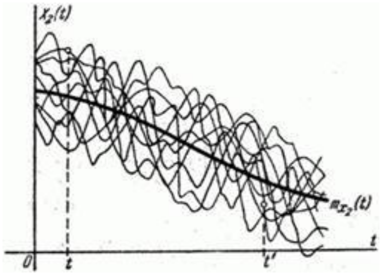
$$\sigma_x(t) = \sqrt{D_x(t)}.$$

Математическое ожидание и дисперсия представляют собой весьма важные характеристики случайной функции; однако для описания основных особенностей случайной функции этих характеристик недостаточно. Чтобы убедиться в этом, рассмотрим две случайные функции $X_1(t)$ и $X_2(t)$, наглядно изображенные семействами реализаций на рис.

У случайных функций $X_1(t)$ и $X_2(t)$ примерно одинаковые математические ожидания и дисперсии; однако характер этих случайных функций резко различен. Для случайной функции $X_1(t)$ характерно плавное, постепенное изменение.



Если, например, в точке t случайная функция $X_1(t)$ приняла значение, заметно превышающее среднее, то весьма вероятно, что и в точке t' она также примет значение больше среднего.



Для случайной функции $X_1(t)$ характерна ярко выраженная зависимость между ее значениями при различных t. Напротив, случайная функция $X_2(t)$ имеет резко колебательный характер с неправильными, беспорядочными колебаниями. Для такой случайной функции характерно быстрое затухание зависимости между ее значениями по мере увеличения расстояния по t между ними.

Очевидно, внутренняя структура обоих случайных процессов совершенно различна, но это различие не улавливается ни математическим ожиданием, ни дисперсией; для его описания необходимо вести специальную характеристику.

Эта характеристика называется корреляционной функцией (иначе - автокорреляционной функцией).

1. 11 Лекция №16 (2 ч.)

Тема: Динамическая система. Оператор динамической системы. Линейные преобразования случайной функции

1.11.1 Вопросы лекции:

1. Понятие динамической системы, примеры. Понятие оператора динамической системы, примеры.

2. Виды линейных операторов. Линейный оператор интегрирования. Линейный оператор дифференцирования.
3. Сложение случайных функций.

1.11.2 Краткое содержание вопросов:

1. Понятие динамической системы, примеры. Понятие оператора динамической системы, примеры.

Имеется некоторая динамическая система A; под «динамической системой» мы понимаем любой прибор, прицел, счетно-решающий механизм, систему автоматического управления и т. п. Эта система может быть механической, электрической или содержать любые другие элементы. Работу системы будем представлять себе следующим образом: на вход системы непрерывно поступают какие-то входные данные; система перерабатывает их и непрерывно выдает некоторый результат. Условимся называть поступающие на вход системы данные: «воздействием», а выдаваемый результат «реакцией» системы на это воздействие.

Рассмотрим самый простой случай: когда на вход системы A подается только одно воздействие, представляющее собой функцию времени $x(t)$: реакция системы на это воздействие есть другая функция времени $y(t)$. Схема работы системы A условно изображена на рис.

Будем говорить, что система A осуществляет над входным воздействием некоторое преобразование, в результате которого функция $x(t)$ преобразуется в другую функцию $y(t)$.

Преобразование A может быть любого вида и любой сложности. В наиболее простых случаях это, например, умножение на заданный множитель (усилители, множительные механизмы), дифференцирование или интегрирование (дифференцирующие или интегрирующие устройства). Однако на практике системы, осуществляющие в чистом виде такие простейшие преобразования, почти не встречаются; как правило, работа системы описывается дифференциальными уравнениями, и преобразование A сводится к решению дифференциального уравнения, связывающего воздействие $x(t)$ с реакцией $y(t)$.

При исследовании динамической системы в первую очередь решается основная задача: по заданному воздействию $x(t)$ определить реакцию системы $y(t)$. Однако для полного исследования системы и оценки ее технических качеств такой элементарный подход является недостаточным. В действительности воздействие $x(t)$ никогда не поступает на вход системы в чистом виде: оно всегда искажено некоторыми случайными ошибками (возмущениями), в результате которых на систему фактически воздействует не заданная функция $x(t)$, а случайная функция $X(t)$; соответственно этому система вырабатывает в качестве реакции случайную функцию $Y(t)$, также отличающуюся от теоретической реакции $y(t)$.

Естественно возникает вопрос: насколько велики будут случайные искажения реакции системы при наличии случайных возмущений на ее входе? И далее: как следует выбрать параметры системы для того, чтобы эти искажения были минимальными?

Из двух поставленных выше задач, естественно, более простой является первая - прямая - задача. Сформулируем ее следующим образом.

На вход динамической системы A поступает случайная функция $X(t)$; система подвергает ее известному преобразованию, в результате чего на выходе системы появляется, случайная функция:

$$Y(t) = A\{X(t)\}.$$

Известны характеристики случайной функции $X(t)$: математическое ожидание и корреляционная функция. Требуется найти аналогичные характеристики случайной функции $Y(t)$. Короче: по заданным характеристикам случайной функции на входе динамической системы найти характеристики случайной функции на выходе.

Поставленная задача может быть решена совершенно точно в одном частном, но весьма важном для практики случае: когда преобразование A принадлежит к классу так называемых линейных преобразований и соответственно система A принадлежит к классу линейных систем.

Понятие оператора динамической системы, примеры.

Понятие оператора является обобщением понятия функции. Когда мы устанавливаем функциональную связь между двумя переменными y и x и пишем:

$$y = f(x).$$

то под символом f мы понимаем правило, по которому заданному значению x приводится в соответствие вполне определенное значение y. Знак f есть символ некоторого преобразования, которому нужно подвергнуть величину x, чтобы получить y. Соответственно виду этого преобразования функции могут быть линейными и нелинейными, алгебраическими, трансцендентными и т. д.

Аналогичные понятия и соответствующая символика применяются в математике и в тех случаях, когда преобразованию подвергаются не величины, а функции. Рассмотрим некоторую функцию $x(t)$ и установим определенное правило A, согласно которому функция $x(t)$ преобразуется в другую функцию $y(t)$. Запишем это преобразование в следующем виде:

$$y(t) = A\{x(t)\}.$$

Правило A, согласно которому функция $x(t)$ преобразуется в функцию $y(t)$, мы будем называть оператором; например, мы будем говорить: оператор дифференцирования, оператор интегрирования, оператор решения дифференциального уравнения и т. д.

Если динамическая система преобразует поступающую на ее вход функцию $x(t)$ в функцию $y(t)$:

$$y(t) = A\{x(t)\},$$

то оператор A называется оператором динамической системы.

В более общем случае на вход системы поступает не одна, а несколько функций; равным образом на выходе системы могут появляться несколько функций; в этом случае оператор системы преобразует одну совокупность функций в другую. Однако в целях простоты изложения мы рассмотрим здесь лишь наиболее элементарный случай преобразования одной функции в другую.

2. Виды линейных операторов. Линейный оператор интегрирования. Линейный оператор дифференцирования.

Преобразования или операторы, применяемые к функциям, могут быть различных типов. Наиболее важным для практики является класс так называемых линейных операторов. Оператор L называется линейным однородным, если он обладает следующими свойствами:

1) к сумме функций оператор может применяться почленно:

$$L\{x_1(t)+x_2(t)\} = L\{x_1(t)\} + L\{x_2(t)\};$$

2) постоянную величину c можно выносить за знак оператора:

$$L\{cx(t)\} = cL\{x(t)\}.$$

Из второго свойства между прочим, следует, что для линейного однородного оператора справедливо свойство

$$L\{0\} = 0,$$

т. е. при нулевом входном воздействии реакция системы равна нулю.

Примеры линейных однородных операторов:

1) оператор дифференцирования:

$$y(t) = \frac{dx(t)}{dt};$$

2) оператор интегрирования:

$$y(t) = \int_0^t x(\tau)d\tau;$$

3) оператор умножения на определенную функцию $\varphi(t)$:

$$y(t) = \varphi(t)x(t),$$

и т. д.

Кроме линейных однородных операторов, существуют еще линейные неоднородные операторы. Оператор L называется линейным неоднородным, если он состоит из линейного однородного оператора с прибавлением некоторой вполне определенной функции $\varphi(t)$:

$$L\{x(t)\} = L_0\{x(t)\} + \varphi(t),$$

где L_0 - линейный однородный оператор.

Примеры линейных неоднородных операторов:

1) $y(t) = \frac{dx(t)}{dt} + \varphi(t)$,

2) $y(t) = \int_0^t x(\tau)\varphi(\tau)d\tau + \varphi_1(t)$,

3) $y(t) = \varphi_1(t)x(t) + \varphi_2(t)$.

где $\varphi(t)$, $\varphi_1(t)$, $\varphi_2(t)$ - вполне определённые функции, а $x(t)$ - преобразуемая оператором функция.

Оператор дифференцирования часто обозначают буквой p: $p = \frac{d}{dt}$, помещаемой в виде множителя перед выражением, подлежащим дифференцированию. При этом запись

$y(t) = px(t)$ равносильна записи $y(t) = \dfrac{dx(t)}{dt}$.

Двойное дифференцирование обозначается множителем p^2:

$$p^2 x(t) = \frac{d^2 x(t)}{dt^2} \text{ и т. д.}$$

Встречающиеся в технике динамические системы часто описываются линейными дифференциальными уравнениями. В этом случае, как нетрудно убедиться, оператор системы является линейным. Динамическая система, оператор которой является линейным, называется линейной динамической системой.

В противоположность линейным операторам и системам рассматриваются системы и операторы нелинейные. Примерами нелинейных операторов могут служить

$$y(t) = x^2(t), \quad y(t) = \int_0^t x^3(\tau)d\tau, \quad y(t) = \sin x(t),$$

а также решение нелинейного дифференциального уравнения, хотя бы

$$y'(t) + \alpha \cos y(t) = x(t).$$

Динамическая система, оператор которой не является линейным, называется нелинейной системой.

На практике линейные системы встречаются очень часто. В связи с линейностью этих систем к анализу их ошибок может быть с большой эффективностью применен аппарат теории случайных функций. Еще чаще, чем линейные системы, на практике встречаются системы не строго линейные, но в известных пределах допускающие линеаризацию.

Пусть на вход линейной системы с оператором L воздействует случайная функция $X(t)$, причем известны ее характеристики: математическое ожидание $m_x(t)$ и корреляционная функция $K_x(t,t')$. Реакция системы представляет собой случайную функцию $Y(t) = L\{X(t)\}$.

Требуется найти характеристики случайной функции $Y(t)$ на выходе системы: $m_y(t)$ и $K_y(t,t')$. Короче: по характеристикам случайной функции на входе линейной системы найти характеристики случайной функции на выходе.

Покажем сначала, что можно ограничиться решением этой задачи только для однородного оператора L. Действительно, пусть оператор L неоднороден и выражается формулой $L\{X(t)\} = L_0\{X(t)\} + \varphi(t)$,

где L_0 - линейный однородный оператор, $\varphi(t)$ - определенная неслучайная функция. Тогда

$$m_y(t) = M\left[L_0\{X(t)\}\right] + \varphi(t),$$

т. е. функция $\varphi(t)$ просто прибавляется к математическому ожиданию случайной функции на выходе линейной системы. Что же касается корреляционной функции, то, как известно, она не меняется от прибавления к случайной функции неслучайного слагаемого. Поэтому в дальнейшем изложении под «линейными операторами» будем разуметь только линейные однородные операторы. Решим задачу об определении характеристик на выходе линейной системы сначала для некоторых частных видов линейных операторов.

Линейный оператор интегрирования.

Дана случайная функция $X(t)$ с математическим ожиданием $m_x(t)$ и корреляционной функцией $K_x(t,t')$. Случайная функция $Y(t)$ связана с $X(t)$ линейным однородным оператором интегрирования:

$$Y(t) = \int_0^t X(\tau)d\tau .$$

Требуется найти характеристики случайной функции $Y(t)$, $m_x(t)$ и $K_x(t,t')$.
Представим интеграл как предел суммы:

$$Y(t) = \int_0^t X(\tau)d\tau = \lim_{\Delta\tau\to 0}\sum_i X(\tau_i)\Delta\tau$$

и применим к равенству операцию математического ожидания. По теореме сложения математических ожиданий имеем:

$$m_y(t) = M[Y(t)] = \lim_{\Delta\tau\to 0}\sum_i M[X(\tau_i)]\Delta\tau =$$

$$= \lim_{\Delta\tau\to 0}\sum_i m_x(\tau_i)\Delta\tau = \int_0^t m_x(\tau)d\tau .$$

Итак,

$$m_y(t) = \int_0^t m_x(\tau)d\tau ,$$

т. е. математическое ожидание интеграла от случайной функции равно интегралу от ее математического ожидания. Иными словами: операцию интегрирования и операцию математического ожидания можно менять местами. Это и естественно, так как операция интегрирования по своей природе не отличается от операции суммирования, которую, как мы раньше убедились, можно менять местами с операцией математического ожидания.

Найдём корреляционную функцию $K_y(t,t')$. Для этого перейдём к центрированным случайным функциям:

$$\overset{\circ}{X}(t) = X(t) - m_x(t), \quad \overset{\circ}{Y}(t) = Y(t) - m_y(t) .$$

Нетрудно убедиться, что

$$\overset{\circ}{Y}(t) = \int_0^t \overset{\circ}{X}(\tau)d\tau .$$

По определению корреляционной функции,

$$K_y(t,t') = M[\overset{\circ}{Y}(t)\overset{\circ}{Y}(t')] ,$$

где

$$\overset{\circ}{Y}(t) = \int_0^t \overset{\circ}{X}(\tau)d\tau ; \quad \overset{\circ}{Y}(t') = \int_0^{t'} \overset{\circ}{X}(\tau')d\tau' .$$

Нетрудно убедиться, что произведение двух интегралов в правой части формулы равно двойному интегралу.
Следовательно,

$$\overset{\circ}{Y}(t)\overset{\circ}{Y}(t') = \int_0^t\int_0^{t'} \overset{\circ}{X}(\tau)\overset{\circ}{X}(\tau')d\tau d\tau'.$$

Применяя к равенству операцию математического оживления и меняя ее в правой части местами с операцией интегрирования, получим:

$$K_y(t,t') = M[\overset{\circ}{Y}(t)\overset{\circ}{Y}(t')] = \int_0^t\int_0^{t'} M[\overset{\circ}{X}(\tau)\overset{\circ}{X}(\tau')]d\tau d\tau',$$

или окончательно:

$$K_y(t,t') = \int_0^t\int_0^{t'} K_x(\tau,\tau')d\tau d\tau'.$$

Таким образом, для того чтобы найти корреляционную функцию интеграла от случайной функции, нужно дважды проинтегрировать корреляционную функцию исходной случайной функции: сначала по одному аргументу, затем - по-другому.

Линейный оператор дифференцирования.

Дана случайная функция $X(t)$ с математическим ожиданием $m_x(t)$ и корреляционной функцией $K_x(t,t')$. Случайная функция $Y(t)$ связана со случайной функцией $X(t)$ линейным однородным оператором дифференцирования:

$$Y(t) = \frac{dX(t)}{dt}.$$

Требуется найти $m_y(t)$ и $K_y(t,t')$.

Представим производную в виде предела: $Y(t) = \lim\limits_{\Delta t\to 0}\dfrac{X(t+\Delta t)-X(t)}{\Delta t}$.

Применяя к равенству операцию математического ожидания, получим:

$$m_y(t) = M[Y(t)] = \lim_{\Delta t\to 0}\frac{m_x(t+\Delta t)-m_x(t)}{\Delta t} = \frac{dm_x(t)}{dt}.$$

Итак,

$$m_y(t) = \frac{dm_x(t)}{dt},$$

т. е. математическое ожидание производной от случайной функции равно производной от ее математического ожидания.

Следовательно, операцию дифференцирования, как и операцию интегрирования тоже можно менять местами с операцией математического ожидания.

Для определения $K_y(t,t')$ перейдём к центрированным случайным функциям $\overset{\circ}{Y}(t)$ и $\overset{\circ}{X}(t)$; очевидно: $\overset{\circ}{Y}(t) = \dfrac{d\overset{\circ}{X}(t)}{dt}$.

По определению

$$K_y(t,t') = M[\overset{\circ}{Y}(t)\overset{\circ}{Y}(t')].$$

Подставим вместо $\overset{\circ}{Y}(t)$ и $\overset{\circ}{Y}(t')$ их выражения:

$$K_y(t,t') = M\left[\frac{d\overset{\circ}{X}(t)}{dt}\frac{d\overset{\circ}{X}(t')}{dt'}\right].$$

Мы доказали, что математическое ожидание производной случайной функции равно производной от математического ожидания, т. е. знаки дифференцирования и математического ожидания можно менять местами. Следовательно,

$$K_y(t,t') = M\left[\frac{\partial^2 \overset{\circ}{X}(t)\overset{\circ}{X}(t')}{\partial t\partial t'}\right] = \frac{\partial^2}{\partial t\partial t'}M[\overset{\circ}{X}(t)\overset{\circ}{X}(t')] = \frac{\partial^2}{\partial t\partial t'}K_x(t,t').$$

Таким образом,

$$K_y(t,t') = \frac{\partial^2 K_x(t,t')}{\partial t\partial t'}.$$

Итак, чтобы найти корреляционную функцию производной, нужно дважды продифференцировать корреляционную функцию исходной случайной функции: сначала по одному аргументу, затем - по другому.

Можно доказать, что такое правило является общим для всех линейных однородных операторов: **если случайная функция $X(t)$ с математическим ожиданием $m_x(t)$ и корреляционной функцией $K_x(t,t')$ преобразуется линейным однородным оператором L в случайную функцию,**

$$Y(t) = L\{X(t)\},$$

то для нахождения математического ожидания случайной функции $Y(t)$ нужно применить тот же оператор к математическому ожиданию случайной функции $X(t)$:

$$m_y(t) = L\{m_x(t)\},$$

а для нахождения корреляционной функции нужно дважды применить тот же оператор к корреляционной функции случайной функции $X(t)$, сначала по одному аргументу, затем - по другому:

$$K_y(t,t') = L^{(t)}L^{(t')}\{K_y(t,t')\}.$$

Во многих задачах практики нас, в конечном счёте, интересует не корреляционная функция $K_y(t,t')$ на выходе линейной системы, а дисперсия $D_y(t)$ характеризующая точность работы системы в условиях наличия случайных возмущений.

Дисперсию $D_y(t)$ можно найти, зная корреляционную функцию:

$$D_y(t) = K_y(t,t).$$

При этом нужно подчеркнуть, что, как правило, для определения дисперсии на выходе линейной системы недостаточно знать дисперсию на ее входе, а существенно важно знать корреляционную функцию.

Действительно, линейная система может совершенно по-разному реагировать на случайные возмущения, поступающие на ее вход, в зависимости от того, какова внутренняя структура этих случайных возмущений; состоят ли они, например, по преимуществу из высокочастотных или низкочастотных колебаний.

Внутренняя же структура случайного процесса описывается не его дисперсией, а корреляционной функцией.

3. Сложение случайных функций.

Во многих задачах практики мы встречаемся с тем, что на вход динамической системы поступает не одна случайная функция $X(t)$, а две или более случайные функции, каждая из которых связана с действием отдельного возмущающего фактора. Возникает задача сложения случайных функций, точнее - задача определения характеристик суммы по характеристикам слагаемых.

Эта задача решается элементарно просто, если две складываемые случайные функции независимы (точнее, некоррелированные) между собой. В общем же случае для ее решения необходимо знание еще одной характеристики - так называемой взаимной корреляционной функции (иначе - корреляционной функции связи).

Взаимной корреляционной функцией двух случайных функций $X(t)$ и $Y(t)$ называется неслучайная функция двух аргументов t и t', которая при каждой паре значений t,t' равна корреляционному моменту соответствующих сечений случайной функции $X(t)$ и случайной функции $Y(t)$:

$$R_{xt}(t,t') = M[\mathring{X}(t)\mathring{Y}(t')].$$

Взаимная корреляционная функция, так же как и обычная корреляционная функция, не изменяется при прибавлении к случайным функциям любых неслучайных слагаемых, а следовательно, и при центрировании случайных функций. Из определения взаимной корреляционной функции вытекает следующее ее свойство:

$$R_{xy}(t,t') = R_{xy}(t',t).$$

Вместо функции $R_{xy}(t,t')$ часто пользуются нормированной взаимной корреляционной функцией:

$$r_{xy}(t,t') = \frac{R_{xy}(t,t')}{\sigma_x(t)\sigma_y(t')}.$$

Если взаимная корреляционная функция равна нулю при всех значениях t,t', то случайные функции $X(t)$ и $Y(t)$ называются некоррелированными (несвязанными).
Зная математические ожидания и корреляционные функции двух случайных функций $X(t)$ и $Y(t)$, а также их взаимную корреляционную функцию, можно найти характеристики суммы этих двух случайных функций:

$$Z(t) = X(t) + Y(t).$$

По теореме сложения математических ожиданий: $m_z(t) = m_x(t) + m_y(t)$,
т. е. при сложении двух случайных функций их математические ожидания складываются.
Для определения корреляционной функции $K_x(t,t')$ перейдём к центрированным случайным функциям $\mathring{Z}(t)$, $\mathring{X}(t)$, $\mathring{Y}(t)$. Очевидно,

$$\mathring{Z}(t) = \mathring{X}(t) + \mathring{Y}(t).$$

По определению корреляционной функции

$$K_z(t,t') = M[\mathring{Z}(t)\mathring{Z}(t')] = M[(\mathring{X}(t)+\mathring{Y}(t))(\mathring{X}(t')+\mathring{Y}(t'))] =$$

$$= M[\mathring{X}(t)\mathring{X}(t')] + M[\mathring{Y}(t)\mathring{Y}(t')] + M[\mathring{X}(t)\mathring{Y}(t')] + M[\mathring{X}(t')\mathring{Y}(t)],$$

или

$$K_z(t,t') = K_x(t,t') + K_y(t,t') + R_{xy}(t,t') + R_{xy}(t',t).$$

В случае, когда случайные функции $X(t)$ и $Y(t)$ некоррелированные, $R_{xy}(t,t')\equiv 0$, и формула принимает вид:

$$K_z(t,t') = K_x(t,t') + K_y(t,t'),$$

т. е. при сложении некоррелированных случайных функций их корреляционные функции складываются. Выведенные формулы могут быть обобщены на случай произвольного числа слагаемых.

1.12 Лекция № 17 (2 ч.)

Тема: Стационарный случайный процесс. Стационарный случайный процесс с эргодическим свойством

1.12.1 Вопросы лекции:

1. Понятие случайного процесса, примеры, виды.
2. Стационарный случайный процесс: особенности и характеристики.
3. Преобразование стационарного случайного процесса стационарной линейной системой.
4. Стационарный процесс с эргодическим свойством и без него, примеры.
5. Характеристики стационарного процесса с эргодическим свойством.

1.12. 2. Краткое содержание вопросов:

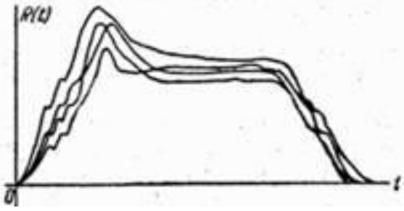
1. Понятие случайного процесса, примеры, виды.

На практике очень часто встречаются случайные процессы, протекающие во времени приблизительно однородно и имеющие вид непрерывных случайных колебаний вокруг некоторого среднего значения, причём ни средняя амплитуда, ни характер этих колебаний не обнаруживают существенных изменений с течением времени. Такие случайные процессы называются стационарными. В качестве примеров стационарных случайных процессов можно привести: 1) колебания самолёта на установившемся режиме горизонтального полёта; 2) колебания напряжения в электрической осветительной сети; 3) случайные шумы в радиоприёмнике; 4) процесс качки корабля и т. п.

Каждый стационарный процесс можно рассматривать как продолжающийся во времени неопределённо долго; при исследовании стационарного процесса в качестве начала отсчёта можно выбрать любой момент времени. Исследуя стационарный процесс на любом участке времени, мы должны получить одни и те же его характеристики. Образно выражаясь, стационарный процесс «не имеет ни начала, ни конца».

В противоположность стационарным случайным процессам можно указать другие, явно нестационарные случайные процессы, например: колебания самолета в режиме пикирования; процесс затухающих колебаний в электрической цепи; процесс горения порохового заряда в реактивной камере. Нестационарный процесс характерен тем, что он имеет определённую тенденцию развития во времени; характеристики такого процесса зависят от начала отсчёта, зависят от времени.

На рис. изображено семейство реализаций явно нестационарного случайного процесса - процесса изменения тяги двигателя реактивного снаряда во времени.



Заметим, что далеко не все нестационарные случайные процессы являются существенно нестационарными на всем протяжении своего развития. Существуют нестационарные процессы, которые (на известных отрезках времени и с известным приближением) могут быть приняты за стационарные.

Вообще, как правило, случайный процесс в любой динамической системе начинается с нестационарной стадии - с так называемого «переходного процесса». После затухания переходного процесса система обычно переходит на установившийся режим, и тогда случайные процессы, протекающие в ней, могут считаться стационарными.

2. Стационарный случайный процесс: особенности и характеристики. Преобразование стационарного случайного процесса стационарной линейной системой.

Стационарные случайные процессы очень часто встречаются в физических и технических задачах. По своей природе эти процессы проще, чем нестационарные, и описываются более простыми характеристиками. Линейные преобразования стационарных случайных процессов также обычно осуществляются проще, чем нестационарных.

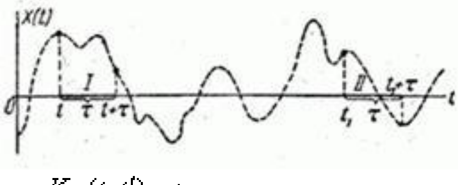
Так как изменение стационарной случайной функции должно протекать однородно по времени, то естественно потребовать, чтобы для стационарной случайной функции математическое ожидание было постоянным:

$$m_x(t) = m_x = \text{const}.$$

Заметим, однако, что это требование не является существенным: мы знаем, что от случайной функции $X(t)$ всегда можно перейти к центрированной случайной функции $\overset{\circ}{X}(t)$, для которой математическое ожидание тождественно равно нулю и, следовательно, удовлетворяет условию. Таким образом, если случайный процесс нестационарен только за счёт переменного математического ожидания, это не мешает нам изучать его как стационарный процесс. Второе условие, которому, очевидно, должна удовлетворять стационарная случайная функция, - это условие постоянства дисперсии:

$$D_x(t) = D_x = \text{const}.$$

Установим, какому условию должна удовлетворять корреляционная функция стационарной случайной функции. Рассмотрим случайную функцию $X(t)$



Положим в выражении $K_x(t,t')$ $t' = t + \tau$ и рассмотрим $K_x(t,t+\tau)$ - корреляционный момент двух сечений случайной функции, разделённых интервалом времени τ. Очевидно, если случайный процесс $X(t)$ действительно стационарен, то этот корреляционный момент не должен зависеть от того, где именно на оси Ot мы взяли участок τ, а должен зависеть только от длины этого участка. Вообще, корреляционная функция стационарного случайного процесса должна зависеть не от положения t первого аргумента на оси абсцисс, а только от промежутка τ между первым и вторым аргументами:

$$K_x(t,t+\tau) = k_x(\tau).$$

Следовательно, корреляционная функция стационарного случайного процесса есть функция не двух, а всего одного аргумента. Это обстоятельство в ряде случаев сильно упрощает операции над стационарными случайными функциями.

Заметим, что условие, требующее от стационарной случайной функции постоянства дисперсии, является частным случаем.

$$D_x(t) = K_x(t,t) = k_x(0) = \text{const}.$$

Таким образом, условие есть единственное существенное условие, которому должна удовлетворять стационарная случайная функция. Поэтому в дальнейшем мы под стационарной случайной функцией будем понимать такую случайную функцию, корреляционная функция которой зависит не от обоих своих аргументов t и t', а только от разности τ между ними. Чтобы не накладывать специальных условий на математическое ожидание, мы будем рассматривать только центрированные случайные функции.

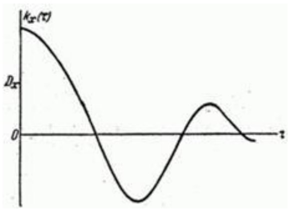
Мы знаем, что корреляционная функция любой случайной функции обладает свойством симметрии:

$$K_x(t,t') = K_x(t',t).$$

Отсюда для стационарного процесса, полагая $t' - t = \tau$,

имеем: $k_x(\tau) = k_x(-\tau)$,

т. е. корреляционная функция $k_x(\tau)$ есть чётная функция своего аргумента. Поэтому обычно корреляционную функцию определяют только для положительных значений аргумента.



На практике, вместо корреляционной функции $k_x(\tau)$, часто пользуются нормированной корреляционной функцией

$$\rho_x(\tau) = \frac{k_x(\tau)}{D_x},$$

где $D_x = k_x(0)$ - постоянная дисперсия стационарного процесса.

4. Стационарный процесс с эргодическим свойством и без него, примеры.

Рассмотрим некоторую стационарную случайную функцию $X(t)$ и предположим, что требуется оценить ее характеристики: математическое ожидание m_x и корреляционную функцию $k_x(\tau)$.

Для этого нужно располагать известным числом реализаций случайной функции $X(t)$. Обрабатывая эти реализации, можно найти оценки для математического ожидания $\tilde{m}_x(t)$ и корреляционной функции $\tilde{K}_x(t,t')$. В связи с ограниченностью числа наблюдений функция $\tilde{m}_x(t)$ не будет строго постоянной; ее придется усреднить и заменить некоторым постоянным $\tilde{m}_x$; аналогично, усредняя значения $\tilde{K}_x(t,t')$ для разных $\tau = t' - t$, получим корреляционную функцию $\tilde{k}_x(\tau)$.

Этот метод обработки, очевидно, является довольно сложным и громоздким. Естественно возникает вопрос: нельзя ли для стационарной случайной функции этот сложный, двухступенчатый процесс обработки заменить более простым, который заранее базируется на предположении, что математическое ожидание не зависит от времени, а корреляционная функция - от начала отсчета?

Кроме того, возникает вопрос: при обработке наблюдений над стационарной случайной функцией является ли существенно необходимым располагать несколькими реализациями? Поскольку случайный процесс является стационарным и протекает однородно по времени, естественно предположить, что одна-единственная реализация достаточной продолжительности может служить достаточным опытным материалом для получения характеристик случайной функции. При более подробном рассмотрении этого вопроса оказывается, что такая возможность существует не для всех случайных процессов: не всегда одна реализация достаточной продолжительности оказывается эквивалентной множеству отдельных реализаций.

Для примера рассмотрим две стационарные случайные функции $X_1(t)$ и $X_2(t)$, представленные совокупностью своих реализаций на рис. 1 и 2.

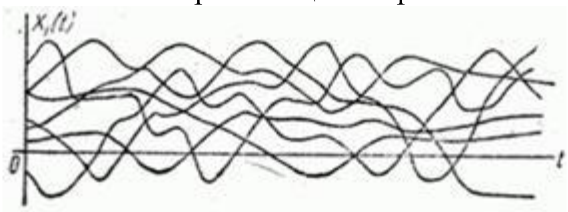


Рис. 1.

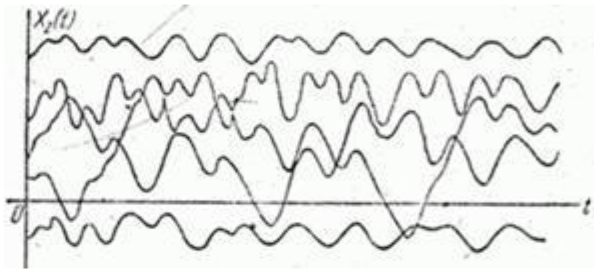


Рис.2.

Для случайной функции $X_1(t)$ характерна следующая особенность: каждая из ее реализаций обладает одними и теми же характерными признаками: средним значением, вокруг которого происходят колебания, и средним размахом этих колебаний. Выберем произвольно одну из таких реализаций и продолжим мысленно опыт, в результате которого она получена, на некоторый участок времени T.

Очевидно, что при достаточно большом T, эта реализация сможет дать нам хорошее представление о свойствах случайной функции в целом.

В частности, усредняя значения этой реализации вдоль оси абсцисс - по времени, мы должны получить приближенное значение математического ожидания случайной функции; усредняя квадраты отклонений от этого среднего, мы должны получить приближенное значение дисперсии, и т. д.

Про такую случайную функцию говорят, что она обладает эргодическим свойством. Эргодическое свойство состоит в том, что каждая отдельная реализация случайной функции является как бы «полномочным представителем» всей совокупности возможных реализаций; одна реализация достаточной продолжительности может заменить при обработке множество реализаций той же общей продолжительности.

Рассмотрим теперь случайную функцию $X_2(t)$. Выберем произвольно одну из ее реализаций, продолжим ее мысленно на достаточно большой участок времени и вычислим ее среднее значение по времени на всем участке наблюдения. Очевидно, это среднее значение для каждой реализации будет свое и может существенно отличаться от математического ожидания случайной функции, построенного как среднее из множества реализаций. Про такую случайную функцию говорят, что она не обладает эргодическим свойством.

Если случайная функция $X(t)$ обладает эргодическим свойством, то для неё среднее по времени (на достаточно большом участке наблюдения) приближённо равно среднему по множеству наблюдений.

То же будет верно и для $X^2(t)$, $X(t)\cdot X(t+\tau)$ и т. д. Следовательно, все характеристики случайной функции (математическое ожидание, дисперсию, корреляционную функцию) можно будет приближённо определять по одной достаточно длинной реализации.

Какие же стационарные случайные функции обладают, а какие не обладают эргодическим свойством?

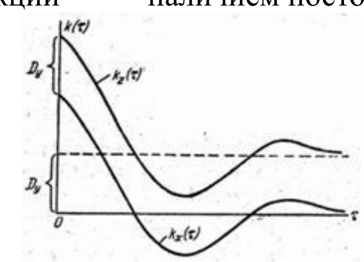
5. Характеристики стационарного процесса с эргодическим свойством.

Пусть для случайного процесса характерно то, что он как бы «разложим» на более элементарные случайные процессы; каждый из них осуществляется с некоторой вероятностью и имеет свои индивидуальные характеристики. Таким образом, разложимость, внутренняя неоднородность случайного процесса, протекающего с некоторой вероятностью по тому или другому типу, есть физическая причина неэргодичности этого процесса.

В частности, неэргодичность случайного процесса может быть связана с наличием в его составе слагаемого в виде обычной случайной величины (т. е. наличие в спектре случайного процесса, помимо непрерывной части, конечной дисперсии при частоте 0). Действительно, рассмотрим случайную функцию $Z(t)=X(t)+Y$,
где $X(t)$ - эргодическая стационарная случайная функция с характеристиками m_x, $k_x(\tau)$; Y - случайная величина с характеристиками m_y и D_y; предположим к тому же, что $X(t)$ и Y некоррелированные. Определим характеристики случайной функции $Z(t)$. Согласно общим правилам сложения случайных функций имеем: $m_z = m_x + m_y$, $k_z(\tau)=k_x(\tau)+D_y$.

Из формул видно, что случайная функция $Z(t)$ является стационарной. Но обладает ли она эргодическим свойством? Очевидно, нет. Каждая ее реализация будет по характеру отличаться от других, будет обладать тем или иным средним по времени значением в зависимости от того, какое значение приняла случайная величина Y.

Об эргодичности или неэргодичности случайного процесса может непосредственно свидетельствовать вид его корреляционной функции. Действительно, рассмотрим корреляционную функцию неэргодической случайной функции. Она отличается от корреляционной функции случайной функции $X(t)$ наличием постоянного слагаемого D_y.



В то время как корреляционная функция $k_x(\tau)$ стремится к нулю при $\tau\to\infty$ (корреляционная связь между значениями случайной функции неограниченно убывает по мере увеличения расстояния между ними), функция $k_z(\tau)$ уже не стремится к нулю при $t\to\infty$, а приближается к постоянному значению D_y.

На практике мы не имеем возможности исследовать случайный процесс и его корреляционную функцию на бесконечном участке времени; участок значений τ, с которым мы имеем дело, всегда ограничен.

Пусть корреляционная функция стационарного случайного процесса при увеличении τ не убывает, а, начиная с некоторого τ, остается приблизительно постоянной. Это обычно есть признак того, что в составе случайной функции имеется слагаемое в виде обычной случайной величины и что процесс не является эргодическим. Стремление же корреляционной функции к нулю при $t \to \infty$ говорит в пользу эргодичности процесса. Во всяком случае, оно достаточно для того, чтобы математическое ожидание функции можно было определять как среднее по времени.

1.13.Лекция 18 (Л-18)

Тема: Определение характеристик эргодических стационарных случайных функций из опыта.

1.13.1 Вопросы лекции:

1. Оценки характеристик стационарной эргодической функции по опытным данным: виды, качество.
2. Методика получения оценок, примеры.
3. Сравнительный анализ результатов оценивания

1.13. 2. Краткое содержание вопросов:

1. Оценки характеристик стационарной эргодической функции по опытным данным: виды, качество.

Рассмотрим стационарную случайную функцию $X(t)$, обладающую эргодическим свойством, и предположим, что в нашем распоряжении имеется всего одна реализация этой случайной функции, но зато на достаточно большом участке времени T. Для эргодической стационарной случайной функции одна реализация достаточно большой продолжительности практически эквивалентна (в смысле объема сведений о случайной функции) множеству реализаций той же общей продолжительности; характеристики случайной функции могут быть приближенно определены не как средние по множеству наблюдений, а как средние по времени t. В частности, при достаточно большом T математическое ожидание m_x может быть приближённо вычислено по формуле

$$m_x \approx \frac{1}{T}\int_0^T x(t)dt .$$

Аналогично может быть приближённо найдена корреляционная функция $k_x(\tau)$ при любом τ. Действительно, корреляционная функция, по определению, представляет собой не что иное, как математическое ожидание случайной функции $\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)$:

$$k_x(\tau) = M[\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)]$$

Это математическое ожидание также, очевидно, может быть приближённо вычислено как среднее по времени. Фиксируем некоторое значение τ и вычислим указанным способом корреляционную функцию $k_x(\tau)$. Для этого удобно предварительно «центрировать» данную реализацию $x(t)$: $\overset{\circ}{x}(t) = x(t) - m_x$.

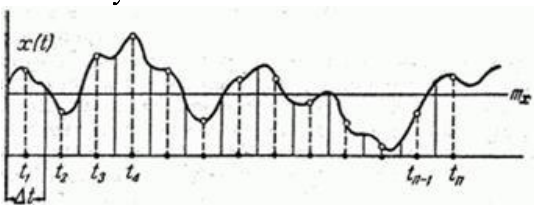
2. Методика получения оценок, примеры.

Вычислим при заданном τ математическое ожидание случайной функции $\mathring{X}(t)\,\mathring{X}(t+\tau)$ как среднее по времени. При этом, очевидно, нам придётся учитывать не весь участок времени от 0 до T, а несколько меньший, так как второй сомножитель $\mathring{X}(t+\tau)$ известен нам не для всех t, а только для тех, для которых $t+\tau \le T$.
Вычисляя среднее по времени указанным выше способом, получим:

$$k_x(\tau) \approx \frac{1}{T-\tau}\int_0^{T-\tau} \mathring{x}(t)\,\mathring{x}(t+\tau)dt$$
.

Вычислив интеграл для ряда значений τ, можно приближённо воспроизвести по точкам весь ход корреляционной функции.

На практике обычно интегралы заменяют конечными суммами. Покажем, как это делается. Разобьём интервал записи случайной функции на n равных частей длиной Δt и обозначим середины полученных участков $t_1, t_2, \ldots, t_n$.



.

Предоставим интеграл как сумму интегралов по элементарным участкам Δt и на каждом из них вынесем функцию $x(t)$ из-под знака интеграла средним значением, соответствующим центру интервала $x(t_i)$. Получим приближённо:

$$m_x = \frac{1}{T}\frac{T}{n}\sum_{i=1}^{n} x(t_i) \quad ,\text{или} \quad m_x = \frac{1}{n}\sum_{i=1}^{n} x(t_i)$$
.

Аналогично можно вычислить корреляционную функцию для значений τ, равных $0, \Delta t, 2\Delta t, \ldots$.

Придадим, например, величине τ значение $\tau = m\Delta t = \dfrac{mT}{n}$

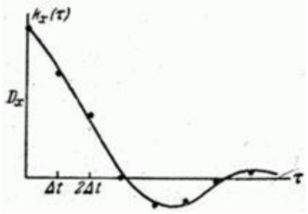
вычислим интеграл, деля интервал интегрирования $T-\tau = T - \dfrac{mT}{n} = \dfrac{n-m}{n}T$

на $n-m$ равных участков длиной Δt и вынося на каждом из них функцию $\mathring{x}(t)\,\mathring{x}(t+\tau)$ за знак интеграла средним значением. Получим:

$$k_x\left(\frac{mT}{n}\right) = \frac{1}{n-m}\sum_{i=1}^{n-m} \mathring{x}(t_i)\,\mathring{x}(t_{i+m})$$
.

Вычисление корреляционной функции производят для $m = 0,1,2,\ldots$ последовательно. Вплоть до таких значений m, при которых корреляционная функция становится практически равной нулю или начинает совершать небольшие нерегулярные колебания около нуля.

Общий ход функции $k_x(\tau)$ воспроизводится по отдельным точкам.



3. Сравнительный анализ результатов оценивания.

Для того чтобы математическое ожидание m_x и корреляционная функция $k_x(\tau)$ были определены с удовлетворительной точностью, нужно, чтобы число точек n было достаточно велико (порядка сотни, а в некоторых случаях даже нескольких сотен).

Выбор длины элементарного участка Δt определяется характером изменения случайной функции. Если случайная функция изменяется сравнительно плавно, участок Δt можно выбирать большим, чем когда она совершает резкие и частые колебания. Чем более высокочастотный состав имеют колебания, образующие случайную функцию, тем чаще нужно располагать опорные точки при обработке.

Ориентировочно можно рекомендовать выбирать элементарный участок Δt так, чтобы на полный период самой высокочастотной гармоники в составе случайной функции приходилось порядка 5-10 опорных точек.

Часто выбор опорных точек вообще не зависит от обрабатывающего, а диктуется темпом работы записывающей аппаратуры.

В этом случае следует вести обработку непосредственно полученного из опыта материала, не пытаясь вставить между наблюдёнными значениями промежуточные, так как это не может повысить точности результата, а излишне осложнит обработку.

1.14 Лекция 19 (Л-19)

Тема: Спектральное разложение стационарной случайной функции.

1.14.1 Вопросы лекции:

1. Характеристики внутренней структуры колебательных процессов.
2. Спектральное разложение стационарной случайной функции на конечном участке времени. Спектр дисперсий.
3. Необходимость перехода от дискретного к непрерывному спектру.
4. Спектральное разложение стационарной случайной функции на бесконечном участке времени. Спектральная плотность, ее свойства.

1.14.2. Краткое содержание вопросов:

1. Характеристики внутренней структуры колебательных процессов.

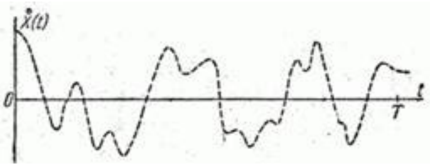
На различных примерах, приведённых ранее, мы наглядно убедились в том, что существует связь между характером корреляционной функции и внутренней структурой соответствующего ей случайного процесса. В зависимости от того, какие частоты и в каких соотношениях преобладают в составе случайной функции, ее корреляционная функция имеет тот или другой вид. Из таких соображений мы непосредственно приходим к понятию о спектральном составе случайной функции.

Понятие «спектра» встречается не только в теории случайных функций; оно широко применяется в математике, физике и технике.

Если какой-либо колебательный процесс представляется в виде суммы гармонических колебаний различных частот (так называемых «гармоник»), то спектром колебательного процесса называется функция, описывающая распределение амплитуд по различным частотам. Спектр показывает, какого рода колебания преобладают в данном процессе, какова его внутренняя структура.

Совершенно аналогичное спектральное описание можно дать и стационарному случайному процессу; вся разница в том, что для случайного процесса амплитуды колебаний будут случайными величинами. Спектр стационарной случайной функции будет описывать распределение дисперсий по различным частотам. Подойдём к понятию о спектре стационарной случайной функции из следующих соображений.

Рассмотрим стационарную случайную функцию $\overset{\circ}{X}(t)$, которую мы наблюдаем на интервале $(0,T)$.



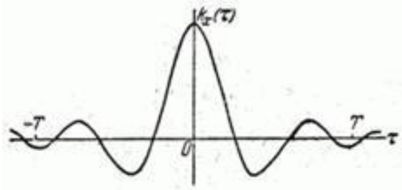
Задана корреляционная функция случайной функции $X(t)$

$$K_x(t,t+\tau)=k_x(\tau).$$

Функция $k_x(\tau)$ есть чётная функция:

$$k_x(\tau)=k_x(-\tau)$$

и, следовательно, на графике изобразится симметричной кривой .



При изменении t и t' от 0 до T аргумент $\tau=t'-t$ изменяется от $-T$ до $+T$.

Мы знаем, что чётную функцию на интервале $(-T,T)$ можно разложить в ряд Фурье, пользуясь только чётными (косинусными) гармониками:

$$k_x(\tau)=\sum_{k=0}^{\infty}D_k\cos\omega_k\tau,$$

где

$$\omega_k=k\omega_1;\quad \omega_1=\frac{2\pi}{2T}=\frac{\pi}{T},$$

а коэффициенты D_k определяются формулами:

$$\left.\begin{aligned}
D_0&=\frac{1}{2T}\int_{-T}^{T}k_x(\tau)d\tau,\\
D_k&=\frac{1}{T}\int_{-T}^{T}k_x(\tau)\cos\omega_k\tau d\tau,\quad \text{при } k\neq 0.
\end{aligned}\right\}$$

Имея в виду, что функции $k_x(\tau)$ и $\cos\omega_k\tau$ чётные, можно преобразовать формулы к виду:

$$\left.\begin{aligned} D_0 &= \frac{1}{T}\int_0^T k_x(\tau)d\tau, \\ D_k &= \frac{2}{T}\int_0^T k_x(\tau)\cos\omega_k\tau d\tau, \qquad \text{при } k \neq 0. \end{aligned}\right\}$$

2. Спектральное разложение стационарной случайной функции на конечном участке времени. Спектр дисперсий.

Перейдём в выражении корреляционной функции $k_x(\tau)$ от аргумента τ снова к двум аргументам t и t'. Для этого положим

$$\cos\omega_k\tau = \cos\omega_k(t'-t) = \cos\omega_k t'\cos\omega_k t + \sin\omega_k t'\sin\omega_k t$$

и подставим это выражение в формулу:

$$k_x(t,t') = \sum_{k=0}^{\infty}(D_k\cos\omega_k t'\cos\omega_k t + D_k\sin\omega_k t'\sin\omega_k t).$$

Мы видим, что выражение есть не что иное, как каноническое разложение корреляционной функции $K_x(t,t')$. Координатными функциями этого канонического разложения являются попеременно косинусы и синусы частот, кратных ω_1:

$$\cos\omega_k t, \sin\omega_k t \qquad (k = 0,1,\ldots).$$

Мы знаем, что по каноническому разложению корреляционной функции можно построить каноническое разложение самой случайной функции с теми же координатными функциями и с дисперсиями, равными коэффициентам D_k в каноническом разложении корреляционной функции.

Следовательно, случайная функция $\overset{\circ}{X}(t)$ может быть представлена в виде канонического разложения:

$$\overset{\circ}{X}(t) = \sum_{k=0}^{\infty}(U_k\cos\omega_k t + V_k\sin\omega_k t),$$

где U_k, V_k - некоррелированные случайные величины с математическими ожиданиями, равными нулю, и дисперсиями, одинаковыми для каждой пары случайных величин с одним и тем же индексом k:

$$D[U_k] = D[V_k] = D_k.$$

Таким образом, мы получили на интервале $(0,T)$ каноническое разложение случайной функции $\overset{\circ}{X}(t)$, координатными функциями которого являются функции $\cos\omega_k t$, $\sin\omega_k t$ при различных ω_k. Разложение такого рода называется спектральным разложением стационарной случайной функции. Спектральное разложение изображает стационарную случайную функцию разложенной на гармонические колебания различных частот:

$$\omega_1, \omega_2, \ldots, \omega_k, \ldots,$$

причём амплитуды этих колебаний являются случайными величинами.

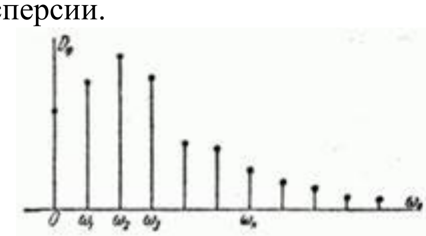
Спектр дисперсий.

Определим дисперсию случайной функции $\overset{\circ}{X}(t)$, заданной спектральным разложением. По теореме о дисперсии линейной функции некоррелированных случайных величин

$$D_x = D[\overset{\circ}{X}(t)] = \sum_{k=0}^{\infty} (\cos^2 \omega_k t + \sin^2 \omega_k t) D_k = \sum_{k=0}^{\infty} D_k .$$

Таким образом, дисперсия стационарной случайной функции равна сумме дисперсий всех гармоник ее спектрального разложения. Формула показывает, что дисперсия функции $\overset{\circ}{X}(t)$ известным образом распределена по различным частотам: одним частотам соответствуют большие дисперсии, другим - меньшие.

Распределение дисперсий по частотам можно проиллюстрировать графически в виде так называемого спектра стационарной случайной функции (точнее - спектра дисперсий). Для этого по оси абсцисс откладываются частоты $\omega_0 = 0, \omega_1, \omega_2, \ldots, \omega_k, \ldots$, а по оси ординат - соответствующие дисперсии.



Очевидно, сумма всех ординат построенного таким образом спектра равна дисперсии случайной функции.

3. Необходимость перехода от дискретного к непрерывному спектру.

Строя спектральное разложение стационарной случайной функции $\overset{\circ}{X}(t)$ на конечном участке времени $(0,T)$, мы получили спектр дисперсий случайной функции в виде ряда отдельных дискретных линий, разделённых равными промежутками (так называемый «прерывистый» или «линейчатый» спектр).

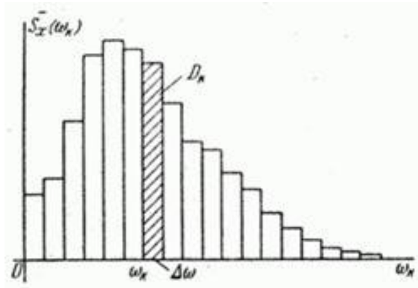
Очевидно, чем больший участок времени мы будем рассматривать, тем полнее будут наши сведения о случайной функции. Естественно поэтому в спектральном разложении попытаться перейти к пределу при $T \to \infty$ и посмотреть, во что при этом обратится спектр случайной функции.

При $T \to \infty$ $\omega_1 = \frac{2\pi}{2T} \to 0$; поэтому расстояния между частотами ω_k, на которых строится спектр, будут при $T \to \infty$ неограниченно уменьшаться. При этом дискретный спектр будет приближаться к непрерывному, в котором каждому сколь угодно малому интервалу частот $\Delta\omega$ будет соответствовать элементарная дисперсия $\Delta D(\omega)$.

Попробуем изобразить непрерывный спектр графически. Для этого мы должны несколько перестроить график дискретного спектра при конечном T. А именно, будем откладывать по оси ординат уже не самую дисперсию D_k (которая безгранично уменьшается при $T \to \infty$), а среднею плотность дисперсии, т. е. дисперсию, приходящуюся на единицу длины данного интервала частот. Обозначим расстояние между соседними частотами $\Delta\omega$:

$$\omega_1 = \frac{2\pi}{2T} = \Delta\omega$$

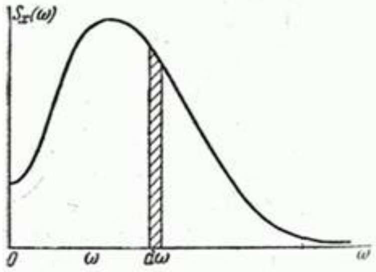
и на каждом отрезке $\Delta\omega$, как на основании, построим прямоугольник с площадью D_k. Получим ступенчатую диаграмму, напоминающую по принципу построения гистограмму статистического распределения.



Высота диаграммы на участке $\Delta\omega$, прилежащем к точке ω_k, равна $S_x(\omega_k) = \dfrac{D_k}{\Delta\omega}$
и представляет собой среднюю плотность дисперсии на этом участке. Суммарная площадь всей диаграммы, очевидно, равна дисперсии случайной функции.

4. Спектральное разложение стационарной случайной функции на бесконечном участке времени. Спектральная плотность, ее свойства.

Будем неограниченно увеличивать интервал T. При этом $\Delta\omega \to 0$, и ступенчатая кривая будет неограниченно приближаться к плавной кривой $S_x(\omega)$. Эта кривая изображает плотность распределения дисперсии по частотам непрерывного спектра, а сама функция $S_x(\omega)$ называется спектральной плотностью дисперсии, или, короче, спектральной плотностью стационарной случайной функции $\overset{\circ}{X}(t)$.



Очевидно, площадь, ограниченная кривой $S_x(\omega)$, по-прежнему должна равняться дисперсии D_x случайной функции $\overset{\circ}{X}(t)$:

$$D_x = \int_0^\infty S_x(\omega)d\omega .$$

Формула есть не что иное, как разложение дисперсии D_x на сумму элементарных слагаемых $S_x(\omega)d\omega$, каждое из которых представляет собой дисперсию, приходящуюся на элементарный участок частот $d\omega$, прилежащий к точке ω.

Таким образом, мы ввели в рассмотрение новую дополнительную характеристику стационарного случайного процесса – спектральную плотность, описывающую частотный состав стационарного процесса. Однако эта характеристика не является самостоятельной; она полностью определяется корреляционной функцией данного процесса.

Подобно тому, как ординаты дискретного спектра D_k выражаются формулами через корреляционную функцию $k_x(\tau)$, спектральная плотность $S_x(\omega)$ также может быть выражена через корреляционную функцию. Выведем это выражение. Для этого перейдем в каноническом разложении корреляционной функции к пределу при $T \to \infty$ и посмотрим, во что оно обратится. Будем исходить из разложения корреляционной функции в ряд Фурье на конечном интервале $(-T, T)$:

$$k_x(\tau) = \sum_{k=0}^{\infty} D_k \cos \omega_k \tau,$$

где дисперсия, соответствующая частоте ω_k, выражается формулой

$$D_k = \frac{2}{T} \int_0^T k_x \cos \omega_k \tau d\tau.$$

Перед тем как переходить к пределу при $T \to \infty$, перейдём от дисперсии D_k к средней плотности дисперсии $\frac{D_k}{\Delta \omega}$. Так как эта плотность вычисляется ещё при конечном значении T и зависит от T, обозначим ее:

$$S_x^{(T)}(\omega_k) = \frac{D_k}{\Delta \omega}.$$

Разделим выражение на $\Delta \omega = \frac{\pi}{T}$; получим:

$$S_x^{(T)}(\omega_k) = \frac{2}{\pi} \int_0^T k_x(\tau) \cos \omega_k \tau d\tau. \qquad D_k = S_x^{(T)}(\omega_k) \Delta \omega$$

Подставим выражение в формулу, получим:

$$k_x(\tau) = \sum_{k=0}^{\infty} S_x^{(T)}(\omega_k) \cos \omega_k \tau \Delta \omega.$$

Очевидно, при этом $\Delta \omega \to 0$; дискретный аргумент ω_k переходит в непрерывно меняющийся аргумент ω; сумма переходит в интеграл по переменной ω; средняя плотность дисперсии $S_x^{(T)}(\omega_k)$ стремится к плотности дисперсии $S_x(\omega)$, и выражение в пределе принимает вид:

$$k_x(\tau) = \int_0^{\infty} S_x(\omega) \cos \omega \tau d\omega,$$

где $S_x(\omega)$ - спектральная плотность стационарной случайной функции.

Спектральная плотность, ее свойства.

Переходя к пределу при $T \to \infty$, получим выражение спектральной плотности через корреляционную функцию:

$$S_x(\omega) = \frac{2}{\pi} \int_0^{\infty} k_x(\tau) \cos \omega \tau d\tau.$$

Выражение этого типа известно в математике под названием интеграла Фурье. Интеграл Фурье есть обобщение разложения в ряд Фурье для случая непериодической функции, рассматриваемой на бесконечном интервале, и представляет собой разложение функции на сумму элементарных гармонических колебаний с непрерывным спектром.

Таким образом, корреляционная функция и спектральная плотность выражаются одна через другую с помощью преобразований Фурье.

На практике вместо спектральной плотности $S_x(\omega)$ часто пользуются нормированной спектральной плотностью:

$$s_x(\omega) = \frac{S_x(\omega)}{D_x},$$

где D_x - дисперсия случайной функции.

Нетрудно убедиться, что нормированная корреляционная функция $\rho_x(\tau)$ и нормированная спектральная плотность $s_x(\omega)$ связаны теми же преобразованиями Фурье:

$$\left.\begin{aligned} \rho_x(\tau) &= \int_0^{\infty} s_x(\omega)\cos\omega\tau d\omega, \\ s_x(\omega) &= \frac{2}{\pi}\int_0^{\infty} \rho_x(\tau)\cos\omega\tau d\tau. \end{aligned}\right\}$$

Полагая в первом из равенств $\tau = 0$ и учитывая, что $\rho_x(0) = 1$, имеем:

$$\int_0^{\infty} s_x(\omega) d\omega = 1,$$

т. е. полная площадь, ограниченная графиком нормированной спектральной плотности, равна единице.

Характер изменения спектральной плотности $s_x(\omega)$ (быстрое или медленное убывание) зависит от параметра τ_0. Полная площадь, ограниченная кривой $s_x(\omega)$, постоянна и равна единице. Изменение τ_0 равносильно изменению масштаба кривой $s_x(\omega)$ по обеим, осям при сохранении ее площади. При увеличении τ_0 масштаб по оси ординат увеличивается, по оси абсцисс - уменьшается; преобладание в спектре случайной функции нулевой частоты становится более ярко выраженным.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ЛАБОРАТОРНЫХ ЗАНЯТИЙ

2.1 Лабораторная работа 1 (ЛР-1)

Тема: Вычисление вероятности СС

2.1.1 Цель работы: ознакомить студентов с методами подсчета вероятности случайного события, выработать навыки подбора соответствующего алгоритма.

2.1.2 Задачи работы:

1. Элементы комбинаторики
2. Непосредственное вычисление вероятности случайного события.
3. Операции над случайными событиями и их свойства. Теоремы о вероятности суммы случайных событий. Теоремы о вероятности суммы произведения событий.

2.1.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.1.4 Описание (ход) работы:

Пример. Сколько раз нужно подбросить два игральных кубика, чтобы вероятность выпадения хотя бы один раз двух шестерок была бы больше $\frac{1}{2}$? (Эта задача впервые поставлена французским математиком и писателем де Мере (1610-1684 гг.), поэтому задача называется его именем).

Решение. Пусть событие A_i - «выпадение двух шестерок при i-м подбрасывании». Так как с каждой из шести граней первого кубика может выпасть любая из шести граней второго кубика, то всего равновозможных попарно несовместных событий 6 · 6 = 36. Только одно из них - выпадение шестерки и на первом и на втором кубике - благоприятствуют событию A_i. Следовательно, $P(A_i)=\frac{1}{36}$, откуда $q=1-p=1-\frac{1}{36}=\frac{35}{36}$.

Подбрасывание игральных кубиков - независимые испытания, поэтому воспользуемся формулой $P(A)=1-q^n$, тогда в данном случае получим: $1-\left(\frac{35}{36}\right)^n>\frac{1}{2}$, или $\left(\frac{35}{36}\right)^n<\frac{1}{2}$.

Решив неравенство, найдем n. Логарифмируя обе части неравенства, получим $n\ln\frac{35}{36}<\ln\frac{1}{2}$, откуда $n>\frac{\ln 2}{\ln 36-\ln 35}=\frac{0,6931}{0,0284}=24,4$.

Итак, чтобы вероятность выпадения двух шестерок была больше $\frac{1}{2}$, достаточно подбросить кубик не менее 25 раз.

Пример. Имеются две урны с шариками трех цветов. В первой находятся 2 голубых, 3 красных, 5 зеленых, а во второй - 4 голубых, 2 красных и 4 зеленых. Из каждой урны извлекают по одному шару и сравнивают их цвета. Найти вероятность того, что цвета вынутых шаров одинаковы (событие A).

Решение. Обозначим событие, состоящее в извлечении из первой урны голубого шара, через B_1, красного - C_1, зеленого - D_1. Аналогичные события для второй урны обозначим соответственно через B_2, C_2, D_2. Событие A наступает в случае B_1B_2, C_1C_2 или D_1D_2. Они несовместны. Для вычисления искомой вероятности события A применим формулы вероятностей суммы несовместных событий и произведения независимых событий

$P(A) = P(B_1B_2+C_1C_2 + D_1D_2) = P(B_1B_2) + P(C_1C_2) + P(D_1D_2)$.

Так как независимы события: B_1 и B_2, C_1 и C_2, D_1 и D_2, то можно пользоваться формулой $P(AB)= P(A)P(B)$ для каждой пары событий:

$P(B_1B_2) = P(B_1)P\{B_2)$,

$P(C_1C_2) = P(C_1)P(C_2)$,

$P(D_1D_2) = P(D_1)P(D_2)$.

Окончательно

$P(A) = P(B_1)P(B_2) + P(C_1)P(C_2) + P(D_1)P(D_2) = 0,2\cdot 0,4 + 0,3\cdot 0,2 + 0,5\cdot 0,4 = 0,34$

2.2 Лабораторная работа 2 (ЛР-2)

Тема: Следствия из основных теорем о вычислении вероятности. Схема повторных испытаний

2.2.1 Цель работы: выработать навыки вычисления вероятности случайного события в схеме повторных испытаний, путем подбора соответствующей формулы.

2.2.2 Задачи работы:

1. Условная вероятность. Формула полной вероятности. Формула Байеса.

2. Схема повторных испытаний. Формула Бернулли. Формула Пуассона. Локальная формулы Лапласа. Интегральная формула Лапласа.
3.Простейший поток событий. 2. Вероятность случайного события с заданной интенсивностью.

2.2.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.2.4 Описание (ход) работы:

Пример. Пусть известно, что при изготовлении некоторого препарата брак (количество упаковок, не соответствующих стандарту) составляет 0,2%. Оценить приближенно вероятность того, что серди 1000 наугад выбранных упаковок окажутся три упаковки, не соответствующие стандарту.
Решение: Выбор каждой очередной упаковки можно рассматривать как независимое испытание. Из условий задачи следует, что n=1000 (т.е. велико) а p=0.002 (т.е. мало) следовательно, А можно считать редким событием. $\lambda=np=1000\cdot0.002=2<10$ Воспользуемся приближенной формулой Пуассона или таблицей.

$$P_n(k)=\frac{\lambda^k e^{-\lambda}}{k!} \quad\Rightarrow\quad P_{1000}(3)=\frac{2^3 e^{-2}}{3!}\approx 0{,}18$$

Пример. Партия электрических лампочек на 20% изготовлена первым заводом, на 30% - вторым, на 50% - третьим. Вероятность выпуска бракованных лампочек соответственно равны: $q_1 = 0{,}01$, $q_2 = 0{,}005$, $q_3 = 0{,}006$. Найти вероятность того, что наудачу взятая из партии лампочка окажется стандартной.

Решение. Введем обозначения: A - событие, состоящее в том, что «из партии взята стандартная лампочка», H_1 - событие, состоящее в том, что «взятая лампочка изготовлена первым заводом», H_2 - событие, состоящее в том, что «взятая лампочка изготовлена вторым заводом», H_3 - событие, состоящее в том, что взятая лампочка изготовлена «третьим заводом». Найдем условные вероятности

$P(A/H_i)$, (i= 1,2,3) по формуле $P(A/H_i)=1-P(\overline{A}/H_i)$, где $\overline{A}$ - событие, противоположное событию A (взята нестандартная лампочка):

$$P(A/H_1)=1-P(\overline{A}/H_1)=1-0{,}01=0{,}99,$$
$$P(A/H_2)=1-P(\overline{A}/H_2)=1-0{,}005=0{,}995,$$
$$P(A/H_3)=1-P(\overline{A}/H_3)=1-0{,}006=0{,}994.$$

Из условия задачи следует, что $P(H_1) = 0{,}2$, $P(H_2) = 0{,}3$, $P(H_3) = 0{,}5$.
Получим по формуле полной вероятности:

$$P(A)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)=$$
$$=0{,}2\cdot0{,}99+0{,}3\cdot0{,}995+0{,}5\cdot0{,}994=0{,}9935$$

Пример. В пяти ящиках находятся одинаковые по весу и размерам шары. В двух ящиках - по 6 зеленых и 4 красных шара (по ящик состава H_1). В двух других ящиках (состава H_2) - по 8 зеленых и 2 красных шара. В одном ящике (состава H_3) - 2 зеленых и 8 красных шаров. Наудачу выбирается ящик, и из него извлекается шар. Извлеченный шар оказался голубым. Какова вероятность того, что зеленый шар извлечен из ящика первого состава?

Решение. Обозначим через A событие, состоящее в том, что и i ящика извлечен голубой шар. Из условия задачи следует, что

$$P(H_1)=\frac{2}{5}=0{,}4;\ P(H_2)=\frac{2}{5}=0{,}4;\ P(H_3)=\frac{1}{5}=0{,}2.$$

Вероятность вынуть голубой шар, если известно, что взят ящик состава $H_1,H_2,...,H_3$ соответственно будут равны:

$$P(A/H_1)=\frac{6}{10}=0{,}6;$$
$$P(A/H_2)=\frac{8}{10}=0{,}8;$$
$$P(A/H_1)=\frac{2}{10}=0{,}2.$$
По формуле полной вероятности находим $P(A) = 0{,}4 \bullet 0{,}6 + 0{,}4 \bullet 0{,}8 + 0{,}2 \bullet 0{,}2 = 0{,}6$.

По формуле Бейеса найдем искомую вероятность

$$P(H_1/A)=\frac{P(H_1)P(A/H_1)}{P(A)}=\frac{0{,}4\cdot 0{,}6}{0{,}6}=0{,}4.$$

Пример. В партии из 400 деталей 80% - стандартных. Найти границы, в которых с вероятностью 0,9544 заключена доля стандартных деталей.

Решение. Воспользуемся формулой, являющейся частным случаем формулы Муавра-Лапласа

$$P(\left|\frac{m}{n}-p\right|\le\varepsilon)=2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right),$$

где m/n - доля числа наступивших событий A в n испытаниях,

n - число испытаний,

p - вероятность наступления события A в одном испытании,

ε - величина отклонения доли m/n от вероятности p,

$q = 1 - p$ - вероятность ненаступления события A в одном испытании.

Для данной задачи A - событие, состоящее в том, что деталь стандартная, $n = 400$; $p = 0{,}8$; $q=0{,}2$; $P = 0{,}9544$, величину ε нужно найти.

Итак: $0{,}9544=2\Phi\left(\frac{\varepsilon\sqrt{400}}{\sqrt{0{,}8\cdot 0{,}2}}\right)\Leftrightarrow\Phi(\varepsilon\cdot 50)=0{,}4772.$

По таблице- приложений значений функции Лапласа $\Phi(x)$ находим, что $\varepsilon\cdot 50 = 2$, следовательно, $\varepsilon = 0{,}04$. Таким образом, $|m/n - 0{,}8| < 0{,}04$ и $0{,}76 < m/n < 0{,}84$.

Пример. Среднее число заявок, поступающих на АТС в одну минуту, равно двум. Найти вероятности того, что за четыре минуты поступит:

а) три вызова;

б) менее трех вызовов;

в) не менее трех вызовов.

Решение, а) По условию $\lambda = 3$, $t = 2$, $k=5$. Воспользуемся формулой: $P_t(k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}$

Подставив данные условия задачи, получим: $P_4(3)=\frac{8^3\cdot e^{-8}}{3!}=\frac{512\cdot 0{,}000335}{6}\approx 0{,}03.$

б) Найдем вероятность того, что за четыре минуты поступит менее трех вызовов, т.е. ни одного вызова, или один вызов, или два вызова. Поскольку эти события несовместны, применим теорему суммы несовместных событий:

$$P_4(k<3)=P_4(0)+P_4(1)+P_4(2)=e^{-8}+8\cdot e^{-8}\cdot\frac{8^2\cdot e^{-8}}{2!}=41\cdot 0{,}000335\approx 0{,}01.$$

в) Найдем вероятность того, что за четыре минуты поступит не менее трех вызовов: так как события «поступило менее трех вызовов» и «поступило не менее трех вызовов» - противоположные, то сумма вероятностей этих событий равна единице: $P_4(k<3) + P_4(k\ge 3) = 1$. Поэтому $P_4(k\ge 3)=1-P_4(k<3)=1-[P_4(0)+P_4(1)+P_4(2)]=1-0{,}01=0{,}99$.

2.3 Лабораторная работа 3 – 4 (ЛР-3-4)

Тема: Случайные величины, законы их распределения, числовые характеристики

2.3.1 Цель работы: выработать навыки работы со случайными величинами, вычисления их числовых характеристик.

2.3.2 Задачи работы:

1.Случайные величины, их классификация.
2. Закон распределения случайной величины. Ряд распределения.
3. Функция распределения. Плотность распределения.
4. Числовые характеристики ДСВ. Числовые характеристики НСВ.
5. Свойства числовых характеристик, их интерпретация

2.3.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.3.4 Описание (ход) работы:

Пример. Компьютер состоит из трех независимо работающих элементов: системного блока, монитора и клавиатуры. При однократном резком повышении напряжения вероятность отказа каждого элемента равна 0,1. Исходя из распределения Бернулли составить закон распределения числа отказавших элементов при скачке напряжения в сети.
Решение. Рассмотрим распределение Бернулли (или биномиальное): вероятность того, что в n испытаниях событие А появится ровно k раз: $P_n(k) = C_n^k p^k q^{n-k}$, или:

X	0	1	…	k	…	n
P	qn	pq^{n-1}	…	$C_n^k p^k q^{n-k}$	…	pn

Вернёмся к задаче.
Возможные значения величины X (число отказов):
x0 =0 – ни один из элементов не отказал;
x1 =1 – отказ одного элемента;
x2 =2 – отказ двух элементов;
x3 =3 – отказ всех элементов.
Так как, по условию, p = 0,1, то q = 1 – p = 0,9. Используя формулу Бернулли, получим

$$P_3(0) = C_3^0 p^0 q^3 = q^3 = 0{,}9^3 = 0{,}729$$

$$P_3(1) = C_3^1 p^1 q^2 = 3pq^2 = 3 \cdot 0{,}1 \cdot 0{,}9^2 = 0{,}243,$$

$$P_3(2) = C_3^2 p^2 q^1 = 3p^2 q = 3 \cdot 0{,}1^2 \cdot 0{,}9 = 0{,}027$$

$$P_3(3) = C_3^3 p^3 q^0 = p^3 = 0{,}1^3 = 0{,}001.$$

Контроль: $0{,}729 + 0{,}243 + 0{,}027 + 0{,}001 = 1$.
Следовательно, искомый закон распределения:

X	0	1	2	3
p	0,729	0,243	0,027	0,001

Задача. Случайная величина X Задана плотностью распределения

$$f(x)=\begin{cases} a\cdot\cos x, & -\pi/2\le x\le \pi/2,\\ 0, & x<-\pi/2,\quad x>\pi/2\end{cases}$$

Требуется:

А) найти значение коэффициента А;

Б) найти функцию распределения;

В) найти вероятность попадания случайной величины на интервал (0, $\pi/2$).

Решение:

А) Воспользуемся свойством 3:

$$\int_{-\infty}^{+\infty} f(x)dx=\int_{-\pi/2}^{\pi/2} a\cdot\cos x\cdot dx = a\cdot\sin x\Big|_{-\pi/2}^{\pi/2}=a\cdot(1-(-1))=2a=1.$$

Отсюда получаем: А=1/2.

Б) Если $x\le\pi/2$, То

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{x}0\cdot dt=0;$$

Если $-\pi/2\le x\le\pi/2$, то

$$F(x)=\int_{-\infty}^{-\pi/2}0\cdot dt+\int_{-\pi/2}^{x}\frac{1}{2}\cdot\cos t\cdot dt=\frac{1}{2}\cdot\sin t\Big|_{-\pi/2}^{x}=\frac{1}{2}\cdot(\sin x+1);$$

Если $x>\pi/2$, то

$$F(x)=\int_{-\infty}^{-\pi/2}0\cdot dt+\int_{-\pi/2}^{\pi/2}\frac{1}{2}\cdot\cos t\cdot dt+\int_{\pi/2}^{x}0\cdot dt=\frac{1}{2}\cdot\sin t\Big|_{-\pi/2}^{\pi/2}=1;$$

Таким образом,

$$F(x)=\begin{cases} 0, & x<-\pi/2,\\ \frac{1}{2}\cdot(\sin x+1), & -\pi/2\le x\le\pi/2,\\ 1, & x>\pi/2.\end{cases}$$

В) По свойству 4:

$$P(0<X<\pi/4)=F(\pi/4)-F(0)=\frac{1}{2}\cdot(\sin(\pi/4)+1)-\frac{1}{2}\cdot(\sin 0+1)=$$

$$=\frac{1}{2}\cdot\left(\frac{\sqrt{2}}{2}+1\right)-\frac{1}{2}=\frac{\sqrt{2}}{4}.$$

Пример. Непрерывная случайная величина задана на интервале $0 < x < 1$ плотностью распределения $f(x) = 2x$, а вне этого интервала $f(x) = 0$. Найти ее числовые характеристики.

Решение. Математическое ожидание:

$$M(X) = \int_0^1 xf(x)dx = \int_0^1 2x^2dx = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}.$$

Дисперсия: $D(X) = \int_0^1 x^2 f(x)dx - (M(x))^2 = \int_0^1 2x^3 dx - \left(\frac{2}{3}\right)^2 = ... = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$.

Среднее квадратическое отклонение: $\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{1}{18}} = 0{,}24$

2.4 Лабораторная работа 5 (ЛР-5)

Тема: Частные виды законов распределения случайных величин

2.4.1 Цель работы: ознакомить студентов с различными видами распределения ДСВ, НСВ, их свойствами и приложениями.

2.4.2 Задачи работы:

1. Биномиальное распределение, его свойства, числовые характеристики.
2. Распределение Пуассона, его свойства, числовые характеристики.
3. Связь распределений ДСВ с нормальным распределением.
4. Равномерное распределение.
5. Показательное распределение.
6. Нормальное распределение, его свойства.

2.4.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.4.4 Описание (ход) работы:

Задача. Два ювелирные заводы производят свадебные кольца в объеме *3:7*. Первый завод производит *95%* колец без дефекта, второй – *90%*. Молодая пара перед свадьбой покупает пару колец. Построить закон распределения, вычислить математическое ожидание и среднее квадратичное отклонение.

Решение. Вероятность события *A* – куплена кольцо оказалась качественной определим по формуле полной вероятности $P(A) = \frac{3}{10}0{,}95 + \frac{7}{10}0{,}9 = 0{,}915.$

Случайная величина *X* – количество колец надлежащего качества среди купленных имеет биномиальное закон распределения с параметрами

$n = 2;\ p = 0{,}915;\ q = 1 - 0{,}915 = 0{,}085.$

Найдем соответствующие вероятности $P(0) = q^2 = 0{,}085^2 \approx 0{,}007;$

$P(1) = C_2^1 pq^{2-1} = 2 \cdot 0{,}915 \cdot 0{,}085 \approx 0{,}155;\quad P(2) = p^2 = 0{,}915^2 \approx 0{,}837;$

Запишем таблицу распределения

$X = k$	0	1	2
P_k	0,007	0,155	0,837

На основе табличных данных вычисляем математическое ожидание

$M(X) = 0 \cdot 0{,}007 + 1 \cdot 0{,}155 + 2 \cdot 0{,}837 = 1{,}829;$

дисперсию

$D(X) = 0^2 \cdot 0{,}007 + 1^2 \cdot 0{,}155 + 2^2 \cdot 0{,}837 = 3{,}503;$

Среднее квадратичное отклонение $\sigma(X) = \sqrt{D(X)} = \sqrt{3{,}503} \approx 1{,}872.$

Как можно убедиться из примеров, биномиальний закон распределения простой как для понимания так и для вычислений.

Задача. В рыбацком городке 99,99% мужчин хотя бы раз в жизни были на рыбалке. Проводят социологические исследования среди 10000 наугад выбранных мужчин. Определить дисперсию D (X) и среднее квадратическое отклонение S (X) случайной величины X — числа мужчин, которые ни разу не были на рыбалке.

Решение. Легко убедиться, что величина X имеет пуассоновский закон распределения. С условия задачи находим $n=10000; p=\frac{100-99,99}{100}=0,0001.$

По формулам находим дисперсию и среднее квадратическое отклонение

$D(X)=np=1; \quad \sigma(X)=\sqrt{np}=\sqrt{1}=1$

Пример. Охотник-любитель стреляет из ружья по неподвижной мишени. Вероятность попасть в мишень при одном выстреле является величиной постоянной и равна 0,65. Стрельба по мишени ведется до первого попадания.

Определить числовые характеристики M (X), D (X), S (X), A(X), E(X) числа израсходованных охотником патронов.

Решение. Случайная величина X подчиняется геометрическому закона распределения поэтому вероятность попадания в каждой попытке постоянна и составляет p=0,65;q=1-p=0,35.

По формулам вычисляем математическое ожидание $M(X)=\frac{1}{p}=\frac{1}{0,65}\approx 1,538;$

Дисперсию $D(X)=\frac{q}{p^2}=\frac{0,35}{0,65^2}\approx 0,828;$

среднее квадратическое отклонение $\sigma(X)=\sqrt{0,828}\approx 0,91;$

асимметрию $A(X)=\frac{2-p}{\sqrt{1-p}}=\frac{2-0,65}{\sqrt{0,35}}\approx 2,282;$

эксцесс $E(X)=6+\frac{p^2}{1-p}=6+\frac{(0,65)^2}{0,35}\approx 7,207.$

Вычисление числовых характеристик для геометрического закона распределения не так сложны, поэтому пользуйтесь приведенным формулам в подобных задачах и получайте только правильные результаты.

Пример. Поезда в метро прибывают на станцию каждые 10 минут. Определить вероятность того, что время ожидания состава не будет больше 4 минуты.

Решение. По условию задачи имеем два интервала

$[\alpha;\beta]=4\ хв;\ [a;b]=10\ хв.$

Согласно формуле, искомая вероятность равна доле этих величин

$P(\alpha\le X\le\beta)=\frac{4}{10}=0,4.$

Пример. Время безотказной работы устройства распределено по закону

$f(t)=0,02e^{-0,02t}, t\ge 0.$

Найти среднее время безотказной работы устройства, вероятность того, что устройство не откажет за среднее время безотказной работы. Найти вероятность отказа за время t= 100 часов.

Решение:

По условию интенсивность отказов m =0,02. Тогда среднее время между двумя отказами, т.е. математическое ожидание M(X)=1/0,02=50часов. Вероятность безотказной работы за этот промежуток времени вычислим по функции надежности:

$R(50)=e^{-0{,}02\cdot 50}=e^{-1}\approx 0{,}37.$

По функции F(t) вычислим вероятность отказа за время t =100 часов:

$F(100)=1-e^{-0{,}02\cdot 100}=1-e^{-2}\approx 0{,}86.$

Пример: Случайная величина X имеет нормальное распределение с математическим ожиданием $a=15$ и средним квадратическим отклонением $\sigma=5$. Найти вероятность того, что случайная величина X примет значение, принадлежащее интервалу $(10,30)$

Решение.

Известно, что вероятность того, что нормально распределенная случайная величина X примет значение, принадлежащее интервалу (α,β), равна:

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right),$$

где a – математическое ожидание, σ – среднее квадратическое отклонение.

По условию $a=15,\ \sigma=5,\ \alpha=10,\ \beta=30$. Следовательно,

$$P(10<X<30)=\Phi\left(\frac{30-15}{5}\right)-\Phi\left(\frac{10-15}{5}\right)=\Phi(3)-\Phi(-1)=\Phi(3)+\Phi(1)=$$

$$=0{,}49865+0{,}3413=0{,}83995.$$

Ответ: $P(10<X<30)=0{,}83995$

2.5 Лабораторная работа 6 -7 (ЛР-6-7)

Тема: Случайный вектор, функция и плотность системы СВ, условные законы распределения, числовые характеристики.

2.5.1 Цель работы: ознакомить студентов с основными понятиями многомерного анализа, выработать навыки вычисления условных числовых характеристик.

2.5.2 Задачи работы:

1. Системы двух случайных величин. Функция и плотность системы СВ, их свойства
2. Условные распределения плотностей ССВ, условные законы распределения.
3. Корреляционный момент ССВ, его свойства.
4. Коэффициент корреляции. Корреляционная матрица

2.5.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.5.4 Описание (ход) работы:

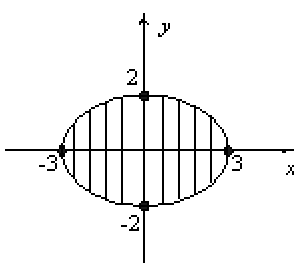


Рис. 2.2.5.

Пример. Известна функция плотности двумерного случайного вектора $(X;Y)$ (рис. 2.2.5):

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{6\pi}, & \text{если } \dfrac{x^2}{9}+\dfrac{y^2}{4}\le 1,\\ 0, & \text{если } \dfrac{x^2}{9}+\dfrac{y^2}{4}>1.\end{cases}$$

Найти плотности распределения $f_X(x)$ и $f_Y(y)$ компонент X и Y.

Решение. Очевидно, что если $|x|>3$, то $f_X(x)=0$. Пусть $|x|\le 3$, тогда:

$$f_X(x)=\frac{1}{6\pi}\int\limits_{-2\sqrt{1-\frac{x^2}{9}}}^{2\sqrt{1-\frac{x^2}{9}}} dy=\frac{2}{6\pi}\int\limits_{0}^{\frac{2}{3}\sqrt{9-x^2}} dy=\frac{2}{9\pi}\sqrt{9-x^2}.$$

Итак, $$f_X(x)=\begin{cases}\dfrac{2}{9\pi}\sqrt{9-x^2}, & \text{если } |x|\le 3,\\ 0, & \text{если } |x|>3.\end{cases}$$

Аналогично $$f_Y(y)=\begin{cases}\dfrac{\sqrt{4-y^2}}{2\pi}, & \text{если } |y|\le 2,\\ 0, & \text{если } |y|>2.\end{cases}$$

Ответ: $$f_X(x)=\begin{cases}\dfrac{2}{9\pi}\sqrt{9-x^2}, & \text{если } |x|\le 3,\\ 0, & \text{если } |x|>3;\end{cases}\quad f_Y(y)=\begin{cases}\dfrac{\sqrt{4-y^2}}{2\pi}, & \text{если } |y|\le 2,\\ 0, & \text{если } |y|>2.\end{cases}$$

Пример. Закон распределения двумерного дискретного случайного вектора $(X;Y)$ задан таблицей:

Y / X	0	1	2	3
1	0,1	0,1	0,1	0,1
2	0,05	0,05	0,05	0,05
3	0,1	0,1	0,1	0,1

Определить, зависимы или независимы компоненты X и Y.

Решение. Составим законы распределения компонент X и Y:

Y / X	0	1	2	3	$p_{i\bullet}$
1	0,1	0,1	0,1	0,1	0,4
2	0,05	0,05	0,05	0,05	0,2
3	0,1	0,1	0,1	0,1	0,4
$p_{\bullet j}$	0,25	0,25	0,25	0,25	1

Проверим теперь выполнение условия $p_{ij}=p_{i\bullet}\cdot p_{\bullet j}$ для всех пар индексов $i=1,2,3$, $j=1,2,3,4$. Очевидно, что это условие выполнено для любых i и j. Значит, компоненты X и Y независимы.

Ответ: компоненты X и Y независимы.

Замечание. В данном случае независимость компонент X и Y можно было установить, внимательно посмотрев на исходную таблицу, задающую закон распределения случайного вектора $(X;Y)$. Из этой таблицы видно, что *закон распределения каждой из компонент не зависит от того, какое значение приняла другая компонента.*

Пример. Система двух непрерывных случайных величин $(X;Y)$ имеет плотность распределения

$$f_{X,Y}(x,y)=\begin{cases} C, & -1\le x<2, 1\le y<3,\\ 0, & \text{иначе}.\end{cases}$$

Найти константу *C*. Определить, зависимы или независимы *X* и *Y*. Составить функцию распределения $F_{X,Y}(x,y)$.

Решение. Из условия нормировки для функции плотности $f_{X,Y}(x,y)$ имеем:

$$1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy = C\int_{-1}^{2}dx\int_{1}^{3}dy=6C,$$

отсюда $C=\dfrac{1}{6}$. Таким образом,

$$f_{X,Y}(x,y)=\begin{cases} \dfrac{1}{6}, & -1\le x<2, 1\le y<3,\\ 0, & \text{иначе}.\end{cases}$$

Найдем функции плотности отдельных компонент:

$$f_X(x)=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy=\frac{1}{6}\int_{1}^{3}dy=\frac{1}{3}, \text{ т.е. } f_X(x)=\begin{cases} \dfrac{1}{3}, & -1\le x<2,\\ 0, & \text{иначе};\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx=\frac{1}{6}\int_{-1}^{2}dx=\frac{1}{2}, \text{ т.е. } f_Y(y)=\begin{cases} \dfrac{1}{2}, & 1\le y<3,\\ 0, & \text{иначе}.\end{cases}$$

Очевидно, что равенство $f_{X,Y}(x,y)=f_X(x)\cdot f_Y(y)$ выполняется для всех точек координатной плоскости. Значит, компоненты *X* и *Y* независимы. Найдем функцию распределения системы $F_{X,Y}(x,y)$. Так как компоненты независимы, значит, $F_{X,Y}(x,y)=F_X(x)\cdot F_Y(y)$. Найдем вначале $F_X(x)$ и $F_Y(y)$:

$$F_X(x)=\begin{cases} 0, & x\le -1,\\ \dfrac{x+1}{3}, & -1<x\le 2,\\ 1, & x>2;\end{cases}\qquad F_Y(y)=\begin{cases} 0, & y\le 1,\\ \dfrac{y-1}{2}, & 1<y\le 3,\\ 1, & y>3.\end{cases}$$

Перемножая $F_X(x)$ и $F_Y(y)$ при $(x;y)\in$ «прямоугольнику» и учитывая, что $F_{X,Y}(+\infty,y)=F_Y(y)$, $F_{X,Y}(x,+\infty)=F_X(x)$, получим:

Y / *X*	$y\le 1$	$1<y\le 3$	$y>3$
$x\le -\mathbf{1}$	0	0	0
$-\mathbf{1}<x\le \mathbf{2}$	0	$\dfrac{(x+1)(y-1)}{6}$	$\dfrac{x+1}{3}$
$x>\mathbf{2}$	0	$\dfrac{y-1}{2}$	1

Пример. Дан двумерный случайный вектор $(X;Y)$, где *X* – время появления в магазине первого покупателя в понедельник, а *Y* – время появления в магазине первого покупателя во вторник. Установлено, что $f_{X,Y}(x,y)=e^{-x-y}$, если $x,y\ge 0$. Найти: $F_X(x\mid Y=y)$, $F_Y(y\mid X=x)$, $f_X(x\mid Y=y)$, $f_Y(y\mid X=x)$. Установить, зависимы или нет случайные величины *X* и *Y*.

Решение. Найдем вначале функцию распределения случайного вектора $(X;Y)$:

$$F_{X,Y}(x,y)=\int_0^y\left(\int_0^x f_{X,Y}(u,v)\,du\right)dv=\int_0^y e^{-v}\,dv\int_0^x e^{-u}\,du=(1-e^{-x})(1-e^{-y}),\quad x,y\ge 0;$$

$F_{X,Y}(x,y)=0$ в остальных случаях.

Тогда по свойству 4 совместной функции распределения $F_{X,Y}(x,y)$ получим:

$F_X(x)=F_{X,Y}(x,+\infty)=1-e^{-x}$ при $x>0$, $F_Y(y)=F_{X,Y}(+\infty,y)=1-e^{-y}$ при $y>0$.

Отсюда $f_X(x)=F_X'(x)=e^{-x}$ при $x>0$, $f_Y(y)=F_Y'(y)=e^{-y}$ при $y>0$.

Найдем теперь условные функции распределения компонент:

$$F_X(x\mid Y=y)=\frac{\int_{-\infty}^x f_{X,Y}(u,y)\,du}{f_Y(y)}=\frac{\int_0^x e^{-u-y}\,du}{f_Y(y)}=e^{y}\int_0^x e^{-u}e^{-y}\,du=1-e^{-x}\quad \text{при } x>0,$$

аналогично
$$F_Y(y\mid X=x)=\frac{\int_{-\infty}^y f_{X,Y}(x,v)\,dv}{f_X(x)}=1-e^{-y}\quad \text{при } y>0.$$

Получим теперь условные плотности компонент:

$$f_X(x\mid Y=y)=\frac{\partial}{\partial x}F_X(x\mid Y=y)=e^{-x}\quad \text{при } x>0,$$

$$f_Y(y\mid X=x)=\frac{\partial}{\partial y}F_Y(y\mid X=x)=e^{-y}\quad \text{при } y>0.$$

Поскольку $f_X(x\mid Y=y)=f_X(x)=e^{-x}$, $f_Y(y\mid X=x)=f_Y(y)=e^{-y}$,

то $f_{X,Y}(x,y)=f_X(x)\cdot f_Y(y)$. Поэтому случайные величины *X* и *Y* независимы. Это означает, что появление в магазине первого покупателя во вторник не зависит от того, когда в магазин пришел первый покупатель в понедельник.

Ответ: при положительных *x* и *y* $F_X(x\mid Y=y)=1-e^{-x}$, $F_Y(y\mid X=x)=1-e^{-y}$, $f_X(x\mid Y=y)=e^{-x}$, $f_Y(y\mid X=x)=e^{-y}$; случайные величины *X* и *Y* независимы.

Пример. Известна плотность совместного распределения непрерывной двумерной случайной величины $(X;Y)$: $f_{X,Y}(x,y)=\frac{1}{\pi}\cdot e^{-\frac{x^2+2xy+5y^2}{2}}$.

Найти: 1) плотности распределения компонент *X* и *Y*; 2) условные плотности распределения компонент *X* и *Y*.

Решение. 1) Найдем вначале плотность распределения компоненты *X*:

$$f_X(x)=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy=\frac{1}{\pi}\cdot\int_{-\infty}^{+\infty} e^{-\frac{x^2+2xy+5y^2}{2}}\,dy.$$

Вынося за знак интеграла множитель $e^{-\frac{x^2}{2}}$, не зависящий от переменной интегрирования y, и дополнив оставшийся показатель степени до полного квадрата, получим:

$$f_X(x)=\frac{1}{\pi}\cdot e^{-\frac{x^2}{2}}\cdot e^{\frac{x^2}{10}}\cdot\sqrt{\frac{2}{5}}\cdot\int\limits_{-\infty}^{+\infty} e^{-\left(y\sqrt{5/2}+x\sqrt{2/5}\right)^2}\,\mathrm{d}\left(y\sqrt{5/2}+x\sqrt{2/5}\right).$$

Учитывая, что интеграл Пуассона $\int\limits_{-\infty}^{+\infty} e^{-u^2}\,\mathrm{d}u=\sqrt{\pi}$, найдем плотность распределения компоненты X: $f_X(x)=\sqrt{\frac{2}{5\pi}}\cdot e^{-0,4x^2}$. Аналогично найдем плотность распределения компоненты Y: $f_Y(y)=\sqrt{\frac{2}{\pi}}\cdot e^{-2y^2}$.

2) Найдем условные плотности распределения компонент X и Y. Выполнив элементарные выкладки, получим:

$$f_X(x\,|\,Y=y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{(x+y)^2}{2}},\quad f_Y(y\,|\,X=x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\frac{\sqrt{5}}{\sqrt{2\pi}}\cdot e^{-\frac{(x+5y)^2}{10}}.$$

Ответ: 1) $f_X(x)=\sqrt{\frac{2}{5\pi}}\cdot e^{-0,4x^2}$, $f_Y(y)=\sqrt{\frac{2}{\pi}}\cdot e^{-2y^2}$;

2) $f_X(x\,|\,Y=y)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{(x+y)^2}{2}}$, $f_Y(y\,|\,X=x)=\frac{\sqrt{5}}{\sqrt{2\pi}}\cdot e^{-\frac{(x+5y)^2}{10}}$.

Пример. Двумерный случайный вектор $(X;Y)$ равномерно распределен на множестве случайных точек Q, задаваемых неравенством $|x|+|y|\le 1$. Выяснить, являются ли случайные величины X и Y: 1) зависимыми; 2) коррелированными.

Решение. Множество точек Q, задаваемых неравенством $|x|+|y|\le 1$, является квадратом. Поскольку двумерный случайный вектор $(X;Y)$ равномерно распределен на множестве Q, его плотность имеет вид

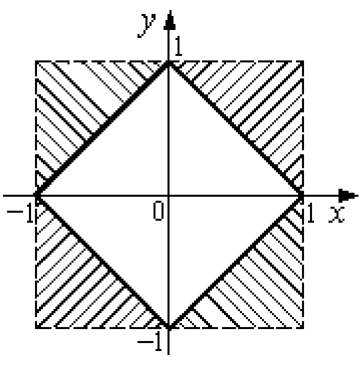
$$f_{X,Y}(x,y)=\begin{cases}C, & (x;y)\in Q;\\ 0, & (x;y)\notin Q.\end{cases}$$

Из условия нормировки найдем константу C:

$$1=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y=\iint\limits_{Q} C\,\mathrm{d}x\,\mathrm{d}y=C\cdot|Q|=2C,$$

где $|Q|$ – площадь квадрата Q, равная 2. Отсюда $C=0{,}5$, а значит,

$$f_{X,Y}(x,y)=\begin{cases}0{,}5, & (x;y)\in Q;\\ 0, & (x;y)\notin Q.\end{cases}$$



1) Найдем вначале плотность распределения компоненты X.

Если $|x|>1$, то, очевидно, $f_{X,Y}(x,y)=0$ для всех $y\in\mathbf{R}$.

Если $|x|\le 1$, то

$$f_X(x)=\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy=\int\limits_{|x|-1}^{1-|x|} 0{,}5\,dy=1-|x|$$, т.е. $$f_X(x)=\begin{cases}1-|x|, & |x|\le 1,\\ 0, & |x|>1.\end{cases}$$

Аналогично находится плотность распределения компоненты *Y*:

$$f_Y(y)=\begin{cases}1-|y|, & |y|\le 1,\\ 0, & |y|>1.\end{cases}$$

Равенство $f_{X,Y}(x,y)=f_X(x)\cdot f_Y(y)$ не выполняется для точек координатной плоскости, принадлежащих заштрихованным областям (рис.), поскольку в этих точках $f_{X,Y}(x,y)=0$, а $f_X(x)\neq 0$ и $f_Y(y)\neq 0$. Суммарная площадь заштрихованных областей равна 2, значит, компоненты *X* и *Y* зависимы.

2) Вычислим математические ожидания компонент *X* и *Y*:

$$m_X=\int\limits_{-\infty}^{+\infty} x\,f_X(x)\,dx=\int\limits_{-1}^{1} x(1-|x|)\,dx=0,$$

т.к. интеграл от нечетной функции в симметричных пределах равен нулю. Аналогично $m_Y=0$.

Определим начальный момент $\alpha_{1,1}$:

$$\alpha_{1,1}=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xy\,f_{X,Y}(x,y)\,dx\,dy=0{,}5\cdot\iint\limits_{\Omega} xy\,dx\,dy=0{,}5\cdot\int\limits_{-1}^{1}dx\int\limits_{|x|-1}^{1-|x|} y\,dy=0.$$

Таким образом, ковариация $K_{X,Y}=\mu_{1,1}=\alpha_{1,1}-\alpha_{1,0}\cdot\alpha_{0,1}=0$.

Значит, компоненты *X* и *Y* некоррелированные.

Ответ: компоненты *X* и *Y* зависимы, но некоррелированны.

Определение. *Условным математическим ожиданием* одной из случайных величин, входящих в двумерный случайный вектор $(X;Y)$, называется ее математическое ожидание, вычисленное при условии, что другая случайная величина приняла определенное значение.

Если случайные величины *X* и *Y* *дискретны*, то условные математические ожидания вычисляются по формулам:

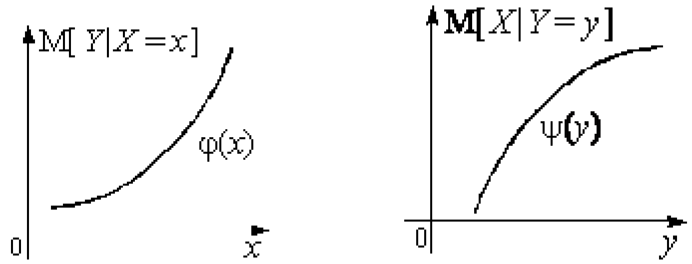
$$\mathrm{M}\left[Y\,|\,X=x_i\right]=\sum_j y_j\,\mathrm{P}\left\{Y=y_j\,|\,X=x_i\right\},\quad \mathrm{M}\left[X\,|\,Y=y_j\right]=\sum_i x_i\,\mathrm{P}\left\{X=x_i\,|\,Y=y_j\right\}.$$

Если случайные величины *X* и *Y* *непрерывны*, то условные математические ожидания вычисляются по формулам:

$$\mathrm{M}\left[Y\,|\,X=x\right]=\int\limits_{-\infty}^{+\infty} y\,f_Y(y\,|\,X=x)\,dy,\quad \mathrm{M}\left[X\,|\,Y=y\right]=\int\limits_{-\infty}^{+\infty} x\,f_X(x\,|\,Y=y)\,dx.$$

Определение. Условное математическое ожидание случайной величины *Y* при заданном значении $X=x$, т.е. $\mathbf{M}\left[Y\,|\,X=x\right]=\varphi(x)$, называется *регрессией Y на x*. Условное математическое ожидание случайной величины *X* при заданном значении $Y=y$, т.е. $\mathbf{M}\left[X\,|\,Y=y\right]=\psi(y)$, называется *регрессией X на y*.

Графики этих зависимостей от *x* и *y* называются *линиями регрессии* *Y* на *x* (рис. *а*) и *X* на *y* (рис. *б*) соответственно.



а *б*

Пример. Закон распределения случайного вектора $(X;Y)$ задан таблицей:

Y / *X*	0	2	5
1	0,1	0	0,2
2	0	0,3	0
4	0,1	0,3	0

Построить регрессии *Y* на *x* и *X* на *y*.

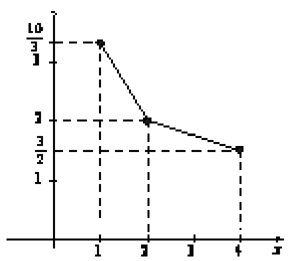
Решение. Найдем законы распределения компонент *X* и *Y*:

Y / *X*	0	2	5	$p_{i\bullet}$
1	0,1	0	0,2	0,3
2	0	0,3	0	0,3
4	0,1	0,3	0	0,4
$p_{\bullet j}$	0,2	0,6	0,2	□

Построим вначале регрессию *Y* на *x*.

1) $P\{Y=0|X=1\}=\dfrac{0,1}{0,3}=\dfrac{1}{3}$, $P\{Y=2|X=1\}=\dfrac{0}{0,3}=0$, $P\{Y=5|X=1\}=\dfrac{0,2}{0,3}=\dfrac{2}{3}$, отсюда $M[Y|X=1]=0\cdot\dfrac{1}{3}+2\cdot 0+5\cdot\dfrac{2}{3}=\dfrac{10}{3}$.

2) $P\{Y=0|X=2\}=\dfrac{0}{0,3}=0$, $P\{Y=2|X=2\}=\dfrac{0,3}{0,3}=1$, $P\{Y=5|X=2\}=\dfrac{0}{0,3}=0$, отсюда $\mathbf{M}[Y|X=2]=\mathbf{0\cdot 0+2\cdot 1+5\cdot 0=2}$.



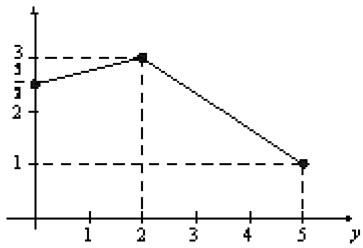
3) $P\{Y=0|X=4\}=\dfrac{0,1}{0,4}=\dfrac{1}{4}$, $P\{Y=2|X=4\}=\dfrac{0,3}{0,4}=\dfrac{3}{4}$, $P\{Y=5|X=4\}=\dfrac{0}{0,4}=0$, отсюда $M[Y|X=4]=0\cdot\dfrac{1}{4}+2\cdot\dfrac{3}{4}+5\cdot 0=\dfrac{3}{2}$.

Графическое изображение регрессии *Y* на *x* показано на рисунке

Построим теперь регрессию *X* на *y*.

1) $\mathrm{P}\{X=1|Y=0\}=\dfrac{0{,}1}{0{,}2}=\dfrac{1}{2}$, $\mathrm{P}\{X=2|Y=0\}=\dfrac{0}{0{,}2}=0$, $\mathrm{P}\{X=4|Y=0\}=\dfrac{0{,}1}{0{,}2}=\dfrac{1}{2}$, отсюда $\mathrm{M}[X|Y=0]=1\cdot\dfrac{1}{2}+2\cdot 0+4\cdot\dfrac{1}{2}=\dfrac{5}{2}$.

2) $\mathrm{P}\{X=1|Y=2\}=\dfrac{0}{0{,}6}=0$, $\mathrm{P}\{X=2|Y=2\}=\dfrac{0{,}3}{0{,}6}=\dfrac{1}{2}$, $\mathrm{P}\{X=4|Y=2\}=\dfrac{0{,}3}{0{,}6}=\dfrac{1}{2}$, отсюда $\mathrm{M}[X|Y=2]=1\cdot 0+2\cdot\dfrac{1}{2}+4\cdot\dfrac{1}{2}=3$.



3) $\mathrm{P}\{X=1|Y=5\}=\dfrac{0{,}2}{0{,}2}=1$, $\mathrm{P}\{X=2|Y=5\}=\dfrac{0}{0{,}2}=0$, $\mathrm{P}\{X=4|Y=5\}=\dfrac{0}{0{,}2}=0$, отсюда $\mathbf{M}[X|Y=5]=1\cdot 1+2\cdot 0+4\cdot 0=1$.

Графическое изображение регрессии X на y показано на рисунке.

Для наглядности значения условного математического ожидания на соединены отрезками прямых.

Замечание 1. Для *независимых* случайных величин линии регрессии Y на x и X на y параллельны координатным осям, т.к. математическое ожидание каждой из них не зависит от того, какое значение приняла другая случайная величина. Линии регрессии могут быть параллельны координатным осям и для *зависимых* случайных величин, если только математическое ожидание каждой из них не зависит от того, какое значение приняла другая случайная величина.

Замечание 2. По аналогии с условными математическими ожиданиями можно рассматривать условные моменты. Например, условные дисперсии $\mathbf{D}[Y|X=x]$, $\mathbf{D}[X|Y=y]$ и т.д.

Пример. Ранее была дана функция плотности $f_{X,Y}(x,y)$:

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{6\pi}, & \dfrac{x^2}{9}+\dfrac{y^2}{4}\le 1,\\[2mm] 0, & \dfrac{x^2}{9}+\dfrac{y^2}{4}>1;\end{cases}$$

и найдены условные плотности распределения компонент X и Y (пример 2.2.21):

$$f_X(x|Y=y)=\begin{cases}\dfrac{1}{3\sqrt{4-y^2}}, & |x|\le\dfrac{3}{2}\sqrt{4-y^2},\\[2mm] 0, & |x|>\dfrac{3}{2}\sqrt{4-y^2};\end{cases}$$

$$f_Y(y|X=x)=\begin{cases}\dfrac{3}{4\sqrt{9-x^2}}, & |y|\le\dfrac{2}{3}\sqrt{9-x^2},\\[2mm] 0, & |y|>\dfrac{2}{3}\sqrt{9-x^2}.\end{cases}$$

Найти регрессии Y на x и X на y, а также условные дисперсии компонент X и Y.

Решение. Условные математические ожидания вычисляются по формулам: $\mathrm{M}[X|Y=y]=\int\limits_{-\infty}^{+\infty} x\, f_X(x|Y=y)\,\mathrm{d}x$, $\mathrm{M}[Y|X=x]=\int\limits_{-\infty}^{+\infty} y\, f_Y(y|X=x)\,\mathrm{d}y$. Поэтому

$$\mathrm{M}[X|Y=y]=\int\limits_{-1{,}5\sqrt{4-y^2}}^{1{,}5\sqrt{4-y^2}} \frac{x}{3\sqrt{4-y^2}}\,\mathrm{d}x=\left\|\begin{matrix}\text{нечетная по } x \text{ функция}\\ \text{в симметричных}\\ \text{пределах}\end{matrix}\right\|=0,$$

$$\mathrm{M}[Y|X=x]=\int\limits_{-\frac{2}{3}\sqrt{9-x^2}}^{\frac{2}{3}\sqrt{9-x^2}} \frac{3y}{4\sqrt{9-x^2}}\,\mathrm{d}y=\left\|\begin{matrix}\text{нечетная по } y \text{ функция}\\ \text{в симметричных}\\ \text{пределах}\end{matrix}\right\|=0.$$

Заметим, что при вычислении условных математических ожиданий можно было воспользоваться тем, что случайная величина *X* при условии $Y=y$ равномерно распределена на отрезке $\left[-\frac{3}{2}\sqrt{4-y^2};\ \frac{3}{2}\sqrt{4-y^2}\right]$, а случайная величина *Y* при условии $X=x$ равномерно распределена на отрезке $\left[-\frac{2}{3}\sqrt{9-x^2};\ \frac{2}{3}\sqrt{9-x^2}\right]$.

Действительно, для равномерно распределенной на отрезке $[a;\,b]$ случайной величины математическое ожидание равно $\frac{a+b}{2}$, а дисперсия равна $\frac{(b-a)^2}{12}$. Отсюда, очевидно, $\mathbf{M}[X|Y=y]=\mathbf{0}$ при $|\,y\,|\le 2$, $\mathbf{M}[Y|X=x]=\mathbf{0}$ при $|\,x\,|\le 3$. Тогда условные дисперсии равны:

$$\mathrm{D}[\,X\,|Y=y]=\frac{\left(1{,}5\sqrt{4-y^2}-\left(-1{,}5\sqrt{4-y^2}\right)\right)^2}{12}=\frac{3(4-y^2)}{4} \text{ при } |\,y\,|\le 2,$$

$$\mathrm{D}[\,Y\,|X=x]=\frac{4(9-x^2)}{27} \text{ при } |\,x\,|\le 3.$$

Ответ: $\mathbf{M}[X|Y=y]=\mathbf{0}$, $\mathrm{D}[\,X\,|Y=y]=\frac{3(4-y^2)}{4}$ при $|\,y\,|\le 2$;

$\mathbf{M}[Y|X=x]=\mathbf{0}$, $\mathrm{D}[\,Y\,|X=x]=\frac{4(9-x^2)}{27}$ при $|\,x\,|\le 3$.

2.6 Лабораторная работа 8 (ЛР-8)

Тема: Первичная обработка данных эксперимента в среде Excel

2.6.1 Цель работы: ознакомить студентов с возможностями применения пакета Excel для первичной обработки статистических данных

2.6.2 Задачи работы:

1. Статистический материал и его первичная обработка.
2. Эмпирические законы распределения. Полигон частот, гистограмма.
3. Числовые характеристики выборки.
4. Точечные оценки выборочных характеристик.

2.6.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.6.4 Описание (ход) работы:

Пример. Построить эмпирическую функцию по данному распределению выборки:

x_i	2	6	10
m_i	12	18	30

Объем выборки n = 12+ 18+ 30 =60. $X_{наим}$= 2, значит при X £ 2,

$$\hat{F}(x) = 12/60$$

X<6 наблюдалось 12 раз, следовательно, при X< 6

$$\hat{F}(x) = 12/60$$.

Значение X<10 наблюдалось 12+18= 30 раз, значит при X<10

$$\hat{F}(x) = 30/60$$

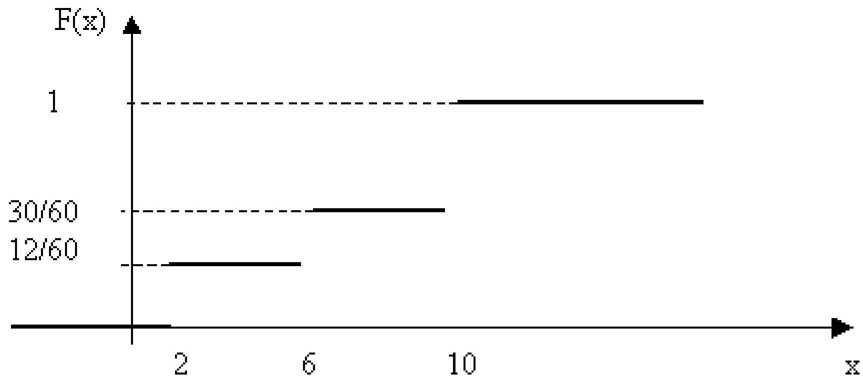
Так как $x_{наиб}$ =10, то при X ³□10

$$\hat{F}(x) = 1$$

Искомая эмпирическая функция имеет вид:

$$\hat{F}(x) = \begin{cases} 0, x \le 2; \\ 12/60, x < 6; \\ 30/60, x < 10; \\ 1, x \ge 10. \end{cases}$$

График строится так же, как и график интегральной функции распределения.

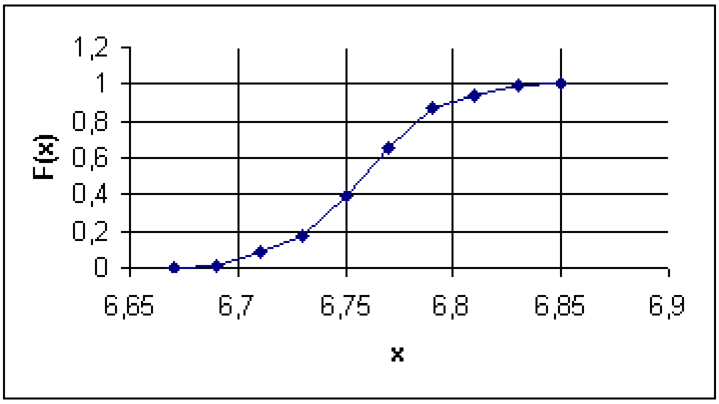


Если результаты наблюдений представлены в виде интервального вариационного ряда, то в качестве x принимают концы частичных интервалов и , пользуясь данным выше определением вычисляют значения эмпирической функции.

Для рассмотренного примера получим таблицу:

x	6,67	6,69	6,71	6,73	6,75	6,77	6,79	6,81	6,83	6,85
$\hat{F}(x)$	0	0,01	0,085	0,17	0,39	0,65	0,87	0,94	0,995	1

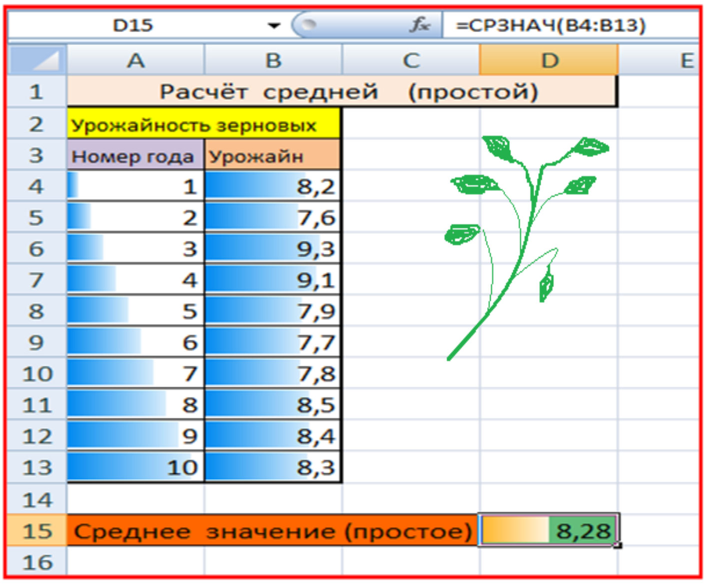
Так как таблица определяет функцию не полностью, то при изображении графика доопределяем функцию, соединяя точки графика, соответствующие концам интервалов, отрезками. График эмпирической функции для интервального вариационного ряда есть непрерывная линия.



1. Расчёт средней арифметической величины (простой) производится по формуле

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad \overline{x} = 8{,}28.$$

Номер значения: x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Величина x_i	8,2	7,6	9,3	9,1	7,9	7,7	7,8	8,5	8,4	8,3

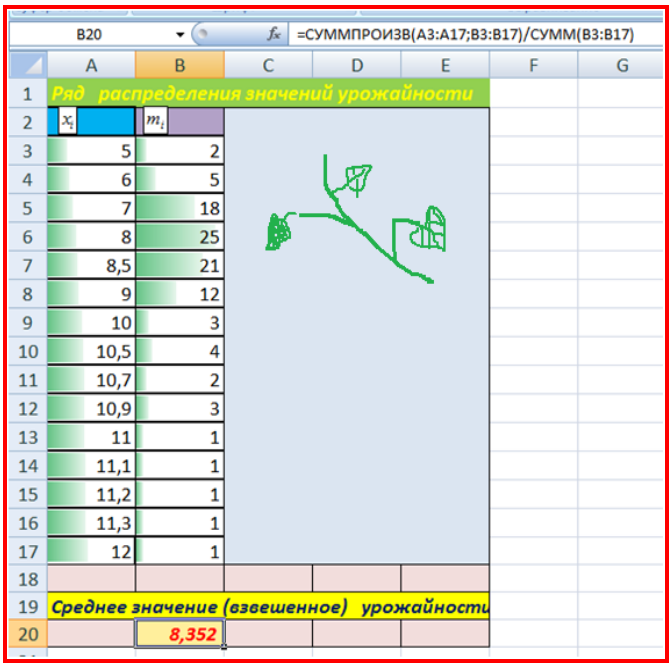


2. Расчёт средней арифметической (взвешенной) производится по формуле

$$\overline{x} = \frac{x_1 \cdot m_1 + x_2 \cdot m_2 + x_3 \cdot m_3 + \ldots + x_n \cdot m_n}{n} = \frac{\sum_{i=1}^{n} x_i \cdot m_i}{n}, \qquad \overline{x} = 8{,}28.$$

x_i	5	6	7	8	8,5	9	10	10,5	10,7	10,9	11	11,1	11,2	11,3	12
m_i	2	5	18	25	21	12	3	4	2	3	1	1	1	1	1

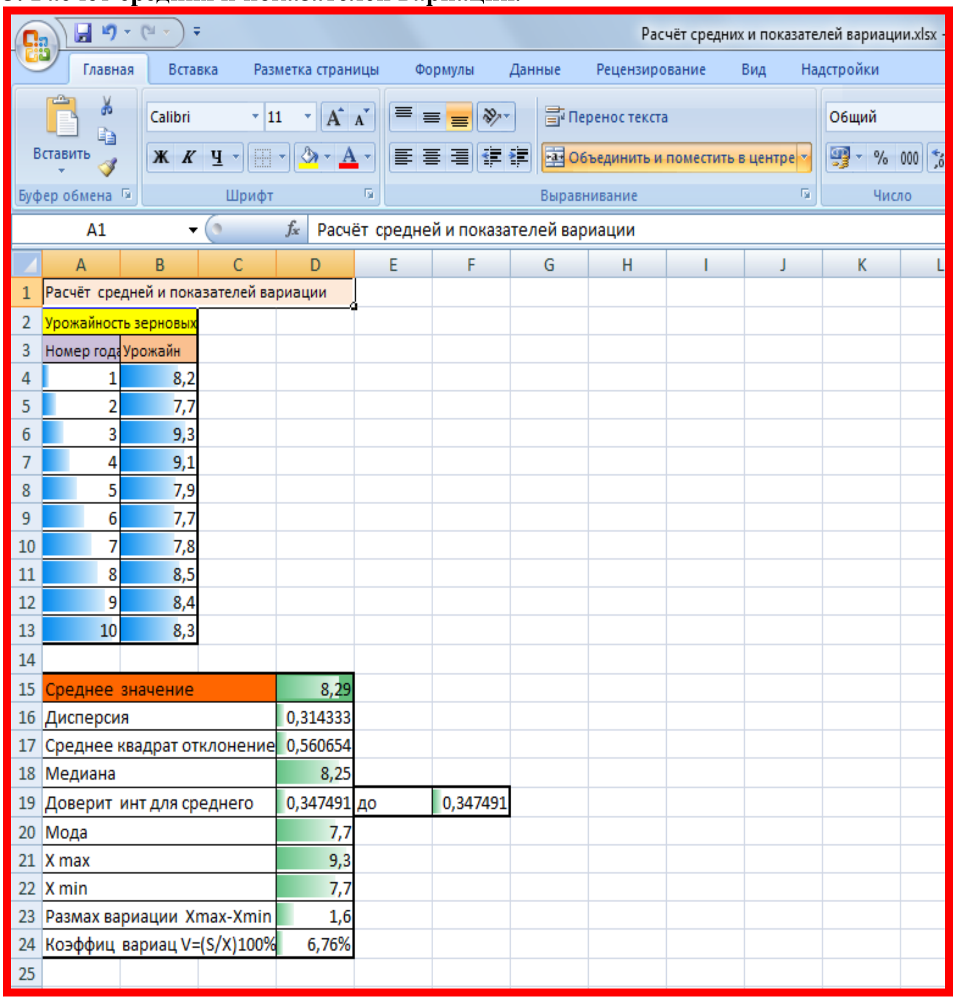
$n = 100$



B20 =СУММПРОИЗВ(A3:A17;B3:B17)/СУММ(B3:B17)

	A	B
1	Ряд распределения значений урожайности	
2	x_i	m_i
3	5	2
4	6	5
5	7	18
6	8	25
7	8,5	21
8	9	12
9	10	3
10	10,5	4
11	10,7	2
12	10,9	3
13	11	1
14	11,1	1
15	11,2	1
16	11,3	1
17	12	1
18		
19	Среднее значение (взвешенное) урожайности	
20		8,352

3. Расчёт средних и показателей вариации.



	A	B	C	D	E	F
1	Расчёт средней и показателей вариации					
2	Урожайность зерновых					
3	Номер года	Урожайн				
4	1	8,2				
5	2	7,7				
6	3	9,3				
7	4	9,1				
8	5	7,9				
9	6	7,7				
10	7	7,8				
11	8	8,5				
12	9	8,4				
13	10	8,3				
14						
15	Среднее значение			8,29		
16	Дисперсия			0,314333		
17	Среднее квадрат отклонение			0,560654		
18	Медиана			8,25		
19	Доверит инт для среднего			0,347491	до	0,347491
20	Мода			7,7		
21	X max			9,3		
22	X min			7,7		
23	Размах вариации Xmax-Xmin			1,6		
24	Коэффиц вариац V=(S/X)100%			6,76%		
25						

2.7 Лабораторная работа 9 -10 (ЛР-9-10)

Тема: Выравнивание рядов

2.7.1 Цель работы: выработать навыки применения статистических критериев к выравниванию рядов.

2.7.2 Задачи работы:

1. Статистические гипотезы и их виды.
2. Критерии согласия.
3. Оценка параметров неизвестного распределения.
4. Выравнивание рядов.

2.7.3 Перечень приборов, материалов, используемых в лабораторной работе:
спецификой дисциплины не предусмотрены

2.7.4 Описание (ход) работы:

Рассмотрим задачу «выравнивания» статистического распределения. Порядок решения этой задачи может быть следующим.

1. На основании статистических данных, оформленных в виде интервальной таблицы частот p^*, строят полигон или гистограмму и по внешнему виду этих графиков выдвигают гипотезу (делают предположение) о возможном теоретическом законе распределения случайной величины (кривой распределения).

Замечание. В некоторых случаях вид теоретической кривой распределения выбирается заранее из соображений, связанных с существом задачи.

2. Выясняют, от каких параметров зависит аналитическое выражение выбранной кривой распределения, и находят статистические оценки этих параметров. В этом случае задача выравнивания статистического распределения переходит в задачу рационального выбора тех значений параметров, при которых соответствие между статистическим и теоретическим распределениями оказывается наилучшим.

Например, если выдвигается гипотеза о нормальном законе распределения $X \sim N(a; \sigma)$, то он зависит только от двух параметров: математического ожидания a и среднего квадратического отклонения σ. Их наилучшими статистическими оценками будут соответственно среднее выборочное $\bar{x}$ и выборочное среднее квадратическое отклонение $\tilde{\sigma}$, т. е.

$$a \approx \bar{x},\ \sigma \approx \tilde{\sigma}.$$

3. С учетом выдвинутой гипотезы о законе распределения случайной величины находят вероятности p_i попадания случайной величины в каждый из интервалов, указанных в статистической таблице распределения; записывают их в третьей строке таблицы и сравнивают полученные значения вероятностей p_i с соответствующими заданными частотами p_i^* (для наглядности можно изобразить графически). Проводя такое сравнение, делается приблизительная оценка степени согласования статистического и теоретического распределений. На этом первый этап решения задачи по определению закона распределения случайной величины заканчивается.

Пример. Для разумного планирования и организации работы ремонтных мастерских специальной техники оказалось необходимым изучить длительность ремонтных операций, производимых мастерскими.

Результаты (сгруппированные по интервалам) соответствующего статистического обследования (фиксированы длительности операций в
100 случаях) представлены в таблице:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
n_i	36	24	16	10	7	4	3

Требуется выровнять это статистическое распределение с помощью показательного закона $f(t) = \lambda e^{-\lambda t}$ (при $t \geq 0$), где λ – длительность операции в единицу времени.

Решение

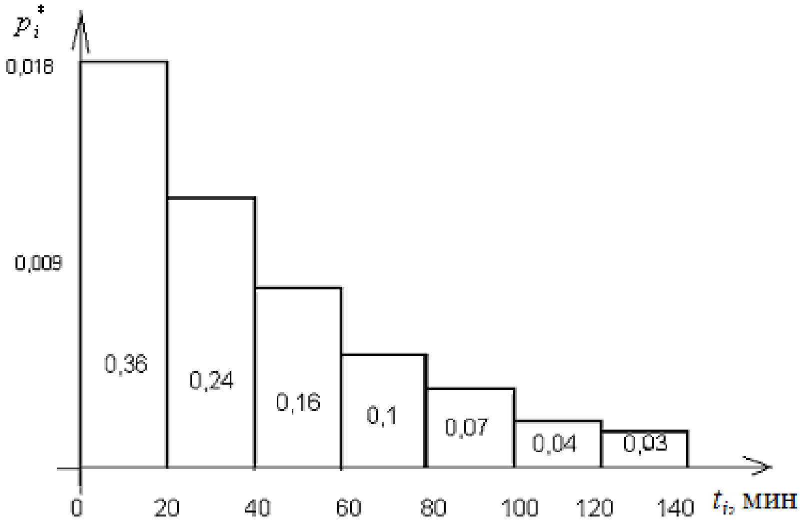
1. По данной таблице абсолютных частот построим таблицу относительных частот и соответствующую ей гистограмму:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
p_i*	0,36	0,24	0,16	0,10	0,07	0,04	0,03

.

$$\sum p_i^* = 0,36 + 0,24 + 0,16 + 0,1 + 0,07 + 0,04 + 0,03$$

Гистограмма относительных частот имеет вид:



Высоты прямоугольников гистограммы равны:

$\Delta_1 = \dfrac{0,36}{20} = 0,018$; $\Delta_2 = \dfrac{0,24}{20} = 0,012$; $\Delta_3 = \dfrac{0,16}{20} = 0,008$; $\Delta_4 = \dfrac{0,10}{20} = 0,005$; $\Delta_5 = \dfrac{0,07}{20} = 0,0035$;

$\Delta_6 = \dfrac{0,04}{20} = 0,002$; $\Delta_7 = \dfrac{0,03}{20} = 0,0015$.

2. По внешнему виду гистограммы выдвигаем гипотезу, что случайная величина T (время ремонта) подчиняется показательному закону

$$f(t) = \lambda e^{-\lambda t},$$

который зависит только от одного параметра λ (длительность операции в единицу времени).

Параметр $\lambda = \dfrac{1}{m_i}$, где m_i – математическое ожидание (среднее время ремонта) случайной величины T.

Следовательно, для выравнивания статистического распределения с помощью кривой показательного распределения найдем статистическую оценку параметра m_i:

$$m_i \approx \bar{t} = 10 \cdot 0{,}36 + 30 \cdot 0{,}24 + 50 \cdot 0{,}16 + 70 \cdot 0{,}1 + 90 \cdot 0{,}07 + 110 \cdot 0{,}04 + 130 \cdot 0{,}03 = 40$$

(числа 10, 30, 50, 70, 90, 110, 130 – это середины интервалов).

Тогда параметр $\lambda = \dfrac{1}{40}$.

3. Запишем теоретический закон распределения в виде функции плотности вероятности с учетом значения $\lambda = \dfrac{1}{40}$:

$$f(t)=\frac{1}{40}e^{-\frac{1}{40}t}.$$

По формуле вероятности попадания случайной величины (распределенной по показательному закону) на заданный интервал (α, β)

$$P(\alpha < X < \beta) = e^{-\lambda\alpha} - e^{-\lambda\beta}$$

найдем теоретические вероятности p_i, попадания случайной величины Т в каждый из семи интервалов и сравним их с соответствующими статистическими частотами p_i^*:

$$p_1 = P(0 < T < 20) = e^{-\frac{0}{40}} - e^{-\frac{20}{40}} = e^0 - e^{-0,5} \approx 1 - 0,6 = 0,4;$$

$$p_2 = P(20 < T < 40) = e^{-\frac{20}{40}} - e^{-\frac{40}{40}} = e^{-0,5} - e^{-1} \approx 0,6 - 0,37 = 0,23;$$

$$p_3 = P(40 < T < 60) = e^{-\frac{40}{40}} - e^{-\frac{60}{40}} = e^{-1} - e^{-1,5} \approx 0,37 - 0,22 = 0,15;$$

$$p_5 = P(60 < T < 80) = e^{-\frac{60}{40}} - e^{-\frac{80}{40}} = e^{-1,5} - e^{-2} \approx 0,22 - 0,14 = 0,08;$$

$$p_5 = P(80 < T < 100) = e^{-\frac{80}{40}} - e^{-\frac{100}{40}} = e^{-2} - e^{-2,5} \approx 0,14 - 0,08 = 0,06;$$

$$p_5 = P(100 < T < 120) = e^{-\frac{100}{40}} - e^{-\frac{120}{40}} = e^{-2,5} - e^{-3} \approx 0,08 - 0,05 = 0,03;$$

$$p_7 = P(120 < T < 140) = e^{-\frac{120}{40}} - e^{-\frac{140}{40}} = e^{-3} - e^{-3,5} \approx 0,05 - 0,03 = 0,02.$$

Для удобства сравнения теоретических вероятностей p_i с частотами p_i^* запишем полученные вероятности p_i в третью строку таблицы:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
p_i^*	0,36	0,24	0,16	0,10	0,07	0,04	0,03
p_i	0,40	0,23	0,15	0,08	0,06	0,03	0,02

Замечаем, что расхождение между опытными частотами p_i^* и теоретическими вероятностями p_i незначительны. Следовательно, вполне допустима гипотеза о показательном законе распределения изучаемой случайной величины Т.

4. Построим на одном графике с гистограммой выравнивающую ее кривую распределения f(t). Для этого вычислим значения

$$f(t)=\frac{1}{40}e^{-\frac{1}{40}t}$$

например, на правых концах интервалов:

$$f(20) = \frac{1}{40}e^{-\frac{1}{40}\cdot 20} = \frac{1}{40}e^{-0,5} \approx 0,015; \quad f(40) = \frac{1}{40}e^{-\frac{1}{40}\cdot 40} = \frac{1}{40}e^{-1} \approx 0,009;$$

$$f(60) = \frac{1}{40}e^{-\frac{1}{40}\cdot 60} = \frac{1}{40}e^{-1,5} \approx 0,006; \quad f(80) = \frac{1}{40}e^{-\frac{1}{40}\cdot 80} = \frac{1}{40}e^{-2} \approx 0,004;$$

$$f(100) = \frac{1}{40}e^{-\frac{1}{40}\cdot 100} = \frac{1}{40}e^{-2,5} \approx 0,002; \quad f(120) = \frac{1}{40}e^{-\frac{1}{40}\cdot 120} = \frac{1}{40}e^{-3} \approx 0,0013;$$

$$f(140) = \frac{1}{40}e^{-\frac{1}{40}\cdot 140} = \frac{1}{40}e^{-3,5} \approx 0,0008.$$

Построим график полученной кривой распределения f(t), в той же системе координат, что и гистограмма относительных частот.

Из рисунка видно, что теоретическая кривая f(t) сохраняет в основном существенные особенности статистического распределения.

2.8 Лабораторная работа 11 (ЛР-11)

Тема: Построение регрессии в среде MathCAD, Excel

2.8.1 Цель работы: ознакомить студентов с возможностью построения регрессии в среде MathCAD, Excel

2.8.2 Задачи работы:

1. Постановка задачи
2. Реализация на основе опытных данных

2.8.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.8.4 Описание (ход) работы:

Задача. Для прогноза возможного объема экспорта на основе ВНП построить и исследовать парную линейную регрессионную модель $\tilde{y} = a + bx + \varepsilon$ зависимости объема экспорта (y, усл. ед.) от ВНП (x, усл. ед.) Использовать построенную модель для прогноза при x_p=2500.

Требуется:

1) ввести данные;

2) построить корреляционное поле зависимости экспорта y от ВНП x ;

3) установить тесноту и вид связи между указанными показателями, т.е. рассчитать ковариацию и корреляцию и проанализировать их;

4) найти точечные и интервальные оценки для коэффициентов регрессии a и b;

5) оценить коэффициент детерминации и провести анализ общего качества уравнения регрессии;

6) указать стандартную ошибку регрессии;

7) оценить статистическую значимость коэффициентов регрессии a и b при уровне значимости $\alpha = 0{,}05$, при необходимости получить новое уравнение регрессии со значимыми коэффициентами;

8) выяснить, выполняются ли условия теоремы Гаусса-Маркова; для этого оценить разброс точек на графике остатков, построить гистограмму остатков, проанализировать их числовые характеристики;

9) дать точечный и интервальный прогнозы объема экспорта по заданному значению ВНП.

Результаты вычислений и анализа оформить в виде отчета (форма отчета прилагается ниже).

Порядок выполнения:

1) В ячейку А1 введите название ВНП, в ячейку В1 – название Экспорт. В ячейки А2, А3, …, А21 введите данные первого столбца выбранного варианта задания, в ячейки В2, В3, …, В21 – данные второго столбца выбранного варианта.

Присвойте листу 1 название «Исходные данные».

2) На листе «Исходные данные» выполните следующие действия:

– на панели инструментов активизируется кнопка Мастер диаграмм (шаг 1 из 4), в одноименном диалоговом окне (рисунок 3.17) среди стандартных типов выбирается Точечная и верхний левый вид диаграммы и нажимается кнопка Далее>;

– открывается диалоговое окно Мастер диаграмм (шаг 2 из 4), в котором во вкладке Диапазон данных в поле Диапазон вводится ссылка на диапазон ячеек A2:B21; нажимается кнопка Далее>;
– открывается диалоговое окно Мастер диаграмм (шаг 3 из 4), в котором во вкладке Заголовки в поле Ось X(категорий) вводится название «ВНП», в поле Ось Y(значений) – название «Экспорт»; во вкладке Легенда снимается флажок Добавить легенду и нажимается кнопка Далее>;
– открывается диалоговое окно Мастер диаграмм (шаг 4 из 4) в поле имеющемся устанавливается флажок и нажимается кнопка Готово.

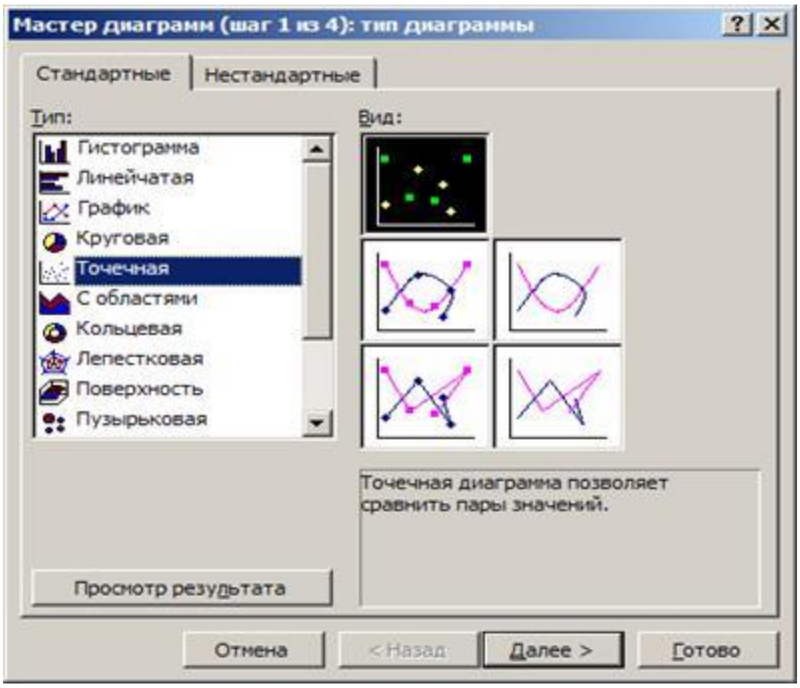


Рис. 1. Диалоговое окно «Мастер диаграмм (шаг 1 из 4)»

3) В меню Сервис выберите дополнение Анализ данных, в предложенных инструментах анализа выделите Ковариация, нажмите кнопку ОК. Установите значения параметров в появившемся диалоговом окне (рисунок 2) следующим образом:

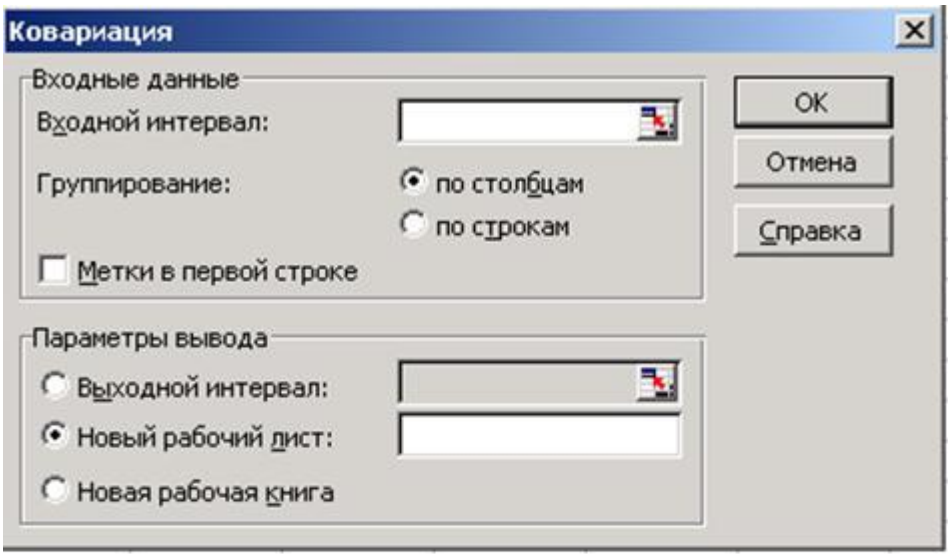


Рис. 2. Диалоговое окно «Ковариация»

· Входной интервал – введите ссылки на ячейки A1:B21 (курсор установите в поле «Входной интервал», указатель мыши поместите в ячейку A1, удерживая нажатой левую клавишу, протяните указатель мыши до ячейки B21);
· Группирование – флажок по столбцам устанавливается автоматически;
· Метки в первой строке – установите флажок щелчком левой кнопки мышки;
· Параметры вывода – установите флажок на Новый рабочий лист, поставив курсор в поле напротив, введите название «Ковариация».
Нажмите ОК.
Вернитесь на лист «Исходные данные». В меню Сервис выберите опцию Анализ данных и выделите Корреляция. Установите в диалоговом окне (рисунок 3) следующим образом значения параметров:

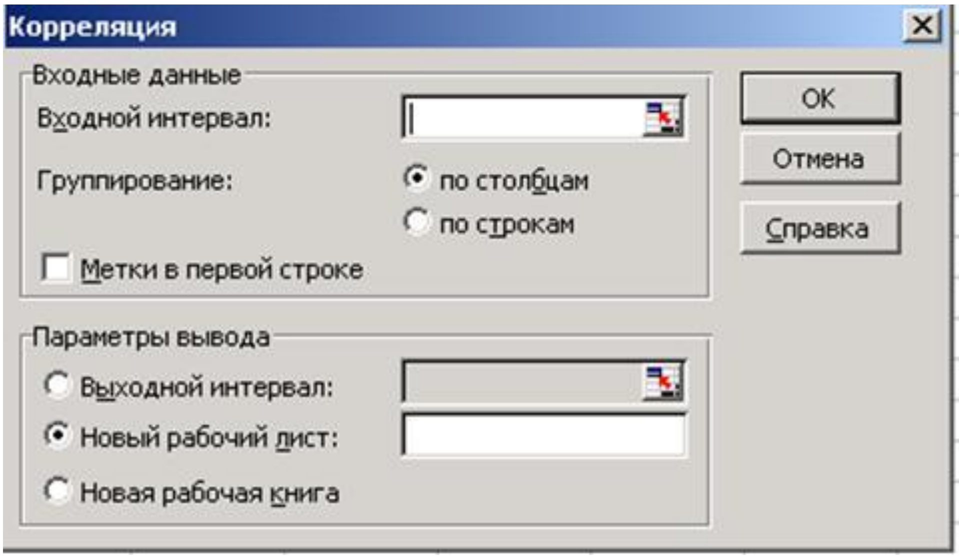


Рис. 3. Диалоговое окно «Корреляция»

· Входной интервал – введите ссылки на ячейки, содержащие исходные данные A1:B21 (курсор установите в поле «Входной интервал», указатель мыши поместите в ячейку A1, удерживая нажатой левую клавишу, протяните указатель мыши до ячейки B21);
· Группирование – установите флажок по столбцам;
· Метки в первой строке – установите флажок;
· Параметры вывода – установите флажок на Новый рабочий лист, введите название «Корреляция».
Нажмите ОК.
Значение линейного коэффициента корреляции находится на листе «Корреляция» в ячейке B3.
4) Вернитесь на лист «Исходные данные». В меню Сервис выберите дополнение Анализ данных укажите Регрессия. Нажмите кнопку ОК. Установите в диалоговом окне (рисунок 3.20) следующим образом значения параметров:
· Входной интервал Y – введите ссылки на ячейки B1:B21;
· Входной интервал X – введите ссылки на ячейки A1:A21;
· Метки – установите флажок;
· Уровень надежности – установите флажок;
· Константа ноль – не активизируйте;
· Параметры вывода – установите флажок на Новый рабочий лист и в поле напротив введите имя «Регрессия»;
· Остатки – установите флажок;

· Стандартизованные остатки – оставьте пустым;
· График остатков – установите флажок;
· График подбора – установите флажок;
· График нормальной вероятности – оставьте пустым.
Нажмите ОК.

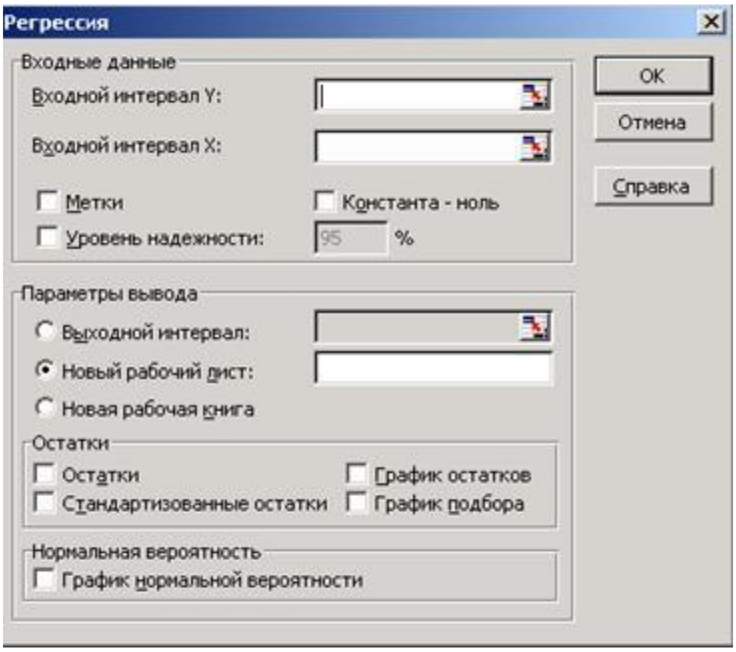


Рис. 4. Диалоговое окно «Регрессия»

Расположите диаграммы рядом (на поле диаграммы нажмите левую кнопку мыши, затем поместите курсор на белое поле и при нажатой левой кнопке передвигайте диаграмму вниз) и растяните их (на поле диаграммы нажмите левую кнопку мыши, нижнюю линию границы диаграммы при нажатой левой клавише протяните вниз).
Точечные оценки коэффициентов регрессии a и b находятся на листе «Регрессия» в ячейках B17 и B18 соответственно. Нижняя и верхняя границы доверительного интервала вычислены на листе «Регрессия» в ячейках F17 и G17 для коэффициента a и в ячейках F18 и G18 для коэффициента b.

5) Значение коэффициента детерминации R^2 находится на листе «Регрессия» в ячейке B5. Наблюдаемое значение F-критерия Фишера $F_{набл}$
находится на листе «Регрессия» в ячейке E12.
Вычислите критическое значение $F_{кр}$в свободной ячейке E15 следующим образом:
– нажмите на $\mathbf{f_x}$ (вставка функций);
– в поле Категория окна Мастер функций выберите статистические, из предложенных ниже функций выделите FРАСПОБР и нажмите ОК.
Откроется окно Аргументы функций. Заполните поля так:
· Вероятность – наберите значение 0,05;
· Степени свободы 1 – установите курсор в поле и выделите ячейку B12 столбца df таблицы «Дисперсионный анализ»;
· Степени свободы 2 – установите курсор в поле и выделите ячейку B13 столбца df таблицы «Дисперсионный анализ».
Нажмите ОК.
6) Значение стандартной ошибки регрессии s находятся на листе «Регрессия» в ячейке B7.

7) Наблюдаемые значения t -статистики $t_{набл}$ коэффициентов регрессии a и b находятся на листе «Регрессия» в ячейках D17 и D18 соответственно.
Вычислите критическое значение $t_{кр}$ в свободной ячейке D19 следующим образом:
– нажмите на $\mathbf{f_x}$ (вставка функций);
– в поле «Категория» окна Мастер функций выберите статистические, из предложенных ниже функций выделите СТЬЮДРАСПОБР и нажмите ОК. Откроется окно «Аргументы функций». Заполните поля:
· Вероятность – наберите значение 0,05;
· Степени свободы – введите 20–1–1, где 20 – число наблюдений, 1 – число факторов (x) в уравнении регрессии, 1 – число свободных членов (a) в уравнении регрессии.
Нажмите ОК.
8) На листе регрессия в меню Сервис выберите Анализ данных, укажите Гистограмма. Нажмите кнопку ОК. Значения параметров в появившемся диалоговом окне (рисунок 3.21) установите следующим образом:
· Входной интервал – введите ссылки на ячейки C24:C44 (столбец Остатки с названием);
· Интервал карманов – не заполняйте;
· Метки – установите флажок;
· Выходной диапазон – введите ссылку на новый рабочий лист «Гистограмма»;
· Парето – оставьте пустым;
· Интегральный процент – оставьте пустым;
· Вывод графика – установите флажок.
Нажмите ОК. Растяните диаграмму вниз.

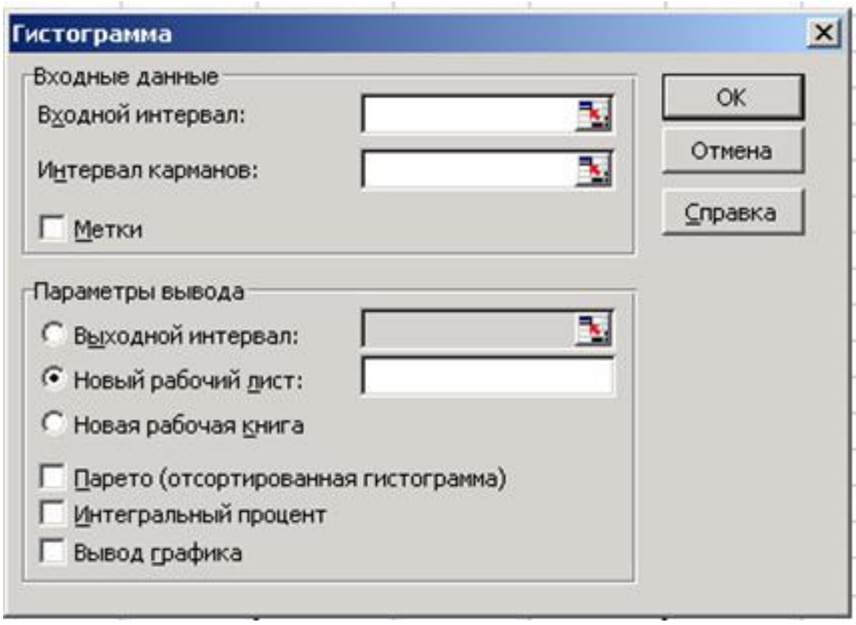


Рис. 5. Диалоговое окно «Гистограмма»

Вернитесь на лист «Регрессия». Выберите в опциях меню Сервис → Анализ данных → Описательная статистика, нажмите ОК. Значения параметров в диалоговом окне (рисунок 6) установите следующим образом:

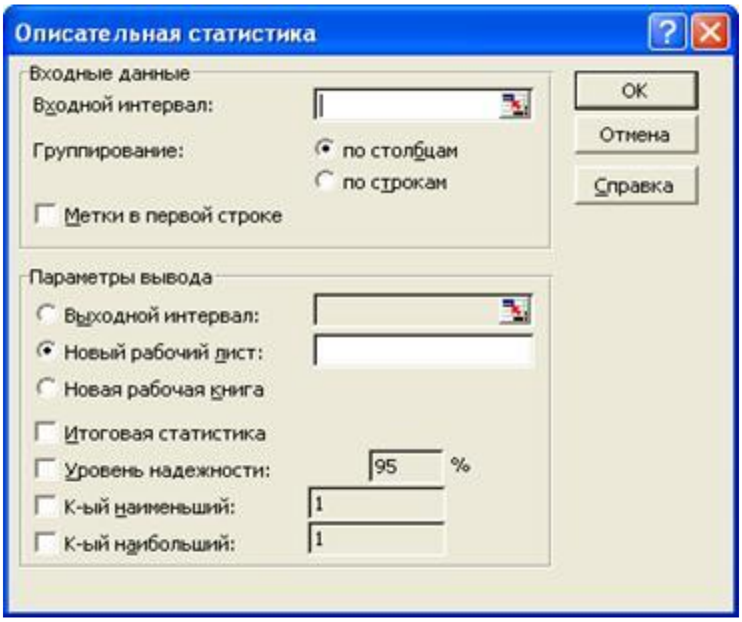


Рис. 6. Диалоговое окно «Описательная статистика»

· Входной интервал – введите ссылки на ячейки C24:C44 (столбец Остатки с названием);
· Группирование – установите флажок по столбцам;
· Метки – установите флажок в первой строке;
· Выходной диапазон – установите флажок на Новый рабочий лист и в поле напротив введите название «Статистика остатков»;
· установите флажки Итоговая статистика, уровень надежности (95%).
Нажмите ОК.
9) Вернитесь на лист «Регрессия» и в пустой ячейке E22 листа введите формулу
= B17+B18*2500 – точечный прогноз.

На листе «Регрессия» в пустой ячейке E23 вычислите значение m_p по формуле

$$m_p = s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

, где s – стандартная ошибка регрессии.
В пустых ячейках E24 и F24 введите формулы
=E22–D19*E23 – левый конец интервала прогноза;
=E22+D19*E23 – правый

2.9 Лабораторная работа 12 (ЛР-12)

Тема: Показатели стохастической зависимости

2.9.1 Цель работы: выработать навыки нахождения коэффициента корреляции, детерминации, определения значимости выборочных характеристик

2.9.2 Задачи работы:

1. Коэффициент детерминации
2. Коэффициент корреляции, его свойства, значимость.

2.9.3 Перечень приборов, материалов, используемых в лабораторной работе:
спецификой дисциплины не предусмотрены

2.9.4 Описание (ход) работы:

Имеются следующие данные разных стран об индексе розничных цен на продукты питания (x) и об индексе промышленного производства (y).

	Индекс розничных цен на продукты питания (x)	**Индекс промышленного производства (y)**
1	100	70
2	105	79
3	108	85
4	113	84
5	118	85
6	118	85
7	110	96
8	115	99
9	119	100
10	118	98
11	120	99
12	124	102
13	129	105
14	132	112

1. Для характеристики зависимости y от x рассчитать параметры следующих функций:
А) линейной;
Б) степенной;
В) равносторонней гиперболы.
2. Для каждой модели рассчитать показатели: тесноты связи и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость параметров регрессии и корреляции.
4. Выполнить прогноз значения индекса промышленного производства y при прогнозном значении индекса розничных цен на продукты питания x=138.

Решение:

1. Для расчёта параметров линейной регрессии

$$\widehat{y_x} = a + b * x$$

Решаем систему нормальных уравнений относительно a и b:

$$\begin{cases} n * a + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum y * x \end{cases}$$

Построим таблицу расчётных данных, как показано в таблице 1.

Таблица 1 Расчетные данные для оценки линейной регрессии

№ п/п	**x**	**y**	**xy**	**x^2**	**y^2**	$\widehat{y_x}$	$\left\lvert\frac{y - \widehat{y_x}}{y}\right\rvert$
1	100	70	7000	10000	4900	74,26340	0,060906
2	105	79	8295	11025	6241	79,92527	0,011712
3	108	85	9180	11664	7225	83,32238	0,019737
4	113	84	9492	12769	7056	88,98425	0,059336
5	118	85	10030	13924	7225	94,64611	0,113484
6	118	85	10030	13924	7225	94,64611	0,113484
7	110	96	10560	12100	9216	85,58713	0,108467
8	115	99	11385	13225	9801	91,24900	0,078293
9	119	100	11900	14161	10000	95,77849	0,042215
10	118	98	11564	13924	9604	94,64611	0,034223
11	120	99	11880	14400	9801	96,91086	0,021102

12	124	102	12648	15376	10404	101,4404	0,005487
13	129	105	13545	16641	11025	107,1022	0,020021
14	132	112	14784	17424	12544	110,4993	0,013399
Итого:	1629	1299	152293	190557	122267	1299,001	0,701866
Среднее значение:	116,3571	92,78571	10878,07	13611,21	8733,357	x	x
σ	8,4988	11,1431	x	x	x	x	x
σ^2	72,23	124,17	x	x	x	x	x

Среднее значение определим по формуле:

$$\bar{x} = \frac{\sum x}{n}$$

Среднее квадратическое отклонение рассчитаем по формуле: $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$
и занесём полученный результат в таблицу 1.

Возведя в квадрат полученное значение получим дисперсию: $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n}$

Параметры уравнения можно определить также и по формулам:

$$b = \frac{\frac{\sum y * x}{n} - \bar{y} * \bar{x}}{\sigma_x^2} = \frac{10878{,}07 - 116{,}3571 * 92{,}78571}{72{,}23} = 1{,}1324$$

$$a = \bar{y} - b * \bar{x} = 92{,}78571 - 1{,}1324 * 116{,}3571 = -38{,}9739$$

Таким образом, уравнение регрессии:

$$\hat{y}_x = -38{,}97 + 1{,}13 * x$$

Следовательно, с увеличением индекса розничных цен на продукты питания на 1, индекс промышленного производства увеличивается в среднем на 1,13.

Рассчитаем линейный коэффициент парной корреляции:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = 1{,}1324 * \frac{8{,}4988}{11{,}1431} = 0{,}86366$$

Связь прямая, достаточно тесная.

Определим коэффициент детерминации:

$$r_{xy}^2 = 0{,}86366^2 = 0{,}7459$$

Вариация результата на 74,59% объясняется вариацией фактора x.

Подставляя в уравнение регрессии фактические значения x, определим теоретические (расчётные) значения $\hat{y}_x$. Так как $\sum y = \sum \hat{y}_x$, следовательно, параметры уравнения определены правильно.

Рассчитаем среднюю ошибку аппроксимации – среднее отклонение расчётных значений от фактических:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y - \hat{y}_x}{y}\right| * 100\% = \frac{1}{14} * 0{,}701866 * 100\% = 5{,}01\%$$

В среднем расчётные значения отклоняются от фактических на 5,01%.

Оценку качества уравнения регрессии проведём с помощью F-теста.

F-тест состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи. Для этого выполняется сравнение фактического $F_{факт}$ и критического (табличного) $F_{табл}$ значений F-критерия Фишера.

$F_{факт}$ определяется по формуле:

$F_{факт} = \frac{r^2}{1-r^2} * \frac{(n-m-1)}{m}$, где n – число единиц совокупности; m – число параметров при переменных x.

$$F_{факт} = \frac{0{,}7459}{1-0{,}7459} * (14-2) = 35{,}2255 \quad F_{табл} = 4{,}75 < F_{факт}$$

Таким образом, H_0 – гипотеза о случайной природе оцениваемых характеристик отклоняется и признаётся их статистическая значимость и надёжность.

Полученные оценки уравнения регрессии позволяют использовать его для прогноза.

Если прогнозное значение индекса розничных цен на продукты питания x = 138, тогда прогнозное значение индекса промышленного производства составит:

$$\widehat{y_x} = -38{,}97 + 1{,}13 * 138 = 116{,}97$$

2. Нелинейная регрессия.

Степенная регрессия имеет вид: $y = a * x^b$

Для определения параметров производят логарифмирование степенной функции:

$$\lg y = \lg a + b * \lg x$$

Для определения параметров логарифмической функции строят систему нормальных уравнений по способу наименьших квадратов:

$$\begin{cases} n * \lg a + b \sum \lg x = \sum \lg y \\ \lg a \sum \lg x + b \sum (\lg x)^2 = \sum \lg y * \lg x \end{cases}$$

Построим таблицу расчётных данных, как показано в таблице 2.

Решая систему нормальных уравнений, определяем параметры логарифмической функции.

$$b = \frac{\frac{\sum \lg y * \lg x}{n} - \lg\bar{y} * \lg\bar{x}}{\sigma^2_{\lg x}} = \frac{4{,}056855 - 1{,}964217 * 2{,}064624}{0{,}001021} = 1{,}4549$$

$$\lg a = \lg \bar{y} - b * \lg \bar{x} = 1{,}964217 - 1{,}4549 * 2{,}064624 = -1{,}0397$$

Таблица 2 Расчетные данные для оценки степенной регрессии

№п/п	x	y	lg x	lg y	lg x*lg y	$(\lg x)^2$	$(\lg y)^2$
1	100	70	2,000000	1,845098	3,690196	4,000000	3,404387
2	105	79	2,021189	1,897627	3,835464	4,085206	3,600989
3	108	85	2,033424	1,929419	3,923326	4,134812	3,722657
4	113	84	2,053078	1,924279	3,950696	4,215131	3,702851
5	118	85	2,071882	1,929419	3,997528	4,292695	3,722657
6	118	85	2,071882	1,929419	3,997528	4,292695	3,722657
7	110	96	2,041393	1,982271	4,046594	4,167284	3,929399
8	115	99	2,060698	1,995635	4,112401	4,246476	3,982560
9	119	100	2,075547	2,000000	4,151094	4,307895	4,000000
10	118	98	2,071882	1,991226	4,125585	4,292695	3,964981
11	120	99	2,079181	1,995635	4,149287	4,322995	3,982560
12	124	102	2,093422	2,008600	4,204847	4,382414	4,034475
13	129	105	2,110590	2,021189	4,265901	4,454589	4,085206
14	132	112	2,120574	2,049218	4,345518	4,496834	4,199295
Итого	1629	1299	28,90474	27,49904	56,79597	59,69172	54,05467
Среднее значение	116,3571	92,78571	2,064624	1,964217	4,056855	4,263694	3,861048
σ	8,4988	11,1431	0,031945	0,053853	x	x	x
σ^2	72,23	124,17	0,001021	0,0029	x	x	x

№п/п	x	y	$\widehat{y_x}$	$(y-\widehat{y_x})^2$	$\left\lvert\frac{y-\widehat{y_x}}{y}\right\rvert$	$(y-\bar{y})^2$
1	100	70	74,16448	17,34292	0,059493	519,1886
2	105	79	79,62057	0,385112	0,007855	190,0458
3	108	85	82,95180	4,195133	0,024096	60,61728
4	113	84	88,59768	21,13866	0,054734	77,1887
5	118	85	94,35840	87,57961	0,110099	60,61728

6	118	85	94,35840	87,57961	0,110099	60,61728
7	110	96	85,19619	116,7223	0,11254	10,33166
8	115	99	90,88834	65,79901	0,081936	38,6174
9	119	100	95,52408	20,03384	0,044759	52,04598
10	118	98	94,35840	13,26127	0,037159	27,18882
11	120	99	96,69423	5,316563	0,023291	38,6174
12	124	102	101,4191	0,337467	0,005695	84,90314
13	129	105	107,4232	5,872099	0,023078	149,1889
14	132	112	111,0772	0,85163	0,00824	369,1889
Итого	1629	1299	1296,632	446,4152	0,703074	1738,357
Среднее значение	116,3571	92,78571	x	x	x	x
σ	8,4988	11,1431	x	x	x	x
σ^2	72,23	124,17	x	x	x	x

Получим линейное уравнение: $\lg y = -1{,}0397 + 1{,}4549 \lg x$

Выполнив его потенцирование, получим:

$\hat{y} = 10^{-1{,}0397} * x^{1{,}4549} = 0{,}0912641 * x^{1{,}4549}$

Подставляя в данное уравнение фактические значения x, получаем теоретические значения результата $\hat{y}$. По ним рассчитаем показатели: тесноты связи – индекс корреляции и среднюю ошибку аппроксимации.

$$\rho_{xy} = \sqrt{1 - \frac{\sum(y - \hat{y}_x)^2}{\sum(y - \bar{y})^2}} = \sqrt{1 - \frac{446{,}4152}{1738{,}357}} = 0{,}8621$$

Связь достаточно тесная.

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y - \hat{y}}{y}\right| * 100\% = \frac{1}{14} * 0{,}703074 * 100\% = 5{,}02\%$$

В среднем расчётные значения отклоняются от фактических на 5,02%.

$$F_{факт} = \frac{\rho^2}{1 - \rho^2} * (n - 2) = \frac{0{,}7432}{1 - 0{,}7432} * (14 - 2) = 34{,}73$$

$$F_{табл} = 4{,}75 < F_{факт}$$

Таким образом, H_0 – гипотеза о случайной природе оцениваемых характеристик отклоняется и признаётся их статистическая значимость и надёжность.

Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение индекса розничных цен на продукты питания x = 138, тогда прогнозное значение индекса промышленного производства составит:

$$\hat{y} = 0{,}0912641 * 138^{1{,}4549} = 118{,}47$$

Уравнение равносторонней гиперболы

$y_x = a + b * \frac{1}{x}$

Для определения параметров этого уравнения используется система нормальных уравнений:

$$\begin{cases} n * a + b\sum\frac{1}{x} = \sum y \\ a\sum\frac{1}{x} + b\sum(\frac{1}{x})^2 = \sum y * \frac{1}{x} \end{cases}$$

Произведем замену переменных $\frac{1}{x} = z$

и получим следующую систему нормальных уравнений:

$$\begin{cases} n*a+b\sum z=\sum y \\ a\sum z+b\sum z^2=\sum y*z \end{cases}$$

Решая систему нормальных уравнений, определяем параметры гиперболы.

Составим таблицу расчётных данных, как показано в таблице 3.

Таблица 3 Расчетные данные для оценки гиперболической зависимости

№п/п	**x**	**y**	**z**	**yz**	z^2	y^2	**№п/п**
1	100	70	0,010000000	0,700000	0,0001000	4900	**1**
2	105	79	0,009523810	0,752381	0,0000907	6241	**2**
3	108	85	0,009259259	0,787037	0,0000857	7225	**3**
4	113	84	0,008849558	0,743363	0,0000783	7056	**4**
5	118	85	0,008474576	0,720339	0,0000718	7225	**5**
6	118	85	0,008474576	0,720339	0,0000718	7225	**6**
7	110	96	0,009090909	0,872727	0,0000826	9216	**7**
8	115	99	0,008695652	0,860870	0,0000756	9801	**8**
9	119	100	0,008403361	0,840336	0,0000706	10000	**9**
10	118	98	0,008474576	0,830508	0,0000718	9604	**10**
11	120	99	0,008333333	0,825000	0,0000694	9801	**11**
12	124	102	0,008064516	0,822581	0,0000650	10404	**12**
13	129	105	0,007751938	0,813953	0,0000601	11025	**13**
14	132	112	0,007575758	0,848485	0,0000574	12544	**14**
Итого:	1629	1299	0,120971823	11,13792	0,0010510	122267	**Итого:**
Среднее значение:	116,3571	92,78571	0,008640844	0,795566	0,0000751	8733,357	**Среднее значение:**
σ	8,4988	11,1431	0,000640820	x	x	x	σ
σ^2	72,23	124,17	0,000000411	x	x	x	σ^2
№п/п	**x**	**y**	$\hat{y}_x$	$\left\lvert\frac{y-\hat{y}_x}{y}\right\rvert$	$(y-\hat{y}_x)^2$	$(y-\bar{y})^2$	
1	100	70	72,3262	0,033231	5,411206	519,1886	
2	105	79	79,49405	0,006254	0,244083	190,0458	
3	108	85	83,47619	0,017927	2,322012	60,61728	
4	113	84	89,64321	0,067181	31,84585	77,1887	
5	118	85	95,28761	0,121031	105,8349	60,61728	
6	118	85	95,28761	0,121031	105,8349	60,61728	
7	110	96	86,01027	0,10406	99,79465	10,33166	
8	115	99	91,95987	0,071112	49,56344	38,6174	
9	119	100	96,35957	0,036404	13,25272	52,04598	
10	118	98	95,28761	0,027677	7,357059	27,18882	
11	120	99	97,41367	0,016024	2,516453	38,6174	
12	124	102	101,46	0,005294	0,291565	84,90314	
13	129	105	106,1651	0,011096	1,357478	149,1889	
14	132	112	108,8171	0,028419	10,1311	369,1889	
Итого:	1629	1299	1298,988	0,666742	435,7575	1738,357	
Среднее значение:	116,3571	92,78571	x	x	x	x	
σ	8,4988	11,1431	x	x	x	x	
σ^2	72,23	124,17	x	x	x	x	

Значения параметров регрессии a и b составили:

$$b=\frac{\frac{\sum y*z}{n}-\bar{y}*\bar{z}}{\sigma_z^2}=\frac{0{,}795566-92{,}78571*0{,}008640844}{0{,}00000041065}=-15052{,}4$$

$$a=\bar{y}-b*\bar{z}=92{,}78571+15052{,}4*0{,}008640844=222{,}851$$

Получено уравнение:

$$\widehat{y_x} = 222{,}851 - 15052{,}4 * \frac{1}{x}$$

Индекс корреляции:

$$\rho = \sqrt{1 - \frac{\sum(y - \widehat{y_x})^2}{\sum(y - \bar{y})^2}} = \sqrt{1 - \frac{435{,}7575}{1738{,}357}} = 0{,}8656$$

Связь достаточно тесная.

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y - \hat{y}}{y}\right| * 100\% = \frac{1}{14} * 0{,}666742 * 100\% = 4{,}76\%$$

В среднем расчётные значения отклоняются от фактических на 4,76%.

$$F_{факт} = \frac{\rho^2}{1 - \rho^2} * (n - 2) = \frac{0{,}7493}{1 - 0{,}7493} * (14 - 2) = 35{,}86$$

$$F_{табл} = 4{,}75 < F_{факт}$$

Таким образом, H_0 – гипотеза о случайной природе оцениваемых характеристик отклоняется и признаётся их статистическая значимость и надёжность.

Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение индекса розничных цен на продукты питания x = 138, тогда прогнозное значение индекса промышленного производства составит:

$$\widehat{y_x} = 222{,}851 - 15052{,}4 * \frac{1}{138} = 113{,}78$$

По уравнению равносторонней гиперболы получена наибольшая оценка тесноты связи по сравнению с линейной и степенной регрессиями. Средняя ошибка аппроксимации остаётся на допустимом уровне.

2.10 Лабораторная работа 13 (ЛР-13).

Тема: Аппроксимация функций в среде MathCAD.

2.10.1 Цель работы: ознакомиться с возможностями аппроксимации опытных данных на примере метода наименьших квадратов; научиться решать задачу численной аппроксимации при работе с таблично заданными функциями.

2.10.2 Задачи работы:

1. Постановка задачи численной аппроксимации.
2. Аппроксимация таблично заданных функций методом наименьших квадратов.

2.10.3 Перечень приборов, материалов, используемых в лабораторной работе:

спецификой дисциплины не предусмотрены

2.10.4 Описание (ход) работы:

1. Производится *n* наблюдений (x_1, y_1), ..., (x_n, y_n) переменных *x* и *y*. Предполагая, что между *x* и *y* существует зависимость вида $y = f(x)$, найти значения параметров *a и b*, наилучшим образом согласованные с экспериментальными данными.

Согласно методу наименьших квадратов параметры функции следует выбирать так, чтобы сумма квадратов невязок была наименьшей.

$$S = \sum_{i=1}^{n}\left(f(x_i) - y_i\right)^2$$

2. Если $f(x)$ — линейная функция, т.е. $y = ax + b$, то $S = \sum_{i=1}^{n}\left(ax_i + b - y_i\right)^2$, а неизвестные параметры a и b определяются из системы нормальных уравнений:

$$\begin{cases} \sum\limits_{i=1}^{n} x_i^2 \cdot a + \sum\limits_{i=1}^{n} x_i \cdot b = \sum\limits_{i=1}^{n} x_i \cdot y_i, \\ \sum\limits_{i-1}^{n} x_i \cdot a + n \cdot b = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

3. Если $f(x)$ — квадратичная функция, т.е. $y = ax^2 + bx + c$, то $S = \sum\limits_{i=1}^{n} \left(ax_i^2 + bx_i + c - y_i\right)^2$, а неизвестные параметры *a*, *b*, *c* определяются из системы нормальных уравнений:

$$\begin{cases} \sum\limits_{i=1}^{n} x_i^4 \cdot a + \sum\limits_{i=1}^{n} x_i^3 \cdot b + \sum\limits_{i=1}^{n} x_i^2 \cdot c = \sum\limits_{i=1}^{n} x_i^2 \cdot y_i, \\ \sum\limits_{i-1}^{n} x_i^3 \cdot a + \sum\limits_{i=1}^{n} x_i^2 \cdot b + \sum\limits_{i=1}^{n} x_i \cdot c = \sum\limits_{i=1}^{n} x_i \cdot y_i, \\ \sum\limits_{i=1}^{n} x_i^2 \cdot a + \sum\limits_{i=1}^{n} x_i \cdot b + n \cdot c = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

4. Линейная регрессия

Для расчета линейной регрессии в MathCAD необходимо воспользоваться следующими операторами:

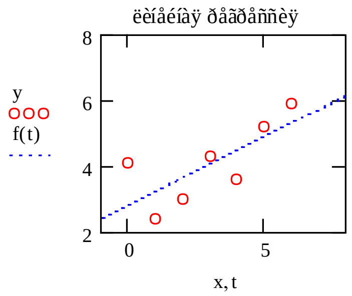
- line (x,y) - вектор из двух элементов (b,a) коэффициентов линейной регрессии $y = b + ax$;
- intercept (x, y) - коэффициент b линейной регрессии;
- slope (x, y) - коэффициент a линейной регрессии;
- x - вектор действительных данных аргумента;
- y - вектор действительных данных значений того же размера.

Пример 1. Линейная регрессия

$x := (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6)^T \qquad y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^T$

$\mathrm{line}(x,y) = \begin{pmatrix} 2.829 \\ 0.414 \end{pmatrix} \qquad \mathrm{intercept}(x,y) = 2.829 \qquad \mathrm{slope}(x,y) = 0.414$

$f(t) := \mathrm{line}(x,y)_0 + \mathrm{line}(x,y)_1 \cdot t$



Пример 2. Имеются следующие данные о расходах на рекламу x (тыс. усл. ед) и сбыте продукции y (тыс. ед):

x_i	**1**	**2**	**3**	**4**	**5**
y_i	**1,6**	**4,0**	**7,4**	**12,0**	**18,0**

Методом наименьших квадратов найти эмпирические формулы прямой $y = ax + b$ и многочлена второй степени $y = ax^2 + bx + c$, аппроксимирующие функцию, заданную таблично. Найти значение многочленов первой и второй степеней в заданных точках, абсолютную погрешность в них и среднеквадратическую погрешность.

Выяснить, какая зависимость предпочтительнее. Построить графики. Для этой же функции построить многочлен первой степени, пользуясь встроенными функциями системы MathCAD для линейной регрессии. Графически сравнить полученные результаты.

Решение:

i	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i\,y_i$	$x_i^2\,y_i$
1	1	1,6	1	1	1	1,6	1,6
2	2	4,0	4	8	16	8,0	16,0
3	3	7,4	9	27	81	22,2	66,6
4	4	12,0	16	64	256	48,0	196,0
5	5	18,0	25	125	625	90,0	450,0
$\sum$	**15**	**43,0**	**55**	**225**	**979**	**169,8**	**680,2**

Система нормальных уравнений имеет вид: $\begin{cases} 55a+15b=169{,}8, \\ 15a+5b=43. \end{cases}$ Её решения a=4,08, b=-3,64. Таким образом, линейная зависимость имеет вид: $y=4{,}08x-3{,}64$.

Система нормальных уравнений имеет вид: $\begin{cases} 979a+225b+55c=680{,}2, \\ 225a+55b+15c=168{,}8, \\ 55a+15b+5c=49{,}0. \end{cases}$ Её решения a=0,3, b=0,48, c=5,06. Таким образом, искомая квадратичная зависимость имеет вид: $y=0{,}3x^2+0{,}48x+5{,}06$.

$\sum\limits_{i=1}^{n}(ax_i+b-y_i)$ - абсолютная погрешность для линейной зависимости

$\sum\limits_{i=1}^{n}(ax_i+b-y_i)^2$ - среднеквадратическая погрешность для линейной зависимости

$[4{,}08\cdot1-3{,}64-1{,}6]^2=[-1{,}16]^2=1{,}3456$ $\quad [4{,}08\cdot2-3{,}64-4]^2=[0{,}52]^2=0{,}2704$

$[4{,}08\cdot3-3{,}64-7{,}4]^2=[1{,}2]^2=1{,}44$ $\quad [4{,}08\cdot4-3{,}64-12]^2=[0{,}68]^2=0{,}4624$

$$[4{,}08\cdot5-3{,}64-18]^2=[-1{,}24]^2=1{,}5376$$

$$\sum_{i=1}^{5}(4{,}08x_i-3{,}64-y_i)=-1{,}16+0{,}52+1{,}2+0{,}68-1{,}24=0$$

$$\sum_{i=1}^{5}(4{,}08x_i-3{,}64-y_i)^2=1{,}3456+0{,}2704+1{,}44+0{,}4624+1{,}5376=5{,}056$$

$\sum\limits_{i=1}^{n}(ax_i^2+bx_i+c-y_i)$ - абсолютная погрешность для квадратичной зависимости

$\sum\limits_{i=1}^{n}(ax_i^2+bx_i+c-y_i)^2$ - среднеквадратическая погрешность для квадратичной зависимости

$[0{,}3\cdot1^2+0{,}48\cdot1+5{,}06-1{,}6]^2=[4{,}24]^2=17{,}9776$ $\quad [0{,}3\cdot2^2+0{,}48\cdot2+5{,}06-4]^2=[3{,}22]^2=10{,}3684$

$[0{,}3\cdot3^2+0{,}48\cdot3+5{,}06-7{,}4]^2=[1{,}8]^2=3{,}24$ $\quad [0{,}3\cdot4^2+0{,}48\cdot4+5{,}06-12]^2=[-0{,}22]^2=0{,}0484$

$$[0{,}3\cdot5^2+0{,}48\cdot5+5{,}06-18]^2=[-3{,}04]^2=9{,}2416$$

$$\sum_{i=1}^{5}(0{,}3x_i^2+0{,}48x_i+5{,}06-y_i)=4{,}24+3{,}22+1{,}8-0{,}22-3{,}04=6$$

$$\sum_{i=1}^{5}(0{,}3x_i^2+0{,}48x_i+5{,}06-y_i)^2=17{,}9776+10{,}3684+3{,}24+0{,}0484+9{,}2416=40{,}876$$

Как видно $S_{лин}<S_{кв}$, следовательно, линейная зависимость предпочтительнее.

$x:=(1\ \ 2\ \ 3\ \ 4\ \ 5)^T$

$y:=(1.6\ \ 4\ \ 7.4\ \ 12\ \ 18)^T$

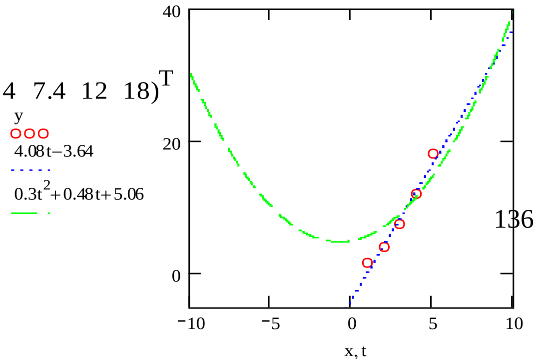


Рис. Изображение в ДСК опытных точек, линейной и квадратичной зависимостей

$x := (1 \;\; 2 \;\; 3 \;\; 4 \;\; 5)^T \qquad y := (1.6 \;\; 4 \;\; 7.4 \;\; 12 \;\; 18)^T$

$$\mathrm{line}(x, y) = \begin{pmatrix} -3.64 \\ 4.08 \end{pmatrix} \qquad \text{или} \qquad \mathrm{intercept}(x, y) = -3.64 \quad \mathrm{slope}(x, y) = 4.08$$

$$f(t) := \mathrm{line}(x, y)_0 + \mathrm{line}(x, y)_1 \cdot t$$

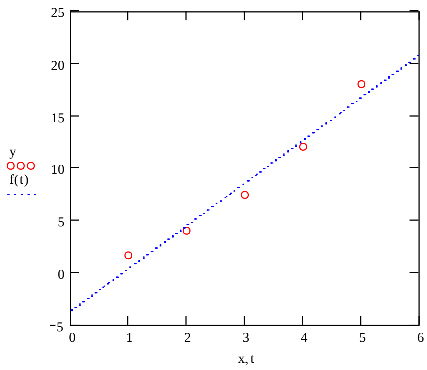


Рис. Изображение в ДСК опытных точек и графика линейной регрессии

2.11 Лабораторная работа 14 (ЛР-14).

Тема: Обработка опытов.

2.11.1 Цель работы: ознакомиться с возможностями математической статистики по обработке опытных данных; научиться решать задачу первичной обработки данных и статистического оценивания, научиться решать задачу оценки неизвестных параметров закона распределения.

2.11.2 Задачи работы:

1. Особенности обработки ограниченного числа опытов.
2. Оценки характеристик выборки.
3. Оценки неизвестных параметров закона распределения.
4. Оценки неизвестных параметров закона распределения.
5. Оценка вероятности по частоте.

2.11.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.11.4 Описание (ход) работы:

№1. В парке учебных машин в течении 30 дней проводились наблюдения над числом выездов в день. Получили следующие результаты:

3	1	3	1	4	5	2	4	0	7	0	2	2	0	2
5	4	3	7	1	4	2	2	1	7	2	1	0	3	6

Проведите первичную обработку выборочных данных и оцените параметры генеральной совокупности (точечно и используя метод доверительных интервалов).

А) К какому группировочному типу (дискретному или непрерывному) относится признак? ___________________________

Б) Составьте ранжированный ряд:

____

В) Перечислите, какие значения принимает признак:________________________

Объем выборки_____ максимальное значение_____ минимальное значение____

Г) Составьте статистический ряд распределения данного признака,

x_i	n_i

⇐ подпишите заголовки столбцов

$x_i \cdot n_i$

$\sum =$

$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 \cdot n_i$

$\sum =$

Д) Постройте полигон частот

Вычислите и подпишите, что означают найденные вами значения

Е) $R=$ ____________________________________

$$\overline{x} = \frac{\sum n_i x_i}{n} =$$ ________________________

$Mo=$

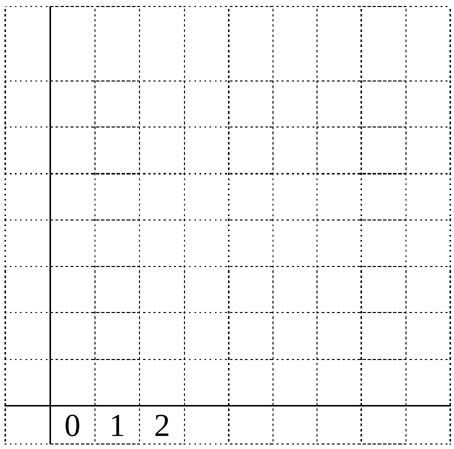
Ж) $$\mathrm{D} = \frac{\sum (x_i - \overline{x})^2 n_i}{n} =$$

____

____

$\sigma = \sqrt{\mathrm{D}} =$

____



Вывод:____

____

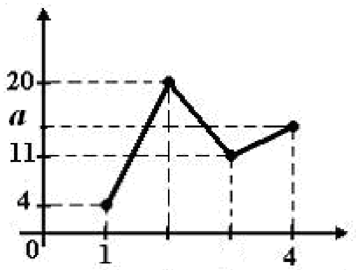
____

№2. Из генеральной совокупности извлечена выборка объема n=50, на рисунке изображен полигон частот изучаемого количественного признака.

Подпишите название осей: ____

____

Постройте соответствующий ряд распределения.



Чему равно значение a? ________________

Чему равно значение R? ________________

№3. В результате взвешивания 50 клубней (гр.), получили следующие результаты:

93	209	135	216	206	80	197	134	145	183
251	53	142	120	147	159	111	185	200	191
96	206	138	213	209	77	200	131	148	180
253	50	145	117	180	156	113	181	203	188
152	150	110	118	140	81	120	135	220	144

Проведите выравнивание статистического ряда, методом моментов оцените параметры предполагаемого распределения.

А) К какому группировочному типу (дискретному или непрерывному) относится признак? ____________________________________

Б) Объем выборки____ максимальное значение____ минимальное значение___

В) На какое количество интервалов лучше разбить данные _________________.
Почему? ____

Рассчитайте ширину классового интервала

____

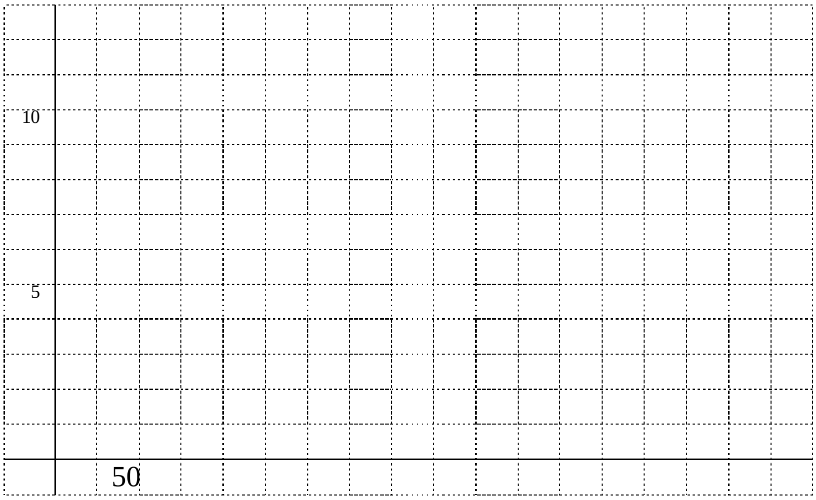
В) Составьте статистическую таблицу распределения данного признака, дайте название колонкам.

$x_i - x_{i+1}$	n_i

$\frac{x_{i+1} + x_i}{2}$	$\frac{x_{i+1} + x_i}{2} n_i$

$\sum =$

Г) Постройте гистограмму. Проведите выравнивание статистического ряда.



Вычислите и подпишите, что означают найденные вами значения

Д) R= ____

$$\overline{x} = \frac{\sum n_i x_i}{n} =$$ ____

*Mo=*____

Вывод:

____

____

2.12 Лабораторная работа 15 (ЛР-15)

Тема: Моделирование случайного процесса.

2.12.1 Цель работы: усвоить основные принципы и методы моделирования случайных процессов

2.12.2 Задачи работы:

1. Выборка случайных данных.
2. Генерация псевдослучайного процесса.
3. Огибающая и фаза нормального случайного процесса.

2.12.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.12.4 Описание (ход) работы:

Описание задачи

У вас есть 10 монет. Вы хотите увеличить свой капитал в два раза. Играя с маклером, вы делаете ставку и бросаете монету. Если выпадает “орел”, маклер выдает вам сумму вашей ставки, в противном случае вы ему отдаете эту сумму. Ставка может быть от 1 до 10 монет. Удвоение начального капитала или банкротство приводит к незамедлительному прекращению этого сеанса игры и расчету. Игра может продолжиться по вашему усмотрению.

Цель моделирования

-Варьируя ставки в игре, выяснить, какая тактика чаще приводит к результату (положительному или отрицательному).

- Предупредить излишне азартных игроков о невозможности обогащения за счет азартных игр и о степени связанного с ними риска.

Анализ объекта

Моделируем игру. Игра – это процесс, в котором участвуют три объекта: игрок, маклер и “случай”, который здесь представлен монетой. Игрок обладает начальным капиталом, который в дальнейшем увеличивается или уменьшается. Другой параметр игрока – величина ставки. Маклер определяет проигрыш или выигрыш игрока, выплачивая выигрыш. Параметром монеты является результат броска – “орел” или “решка”. Случай характеризуется угадываем того, на какую сторону ляжет монета, и имеет два значения -– “угадал” (1) или “не угадал” (0).

При этом вероятность выпадения той или иной стороны 0,5.

Разработка модели

Информационная модель

Объект “игрок” имеет *управляемые параметры*: Ставка – количество поставленных на бросок монет; Наличность – количество монет у игрока после очередного броска.

имеет *неуправляемые параметры (константы)*:Начальный капитал: 10 монет.

Действия над объектом: Выбор ставки; Вычисление наличности; Продолжение игры.

Объект “маклер” имеет *управляемые параметры*: Бросок – определение выигрыша или проигрыша после очередного броска; Выигрыш – прекращение игры после увеличения капитала игрока вдвое или больше; Проигрыш – прекращение игры после банкротства игрока.

имеет *неуправляемые параметры (константы)*: Нет.

Действия над объектом: Выплата проигранного; Прекращение игры по банкротству.

Объект “монета (случай)” имеет *управляемые параметры*: Положение при приземлении (“орел” или “решка”).

имеет *неуправляемые параметры (константы)*: Количество возможных вариантов падения: 2; Вероятность угадывания результата 1/2.

Действия над объектом: Подбрасывание монеты; Определение результата падения.

Логико-математическая модель

Имитировать угадывание результата броска монеты можно с помощью функции СЛЧИС(). Эта функция выдает случайные числа x в диапазоне $0<=x<1$. Поскольку вероятность выпадения той или иной стороны S, то будем считать: если СЛЧИС<0,5 , то результат “угадал” (1), в противном случае – “не угадал” (0). В модели используется логическая функция ЕСЛИ (Условие; Знач Истина; Знач Ложь).

Функция угадывания результата при броске имеет вид:

Бросок = ЕСЛИ(СЛЧИС<0,5;1;0),

здесь “1” на выходе функции означает, что игрок угадал, а “0” – не угадал.

Функция изменения наличности игрока:

Наличность = ЕСЛИ(Бросок=1; Наличность + Ставка; Наличность – Ставка).

Функция определения выигрыша:

Выигрыш = ЕСЛИ(Наличность< 2*Начкапитал; “-“; “банк”),

здесь выдается сообщение “банк” при увеличении начального капитала вдвое или больше, что является условием прекращения игры.

Функция определения проигрыша:

Проигрыш = ЕСЛИ(Наличность>0; “-“; “банкрот”),

здесь выдается сообщение “банкрот” по окончании наличности, что также является условием прекращения игры.

Компьютерная модель

Для моделирования выберем среду электронной таблицы. В этой среде информационная и математическая модели объединяются в таблицу, которая содержит две области: исходные данные – константы и управляемые параметры; расчетные данные (результаты).

Задание

Ввести в таблицу исходные данные, а в расчетную часть – следующие формулы в ячейки:
A8: =ЕСЛИ(СЛЧИС()<0,5; 0; 1)
B8: =ЕСЛИ(A8=1; $B$5+$D$5; $B$5-$D$5)
C8: =ЕСЛИ(B8<2*$B$5; “-”; “банк”)
D8: =ЕСЛИ(B8>0; “-”; “банкрот”)
B9: =ЕСЛИ(A9=1; B8+$D$5; B8-$D$5)
При записи формул вставки стандартных функций ЕСЛИ(…) и СЛЧИС(…).

Компьютерный эксперимент

План моделирования

1. Проверить правильность ввода формул.
2. Провести расчеты.
3. Результаты расчетов проанализировать.

Технология моделирования

1. Ввести в таблицу контрольные данные и расчетные формулы в первую строку.
2. Скопировать формулы в нижестоящие ячейки в обозримом пространстве экрана (20 попыток)

Появление в столбце “Выигрыш” сообщения “банк” означает удвоение наличности. Появление в столбце “Проигрыш” сообщения “банкрот” - нулевую наличность. И то и другое приводит к концу сеанса игры. Следующий сеанс игры проводится в тех же ячейках путем обновления данных 1-го столбца, для чего столбец следует выделить и выбрать команду **Заменить** по столбцам меню **Правка**.

Приведите статистику выигрышей и проигрышей, повторяя игру многократно, и сделайте выводы.

2.13 Лабораторная работа 16 (ЛР-16)

Тема: Характеристики случайной функции.

2.13.1 Цель работы: усвоить вероятностный смысл и назначение характеристик случайной функции; изучить методы оценки характеристик по опытным данным. проанализировать влияние линейных операторов на характеристики входного воздействия; изучить методы определения характеристик преобразованных случайных функций по характеристикам исходных.

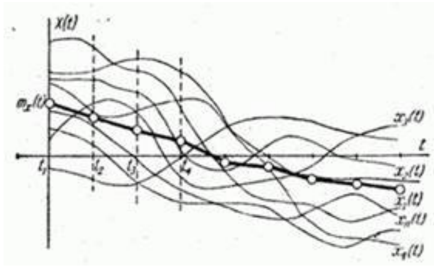
2.13.2 Задачи работы:

1. Свойства и вероятностный смысл характеристик.
2. Определение характеристик случайной функции из опыта.
3. Линейные преобразования случайных функций.
4. Методы определения характеристик преобразованных случайных функций по характеристикам исходных случайных функций.

2.13.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.13.4 Описание (ход) работы:

Пусть над случайной функцией $X(t)$ произведено n независимых опытов (наблюдений) и в результате получено n реализаций случайной функции .



Требуется найти оценки для характеристик случайной функции: ее математического ожидания $m_x(t)$, дисперсии $D_x(t)$ и корреляционной функции $K_x(t,t')$.

Для этого рассмотрим ряд сечений случайной функции для моментов времени $t_1, t_2, \ldots, t_m$ и зарегистрируем значения, принятые функцией $X(t)$ в эти моменты времени. Каждому из моментов $t_1, t_2, \ldots, t_m$ будет соответствовать n значений случайной функции.

Значения $t_1, t_2, \ldots, t_m$ обычно задаются равноотстоящими точками: величина интервала между соседними значениями выбирается в зависимости от вида экспериментальных кривых так, чтобы по выбранным точкам можно было восстановить основной ход кривых. Зарегистрированные значения $X(t)$ заносятся в таблицу, каждая строка которой соответствует определенной реализации, а число столбцов равно числу опорных значений аргумента.

Таблица 1

t / $X(t)$	t_1	t_2	...	t_k	...	t_l	...	t_m
$x_1(t)$	$x_1(t_1)$	$x_1(t_2)$	...	$x_1(t_k)$	...	$x_1(t_l)$	...	$x_1(t_m)$
$x_1(t)$	$x_2(t_1)$	$x_2(t_2)$	...	$x_2(t_k)$	...	$x_2(t_l)$	...	$x_2(t_m)$
...	...	...	...	...	...	...	...	...
$x_i(t)$	$x_i(t_1)$	$x_i(t_2)$	...	$x_i(t_k)$	...	$x_i(t_l)$	...	$x_i(t_m)$
...	...	...	...	...	...	...	...	...
$x_n(t)$	$x_n(t_1)$	$x_n(t_2)$	...	$x_n(t_k)$	...	$x_n(t_l)$	...	$x_n(t_m)$

В таблице 1 в i-й строке помещены значения случайной функции, наблюденной в i-й реализации (i-м опыте) при значениях аргумента $t_1, t_2, \ldots, t_m$. Символом $x_i(t_k)$ обозначено значение, соответствующее i-й реализации в момент t_k.Полученный материал представляет собой не что иное, как результаты n опытов над системой m случайных величин

$X(t_1), X(t_2), \ldots, X(t_m)$,

Прежде всего находятся оценки для математических ожиданий по формуле

$$\widetilde{m}_k(t_k) = \frac{\sum_{i=1}^{n} x_i(t_k)}{n},$$

затем - для дисперсий

$$\widetilde{D}_x(t_k) = \frac{\sum_{i=1}^{n}\left[x_i(t_k) - \widetilde{m}_x(t_k)\right]^2}{n-1}$$

и, наконец, для корреляционных моментов

$$\widetilde{K}_x(t_k, t_l) = \frac{\sum_{i=1}^{n}\left[x_i(t_k) - \widetilde{m}_x(t_k)\right]\left[x_i(t_l) - \widetilde{m}_x(t_l)\right]}{n-1}.$$

После того, как эти характеристики вычислены, можно, пользуясь рядом значений $\widetilde{m}_x(t_1), \widetilde{m}_x(t_2), \ldots, \widetilde{m}_x(t_m)$, построить зависимость $\widetilde{m}_x(t)$. Аналогично строится зависимость $\widetilde{D}_x(t)$. Функция двух аргументов $\widetilde{K}_x(t,t')$ воспроизводится по ее значениям в прямоугольной сетке точек. В случае надобности все эти функции аппроксимируются какими-либо аналитическими выражениями.

Пусть на вход линейной системы с оператором L воздействует случайная функция $X(t)$, причем известны ее характеристики: математическое ожидание $m_x(t)$ и корреляционная функция $K_x(t,t')$. Реакция системы представляет собой случайную функцию $Y(t) = L\{X(t)\}$. Требуется найти характеристики случайной функции $Y(t)$ на выходе системы: $m_y(t)$ и $K_y(t,t')$.

Случайная функция $Y(t)$ связана с $X(t)$ линейным однородным оператором интегрирования:

$Y(t) = \int\limits_0^t X(\tau)d\tau$, тогда $m_y(t) = \int\limits_0^t m_x(\tau)d\tau$

т. е. математическое ожидание интеграла от случайной функции равно интегралу от ее математического ожидания

$$K_y(t,t') = \int\limits_0^t\int\limits_0^{t'} K_x(\tau,\tau')d\tau d\tau'.$$

Таким образом, для того чтобы **найти корреляционную функцию интеграла от случайной функции, нужно дважды проинтегрировать корреляционную функцию исходной случайной функции: сначала по одному аргументу, затем - по другому.**

Для знакомства с методами определения характеристик преобразованных случайных функций по характеристикам исходных случайных функций, выполним несколько теоретических упражнений.

1. Если задана СФ $X(t)=Ue^{3t}\cos 2t$, где U–случайная величина и M(U)=5, D(U)=1, то математическое ожидание интеграла исходной функции имеет вид…

1) $5e^{3t}\cos 2t$; 2) $e^{3t}\cos 2t$; 3) $15e^{3t}\cos 2t$; 4) 1;

2. Задано математическое ожидание СФ $m_x(t)=t^2+7$, тогда математическое ожидание СФ $Y(t)= t X'(t)+ t^3$ имеет вид…

1) t^2+7; 2) $2t+7$; 3) $t^2(t+2)$; 4) 7;

3. Задана корреляционная функция $k_x(\tau) = 2e^{-0,5\tau^2}$ стационарной случайной функции $X(t)$. Дисперсия этой функции после воздействия оператора дифференцирования равна….

Дисперсия этой функции после воздействия оператора интегрирования равна….

2.14 Лабораторная работа 17 (ЛР-17)

Тема: Динамические системы.

2.14.1 Цель работы: изучить основные понятия, связанные с теорией динамических систем; рассмотреть классические модели динамических систем

2.14.2 Задачи работы:

1. Постановка задачи.
2. Фазовый портрет динамической системы.
3. Классические динамические системы: автоколебания, аттрактор, брюсселятор

2.14.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

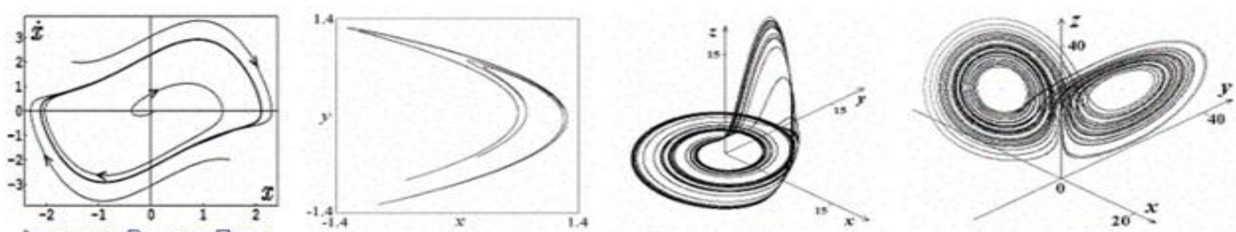
2.14.4 Описание (ход) работы:

Динамические системы можно классифицировать в зависимости от вида оператора отображения и структуры фазового пространства. Если оператор предусматривает исключительно линейные преобразования начального состояния, то он называется линейным. Линейный оператор обладает свойством суперпозиции: T[x(t)+y(t)]=Tx(t)+Ty(t). Если оператор нелинейный, то и соответствующая динамическая система называется нелинейной. Различают непрерывные и дискретные операторы и соответственно системы с непрерывным и дискретным временем.

Способы задания оператора Т также могут различаться. Оператор Т можно задать в виде дифференциального или интегрального преобразования, в виде матрицы или таблицы, в виде графика или функции и т.д.

В зависимости от того, какой ряд значений могут принимать фазовые координаты, определяющие состояние системы, различают непрерывные и дискретные фазовые пространства.

Поведение динамических систем изучают в «пространстве состояний». Точка в этом пространстве однозначно задает состояние системы. В простейшем случае, например, для маятника – это плоскость (координата, скорость). Притягивающие объекты в фазовом пространстве – аттракторы – определяют свойства установившегося с течением времени колебательного процесса в системе. Аттрактор (от английского to attract - притягивать) может иметь вид простой замкнутой кривой. Это предельный цикл, являющийся образом автоколебаний.



Аттрактор Ван-дер-Поля Аттрактор Хенона Аттрактор Рёсслера Аттрактор Лоренца

Благодаря посещению различных областей в трехмерном пространстве фазовая траектория может «запутаться», что приводит к возникновению хаотических режимов. Основной атрибут хаоса – наличие очень сильной зависимости режима от начальных условий. В результате даже очень малое различие в начальных состояниях системы со временем приведет к существенному различию в поведении.

Задание

1. Дана функция $f(z)=z^2+c$, показать, что это множество не симметрично оси ординат. (Подсказка: c=-1 принадлежит множеству, c=1 не принадлежит множеству, далее исследовать орбиту нуля при обоих отображениях).

2. Дана функция $f(z)=z^2+0,1+0,1i$ найти неподвижные точки для этой функции и выяснить их характер двумя способами: 1) аналитический способ (через производную); 2) компьютерный эксперимент.

3. (Компьютерный эксперимент). Используйте компьютер для получения изображений множества Мандельброта для $f(z)=z^3+c$. Покажите, что если |c|>2, то $z \to \infty$.

4. Построить фазовые портреты динамических систем, заданные следующими операторами

1. $$y' = \frac{x-4y}{2y-3x}.$$

2. $$y' = \frac{y-2x}{y}.$$

Построение фазовых портретов динамических систем

1. Порядок построения фазового портрета линейной динамической системы

Для построения фазового портрета динамической системы

$$\begin{cases} \dfrac{dx}{dt} = ax + by, \\ \dfrac{dy}{dt} = cx + dy \end{cases} \tag{1}$$

необходимо выполнить следующие действия.

1. Выписать матрицу коэффициентов системы (1)

   $$\mathrm{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

   найти ее след $\operatorname{tr}\mathrm{M}$ и определитель $\det \mathrm{M}$.

2. Используя рис. 1, определить тип особой точки.

3. Найти уравнения особых направлений $\frac{dx}{dt} = 0$ и $\frac{dy}{dt} = 0$.

   $$y = -\frac{a}{b}x,$$

   $$y = -\frac{c}{d}x.$$

4. Если особая точка является седлом или узлом, то найти асимптоты, используя подстановку $y = kx$.

5. Определить направление фазовых траекторий.

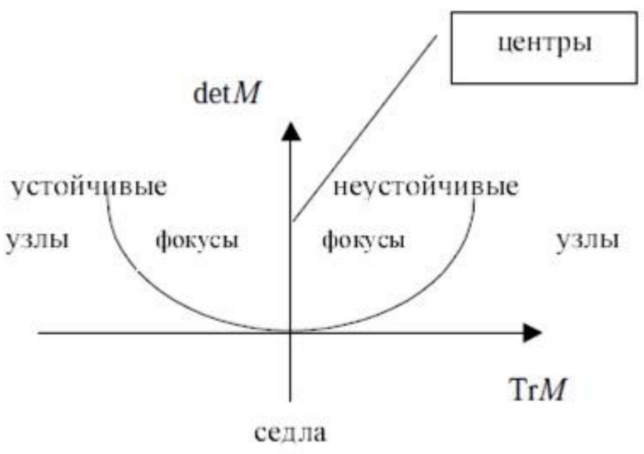


Рис. 1. Зависимость типа особой точки от определителя и следа матрицы коэффициентов динамической системы

2.15 Лабораторная работа 18 (ЛР-18)

Тема: Характеристики стационарной случайной функции. Стационарные случайные функции с эргодическим свойством.

2.15.1 Цель работы: усвоить вероятностный смысл, назначение и особенности характеристик стационарной случайной функции; изучить методы оценки характеристик по опытным данным. Проанализировать особенности процесса, обладающего эргодическим свойством, научиться определять характеристики таких процессов по одной реализации

2.15.2 Задачи работы:

1. Линейные преобразования стационарных случайных функций.
2. Методы определения характеристик преобразованных стационарных случайных функций по характеристикам исходных стационарных случайных функций.
3. Применение теории стационарных случайных процессов к решению задач, связанных с анализом и синтезом динамических систем.
4. Определение характеристик эргодической стационарной функции по одной реализации.

2.15.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.15.4 Описание (ход) работы:

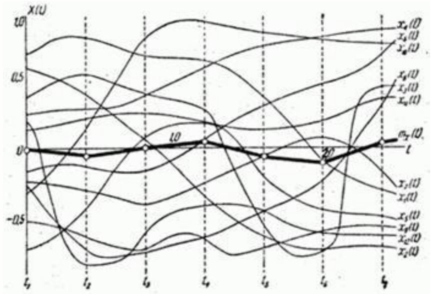
Случайная функция $X(t)$ называется стационарной, если все ее вероятностные характеристики не зависят от t (точнее, не меняются при любом сдвиге аргументов, от которых они зависят, по оси t). Так как изменение стационарной случайной функции должно протекать однородно по времени, то математическое ожидание должно быть постоянным.

Это требование не является существенным: мы знаем, что от случайной функции $X(t)$ всегда можно перейти к $\overset{\circ}{X}(t)$, для которой математическое ожидание тождественно равно нулю. Таким образом, если случайный процесс нестационарен только за счет переменного математического ожидания, это не мешает нам изучать его как стационарный процесс. Вообще, корреляционная функция стационарного случайного процесса должна зависеть не от положения t первого аргумента на оси абсцисс, а только от промежутка τ между первым и вторым аргументами.

Поэтому в дальнейшем мы под стационарной случайной функцией будем понимать такую случайную функцию, корреляционная функция которой зависит не от обоих своих аргументов t и t', а только от разности τ между ними.

В качестве примеров рассмотрим два образца приблизительно стационарных случайных процессов и построим их характеристики.

Пример 1. Случайная функция $X(t)$ задана совокупностью 12 реализаций (рис. 1). а) Найти ее характеристики $m_x(t)$, $K_x(t,t')$, $D_x(t)$ и нормированную корреляционную функцию $r_x(t,t')$. б) Приближенно рассматривая случайную функцию $X(t)$ как стационарную, найти ее характеристики.



Решение. Так как случайная функция $X(t)$ меняется сравнительно плавно, можно брать сечения не очень часто, например через 0,4 сек. Тогда случайная функция будет сведена к системе семи случайных величин, отвечающих сечениям $t = 0;\ 0,4;\ 0,8;\ 1,2;\ 1,6;\ 2,0;\ 2,4$. Намечая эти сечения на графике и снимая с графика значения случайной функции в этих сечениях, получим таблицу (табл. 1).

Таблица 1

t № реализации	0	0,4	0,8	1,2	1,6	2,0	2,4
1	0,64	0,74	0,62	0,59	0,35	-0,09	0,39
2	0,54	0,37	0,06	-0,32	-0,60	-0,69	-0,67
3	0,34	0,50	0,37	0,26	-0,52	-0,72	0,42
4	0,23	0,26	0,35	0,55	0,69	0,75	0,80
5	0,12	0,20	0,24	0,18	-0,20	-0,42	-0,46
6	-0,16	-0,12	-0,15	0,05	0,29	0,43	0,63
7	-0,22	-0,29	-0,38	-0,24	-0,06	0,07	-0,16
8	-0,26	-0,69	-0,70	-0,61	-0,43	-0,22	0,29
9	-0,50	-0,60	-0,68	-0,62	-0,68	-0,56	-0,54
10	-0,30	0,13	0,75	0,84	0,78	0,73	0,71
11	-0,69	-0,40	0,08	0,16	0,12	0,18	0,33
12	0,18	-0,79	-0,56	-0,39	-0,42	-0,58	-0,53

Таблицу рекомендуется заполнять по строчкам, передвигаясь все время вдоль одной реализации.

Далее находим оценки для характеристик случайных величин $X(0), X(0,4), \ldots, X(2,4)$. Суммируя значения по столбцам и деля сумму на число реализаций $n = 12$, найдем приближенно зависимость математического ожидания от времени:

t	0	0,4	0,8	1,2	1,6	2,0	2,4
$\tilde{m}_X(t)$	-0,007	-0,057	0,000	0,037	-0,057	-0,093	0,036

Далее находим оценки для элементов корреляционной матрицы: дисперсий и корреляционных моментов. Вычисления удобнее всего производить по следующей схеме. Для вычисления статистической дисперсии суммируются квадраты чисел, стоящих в соответствующем столбце; сумма делится на $n = 12$; из результата вычитается квадрат соответствующего математического ожидания. Для получения несмещенной оценки результат множится на поправку $\frac{n}{n-1} = \frac{12}{11}$. Аналогично оцениваются корреляционные моменты.

Для вычисления статистического момента, отвечающего двум заданным сечениям, перемножаются числа, стоящие в соответствующих столбцах; произведении складываются алгебраически; полученная сумма делится на $n = 12$; из результата вычитается произведение соответствующих математических ожиданий; для получения несмещенной оценки корреляционного момента результат множится на $\frac{n}{n-1}$.

t	0	0,4	0,8	1,2	1,6	2,0	2,4
$\tilde{D}_x(t)$	0,1632	0,2385	0,2356	0,2207	0,2407	0,2691	0,278

Извлекая из этих величин квадратные корни, найдем зависимость среднего квадратического отклонения $\tilde{\sigma}_x$ от времени:

t	0	0,4	0,8	1,2	1,6	2,0	2,4
$\tilde{\sigma}_x(t)$	0,404	0,488	0,485	0,470	0,491	0,519	0,536

Деля полученные значения на произведения соответствующих средних квадратических отклонений, получим таблицу значений нормированной корреляционной функции $\tilde{r}_x(t,t')$ (табл. 2).

Таблица 2

t / t'	0	0,4	0,8	1,2	1,6	2,0	2,4
0	1	0,700	0,405	0,241	-0,053	-0,306	-0,299
0,4		1	0,856	0,707	0,345	0,090	0,095
0,8			1	0,943	0,643	0,390	0,344
1,2				1	0,829	0,612	0,524
1,6					1	0,923	0,650
2,0						1	0,760
2,4							1

Если судить непосредственно по данным, полученным в результате обработки, то можно прийти к выводу, что случайная функция $X(t)$ стационарной не является: ее математическое ожидание не вполне постоянно; дисперсия также несколько меняется со временем; значения нормированной корреляционной функции вдоль параллелей главной диагонали также не вполне постоянны. Однако, принимая во внимание весьма ограниченное число обработанных реализаций ($n=12$) и в связи с этим наличие большого элемента случайности в полученных оценках, эти видимые отступления от стационарности вряд ли можно считать значимыми, тем более, что они не носят сколько-нибудь закономерного характера. Поэтому вполне целесообразной будет приближенная замена функции $X(t)$ стационарной. Для приведения функции к стационарной прежде всего осредним по времени оценки для математического ожидания:

$$\tilde{m}_x = \frac{\tilde{m}_x(0)+\tilde{m}_x(0,4)+\ldots+\tilde{m}_x(2,4)}{7} \approx -0,22.$$

Аналогичным образом осредним оценки для дисперсии:

$$\tilde{D}_x = \frac{\tilde{D}_x(0)+\tilde{D}_x(0,4)+\ldots+\tilde{D}_x(2,4)}{7} \approx 0,236.$$

Извлекая корень, найдем осредненную оценку с. к. о.:

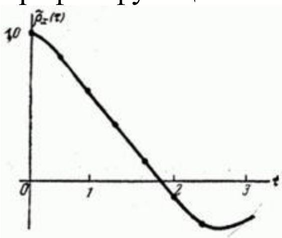
$$\tilde{\sigma}_x \approx 0,486.$$

Перейдем к построению нормированной корреляционной функции того стационарного процесса, которым можно заменить случайную функцию $X(t)$. Для стационарного процесса корреляционная функция (а значит, и нормированная корреляционная функция) зависит только от $\tau = t' - t$; следовательно, при постоянном τ корреляционная функция должна быть постоянной. В таблице 2 постоянному τ соответствуют: главная диагональ ($\tau = 0$) и параллели этой диагонали ($\tau = 0,4$; $\tau = 0,8$; $\tau = 1,2$ и т. д.). Осредняя оценки нормированной корреляционной функции вдоль этих параллелей главной диагонали, получим значения функции $\rho_x(\tau)$:

t	0	0,4	0,8	1,2	1,6	2,0	2,4

$\tilde{\rho}_x(\tau)$	1,00	0,84	0,60	0,38	0,13	-0,10	-0,30

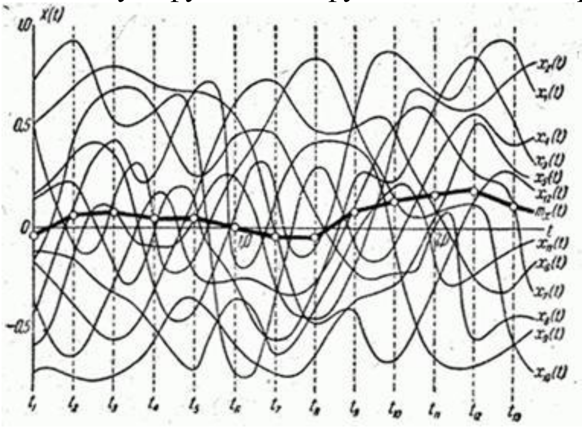
График функции $\tilde{\rho}_x(\tau)$ представлен на рис. 3



При рассмотрении рис. обращает на себя внимание наличие для некоторых τ отрицательных значений корреляционной функции. Это указывает на то, что в структуре случайной функции имеется некоторый элемент периодичности, в связи с чем на расстоянии по времени, равном примерно половине периода основных колебаний, наблюдается отрицательная корреляция между значениями случайной функции: положительным отклонениям от среднего в одном сечении соответствуют отрицательные отклонения через определенный промежуток времени, и наоборот.

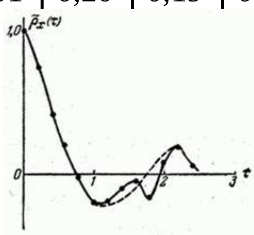
Такой характер корреляционной функции, с переходом на отрицательные значения, очень часто встречается на практике. Обычно в таких случаях по мере увеличения τ амплитуда колебаний корреляционной функции уменьшается и при дальнейшем увеличении τ корреляционная функция стремится к нулю.

Пример 2. Случайная функция $X(t)$ задана совокупностью 12 своих реализаций (рис. 4). Приближенно заменив функцию $X(t)$ стационарной, сравнить ее нормированную корреляционную функцию с функцией $\tilde{\rho}_x(\tau)$ предыдущего примера.



Решение. Так как случайная функция $X(t)$ отличается значительно менее плавным ходом по сравнению с функцией $X(t)$ предыдущего примера, промежуток между сечениями уже нельзя брать равным 0,4 сек, как в предыдущем примере, а следует взять по крайней мере вдвое меньше (например, 0,2 сек.). В результате обработки получаем оценку для нормированной корреляционной функции $\tilde{\rho}_x(\tau)$:

τ	0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	1,8	2,0	2,2	2,4
$\tilde{\rho}_x(\tau)$	1,00	0,73	0,41	0,22	-0,01	-0,20	-0,19	-0,10	-0,06	-0,15	0,08	0,19	0,05



Из сравнения графиков видно, что корреляционная функция, изображенная убывает значительно быстрее. Это и естественно, так как характер изменения функции $X(t)$ в примере 1 гораздо более плавный и постепенный, чем в примере 2; в связи с этим корреляция между значениями случайной функции в примере 1 убывает медленнее.

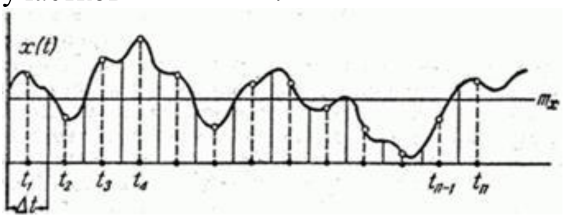
При рассмотрении бросаются в глаза незакономерные колебания ции $\tilde{\rho}_x(\tau)$ для больших значений τ. Так как при больших значениях τ точки графика получены осреднением сравнительно очень небольшого числа данных, их нельзя считать надежными. В подобных случаях имеет смысл сгладить корреляционную функцию.

Рассмотрим стационарную случайную функцию $X(t)$, обладающую эргодическим свойством, и предположим, что в нашем распоряжении имеется всего одна реализация этой случайной функции, но зато на достаточно большом участке времени T. Для эргодической стационарной случайной функции одна реализация достаточно большой продолжительности практически эквивалентна (в смысле объема сведений о случайной функции) множеству реализаций той же общей продолжительности; характеристики случайной функции могут быть приближенно определены не как средние по множеству наблюдений, а как средние по времени t. В частности, при достаточно большом T математическое ожидание m_x может быть приближенно вычислено по формуле $$m_x \approx \frac{1}{T}\int_0^T x(t)dt$$.

Аналогично может быть приближенно найдена корреляционная функция $k_x(\tau)$ при любом τ.

$$k_x(\tau) \approx \frac{1}{T-\tau}\int_0^{T-\tau} \overset{\circ}{x}(t)\,\overset{\circ}{x}(t+\tau)dt$$.

Вычислив интеграл для ряда значений τ, можно приближенно воспроизвести по точкам весь ход корреляционной функции. На практике обычно интегралы заменяют конечными суммами. Покажем, как это делается. Разобьем интервал записи случайной функции на n равных частей длиной Δt и обозначим середины полученных участков $t_1, t_2, \ldots, t_n$.



Предоставим интеграл как сумму интегралов по элементарным участкам Δt и на каждом из них вынесем функцию $x(t)$ из-под знака интеграла средним значением, соответствующим центру интервала $x(t_i)$. Получим приближенно: $$m_x = \frac{1}{T}\frac{T}{n}\sum_{i=1}^{n} x(t_i)$$, или

$$m_x = \frac{1}{n}\sum_{i=1}^{n} x(t_i)$$.

Аналогично можно вычислить корреляционную функцию для значений τ, равных $0, \Delta t, 2\Delta t, \ldots$. Придадим, например, величине τ значение $$\tau = m\Delta t = \frac{mT}{n}$$

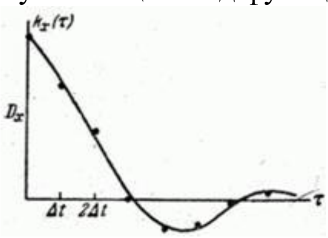
вычислим интеграл , деля интервал интегрирования $T-\tau=T-\frac{mT}{n}=\frac{n-m}{n}T$

на $n-m$ равных участков длиной Δt и вынося на каждом из них функцию $\overset{\circ}{x}(t)\,\overset{\circ}{x}(t+\tau)$ за знак интеграла средним значением. Получим:

$$k_x\left(\frac{mT}{n}\right)=\frac{n}{(n-m)T}\frac{T}{n}\sum_{i=1}^{n-m}\overset{\circ}{x}(t_i)\,\overset{\circ}{x}(t_{i+m})$$
,

или окончательно
$$k_x\left(\frac{mT}{n}\right)=\frac{1}{n-m}\sum_{i=1}^{n-m}\overset{\circ}{x}(t_i)\,\overset{\circ}{x}(t_{i+m})$$
.

Вычисление корреляционной функции производят для $m=0,1,2,\ldots$ последовательно. Вплоть до таких значений m, при которых корреляционная функция становится практически равной нулю или начинает совершать небольшие нерегулярные колебания около нуля. Общий ход функции $k_x(\tau)$ воспроизводится по отдельным точкам.



Для того чтобы математическое ние m_x и корреляционная функция $k_x(\tau)$ были определены с удовлетворительной точностью, нужно, чтобы число точек n было достаточно велико (порядка сотни, а в некоторых случаях даже нескольких сотен). Выбор длины элементарного участка Δt определяется характером изменения случайной функции. Если случайная функция изменяется сравнительно плавно, участок Δt можно выбирать большим, чем когда она совершает резкие и частые колебания.

Чем более высокочастотный состав имеют колебания, образующие случайную функцию, тем чаще нужно располагать опорные точки при обработке. Ориентировочно можно рекомендовать выбирать элементарный участок Δt так, чтобы на полный период самой высокочастотной гармоники в составе случайной функции приходилось порядка 5-10 опорных точек.

Пример. В условиях работы экспериментального механизма произведена запись отклонений. Отклонения регистрировались на участке времени 200 сек с интервалом 2 сек. Результаты приведены в таблице 1.

Таблица 1

t (сек)	отклонение $N(t)$	t (сек)	отклонение $N(t)$	t (сек)	отклонение $N(t)$	t (сек)	отклонение $N(t)$
0	1,0	50	1,0	100	1,2	150	0,8
2	1,3	52	1,1	102	1,4	152	0,6
4	1,1	54	1,5	104	0,8	154	0,9
6	0,7	56	1,0	106	0,9	156	1,2
8	0,7	58	0,8	108	1,0	158	1,3

10	1,1	60	1,1	110	0,8	160	0,9
12	1,3	62	1,1	112	0,8	162	1,3
14	0,8	64	1,2	114	1,4	164	1,5
16	0,8	66	1,0	116	1,6	166	1,2
18	0,4	68	0,8	118	1,7	168	1,4
20	0,3	70	0,8	120	1,3	170	1,4
22	0,3	72	1,2	122	1,6	172	0,8
24	0,6	74	0,7	124	0,8	174	0,8
26	0,3	76	0,7	126	1,2	176	1,3
28	0,5	78	1,1	128	0,6	178	1,0
30	0,5	80	1,2	130	1,0	180	0,7
32	0,7	82	1,0	132	0,3	182	1,1
34	0,8	84	0,6	134	0,8	184	0,9
36	0,6	86	0,9	136	0,7	186	0,9
38	1,0	88	0,8	138	0,9	188	1,1
40	0,5	90	0,8	140	1,3	190	1,2
42	1,0	92	0,9	142	1,5	192	1,3
44	0,9	94	0,9	144	1,1	194	1,3
46	1,4	96	0,6	146	0,7	196	1,6
48	1,4	98	0,4	148	1,0	198	1,5

Считая процесс изменения отклонений стационарным, определить приближенно математическое ожидание m_N, дисперсию D_N и нормированную корреляционную функцию $\rho_N(\tau)$. Аппроксимировать $\rho_N(\tau)$ какой-либо аналитической функцией, найти и построить спектральную плотность случайного процесса.

Решение. По формуле имеем: $m_N = \frac{\sum_{i=1}^{100} N(t_i)}{100} \approx 0,98$.

Центрируем случайную функцию.

Таблица 2

t (сек)	$\overset{\circ}{N}(t)$	t (сек)	$\overset{\circ}{N}(t)$	t (сек)	$\overset{\circ}{N}(t)$	t (сек)	$\overset{\circ}{N}(t)$
0	0,02	50	0,02	100	0,22	150	-0,18
2	0,32	52	0,12	102	0,42	152	-0,38
4	0,12	54	0,52	104	-0,18	154	-0,08
6	-0,28	56	0,02	106	-0,08	156	0,22
8	-0,28	58	-0,18	108	0,02	158	0,32
10	0,12	60	0,12	110	-0,18	160	-0,08
12	0,32	62	0,12	112	-0,18	162	0,32
14	-0,18	64	0,22	114	0,42	164	0,52
16	-0,18	66	0,02	116	0,62	166	0,22
18	-0,58	68	-0,18	118	0,72	168	0,42
20	-0,68	70	-0,18	120	0,32	170	0,42
22	-0,68	72	0,22	122	0,62	172	-0,18
24	-0,38	74	-0,28	124	-0,18	174	-0,18
26	-0,68	76	-0,28	126	0,22	176	0,32
28	-0,48	78	0,12	128	-0,38	178	0,02
30	-0,48	80	0,52	130	0,02	180	-0,28
32	-0,28	82	0,02	132	-0,38	182	0,12

34	-0,18	84	-0,38	134	-0,18	184	-0,08
36	-0,38	86	-0,08	136	-0,28	186	-0,08
38	0,02	88	-0,18	138	-0,08	188	0,12
40	-0,48	90	-0,18	140	0,32	190	0,22
42	0,02	92	-0,08	142	0,52	192	0,32
44	-0,08	94	-0,08	144	0,12	194	0,32
46	0,42	96	-0,38	146	-0,28	196	0,62
48	0,42	98	-0,58	148	0,02	198	0,52

Возводя в квадрат все значения $\overset{\circ}{N}(t)$ и деля сумму на $n=100$ получим приближенно дисперсию случайной функции $N(t)$:

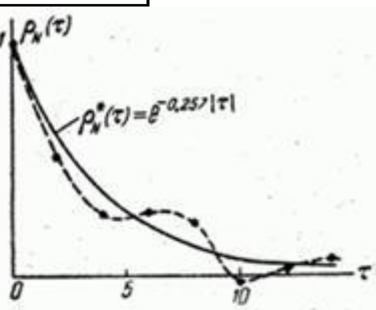
$$D_N = \frac{\sum_{i=1}^{100}\left[\overset{\circ}{N}(t_i)\right]^2}{100} \approx 0{,}1045$$

и среднее квадратическое отклонение: $\sigma_N \approx 0,323$.

Перемножая значения $\overset{\circ}{N}(t)$, разделенные интервалом $\tau = 2,4,6,\ldots$, и деля сумму произведений соответственно на $n-1=99$; $n-2=98$; $n-3=97$; …, получим значения корреляционной функции $k_N(\tau)$. Нормируя корреляционную функцию делением на $D_N = 0{,}1045$, получим таблицу значений функции $\rho_N(\tau)$.

Таблица 3

τ	$\rho_N(\tau)$	$\rho_N^*(\tau) = e^{-a\lvert\tau\rvert}$
отклонение 0	1,000	1,000
2	0,505	0,598
4	0,276	0,358
6	0,277	0,214
8	0,231	0,128
10	-0,015	0,077
12	0,014	0,046
14	0,071	0,027



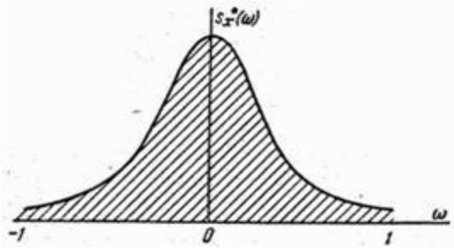
Не вполне гладкий ход корреляционной функции может быть объяснен недостаточным объемом экспериментальных данных (недостаточной продолжительностью опыта), в связи с чем случайные неровности в ходе функции не успевают сгладиться. Вычисление $\rho_N(\tau)$ продолжено до таких значений при которых фактически корреляционная связь пропадает.

Для того чтобы сгладить явно незакономерные колебания экспериментально найденной функции $\rho_N(\tau)$, заменим ее приближенно функцией вида: $\rho_N^*(\tau) = e^{-a\lvert\tau\rvert}$,где параметр a подберем методом наименьших квадратов

Применяя этот метод, находим $a = 0,257$. Вычисляя значения функции $\rho_N^*(\tau)$ при $\tau = 0,2,4,\ldots$, построим график сглаживающей кривой. На рис. он проведен сплошной линией. В последнем столбце таблицы 3 приведены значения функции $\rho_N^*(\tau)$. Пользуясь приближенным выражением корреляционной функции, получим нормированную спектральную плотность случайного процесса в виде:

$$S_x^*(\omega) = \frac{a}{\pi(a^2+\omega^2)} = \frac{0,257}{\pi(0,257^2+\omega^2)}.$$

График нормированной спектральной плотности представлен на рис.



2.16 Лабораторная работа 19 (ЛР-19)

Тема: Метод канонических разложений случайных функций.

2.16.1 Цель работы: изучить теоретические основания метода канонических разложений, выявить и проанализировать особенности метода в ходе решения практических задач

2.16.2 Задачи работы:

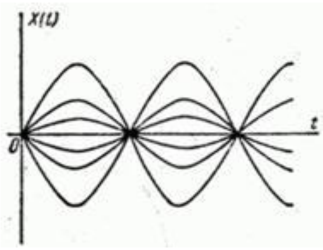
1. Представление случайной функции в виде суммы элементарных случайных функций.
2. Каноническое разложение случайной функции.
3. Линейные преобразования случайных функций, заданных каноническими разложениями.

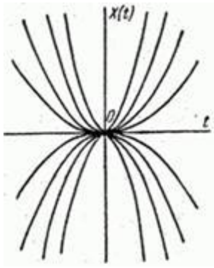
2.16.3 Перечень приборов, материалов, используемых в лабораторной работе: спецификой дисциплины не предусмотрены

2.16.4 Описание (ход) работы:

Элементарной случайной функцией называется функция вида:
$X(t) = V\varphi(t)$, где V - обычная случайная величина, $\varphi(t)$ - обычная (неслучайная) функция. Элементарная случайная функция является наиболее простым типом случайной функции. Все возможные реализации элементарной случайной функции $X(t)$ могут быть получены из графика функции $x = \varphi(t)$ простым измерением масштаба по оси ординат

При этом ось абсцисс $(x = 0)$ также представляет собой одну из возможных реализаций случайной функции $X(t)$, осуществляющуюся, когда случайная величина V принимает значение 0 (если это значение принадлежит к числу возможных значений величины V). В качестве примеров элементарных случайных функций приведем функции $X(t) = V\sin t$ и $X(t) = Vt^2$.





Элементарная случайная функция характерна тем, что в ней разделены две особенности случайной функции: случайность вся сосредоточена в коэффициенте V, а зависимость от времени - в обычной функции $\varphi(t)$.

Когда элементарная случайная функция поступает на вход линейной системы, то задача ее преобразования сводится к простой задаче преобразования одной неслучайной функции $\varphi(t)$. Отсюда возникает идея: представить случайную функцию на входе - точно или приближенно - в виде суммы элементарных случайных функций и только затем подвергать преобразованию. Такая идея разложения случайной функции на сумму элементарных случайных функций и лежит в основе **метода канонических разложений**.

Пусть имеется случайная функция: $X(t) = m_x(t) + \overset{\circ}{X}(t)$. Допустим, что нам удалось - точно или приближенно - представить ее в виде суммы

$X(t) = m_x(t) + \sum\limits_{i=1}^{m} V_i \varphi_i(t)$, где V_i - случайные величины с математическими ожиданиями, равными нулю;

$\varphi_i(t)$ - неслучайные функции; $m_x(t)$ - математическое ожидание функции $X(t)$.

Обозначая $Y(t)$ реакцию системы на случайное воздействие $X(t)$, имеем:

$Y(t) = L\{X(t)\} = L\{m_x(t)\} + \sum\limits_{i=1}^{m} V_i L\{\varphi_i(t)\}$. Придадим выражению несколько иную форму.

$Y(t) = m_y(t) + \sum\limits_{i=1}^{m} V_i \psi_i(t)$. Выражение представляет собой не что иное, как разложение случайной функции $Y(t)$ по элементарным функциям. Коэффициентами этого разложения являются те же случайные величины $V_1, V_2, \ldots, V_m$, а математическое ожидание и координатные функции получены из математического ожидания и координатных функций исходной случайной функции тем же линейным преобразованием L, какому подвергается случайная функция $X(t)$.

Если случайная функция $X(t)$, заданная разложением по элементарным функциям, подвергается линейному преобразованию Z, то коэффициенты разложения остаются неизменными, а математическое ожидание и координатные функции подвергаются тому же линейному преобразованию L.

Рассмотрим случайную функцию $X(t)$, заданную разложением

$X(t) = m_x(t) + \sum\limits_{i=1}^{m} V_i \varphi_i(t)$, где коэффициенты $V_1, V_2, \ldots, V_m$ представляют собой систему случайных величин с математическими ожиданиями, равными нулю и с корреляционной матрицей $\|K_{ij}\|$. Найдем корреляционную функцию и дисперсию случайной функции $X(t)$.

$$K_x(t,t') = \sum_{i=1}^{m} \varphi_i(t)\varphi_i(t')D_i + \sum_{i \neq j} \varphi_i(t)\varphi_j(t')K_{ij}$$

. Полагая в выражении $t' = t$ получим дисперсию случайной функции $X(t)$:

$$D_x(t) = \sum_{i=1}^{m} [\varphi_i(t)]^2 D_i + \sum_{i \neq j} \varphi_i(t)\varphi_j(t)K_{ij}$$

.

Очевидно, эти выражения приобретают особенно простой вид, когда все коэффициенты V_i разложения некоррелированны, т. е. $K_{ij} = 0$ при $i \neq j$. В этом случае разложение случайной функции называется «каноническим».

$$K_x(t,t') = \sum_{i=1}^{m} \varphi_i(t)\varphi_i(t')D_i \qquad . \qquad D_x(t) = \sum_{i=1}^{m} [\varphi_i(t)]^2 D_i$$

Рассмотрим несколько подробнее применение метода канонических разложений к определению реакции динамической системы на случайное входное воздействие, когда работа системы описывается линейным дифференциальным уравнением, в общем случае - с переменными коэффициентами. Запишем это уравнение в операторной форме:

$$A_n(p,t)Y(t) = B_m(p,t)X(t)$$

.

Согласно вышеизложенным правилам линейных преобразований случайных функций математические ожидания воздействия и реакции должны удовлетворять тому же уравнению:

$$A_n(p,t)m_y(t) = B_m(p,t)m_x(t)$$

. Аналогично каждая из координатных функций должна удовлетворять тому же дифференциальному уравнению:

$$A_n(p,t)\psi_i(t) = B_m(p,t)\varphi_i(t), \quad (i = 1, 2, \ldots, k).$$

Таким образом, задача определения реакции линейной динамической системы на случайное воздействие свелась к обычной математической задаче ния $k+1$ обыкновенных дифференциальных уравнений, содержащих обычные, не случайные функции. Остается осветить вопрос о начальных условиях, при которых следует интегрировать уравнения.

Сначала рассмотрим наиболее простой случай, когда начальные условия для данной динамической системы являются неслучайными. В этом случае при $t = 0$ должны выполняться условия:

$$\left.\begin{array}{l} Y(0) = y_0, \\ Y'(0) = y_1, \\ \ldots\ldots\ldots\ldots \\ Y^{(r)}(0) = y_r, \\ \ldots\ldots\ldots\ldots\ldots \\ Y^{(n-1)}(0) = y_{n-1}, \end{array}\right\}$$

где $y_0, y_1, \ldots, y_{n-1}$ - неслучайные числа.

Следует заметить, что на практике весьма часто встречаются случаи, когда для моментов времени, достаточно удаленных от начала случайного процесса, начальные условия уже не оказывают влияния на его течение: вызванные ими переходные процессы успевают затухнуть. Системы, обладающие таким свойством, называются асимптотически устойчивыми. Если нас интересует реакция асимптотически устойчивой динамической системы на участках времени, достаточно удаленных от начала, то можно ограничиться исследованием решения $Y_I(t)$, полученного при нулевых начальных условиях.

Для закрепления теоретического материала предлагаются следующие задачи:

1. На вход стационарной линейной динамической системы, описываемой данным дифференциальным уравнением, подается стационарный случайный процесс X(t) с математическим ожиданием m_X и спектральной плотностью $S_X(\omega)$. Найти математическое ожидание и дисперсию случайного процесса Y(t) на выходе системы в установившемся режиме: $y''+4y'+4y = 3x'$, $m_X = 14$, $S_X(\omega) = 2(\sin 4\omega)/\omega$.

2. На вход стационарной линейной динамической системы, описываемой данным дифференциальным уравнением, подается стационарный случайный процесс X(t) с корреляционной функцией $k_X(\tau)$. Найти спектральную плотность $S_Y(\omega)$ случайного процесса Y(t) на выходе системы в установившемся режиме: $y''+6y'+5y = x''+7x'+10x$, $k_X(\tau) = 18/(9+\tau\ 2)$

3. Является ли фильтром следующее линейное преобразование: $\eta(n) = (\xi(n+1)+\xi(n-1))/2$, $n \in Z$? Найти спектральную плотность $s\eta(\lambda)$, если ξ — стандартный белый шум. Какое физически реализуемое преобразование дает «на выходе» ту же спектральную плотность?

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

3.1 Практическое занятие 1 - 2 (ПЗ-1-2)

Тема: Случайные события, их вероятность. Основные теоремы теории вероятностей. Условная вероятность. Следствия основных теорем теории вероятностей. Схема повторных испытаний

3.1.1 Задание для работы:

1. Элементы комбинаторики. Непосредственное вычисление вероятности случайного события. Операции над случайными событиями и их свойства.

2. Теоремы о вероятности суммы случайных событий. Теоремы о вероятности суммы произведения событий.

3. Условная вероятность. Формула полной вероятности. Формула Байеса.

4. Схема повторных испытаний. Формула Бернулли. Формула Пуассона. Локальная формулы Лапласа. Интегральная формула Лапласа.

5. Простейший поток событий. Вероятность случайного события с заданной интенсивностью.

3.1.2 Краткое описание проводимого занятия

1. Элементы комбинаторики. Непосредственное вычисление вероятности случайного события. Операции над случайными событиями и их свойства.

Элементы комбинаторики

Комбинаторика изучает способы подсчета числа элементов в конечных множествах. Формулы комбинаторики используются при непосредственном вычислении вероятностей.

Приведем некоторые сведения.

Соединениями называют различные группы предметов, составленные из каких-либо объектов.

Элементами называются объекты, из которых составлены соединения. Рассмотрим следующие три вида соединений: перестановки, размещения и сочетания.

Перестановками из n элементов называют *соединения,* содержащие все n элементов и отличающиеся между собой лишь порядком элементов. Число перестановок из n элементов находится по формуле $P_n = n!$,

где $n!$ - произведение натуральных чисел от 1 до n включительно, т.е. $n!=1\cdot 2\cdot 3\cdot\ldots\cdot n$. Например, $P_6 = 6! = 1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6 = 720$.

Размещениями из n элементов по k в каждом ($n \geq k$) называются *такие соединения,* в каждый из которых входит k элементов, взятых из данных n элементов, и отличающихся друг от друга либо самими элементами, либо порядком их расположения.

Число размещений из n элементов по k находят по формуле

$$A_n^k = n(n-1)(n-2)\ldots(n-k+1) \text{ или, } A_n^k = \frac{n!}{(n-k)!}$$

Например, $A_6^4 = 6\cdot 5\cdot 4\cdot 3 = \dfrac{6!}{(6-4)!} = \dfrac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}{1\cdot 2} = 360$

Сочетаниями из n элементов по k ($n>k$) называют *соединения,* в каждый из которых входит k элементов, взятых из данных n элементов и отличающихся друг от друга, по крайней мере, одним элементом. Число сочетаний из n элементов по k находят по формуле:

$$C_n^k = \frac{A_n^k}{P_k} \text{ или } C_n^k = \frac{n!}{(n-k)!k!}.$$

Для упрощения вычислений при $k > \dfrac{1}{2}n$ полезно использовать следующее свойство сочетаний: $C_n^k = C_n^{n-k}$.

Замечания:

1) по определению $C_n^0 = 1$;

2) для определения числа сочетаний справедливы равенства

$$C_n^m = C_n^{n-m},\ C_{n+1}^{m+1} = C_n^m + C_n^{m+1},\ C_n^0 + C_n^1 + \ldots + C_n^n = 2^n$$

3) В записанных выше формулах комбинаторики предполагалось, что все n элементов различны. Если же некоторые элементы в соединениях повторяются, то в этом случае соединения с повторениями вычисляются по другим формулам.

Пусть среди n элементов рассматриваемого множества есть n_1 элементов одного вида, n_2 элементов другого вида и т.д. Число перестановок с повторениями определяется по формуле $P_n(n_1, n_2, \ldots, n_k) = \dfrac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_k!}$,

где $n_1 + n_2 + \ldots + n_k = n$.

Число размещений по m элементов с повторениями из n элементов равно n^m, т.е. $\left(A_n^m\right)_{\text{повт.}} = n^m$.

Число сочетаний с повторениями из n элементов по m элементов равно числу сочетаний без повторений из ($n+m-1$) элементов по m, т.е.

$$\left(C_n^m\right)_{\text{повт.}} = C_{n+m-1}^m$$

4) При решении задач комбинаторики можно использовать следующие правила:

правило суммы. Если некоторый объект *A* может быть выбран из множества объектов *m* способами, а объект *B* может быть выбран *n* способами, то выбрать либо *A,* либо *B* можно *(m + n)* способами.

правило произведения. Если объект *A* можно выбрать из множества объектов *m* способами и после каждого такого выбора объект *B* можно выбрать *n* способами, то пара объектов *(A, B)* в указанном порядке может быть выбрана *m · n* способами.

Непосредственное вычисление вероятности случайного события

Пример 1. В урне 10 одинаковых по размерам и весу шаров, из которых 4 красных и 6 голубых. Из урны извлекается один шар. Какова вероятность того, что извлеченный шар окажется голубым?

Решение. Событие, состоящее в том, что «извлеченный шар оказался голубым», обозначим буквой А. Данное испытание имеет 10 равновозможных элементарных исходов, из которых 6 благоприятствуют появлению события А. По формуле классической вероятности события получим: $P(A)=\frac{6}{10}=0{,}6$.

Пример 2. Среди 25 студентов группы, в которой 10 девушек, разыгрывается 5 билетов лотереи. Найти вероятность того, что среди обладателей билетов окажутся две девушки.

Решение. Пусть А - событие, состоящее в том, что среди обладателей билетов окажутся две девушки. Найдем числа m, n.

Число всех равновозможных случаев распределения 5 билетов среди 25 студентов равно числу сочетаний из 25 элементов по 5, т.е. $C_{15}^{5}\cdot$ Число групп по трое юношей из 15, которые могут получить билеты, равно C_{15}^{3}. Каждая такая тройка может сочетаться с любой парой из десяти девушек, а число таких пар равно C_{10}^{2}. Следовательно, число групп по 5 студентов, образованных из групп в 25 студентов, в каждую из которых будут входить трое

юношей и две девушки, равно произведению $C_{15}^{3}\cdot C_{10}^{2}$. Это произведение равно числу благоприятствующих случаев распределения пяти билетов среди студентов группы так, чтобы три билета получили юноши и два билета - девушки.

В соответствии с формулой $P(A)=\frac{m}{n}$ находим искомую вероятность

$$P(A)=\frac{C_{15}^{3}\cdot C_{10}^{2}}{C_{25}^{5}}=\frac{15!}{3!\cdot 12!}\cdot\frac{10!}{2!\cdot 8!}:\frac{25!}{5!\cdot 20!}=\frac{20!\cdot 15!\cdot 10!\cdot 5!}{25!\cdot 12!\cdot 8!\cdot 3!\cdot 2!}=\frac{13\cdot 14\cdot 15\cdot 9\cdot 10\cdot 4\cdot 5}{25\cdot 24\cdot 23\cdot 22\cdot 21\cdot 2}=\frac{13\cdot 5\cdot 3}{23\cdot 22}=\frac{195}{506}\approx 0{,}385$$

Пример 3. В круг вписан квадрат (рис.2). В круг наудачу бросается точка. Какова вероятность того, что эта точка попадет в квадрат?

Рис. 2.

Решение. Введем обозначения: R - радиус круга, a - сторона вписанного квадрата, A - событие, состоящее в том, что точка попала в квадрат, S - площадь круга, S_1 - площадь вписанного квадрата. Известно, что площадь круга $S=\pi R^2$.

Сторона вписанного квадрата через радиус описанной окружности выражается формулой $a=\sqrt{2}R$, поэтому площадь квадрата $S_1 = 2R^2$

Полагая в формуле $P(A)=\frac{S_g}{S_G}$ $S_g = S_1$, $S_G = S$,

находим искомую вероятность $P(A)=\frac{2R^2}{\pi R^2}=\frac{2}{\pi}\approx 0{,}637$.

Замечание. Выражение стороны квадрата через радиус окружности можно получить следующим образом. Из треугольника ΔKMN по теореме Пифагора будем иметь: $KN^2 + NM^2 = KM^2$, т.е.

$a^2 + a^2 = (2R)^2,\ 2a^2 = 4R^2,\ a^2 = 2R^2,\ a = \sqrt{2}R.$

2. Теоремы о вероятности суммы случайных событий. Теоремы о вероятности суммы произведения событий.

Пример 1. Подбрасываются два игральных кубика. Найти вероятность события A, состоящего в том, что - «сумма выпавших очков не превосходит четырех».

Решение. Событие A - событие, состоящее в том, что есть сумма трех несовместных событий B_2, B_3, B_4. Тогда сумма очков равна соответственно 2, 3, 4. Поскольку $P(B_2) = \frac{1}{36}$, $P(B_3) = \frac{2}{36}$, $P(B_4) = \frac{3}{36}$, по теореме сложения вероятностей несовместных событий получим

$$P(A) = P(B_2) + P(B_3) + P(B_4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = \frac{1}{6}.$$

Замечание. Этот же результат можно было получить, используя непосредственный подсчет вероятности. Действительно, событию A благоприятствуют 6 элементарных исходов: (1,1), (1,2), (2,1), (1,3), (3,1), (2,2). Всего же элементарных исходов, образующих полную группу событий, n= 36, поэтому $P(A) = \frac{6}{36} = \frac{1}{6}$.

Пример 2. Три станка работают независимо. Вероятность того, что в течение смены станок (любой) потребует наладки равна 0,1. Найти вероятность того, что в течение смены хотя бы один станок из трех потребует внимания наладчика.

Решение. Пусть A_k - событие, заключающееся в том, что k- тый по счету станок потребует наладки в течение смены (k = 1, 2,3). Тогда событие $A_1 + A_2 + A_3$ заключается в том, что в течение смены наладки потребует хотя бы один из трех станков. Сначала вычислим вероятность противоположного события $\overline{A_1 + A_2 + A_3}$, заключающегося в том, что все три станка всю смену проработают безотказно. Поскольку $\overline{A_1 + A_2 + A_3} = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$, причем события $\overline{A_1}, \overline{A_2}, \overline{A_3}$ независимы, то $P(\overline{A_1 + A_2 + A_3}) = P(\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}) = P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(\overline{A_3})$ по теореме умножения вероятностей для независимых событий. По условию $P(A_k)$=0,1,тогда вероятность противоположного события $P(\overline{A_k}) = 1 - P(A_k) = 0{,}9$. Итак, $P(A_1 + A_2 + A_3) = 1 - P(\overline{A_1 + A_2 + A_3})$ и искомая вероятность события будет $P(A_1 + A_2 + A_3) = 1 - 0{,}9 \cdot 0{,}9 \cdot 0{,}9 = 0{,}271.$

Пример 3. Сколько раз нужно подбросить два игральных кубика, чтобы вероятность выпадения хотя бы один раз двух шестерок была бы больше $\frac{1}{2}$? (Эта задача впервые поставлена французским математиком и писателем де Мере (1610-1684 гг.), поэтому задача называется его именем).

Решение. Пусть событие A_i - «выпадение двух шестерок при i-м подбрасывании». Так как с каждой из шести граней первого кубика может выпасть любая из шести граней второго кубика, то всего равновозможных попарно несовместных событий $6 \cdot 6 = 36$. Только одно из них - выпадение шестерки и на первом и на втором кубике - благоприятствуют событию A_i. Следовательно, $P(A_i) = \frac{1}{36}$, откуда $q = 1 - p = 1 - \frac{1}{36} = \frac{35}{36}$.

Подбрасывание игральных кубиков - независимые испытания, поэтому воспользуемся формулой $P(A)=1-q^n$, тогда в данном случае получим: $1-\left(\frac{35}{36}\right)^n>\frac{1}{2}$, или $\left(\frac{35}{36}\right)^n<\frac{1}{2}$.

Решив неравенство, найдем *n*. Логарифмируя обе части неравенства, получим $n\ln\frac{35}{36}<\ln\frac{1}{2}$, откуда $n>\frac{\ln 2}{\ln 36-\ln 35}=\frac{0{,}6931}{0{,}0284}=24{,}4$.

Итак, чтобы вероятность выпадения двух шестерок была больше $\frac{1}{2}$, достаточно подбросить кубик не менее 25 раз.

3. Условная вероятность. Формула полной вероятности. Формула Байеса.

Пример 1. Слово *папаха* составлено из букв разрезной азбуки. Карточки с буквами тщательно перемешаны. Четыре карточки извлекаются по очереди и раскладываются в ряд. Какова вероятность получить при этом слово *папа?*

Решение. Обозначим через *A, B, C, D* соответственно события, состоящие в том, что: извлечена первая, вторая, третья и четвертая буква слова *папа* из набора в 6 букв: *а, а, а, п, п, х.* Найдем вероятности событий: *A, B/A, C/AB , D/ABC.*

$$P(A)=\frac{2}{6}=\frac{1}{3};\quad P(B/A)=\frac{3}{5};\quad P(C/AB)=\frac{1}{4};\quad P(D/ABC)=\frac{2}{3}.$$

В соответствии с формулой вероятности произведения зависимых событий при $n=4$ будем иметь:

$$P(ABCD) = P(A)P(B/A)P(C/AB)P(D/ABC) = \frac{1}{3}\cdot\frac{3}{5}\cdot\frac{1}{4}\cdot\frac{2}{3}=\frac{1}{30}.$$

Пример 2. В пяти ящиках находятся одинаковые по размерам и весу шары. В двух ящиках - по 6 голубых и 4 красных шара (это ящик состава H_1). В двух других ящиках (состава H_2) - по 8 голубых и 2 красных шара. И в пятом ящике (состава H_3) - 8 красных и 2 голубых шара. Наудачу выбирается ящик, и из него извлекается шар. Какова вероятность того, что извлеченный шар оказался красным?

Решение. Событие, состоящее в том, что «извлечен красный шар» обозначим через *A*. Из условия задачи следует, что $P(H_1)=\frac{2}{5}=0{,}4$, $P(H_2)=\frac{2}{5}=0{,}4$, $P(H_3)=\frac{1}{5}=0{,}2$.

Вероятность вынуть красный шар, если известно, что взят ящик первого состава H_2, будет определяться так:

$$P(A/H_1)=\frac{4}{10}=0{,}4$$

Вероятность извлечь красный шар, если известно, что взят ящик второго состава H_2, будет

$P(A/H_2)=\frac{2}{10}=0{,}2$. Вероятность извлечь красный шар, если известно, что взят ящик третьего состава H_3, будет $P(A/H_3)=\frac{8}{10}=0{,}8$.

При $n = 3$ находим искомую вероятность

$$P(A)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)=0{,}4\cdot 0{,}4+0{,}4\cdot 0{,}2+0{,}2\cdot 0{,}8=0{,}4$$

4. Схема повторных испытаний. Формула Бернулли. Формула Пуассона. Локальная формулы Лапласа. Интегральная формула Лапласа.

Пример 1. Частица находится на прямой в начале координат. Под действием случайных толчков частица каждую секунду перемещается вправо (с вероятностью $\frac{1}{3}$) или влево (с вероятностью $\frac{2}{3}$) на единицу масштаба. Найти вероятность того, что через 4 секунды частица вернется в начало координат.

Решение. Через 4 секунды частица вернется в начало координат в том случае, если она переместится ровно два раза вправо (и, значит, два раза влево). По формуле Бернулли найдем вероятность того, что из четырех независимых перемещений частицы ровно два перемещения будут вправо:

$$n = 4 \quad k = 2 \quad p = \frac{1}{3} \quad q = \frac{2}{3}. \quad P4(2) = C_4^2\left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^2 = 6 \cdot \frac{1}{9} \cdot \frac{4}{9} = \frac{24}{81} \approx 0,296.$$

Пример 2. Вероятность того, что при броске мяча баскетболист попадёт в корзину, равна 0,3. Найти наивероятнейшее число попаданий при 8 бросках и соответствующую вероятность.

А это уже если и не Терминатор, то, как минимум, хладнокровный спортсмен =)

Решение: для оценки наивероятнейшего числа попаданий используем двойное неравенство $np - q \le m_0 < np + p$. В данном случае:

$n = 8$ – всего бросков;

$p = 0,3$ – вероятность попадания в корзину при каждом броске;

$q = 1 - p = 1 - 0,3 = 0,7$ – вероятность промаха при каждом броске.

Таким образом, наивероятнейшее количество попаданий при 8 бросках находится в следующих пределах:

$8 \cdot 0,3 - 0,7 \le m_0 < 8 \cdot 0,3 + 0,3$

$2,4 - 0,7 \le m_0 < 2,4 + 0,3$

$1,7 \le m_0 < 2,7$

Поскольку левая граница – дробное число *(пункт №1)*, то существует единственное наивероятнейшее значение, и, очевидно, что оно равно $m_0 = 2$.

Используя формулу Бернулли $P_n^m = C_n^m p^m q^{n-m}$, вычислим вероятность того, что при 8 бросках будет ровно 2 попадания:

$$P_8^2 = C_8^2 \cdot (0,3)^2 \cdot (0,7)^6 = \frac{8!}{6! \cdot 2!} \cdot (0,3)^2 \cdot (0,7)^6 = \frac{7 \cdot 8}{2} \cdot (0,3)^2 \cdot (0,7)^6 = 0,29647548$$

Ответ: $m_0 = 2$ – наивероятнейшее количество попаданий при 8 бросках,

$P_8^2 \approx 0,2965$ – соответствующая вероятность.

Пример 3. Станок изготавливает за смену 100000 деталей. Вероятность изготовления бракованной детали p=0,0001. Найти вероятность того, что за смену будет изготовлено 5 бракованных деталей.

Решение

Обозначим n=100000, k=5, p=0,0001. События, состоящие в том, что отдельная деталь бракована, независимы, число испытаний велико, а вероятность мала, поэтому воспользуемся распределением Пуассона

$$P_n(k) = \lambda^k \cdot \frac{e^{-\lambda}}{k!}; \quad \lambda = np.$$

$$\lambda = np = 100\,000 \cdot 0,0001 = 10; \quad P_{100000}(5) = 10^5 \cdot \frac{e^{-10}}{5!} \approx 0,0378.$$

Локальная формулы Лапласа. Интегральная формула Лапласа.

Пример 1. К электросети подключено 36 приборов, каждый мощностью 5 киловатт и потребляет в данный момент энергию с вероятностью 0,2. Найти вероятность того, что потребляемая в данный момент мощность:

а) составит ровно 50 киловатт;

б) превзойдет 50 киловатт.

Решение. В случае а) надо найти вероятность того, что из 36 приборов работают ровно 10. Применим локальную теорему Лапласа: $n = 36 \quad k = 10 \quad p = 0{,}2 \quad q = 0{,}8$.

$$\sqrt{npq} = 24 \quad x = \frac{k - np}{\sqrt{npq}} = 1{,}4. \quad P_{36}(10) = \frac{1}{\sqrt{npq}}\varphi(x) = \frac{1}{2{,}4}\cdot\varphi(1{,}4) = 0{,}0624.$$

Значение функции локальной функции Лапласа *φ(x)* взято из таблицы приложений.

В случае б) находим вероятность $P_{36}(k \geq 10)$ того, что работают более десяти приборов. Применяем для решения этой части задачи интегральную теорему Лапласа. Находим сначала значения x_1, x_2:

$$x_1 = \frac{k - np}{\sqrt{npq}} = 1{,}4, \quad x_2 = \frac{n - np}{\sqrt{npq}} = 12/$$

Тогда искомая вероятность будет:

$$P_{36}(k \geq 10) = \Phi(x_2) - \Phi(x_1) = 0{,}5 - 0{,}4192 = 0{,}08808,$$

Значения функции Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{\frac{t^2}{2}}dt$ взяты из таблицы приложений.

Пример 2. В нерестовике содержится 200 рыб - производителей вида *А*. Вероятность отдачи икры в искусственных условиях рыбы вида *А* равна $\frac{3}{4}$. Требуется найти вероятность того, что

икру отдадут 150 рыб.

Решение. Вероятность того, что ровно 150 рыб из 200 отдадут икру, найдем, используя локальную теорему Лапласа

$$P_n(k) = \frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}e^{\frac{x^2}{2}} = \frac{1}{\sqrt{npq}}\cdot\varphi(x), \text{ где } x = \frac{k - np}{\sqrt{npq}}.$$

Значение функции *φ(x)* возьмем из таблицы. Находим:

n = 200, npq = 200 · 0,75 · 0,25 = 150 · 0,25 = 0,375 .

$m = к = np$ = 150 $\qquad \sqrt{npq} = 6{,}12$.

p = 0,75 $\qquad x = \frac{150 - 150}{6{,}12} = 0$.

q = 0,25.

Получим: $P_{200}(150) = \frac{1}{6{,}12}\cdot\varphi(0) = \frac{0{,}3989}{6{,}12} \approx 0{,}07$.

5. Простейший поток событий. Вероятность случайного события с заданной интенсивностью.

Пример 1. Среднее число заявок, поступающих на предприятие бытового обслуживания за 1 час, равно трем. Найти вероятность того, что за 2 часа поступит 5 заявок. Предполагается, что поток заявок - простейший.

Решение. По условию λ = 3, $t = 2$, k = 5. Воспользуемся формулой

$$P_t(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Искомая вероятность того, что за 2 часа поступит 5 заявок, равна $P_2(5) = \frac{(6)^5 \cdot 0{,}00248}{120} \approx 0{,}268$.

3.1.3 Результаты и выводы: В результате проведенного занятия студенты:

- освоили основные понятия комбинаторики, теории случайных событий, классификацию случайных событий; усвоили основные правила, применяемые в теории случайных событий; выработали навыки по вычислению вероятностей случайных событий, их суммы, произведения;
- освоили понятие условная вероятность, формулу полной вероятности, формулы Байеса; усвоили основные правила применения формул Байеса.
- освоили основные понятия, связанные со схемой повторных испытаний; усвоили основные формулы, применяемые при решении задач по схеме повторных испытаний;
- выработали навыки по вычислению вероятностей случайных событий по формулам Бернулли, Пуассона, Лапласа;
- освоили понятие случайный поток, его свойства; усвоили основные правила формулы вероятности случайного события с заданной интенсивностью.

3.2 Практическое занятие 3 (ПЗ-3)

Тема: Случайные величины. Функция и плотность распределения СВ. Числовые характеристики случайной величины

3.2.1 Задание для работы:

1.Случайные величины, их классификация. Закон распределения случайной величины. Ряд распределения.

2. Функция распределения. Плотность распределения.

3. Числовые характеристики ДСВ. Числовые характеристики НСВ. Свойства числовых характеристик, их интерпретация.

3.2.2 Краткое описание проводимого занятия

1.Случайные величины, их классификация. Закон распределения случайной величины. Ряд распределения.

Пример1. Найти закон распределения дискретной случайной величины X – числа появлений «орла» при двух бросаниях монеты.

Решение. Возможные значения случайной величины: 0, 1, 2. Вероятности этих значений находим по формуле Бернулли:

$$p_0 = P(X=0) = P_2(0) = C_2^0 p^0 q^2 = \frac{2!}{0!\,2!} \cdot 0.5^0 \cdot 0.5^2 = 0.25;$$

$$p_1 = P(X=1) = P_2(1) = C_2^1 p^1 q^1 = \frac{2!}{1!\,1!} \cdot 0.5^1 \cdot 0.5^1 = 0.50;$$

$$p_2 = P(X=2) = P_2(2) = C_2^2 p^2 q^0 = \frac{2!}{2!\,0!} \cdot 0.5^2 \cdot 0.5^0 = 0.25.$$

Записываем ряд распределения:

X	0	1	2
P	0.25	0.50	0.25

Пример 2. В лотерее выпущено 100 билетов. Разыгрывался один выигрыш в 50 у.е. и десять выигрышей по 10 у.е. Найти закон распределения величины X – стоимости возможного выигрыша.

Решение. Возможные значения величины X: x1 = 0; x2 = 10 и x3 = 50. Так как «пустых» билетов – 89, то p1 = 0,89, вероятность выигрыша 10 у.е. (10 билетов) – p2 = 0,10 и для выигрыша 50 у.е. – p3 = 0,01. Таким образом:

X	0	10	50
P	0,89	0,10	0,01

Легко проконтролировать: $p_1 + p_2 + p_3 = 0{,}89 + 0{,}10 + 0{,}01 = 1$

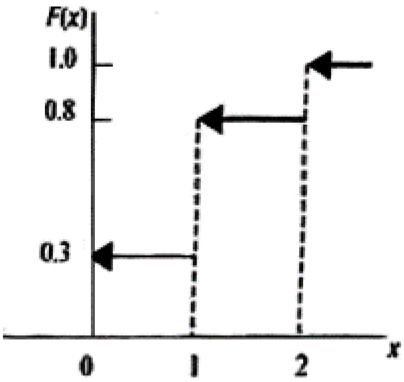
2. Функция распределения. Плотность распределения.

Функцией распределения случайной величины X Называется функция F(X), Определенная на всей числовой оси следующим образом: F(X)= P(X < x),

Т. е. F(X) есть вероятность того, что случайная величина X Примет значение меньшее, чем X.

Функцию распределения можно представить графически. Для дискретной случайной величины график имеет ступенчатый вид. Построим, например, график функции распределения случайной величины, заданной следующим рядом (рис. 1):

X	0	1	2
P	0.3	0.5	0.2



$$F(x) = \begin{cases} 0, & x \le 0; \\ 0.3, & 0 < x \le 0; \\ 0.8, & 1 < x \le 2; \\ 1, & x > 2 \end{cases}$$

Рис.1. График функции распределения
дискретной случайной величины

Скачки функции происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений. В точках разрыва функция F(X) непрерывна слева.

Задача 1. Дискретная случайная величина X задана законом распределения:

X	2	5	8
P	0,6	0,1	0,3

Найти функцию распределения F(x) и построить ее график.

Решение. Так как функция распределения,

$F(x) = \sum_{i=1}^{k} p_i$ для $x_k < x \le x_{k+1}$, то

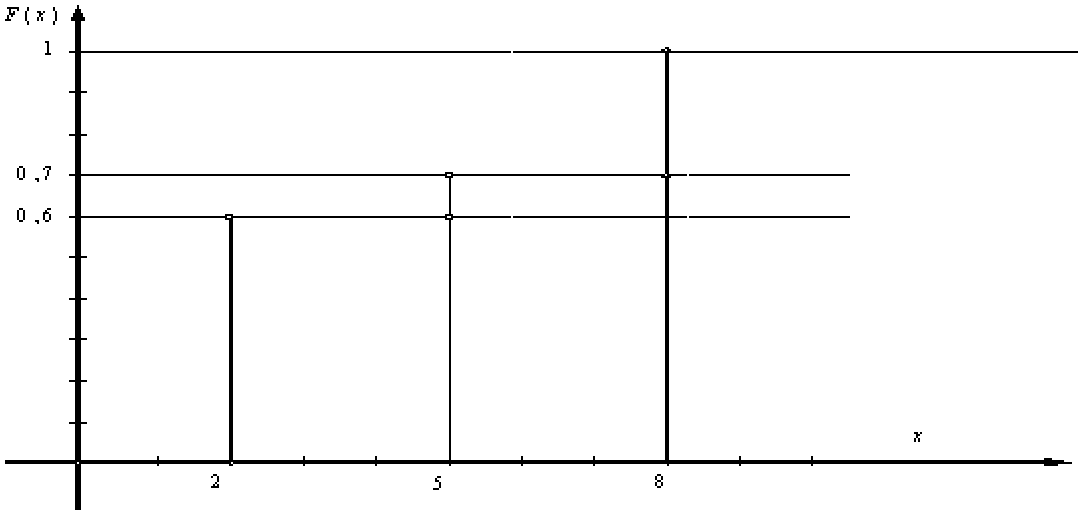
при $x \le 2$ $F(x) = 0$;

при $2 < x \le 5$ $F(x) = 0 + 0{,}6 = 0{,}6$;

при $5 < x \le 8$ $F(x) = 0 + 0{,}6 + 0{,}1 = 0{,}7$;

при $x > 8$ $F(x) = 0 + 0{,}6 + 0{,}1 + 0{,}3 = 1$;

Соответствующий график:



Задача 2. Непрерывная случайная величина X задана дифференциальной функцией распределения:

$$f(x) = \begin{cases} 0 & \text{при } x \le 0 \\ 2e^{-2x} & \text{при } x > 0 \end{cases}.$$

Найти вероятность попадания X в интервал

$1 < x < 2$.

Решение. Заметим, что это частный случай показательного закона распределения.

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

$$P(1 < X < 2) = \int_1^2 2e^{-2x}dx = \ldots = \frac{1}{e^2} - \frac{1}{e^4} = \frac{e^2 - 1}{e^4} \approx 0{,}12.$$

3. Числовые характеристики ДСВ. Числовые характеристики НСВ. Свойства числовых характеристик, их интерпретация.

Пример 1. Найти математическое ожидание дискретной случайной величины, заданной рядом распределения:

X	0	1	2	3
P	0.2	0.4	0.3	0.1

Решение.

$M(X) = \sum_{i=1}^{4} p_i x_i$, $=0\times0.2 + 1\times0.4 + 2\times0.3 + 3\times0.1=1.3$.

Пример 2. Найти математическое ожидание случайной величины, заданной плотностью распределения:

$$f(x)=\begin{cases} x/2, & 0 \le x \le 2, \\ 0, & x<0, \quad x>2 \end{cases}.$$

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^2 x\cdot x/2\cdot dx=\frac{1}{2}\int_0^2 x^2\cdot dx=\frac{x^3}{6}\Big|_0^2=\frac{8}{6}=1.33.$$

Решение.

Пример 3. Найти дисперсию случайной величины, заданной рядом распределения

X	0	1	2	3
P	0.2	0.4	0.3	0,1

Решение. Чтобы вычислить дисперсию, необходимо знать математическое ожидание. Для данной случайной величины выше было найдено: M=1.3. Вычисляем дисперсию по формуле (3.5):

$$D(X)=M\left(X^2\right)-m^2=\sum_{i=1}^{4}x_i^2p_i-m^2=.$$

$$=0^2\cdot 0.2\ +\ 1^2\cdot 0.4+\ 2^2\cdot 0.3\ +\ 3^2\ \cdot 0.1-1.3^2\ =0.81.$$

Пример 4. Случайная величина задана плотностью распределения

$$f(x)=\begin{cases} (1/2)\cdot\cos x, & |x|\le \pi/2, \\ 0, & |x|>\pi/2. \end{cases}$$

Найти дисперсию и среднее квадратическое отклонение.

Решение. Находим сначала математическое ожидание:

$$m=\int_{-\infty}^{+\infty} xf(x)dx=\frac{1}{2}\cdot\int_{-\pi/2}^{\pi/2} x\cdot\cos x\cdot dx=0$$

(как интеграл от нечетной функции по симметричному промежутку).

Теперь вычисляем дисперсию и среднее квадратическое отклонение:

$$D(X)=\int_{-\infty}^{\infty}(x-m)^2f(x)dx=\frac{1}{2}\cdot\int_{-\pi/2}^{\pi/2}x^2\cdot\cos x\cdot dx=$$

$$=\int_0^{\pi/2}x^2\cdot\cos x\cdot dx=x^2\cdot\sin x\Big|_0^{\pi/2}-2\cdot\int_0^{\pi/2}x\cdot\sin x\cdot dx=$$

$$=x^2\cdot\sin x\Big|_0^{\pi/2}+2x\cdot\cos x\Big|_0^{\pi/2}-2\cdot\int_0^{\pi/2}x\cdot\cos x\cdot dx=$$

$$=x^2\cdot\sin x\Big|_0^{\pi/2}+2x\cdot\cos x\Big|_0^{\pi/2}-2\sin x\Big|_0^{\pi/2}=\frac{\pi^2}{4}-2=\frac{\pi^2-8}{4};$$

$$\sigma(X)=\sqrt{D(X)}=\frac{\sqrt{\pi^2-8}}{2}.$$

3.2.3 Результаты и выводы: В результате проведенного занятия студенты:

- освоили основные понятия, связанные теорией случайных величин;

- усвоили классификацию СВ, алгоритм построения функции распределения ДСВ, нахождения плотности распределения НСВ;
- выработали навыки по вычислению вероятности попадания СВ в интервал;
- освоили понятие числовых характеристик СВ, их свойства, интерпретацию;
- выработали навыки вычисления числовых характеристик СВ.

3.3 Практическое занятие 4 (ПЗ-4)

Тема: Некоторые распределения ДСВ. Некоторые распределения НСВ

3.3.1 Задание для работы:

1. Биномиальное распределение, его свойства, числовые характеристики.
2. Распределение Пуассона, его свойства, числовые характеристики. Связь распределений ДСВ с нормальным распределением.
3. Равномерное распределение. Показательное распределение.
4. Нормальное распределение, его свойства.

3.3.2 Краткое описание проводимого занятия

1. Биномиальное распределение, его свойства, числовые характеристики.

1. Математическое ожидание случайной величины через образующую функцию для биномиального распределения вычисляем по формуле $M(X)=np.$
2. Дисперсия $D(X)=npq.$
Имея дисперсию нетрудно установить среднее математическое отклонение $\sigma(X)=\sqrt{npq}.$
3. Коэффициент асимметрии A(X) и эксцесс E(X) для биномиального распределения определяют по формулам $A(X)=\dfrac{q-p}{\sqrt{npq}}=\dfrac{1-2p}{\sqrt{npq}};\quad E(X)=\dfrac{1-6pq}{npq}.$

Задача 1. В партии однотипных деталей стандартные составляют *97%*. Наугад из партии берут *400* деталей. Определить математическое ожидание, дисперсию и среднее квадратическое отклонение *M(X), D(X), S(X)* для дискретной случайной величины *X* — появления числа стандартных деталей среди 400 наугад взятых.

Решение. Целочисленных случайная величина *X* имеет биномиальное закон распределения вероятностей, которая может принимать значения $X = k = 0, 1, 2, \ldots, 400$.Вероятности возможных значений для данной задачи определяются по формуле Бернулли и составляют $P_k=P(X=k)=C_{400}^k p^k q^{400-k}$, где $p = 0,97$ — вероятность появления стандартной детали, $q = 1 - p = 1 - 0,97 = 0,03$ — вероятность появления нестандартной детали. Согласно приведенным выше формулам определяем нужные величины:

$$M(X)=np=400\cdot 0,97=388;\quad D(X)=npq=400\cdot 0,97\cdot 0,03=11,64;\quad \sigma(X)=\sqrt{npq}=\sqrt{11,64}\approx 3,41.$$

Задача 2. Два ювелирные заводы производят свадебные кольца в объеме *3:7*. Первый завод производит *95%* колец без дефекта, второй – *90%*. Молодая пара перед свадьбой покупает пару колец. Построить закон распределения, вычислить математическое ожидание и среднее квадратичное отклонение.

Решение. Вероятность события A – куплена кольцо оказалась качественной определим по формуле полной вероятности $P(A)=\frac{3}{10}0{,}95+\frac{7}{10}0{,}9=0{,}915.$

Случайная величина X – количество колец надлежащего качества среди купленных имеет биномиальное закон распределения с параметрами

$n=2;\ p=0{,}915;\ q=1-0{,}915=0{,}085.$

Найдем соответствующие вероятности $P(0)=q^2=0{,}085^2\approx 0{,}007;$

$P(1)=C_2^1pq^{2-1}=2\cdot 0{,}915\cdot 0{,}085\approx 0{,}155;\quad P(2)=p^2=0{,}915^2\approx 0{,}837;$

Запишем таблицу распределения

$X=k$	0	1	2
P_k	0,007	0,155	0,837

На основе табличных данных вычисляем математическое ожидание

$M(X)=0\cdot 0{,}007+1\cdot 0{,}155+2\cdot 0{,}837=1{,}829;$

дисперсию

$D(X)=0^2\cdot 0{,}007+1^2\cdot 0{,}155+2^2\cdot 0{,}837=3{,}503;$

Среднее квадратичное отклонение $\sigma(X)=\sqrt{D(X)}=\sqrt{3{,}503}\approx 1{,}872.$

Как можно убедиться из примеров, биномиальний закон распределения простой как для понимания так и для вычислений.

2. Распределение Пуассона, его свойства, числовые характеристики. Связь распределений ДСВ с нормальным распределением.

Дискретная случайная величина X имеет закон распределения Пуассона, если вероятности ее возможных значений $P_k=P(X=k)=\frac{a^k}{k!}e^{-a},k=\ 0,\ 1,\ 2\ ,3,\ \ldots,n,$ вычисляется по формуле Пуассона, где a=np<10. Как правило, Пуассоновское распределение касается вероятности появления благоприятного события в большом количестве экспериментов, если в одном - вероятность успешного завершения стремится к нулю.

1. Математическое ожидание определяется по формуле $M(X)=a=np.$

2. Дисперсия определяется по формуле $D(X)=a;$

Среднее квадратическое отклонение $\sigma(X)=\sqrt{a}.$

Следовательно, для пуассоновского закона распределения вероятностей математическое ожидание и дисперсия равны произведению количества опытов на вероятность благоприятной события $M(X)=D(X)=a.$

На практике, если математическое ожидание и дисперсия близкие по значению то принимают гипотезу, что исследуемая величина имеет закон распределения Пуассона.

3. Асимметрия и эксцесс для пуассоновский закон также уровни и вычисляются по формулам $A(X)=\frac{1}{\sqrt{a}};E(X)=\frac{1}{\sqrt{a}}.$

Задача 1. Микропроцессор имеет 10000 транзисторов, работающих независимо друг от друга. Вероятность того, что транзистор выйдет из строя во время работы прибора, является величиной маловероятной и составляет 0,0007. Определить математическое ожидание М (X) и среднее квадратическое отклонение S (X) случайной величины X — числа транзисторов, выйдут из строя во время работы процессора.

Решение. Задача удовлетворяет всем законам пуассоновский распределения:
количество испытаний n=10000 велика;
вероятность p=0,0007 близка к нулю;
их произведение a=np=7<10.
На основе данных вычисляем заданные величины

$$M(X) = np = 10000 \cdot 0,0007 = 7; \quad D(X) = M(X) = np = 7; \quad \sigma(X) = \sqrt{np} = \sqrt{7} \approx 2,64$$

Геометрический закон распределения имеет место в таких науках как микробиология, генетика, физика. На практике эксперимент или опыт осуществляют до первого появления успешной события A. Число проведенных попыток будет целочисленной случайной величиной 1,2,.... Вероятность появления события A в каждом опыте не зависит от предыдущих и составляет p, q=1-p. Вероятности возможных значений случайной величины X определяется зависимостью

$$P_k = P(X = k) = pq^{k-1}, \quad (k = 1, 2, 3, \ldots, n);$$

Есть во всех предыдущих опытах кроме k-го эксперимент дал плохой результат и только в k-му был успешным. Данную формулу вероятностей называют геометрическим законом распределения, поскольку правая его часть совпадает с выражением общего элемента геометрической прогрессии.

1. Математическое ожидание $M(X) = \frac{1}{p}.$

2. Дисперсию и среднее квадратическое отклонение по формулам $D(X) = \frac{q}{p^2}; \ \sigma(X) = \frac{\sqrt{q}}{p}.$

3. Коэффициент асимметрии и эксцесса для геометрического распределения определяют по формуле

$$A(X) = \frac{2-p}{\sqrt{1-p}} = \frac{2-p}{\sqrt{q}}. \quad E(X) = 6 + \frac{p^2}{1-p} = 6 + \frac{p^2}{q}.$$

Пример 1. Игральная кость подбрасывается до первого появления цифры 1. Определить все числовые характеристики M (X), D (X), S (X), A(X), E(X) для случайной величины X числа осуществляемых подбрасываний.

Решение. По условию задачи случайная величина X является целочисленной с геометрическим закон распределения вероятностей. Вероятность успешного подбрасывания величина постоянная и равна единице разделенной на количество граней кубика

$$p = \frac{1}{6}; q = \frac{5}{6}.$$

Имея p,q необходимые числовые характеристики X находим по приведенным выше формулам

$$M(X) = \frac{1}{p} = \frac{1}{\frac{1}{6}} = 6; \quad D(X) = \frac{q}{p^2} = \frac{\frac{5}{6}}{\frac{1}{36}} = 30; \quad \sigma(X) = \sqrt{30} \approx 5,48;$$

$$A(X) = \frac{2-p}{\sqrt{1-p}} = \frac{2-\frac{1}{6}}{\sqrt{\frac{5}{6}}} \approx 1,67; \quad E(X) = 6 + \frac{p^2}{1-p} = 6 + \frac{\left(\frac{1}{6}\right)^2}{\frac{5}{6}} \approx 6,033.$$

Гипергеометрический закон распределения вероятностей столь тяжелый при первом ознакомлении, что лучше всего его объяснять на конкретном примере.

Пусть задано некоторое множество однотипных элементов, число которых равно N; из них K элементов имеют, например, признак A (цвет, стандартность, наполнения), а остальные N-K элементов - признак B. С этого множества наугад берут n элементов. Случайная величина X – число элементов с признаком вида A, что случается среди n наугад взятых элементов. Тогда X принимает значения k=0,1,2,...,min(n,K) , а вероятность их появления определяется гипергеометрическим законом распределения

$$P_k = P(X = k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}.$$

$k \le n; n \le N;$

Числовые характеристики этого закона вычисляются по приведенным ниже формулам:

1. Математическое ожидание

$$M(X) = \frac{1}{C_N^n}\sum_{k=0}^{m} kC_K^k C_{N-K}^{n-m} = n\frac{M}{N}.$$

2. Дисперсия и среднее квадратическое отклонение

$$D(X) = \frac{1}{C_N^n}\sum_{k=0}^{m} k^2 C_K^k C_{N-K}^{n-k} - M^2(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)};$$

$$\sigma(X) = \sqrt{D(X)} = \frac{1}{N}\sqrt{\frac{nM(N-M)(N-n)}{(N-1)}}.$$

3. Для асимметрии

$$A(X) = \frac{(N-2M)(N-2n)\sqrt{N-1}}{nM(N-M)(N-2)\sqrt{N-n}};$$

и эксцесса

$$E(X) = \frac{N^2(N-1)(N-2n)}{n(N-2)(N-3)(N-n)} \times$$
$$\times\left(\frac{N(N+1)-6N(N-n)}{M(N-M)} + \frac{3n(N-n)(N+6)}{N^2} - 6\right).$$

Рассмотрим несколько примеров на применение приведенных выше формул

Пример 1. В ящике содержится 10 однотипных деталей, из них 7 стандартных, а остальные являются бракованными. Наугад из ящика берут m деталей. Построить законы распределения целочисленной случайной величины X — появление числа стандартных деталей среди m наугад взятых и вычислить математическое ожидание M(X), дисперсию D (X), и среднее математическое отклонение S(X), если m = 3

Решение. Построим гипергеометрические законы распределения:

Имеем следующие начальные условия для случая выбора трех деталей

n = 3; N=10; K= 7; N-K= 3; k = 0, 1, 2, 3.

В табличной форме гипергеометрический закон для этих данных имеет вид

$X = x_k = k$	0	1	2	3
$P_k = \frac{C_7^k C_3^{3-k}}{C_{10}^3}$	$\frac{C_7^0 C_3^3}{C_{10}^3}$	$\frac{C_7^1 C_3^2}{C_{10}^3}$	$\frac{C_7^2 C_3^1}{C_{10}^3}$	$\frac{C_7^3 C_3^0}{C_{10}^3}$

или после вычисления сочетаний

$$\frac{C_7^0 C_3^3}{C_{10}^3} = \frac{1\cdot 1}{\frac{10!}{7!3!}} = \frac{3!}{10\cdot 9\cdot 8} = \frac{3\cdot 2}{10\cdot 3\cdot 3\cdot 4\cdot 2} = \frac{1}{120};$$

$$\frac{C_7^1 C_3^2}{C_{10}^3} = \frac{7!}{6!1!}\cdot\frac{3!}{1!2!}\cdot\frac{7!3!}{10!} = \frac{7\cdot 3\cdot 3\cdot 2}{10\cdot 9\cdot 8} = \frac{7\cdot 3\cdot 3\cdot 2}{10\cdot 3\cdot 3\cdot 4\cdot 2} = \frac{7}{40};$$

$$\frac{C_7^2 C_3^1}{C_{10}^3} = \frac{7!}{5!2!}\cdot\frac{3!}{2!1!}\cdot\frac{7!3!}{10!} = \frac{7\cdot 6\cdot 3\cdot 3\cdot 2}{2\cdot 10\cdot 9\cdot 8} = \frac{7\cdot 3\cdot 2\cdot 3\cdot 2}{2\cdot 10\cdot 3\cdot 3\cdot 4\cdot 2} = \frac{7}{40};$$

$$\frac{C_7^3 C_3^0}{C_{10}^3}=\frac{7!}{4!3!}\cdot 1\cdot\frac{7!3!}{10!}=\frac{7\cdot 6\cdot 5\cdot 3\cdot 2}{3\cdot 2\cdot 10\cdot 9\cdot 8}=\frac{7\cdot 3\cdot 2\cdot 5\cdot 3\cdot 2}{3\cdot 2\cdot 5\cdot 2\cdot 3\cdot 3\cdot 8}=\frac{7}{24};$$

в виде таблицы вероятностей

k	0	1	2	3
$P_k=\frac{C_7^k C_3^k}{120}$	$\frac{1}{120}$	$\frac{21}{120}$	$\frac{63}{120}$	$\frac{35}{120}$

Условие нормирования

$$\sum P_k=\frac{1+21+63+35}{120}=\frac{120}{120}=1.$$

выполняется, следовательно все верно посчитано. Не ленитесь проверять его, оно намного скорее укажет Вам на присутствие ошибки при неправильной правой части. Вычисляем числовые характеристики:

Математическое ожидание

$$M(X)=\sum kp_k=0\frac{1}{120}+1\frac{21}{120}+2\frac{63}{120}+3\frac{35}{120}=\frac{21+126+105}{120}=\frac{252}{120}=2{,}1;$$

Дисперсию

$$M(X^2)=\sum k^2 p_k=1\frac{21}{120}+4\frac{63}{120}+9\frac{35}{120}=\frac{21+252+315}{120}=\frac{588}{120}=4{,}9;$$

$$D(X)=M(X^2)-M^2(X)=4{,}9-(2{,}1)^2=4{,}9-4{,}41=0{,}49;$$

Среднее квадратичное отклонение $\sigma(X)=\sqrt{0{,}49}=0{,}7.$

3. Равномерное распределение. Показательное распределение.

СВ имеет равномерное распределение, если некотором интервале плотность вероятностей принимает постоянное значение

$$f(x)=\begin{cases}0, & x\le a;\\ \frac{1}{b-a}, & a<x\le b;\\ 0, & x>b.\end{cases}$$

Функция распределения вероятностей для равномерного закона определяется интегрированием

$$F(x)=\int_{-\infty}^{x} f(x)=\begin{cases}0, & x\le a;\\ \frac{x-a}{b-a}, & a<x\le b;\\ b, & x>b.\end{cases}$$

Математическое ожидание в таких случаях определяют зависимости $M(X)=\frac{a+b}{2};$

дисперсию по формуле $D(X)=\frac{(b-a)^2}{12};$

и среднее квадратическое отклонение через корень $S(X)=\sqrt{D(X)}=\frac{b-a}{2\sqrt{3}}.$

Вероятность попадания случайной величины X в некоторый интервал $[\alpha;\beta]$, содержащийся внутри интервала $[\alpha;\beta]\in[a;b]$ определяется по формуле $P(\alpha\le X\le\beta)=\frac{\beta-\alpha}{b-a}.$

Пример 1. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение M (X), D (X), S (X), Найти математическое ожидание, дисперсию и среднее квадратическое отклонение X имеет равномерный закон распределения и возможные значения ее значение лежит в диапазоне 1..50:

$$X_k = k = 1,\ 2,\ 3,\ \ldots,\ 50.$$

Решение. По условию задачи имеем следующие данные n = 50, p = 1/50=0,02.
Согласно формулам вычисляем математическое ожидание

$$M(X) = \frac{n+1}{2} = \frac{50+1}{2} = 25{,}5$$

дисперсия

$$D(X) = \frac{n^2-1}{12} = \frac{50^2-1}{12} = 208{,}25$$

среднее квадратическое отклонение

$$\sigma(X) = \frac{\sqrt{n^2-1}}{2\sqrt{3}} = \frac{\sqrt{2499}}{2\sqrt{3}} \approx 14{,}43.$$

Показательное распределение.

Показатель	Показательный закон распределения
Определение	***Показательным (экспоненциальным) называется*** распределение вероятностей непрерывной случайной величины X, которое описывается плотностью, имеющей вид $f(x)=\begin{cases}0 & \text{при } x<0,\\ \lambda e^{-\lambda x} & \text{при } x\ge 0,\end{cases}$ где λ – постоянная положительная величина
Функция распределения	$F(x)=\begin{cases}0 & \text{при } x<0,\\ 1-e^{-\lambda x} & \text{при } x\ge 0.\end{cases}$
Вероятность попадания в интервал	$P(a<X<b) = \int_a^b \lambda e^{-\lambda x}dx = e^{-\lambda a} - e^{-\lambda b}$
Математическое ожидание	$M(X)=\frac{1}{\lambda}$
Дисперсия	$D(X)=\frac{1}{\lambda^2}$
Среднее квадратическое отклонение	$\sigma(X)=\frac{1}{\lambda}$

Пример 1. Показательное распределение задано при x ≥ 0 плотностью f(x) = 5e – 5x. Требуется: а) записать выражение для функции распределения; б) найти вероятность того, что в результате испытания X попадает в интервал (1;4); в) найти вероятность того, что в результате испытания X ≥ 2 ; г) вычислить M(X), D(X), σ(X).

Решение.

1. Поскольку по условию задано показательное распределение, то из формулы плотности распределения вероятностей случайной величины X получаем λ = 5. Тогда функция распределения будет иметь вид:

$$F(x)=\begin{cases}0 & \text{при } x<0,\\ 1-e^{-5x} & \text{при } x\ge 0;\end{cases}$$

2. Вероятность того, что в результате испытания X попадает в интервал (1;4) будем находить по формуле: P(a < X < b) = e−λa − e−λb.
P(1 < X < 4) = e−5*1 − e−5*4 = e−5 − e−20.

3. Вероятность того, что в результате испытания $X \geq 2$ будем находить по фор муле: $P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}$ при a=2, b=∞.

$P(X\geq 2) = P(1< X < 4) = e^{-\lambda*2} - e^{-\lambda*\infty} = e^{-2\lambda} - e^{-\infty} = e^{-2\lambda} - 0 = e^{-10}$ (т.к. предел e^{-x} при
x стремящемся к ∞ равен нулю).

4. Находим для показательного распределения:
математическое ожидание по формуле $M(X) =1/\lambda = 1/5 = 0,2$;
дисперсию по формуле $D(X) = 1/ \lambda 2= 1/25 = 0,04$;
среднее квадратическое отклонение по формуле $\sigma(X) = 1/\lambda = 1/5 = 1,2$.

Пример 2 . Непрерывная случайная величина X распределена по показательному закону
$f(x) = 5e^{-5x}$ При $x \geq 0$; $f(x) = 0$ при $x < 0$. Найти математическое ожидание, среднее квадратическое отклонение и дисперсию X.

Решение: По условию, $\lambda = 5$. Следовательно,
$M(X) = \sigma(X) = 1/\lambda = 1/5 = 0.2$,
$D(X) = 1/\lambda^2 = 1/5^2 = 0,04$

4. Нормальное распределение, его свойства.

Дадим понятие нормального закона распределения, функции распределения такого закона, порядка вычисления вероятности попадания случайной величины X в определенный интервал.

№	Показатель	Нормальный закон распределения	Примечание
1	Определение	***Нормальным называется*** распределение вероятностей непрерывной случайной величины X, плотность которого имеет вид	
		$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{\frac{-(x-m_x)^2}{2\sigma_x^2}}$	где m_x – математическое ожидание случайной величины X, σ_x – среднее квадратическое отклонение
2	***Функция распределения***	$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt$	$u = \frac{x - m_x}{\sigma_x}$
3	***Вероятность*** попадания в интервал (a;b)	$P(a \leq X \leq b) = \Phi(\frac{b - m_x}{\sigma_x}) - \Phi($	$\Phi(u) = F(u) - 1/2 = \frac{1}{\sqrt{2\pi}} \int_0^u e$ - интегральная функция Лапласа
4	***Вероятность*** того, что абсолютная величина отклонения меньше положи-	$P(\|X- m_x\|< \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$	при m_x= 0 $P(\|X\| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$

	тельного числа δ		

Задача. Математическое ожидание нормально распределенной случайной величины $a = 10$, а среднее квадратическое отклонение - $\sigma = 2$. Найти вероятность того, что в результате испытания X примет значение из интервала $12 < X < 14$ и записать закон распределения.

Решение. Запишем вначале закон распределения. Общая формула имеет вид: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}$.

Подставляя $a = 10$ и $\sigma = 2$, получим: $f(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$.

Вероятность того, что X примет значение из интервала $\alpha < X < \beta$ имеет вид:

$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$, где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$ – функция Лапласа.

Значения этой функции находятся с помощью таблицы.

В нашем случае: $P(12 < X < 14) = \Phi\left(\frac{14 - 10}{2}\right) - \Phi\left(\frac{12 - 10}{2}\right) = \Phi(2) - \Phi(1)$.

По таблице находим: $\Phi(2) = 0{,}47725; \quad \Phi(1) = 0{,}34134$, следовательно: $P(12 < X < 14) = 0{,}47725 - 0{,}34134 = 0{,}13591$.

Задача. Длина X некоторой детали представляет собой случайную величину, распределенную по нормальному закону распределения, и имеет среднее значение 20 мм и среднее квадратическое отклонение – 0,2 мм.

Необходимо:

а) записать выражение плотности распределения;

б) найти вероятность того, что длина детали будет заключена между 19,7 и 20,3 мм;

в) найти вероятность того, что величина отклонения не превышает 0,1 мм;

г) определить, какой процент составляют детали, отклонение которых от среднего значения не превышает 0,1 мм;

д) найти, каким должно быть задано отклонение, чтобы процент деталей, отклонение которых от среднего не превышает заданного, повысился до 54%;

е) найти интервал, симметричный относительно среднего значения, в котором будет находиться X с вероятностью 0,95.

Решение. а) Плотность вероятности случайной величины X, распределенной по нормальному закону находим

$$f(x) = \frac{1}{0{,}2\sqrt{2\pi}} \cdot e^{-\frac{(x-20)^2}{0{,}08}}$$

при условии, что m_x=20, σ =0,2.

б) Для нормального распределения случайной величины вероятность попасть в интервал (19,7; 20,3) определяется

Ф((20,3-20)/0,2) – Ф((19,7-20)/0,2) = Ф(0,3/0,2) – Ф(-0,3/0,2) = 2Ф(0,3/0,2) = 2Ф(1,5) = 2*0,4332 = 0,8664.

http://www.ekonomika-st.ru/drugie/metodi/t-ver-1-4-0.htmlЗначение Ф(1,5) = 0,4332 мы нашли в приложениях, в таблице значений интегральной функции Лапласа Ф(x) (

в) Вероятность того, что абсолютная величина отклонения меньше положительного числа 0,1 найдем

$P(|X-20| < 0{,}1) = 2\Phi(0{,}1/0{,}2) = 2\Phi(0{,}5) = 2*0{,}1915 = 0{,}383$.

http://www.ekonomika-st.ru/drugie/metodi/t-ver-1-4-0.htmlЗначение $\Phi(0{,}5) = 0{,}1915$ мы нашли в приложениях, в таблице значений интегральной функции Лапласа $\Phi(x)$

г) Поскольку вероятность отклонения, меньшего 0,1 мм, равна 0,383, то отсюда следует, что в среднем 38,3 детали из 100 окажутся с таким отклонением, т.е. 38,3%.

д) Поскольку процент деталей, отклонение которых от среднего не превышает заданного, повысился до 54%, то $P(|X-20| < \delta) = 0{,}54$. Отсюда следует, что $2\Phi(\delta/\sigma) = 0{,}54$, а значит $\Phi(\delta/\sigma) = 0{,}27$.

http://www.ekonomika-st.ru/drugie/metodi/t-ver-1-4-0.htmlИспользуя приложение, находим $\delta/\sigma = 0{,}74$. Отсюда $\delta = 0{,}74*\sigma =$

$0{,}74*0{,}2 = 0{,}148$ мм.

е) Поскольку искомый интервал симметричен относительно среднего значения $m_x = 20$, то его можно определить как множество значений X, удовлетворяющих неравенству

$20 - \delta < X < 20 + \delta$ или $|x - 20| < \delta$.

По условию вероятность нахождения X в искомом интервале равна 0,95, значит

$P(|x - 20| < \delta) = 0{,}95$. С другой стороны $P(|x - 20| < \delta) = 2\Phi(\delta/\sigma)$, следовательно

$2\Phi(\delta/\sigma) = 0{,}95$, а значит $\Phi(\delta/\sigma) = 0{,}475$.

http://www.ekonomika-st.ru/drugie/metodi/t-ver-1-4-0.htmlИспользуя приложение находим $\delta/\sigma = 1{,}96$. Отсюда $\delta = 1{,}96*\sigma = 1{,}96*0{,}2 = 0{,}392$.

Искомый интервал: (20 – 0,392; 20 + 0,392) или (19,608; 20,392).

3.3.3 Результаты и выводы: В результате проведенного занятия студенты:
- освоили основные понятия, связанные с определением закона распределения СВ;
- усвоили способы задания СВ, распределенных биномиально, по закону Пуассона, геометрически, гипергеометрически;
- выработали навыки по вычислению числовых характеристик СВ, распределенных биномиально, по закону Пуассона, геометрически, гипергеометрически.
- освоили основные понятия, связанные с определением закона распределения НСВ;
- усвоили способы задания СВ, распределенных равномерно, показательно, нормально;
- выработали навыки по вычислению числовых характеристик СВ, распределенных равномерно, показательно, нормально; усвоили основные свойства СВ, распределенных нормально, свойства кривой Гаусса, правило трех сигм.

3.4 Практическое занятие 5-6 (ПЗ-5-6)

Тема: Случайный вектор. Распределение многомерной СВ. Условные законы распределения. Числовые характеристики случайного вектора. Коэффициент корреляции.

3.4.1 Задание для работы:

1. Системы двух случайных величин. Функция и плотность системы СВ, их свойства
2.Условные распределения плотностей ССВ, условные законы распределения.
3. Корреляционный момент ССВ, его свойства. Коэффициент корреляции. Корреляционная матрица

3.4.2 Краткое описание проводимого занятия

1. Системы двух случайных величин. Функция и плотность системы СВ, их свойства

Пример 1. Дана двумерная функция распределения: $F_{X,Y}(x,y)=\sin x\sin y$, где $0\le x\le\frac{\pi}{2}$, $0\le y\le\frac{\pi}{2}$. Найти вероятность попадания случайной точки $(X;Y)$ в прямоугольник, ограниченный прямыми $x=\frac{\pi}{6}$, $x=\frac{\pi}{2}$, $y=\frac{\pi}{4}$, $y=\frac{\pi}{3}$.

Решение. $\mathrm{P}\left\{\frac{\pi}{6}\le X<\frac{\pi}{2},\frac{\pi}{4}\le Y<\frac{\pi}{3}\right\}=\left(\sin\frac{\pi}{2}\sin\frac{\pi}{3}-\sin\frac{\pi}{6}\sin\frac{\pi}{3}\right)-$

$$-\left(\sin\frac{\pi}{2}\sin\frac{\pi}{4}-\sin\frac{\pi}{6}\sin\frac{\pi}{4}\right)=\left(\frac{\sqrt{3}}{2}-\frac{\sqrt{3}}{4}\right)-\left(\frac{\sqrt{2}}{2}-\frac{\sqrt{2}}{4}\right)=\frac{\sqrt{3}-\sqrt{2}}{4}.$$

Ответ: $0{,}25\cdot(\sqrt{3}-\sqrt{2})$.

Пример 2. Закон распределения двумерного дискретного случайного вектора $(X;Y)$ задан таблицей:

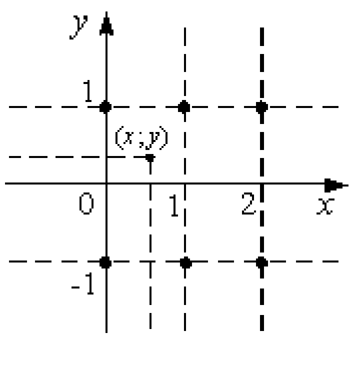
Y / *X*	–1	1
0	0,1	0,06
1	0,3	0,18
2	0,2	0,16

Найти: одномерные законы распределения компонент *X* и *Y*; вероятность $\mathrm{P}\{X\ge Y\}$. Составить функцию распределения $F_{X,Y}(x,y)$.

Решение. 1) Одномерные законы $p_{i\bullet}$ и $p_{\bullet j}$ распределения компонент *X* и *Y* соответственно построены в таблице:

Y / *X*	–1	1	$p_{i\bullet}$
0	0,1	0,06	0,16
1	0,3	0,18	0,48
2	0,2	0,16	0,36
$p_{\bullet j}$	0,6	0,4	1

2) $\mathrm{P}\{X\ge Y\}=1-\mathrm{P}\{X<Y\}=1-\mathrm{P}\{X=0,Y=1\}=1-0{,}06=0{,}94$.



3) Согласно определению функции распределения $F_{X,Y}(x,y)=\mathrm{P}\{X<x,Y<y\}$. Напомним, что геометрически значение $F_{X,Y}(x,y)$ – это вероятность попадания случайной точки $(X;Y)$ в бесконечный квадрант с вершиной $(x;y)$. Для вершины этого квадранта, согласно условию задачи, есть двенадцать областей, образованных тремя вертикальными прямыми $x=0$, $x=1$, $x=2$ и двумя горизонтальными прямыми $y=-1$, $y=1$.

На рис. показан случай, когда вершина $(x;y)$ находится внутри прямоугольника $0<x\le1$, $-1<y\le1$. При этом внутри квадранта находится толь-

ко одна точка с координатами $(0;-1)$, в которой имеется ненулевая вероятность, равная 0,1. Функцию распределения $F_{X,Y}(x,y)$ удобно задавать в виде таблицы (ее значение для случая, когда вершина $(x;y)$ квадранта находится внутри ка $\mathbf{0}<x\le\mathbf{1}$, $-1<y\le 1$ выделено жирным шрифтом):

y / x	$y\le -1$	$-1<y\le 1$	$y>1$
$x\le\mathbf{0}$	0	0	0
$\mathbf{0}<x\le\mathbf{1}$	0	**0,1**	0,16
$\mathbf{1}<x\le\mathbf{2}$	0	0,4	0,64
$x>\mathbf{2}$	0	0,6	1

Пример 3. Известна функция распределения $F_{X,Y}(x,y)$ двумерного дискретного случайного вектора $(X;Y)$:

y / x	$y\le 1$	$1<y\le 2$	$2<y\le 3$	$y>3$
$x\le\mathbf{0}$	0	0	0	0
$\mathbf{0}<x\le\mathbf{1}$	0	0,5	0,5	0,5
$\mathbf{1}<x\le\mathbf{2}$	0	0,5	0,75	0,75
$\mathbf{2}<x\le\mathbf{3}$	0	0,5	0,75	0,875
$x>\mathbf{3}$	0	0,5	0,75	1

Составить функции распределения $F_X(x)$ и $F_Y(y)$ компонент *X* и *Y*, а затем построить их законы распределения.

Решение. Учитывая, что $F_{X,Y}(x,+\infty)=F_X(x)$, $F_{X,Y}(+\infty,y)=F_Y(y)$, получим («проходя» соответственно по последнему столбцу и последней строке таблицы):

$$F_X(x)=\begin{cases}0, & x\le 0,\\ 0{,}5, & 0<x\le 1,\\ 0{,}75, & 1<x\le 2,\\ 0{,}875, & 2<x\le 3,\\ 1, & x>3;\end{cases}\qquad F_Y(y)=\begin{cases}0, & y\le 1,\\ 0{,}5, & 1<y\le 2,\\ 0{,}75, & 2<y\le 3,\\ 1, & y>3.\end{cases}$$

Значит, для случайной величины *X* функция распределения испытывает «скачки» в точках $x=\mathbf{0};\mathbf{1};\mathbf{2};\mathbf{3}$, для случайной величины *Y* – в точках $y=\mathbf{1};\mathbf{2};\mathbf{3}$. Поэтому законы распределения компонент выглядят следующим образом:

X	0	1	2	3
$p_{i\bullet}$	0,5	0,25	0,125	0,125

Y	1	2	3
$p_{\bullet j}$	0,5	0,25	0,25

Пример 4. Известна функция плотности двумерного случайного вектора $(X;Y)$:

$$f_{X,Y}(x,y)=\frac{1}{\pi^2(1+x^2)(1+y^2)},\ x,y\in\mathbf{R}.$$

Составить функцию распределения $F_{X,Y}(x,y)$.

Решение. По определению функции распределения

$$F_{X,Y}(x,y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y}f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v,$$

поэтому:

$$F_{X,Y}(x,y)=\frac{1}{\pi^2}\int\limits_{-\infty}^{y}\mathrm{d}v\int\limits_{-\infty}^{x}\frac{1}{(1+u^2)(1+v^2)}\mathrm{d}u=\frac{1}{\pi^2}\operatorname{arctg}u\Big|_{-\infty}^{x}\cdot\operatorname{arctg}v\Big|_{-\infty}^{y}=$$

$$=\frac{1}{\pi^2}\left(\operatorname{arctg}x+\frac{\pi}{2}\right)\left(\operatorname{arctg}y+\frac{\pi}{2}\right).$$

Ответ: $F_{X,Y}(x,y)=\frac{1}{\pi^2}\left(\operatorname{arctg}x+\frac{\pi}{2}\right)\left(\operatorname{arctg}y+\frac{\pi}{2}\right)$.

Пример 5 Найти плотность распределения двумерного случайного вектора $(X;Y)$, если известна функция распределения

$$F_{X,Y}(x,y)=\begin{cases}1+\mathrm{e}^{-2x-3y}-\mathrm{e}^{-2x}-\mathrm{e}^{-3y}, & x\ge 0,\ y\ge 0,\\ 0, & \text{иначе}.\end{cases}$$

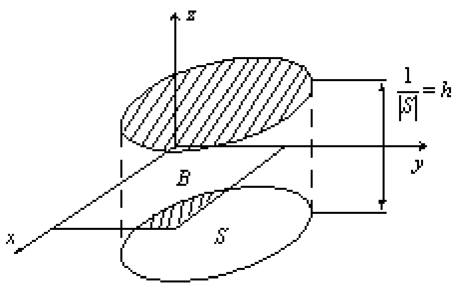
Решение. Согласно свойству 2 $f_{X,Y}(x,y)=\frac{\partial^2 F_{X,Y}(x,y)}{\partial x\,\partial y}$ во всех точках непрерывности функции $f_{X,Y}(x,y)$. Поэтому

$$f_{X,Y}(x,y)=\begin{cases}6\cdot\mathrm{e}^{-2x-3y}, & x\ge 0,\ y\ge 0,\\ 0, & \text{иначе}.\end{cases}$$

Ответ: $f_{X,Y}(x,y)=\begin{cases}6\cdot\mathrm{e}^{-2x-3y}, & x\ge 0,\ y\ge 0,\\ 0, & \text{иначе}.\end{cases}$

Пример 6 (*двумерное равномерное распределение*). Плотность $f_{X,Y}(x,y)$ равномерного распределения на области $S\subset\mathbf{R}^2$ конечной двумерной площади $|S|$:

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{|S|}, & (x;y)\in S,\\ 0, & (x;y)\notin S.\end{cases}$$



Вероятность $\mathrm{P}\{(X;\,Y)\in B\}$ в этом случае определяется отношением площадей $B\cap S$ и S (рис.):

$$\mathrm{P}\{(X;\,Y)\in B\}=\frac{|B\cap S|}{|S|}.$$

Замечание. По последней формуле вычисляются так называемые геометрические вероятности.

Пример 7. Двумерный случайный вектор $(X;Y)$ подчинен закону распределения с плотностью

$$f_{X,Y}(x,y)=\begin{cases}a(x+y), & (x;y)\in D;\\ 0, & (x;y)\notin D.\end{cases}$$

Область D – треугольник, ограниченный прямыми $x+y-3=0$, $x=0$, $y=0$. Найти коэффициент a.

Решение. Согласно условию нормировки $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy = 1$. Поскольку только в области *D* подынтегральная функция $f_{X,Y}(x,y)$ отлична от нуля, то имеем уравнение

$$\iint\limits_D f_{X,Y}(x,y)\,dx\,dy = 1, \text{ или } \int\limits_0^3 dx \int\limits_0^{3-x} a(x+y)\,dy = 1.$$

Тогда

$$\int\limits_0^3 dx \int\limits_0^{3-x} a(x+y)\,dy = a\int\limits_0^3 dx\left(xy+\frac{y^2}{2}\right)\Bigg|_0^{3-x} = a\int\limits_0^3 (x(3-x)+0{,}5(3-x)^2)\,dx =$$

$$= a\int\limits_0^3 (3x - x^2 + 4{,}5 - 3x + 0{,}5x^2)\,dx = a\left(\frac{9x}{2}-\frac{x^3}{6}\right)\Bigg|_0^3 = a\left(\frac{27}{2}-\frac{9}{2}\right) = 9a.$$

Отсюда, решая уравнение $9a = 1$, получим $a = 1/9$.

Ответ: $a = 1/9$.

2.Условные распределения плотностей ССВ, условные законы распределения.

Если случайные величины, образующие систему, зависимы, то для нахождения закона распределения системы недостаточно знать законы распределения отдельных величин, входящих в систему. Требуется еще знать так называемый условный закон распределения одной из них.

Определение. *Условным законом распределения* одной из величин $(X;Y)$, входящих в систему, называется ее закон распределения, вычисленный при условии, что другая случайная величина приняла определенное значение.

Начнем с наиболее простого случая. Пусть случайная величина *Y* является дискретной.

Определение. *Условной функцией распределения* $F_X(x\,|\,Y=y_j)$ случайной величины *X* при условии $Y = y_j$ называется условная вероятность события $\{X<x\}$ при условии события $\{Y=y_j\}$, т.е.

$$F_X(x\,|\,Y=y_j) = \mathrm{P}\{X<x\,|\,Y=y_j\} = \frac{\mathrm{P}\{X<x, Y=y_j\}}{\mathrm{P}\{Y=y_j\}}.$$

Аналогично определяется условная функция распределения $F_Y(y\,|\,X=x_i)$ случайной величины *Y* при условии $X = x_i$ (когда случайная величина *X* является дискретной):

$$F_Y(y\,|\,X=x_i) = \mathrm{P}\{Y<y\,|\,X=x_i\} = \frac{\mathrm{P}\{Y<y, X=x_i\}}{\mathrm{P}\{X=x_i\}}.$$

Замечание 1. Условная функция распределения обладает всеми свойствами, которые присущи обычной (безусловной) функции распределения.

Замечание 2. Если случайная величина *X* также *дискретная*, причем $\mathrm{P}\{X=x_i, Y=y_j\} = p_{ij}$, то удобно рассматривать условную вероятность случайной величине *X* принять значение x_i при условии $Y = y_j$:

$$\mathrm{P}\{X=x_i\,|\,Y=y_j\} = \frac{\mathrm{P}\{X=x_i, Y=y_j\}}{\mathrm{P}\{Y=y_j\}} = \frac{p_{ij}}{p_{\bullet j}}.$$

Обычно условное распределение $\{X = x_i | Y = y_j\}$ описывают с помощью таблицы. Ясно, что элементы второй строки этой таблицы получаются по формулам $\dfrac{p_{ij}}{p_{\bullet j}}$.

Аналогично определяется условная вероятность случайной величине *Y* принять значение y_j при условии $X = x_i$:

$$\mathrm{P}\left\{Y = y_j \middle| X = x_i\right\} = \frac{\mathrm{P}\{X = x_i, Y = y_j\}}{\mathrm{P}\{X = x_i\}} = \frac{p_{ij}}{p_{i\bullet}}$$

Пример 1. Закон распределения случайного вектора $(X;Y)$ задан таблицей:

Y / *X*	1	2
–1	0,3	0,25
0	0,1	0,05
1	0,2	0,1

Описать условные законы распределения: 1) случайной величины *X* при условии $Y = 1$; 2) случайной величины *Y* при условии $X = -1$.

Решение. Найдем безусловные законы распределения компонент *X* и *Y*:

Y / *X*	1	2	$p_{i\bullet}$
–1	0,3	0,25	0,55
0	0,1	0,05	0,15
1	0,2	0,1	0,3
$p_{\bullet j}$	0,6	0,4	1

1) Условные вероятности случайной величине *X* принять значения x_i ($i = \mathbf{1, 2, 3}$) при условии $Y = 1$ вычисляются по формулам:

$$\mathrm{P}\left\{X = -1 \middle| Y = 1\right\} = \frac{\mathrm{P}\{X = -1, Y = 1\}}{\mathrm{P}\{Y = 1\}} = \frac{p_{11}}{p_{\bullet 1}} = \frac{0,3}{0,6} = 0,5,$$

$$\mathrm{P}\left\{X = 0 \middle| Y = 1\right\} = \frac{\mathrm{P}\{X = 0, Y = 1\}}{\mathrm{P}\{Y = 1\}} = \frac{p_{21}}{p_{\bullet 1}} = \frac{0,1}{0,6} = \frac{1}{6},$$

$$\mathrm{P}\left\{X = 1 \middle| Y = 1\right\} = \frac{\mathrm{P}\{X = 1, Y = 1\}}{\mathrm{P}\{Y = 1\}} = \frac{p_{31}}{p_{\bullet 1}} = \frac{0,2}{0,6} = \frac{1}{3}.$$

Тогда условный закон распределения случайной величины *X* при условии $Y = 1$ имеет вид:

X	–1	0	1	Контроль:
$\mathbf{P}\left\{X = x_i \middle| Y = \mathbf{1}\right\}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}+\frac{1}{6}+\frac{1}{3}=1$

2) Условные вероятности случайной величине *Y* принять значения y_j ($j = \mathbf{1, 2}$) при условии $X = -1$ вычисляются по формулам:

$$\mathrm{P}\left\{Y = 1 \middle| X = -1\right\} = \frac{\mathrm{P}\{X = -1, Y = 1\}}{\mathrm{P}\{X = -1\}} = \frac{p_{11}}{p_{1\bullet}} = \frac{0,3}{0,55} = \frac{6}{11},$$

$$\mathrm{P}\left\{Y = 2 \middle| X = -1\right\} = \frac{\mathrm{P}\{X = -1, Y = 2\}}{\mathrm{P}\{X = -1\}} = \frac{p_{12}}{p_{1\bullet}} = \frac{0,25}{0,55} = \frac{5}{11}.$$

Тогда условный закон распределения случайной величины Y при условии $X=-1$ имеет вид:

Y	1	2	Контроль:
$\mathbf{P}\{Y=y_j \mid X=-1\}$	$\frac{6}{11}$	$\frac{5}{11}$	$\frac{6}{11}+\frac{5}{11}=1$

В общем случае условную функцию распределения случайной величины X при условии $Y=y$ также естественно было бы определить формулой

$$F_X(x \mid Y=y)=\frac{\mathrm{P}\{X<x, Y=y\}}{\mathrm{P}\{Y=y\}}.$$

Однако это не всегда возможно (например, для непрерывной случайной величины Y событие $\{Y=y\}$ имеет нулевую вероятность). Во избежание этих неприятностей,

вместо события $\{Y=y\}$ рассматривается событие $\{y \le Y < y+\Delta\}$ и D устремляется к нулю. Тогда

$$\mathbf{P}\{X<x \mid y \le Y < y+\Delta\}=\frac{\mathrm{P}\{X<x, y \le Y < y+\Delta\}}{\mathrm{P}\{y \le Y < y+\Delta\}}=\frac{F_{X,Y}(x,y+\Delta)-F_{X,Y}(x,y)}{F_Y(y+\Delta)-F_Y(y)}.$$

Тогда *условной функцией распределения* $F_X(x \mid Y=y)$ называется предел

$$F_X(x \mid Y=y)=\lim_{\Delta\to 0} F_X(x \mid y \le Y < y+\Delta).$$

Оказывается, такой предел всегда существует. Аналогично

$$F_Y(y \mid X=x)=\lim_{\Delta\to 0} F_Y(y \mid x \le X < x+\Delta)$$

Если случайная величина Y *непрерывна*, то условную функцию распределения можно определить следующим выражением:

$$F_X(x \mid Y=y)=\frac{\frac{\partial}{\partial y}F_{X,Y}(x,y)}{f_Y(y)}.$$ Аналогично $$F_Y(y \mid X=x)=\frac{\frac{\partial}{\partial x}F_{X,Y}(x,y)}{f_X(x)}.$$

В наиболее важных для приложений случаях вектор $(X;Y)$ представляет собой двумерную непрерывную случайную величину с совместной плотностью $f_{X,Y}(x,y)$. Тогда

$$F_X(x \mid Y=y)=\frac{\int_{-\infty}^{x} f_{X,Y}(u,y)\,\mathrm{d}u}{f_Y(y)},\quad F_Y(y \mid X=x)=\frac{\int_{-\infty}^{y} f_{X,Y}(x,v)\,\mathrm{d}v}{f_X(x)}.$$

Нетрудно видеть, что условная функция распределения $F_X(x \mid Y=y)$ имеет производную по x, т.е. существует *условная плотность распределения* случайной величины X при условии $Y=y$:

$$f_X(x \mid Y=y)=\frac{\partial}{\partial x}F_X(x \mid Y=y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Аналогично

$$f_Y(y \mid X=x)=\frac{\partial}{\partial y}F_Y(y \mid X=x)=\frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Пример 2. Дана функция плотности $f_{X,Y}(x,y)$:

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{6\pi}, & \dfrac{x^2}{9}+\dfrac{y^2}{4}\le 1,\\ 0, & \dfrac{x^2}{9}+\dfrac{y^2}{4}>1;\end{cases}$$

и найдены безусловные плотности распределения компонент *X* и *Y*:

$$f_X(x)=\begin{cases}\dfrac{2}{9\pi}\sqrt{9-x^2}, & |x|\le 3,\\ 0, & |x|>3;\end{cases}\qquad f_Y(y)=\begin{cases}\dfrac{\sqrt{4-y^2}}{2\pi}, & |y|\le 2,\\ 0, & |y|>2.\end{cases}$$

Найти условные плотности распределения компонент *X* и *Y*.

Решение. Условные плотности распределения компонент *X* и *Y* находятся по формулам $f_X(x|Y=y)=\dfrac{f_{X,Y}(x,y)}{f_Y(y)}$, $f_Y(y|X=x)=\dfrac{f_{X,Y}(x,y)}{f_X(x)}$.

Поэтому

$$f_X(x|Y=y)=\begin{cases}\dfrac{1}{3\sqrt{4-y^2}}, & |x|\le\dfrac{3}{2}\sqrt{4-y^2},\\ 0, & |x|>\dfrac{3}{2}\sqrt{4-y^2};\end{cases}$$

$$f_Y(y|X=x)=\begin{cases}\dfrac{3}{4\sqrt{9-x^2}}, & |y|\le\dfrac{2}{3}\sqrt{9-x^2},\\ 0, & |y|>\dfrac{2}{3}\sqrt{9-x^2}.\end{cases}$$

Таким образом, случайная величина *X* при условии $Y=y$ *равномерно* распределена на отрезке $\left[-\dfrac{3}{2}\sqrt{4-y^2};\ \dfrac{3}{2}\sqrt{4-y^2}\right]$, а случайная величина *Y* при условии $X=x$ *равномерно* распределена на отрезке $\left[-\dfrac{2}{3}\sqrt{9-x^2};\ \dfrac{2}{3}\sqrt{9-x^2}\right]$. Условная плотность $f_X(x|Y=y)$ не определена при $|y|>2$, а условная плотность $f_Y(y|X=x)$ не определена при $|x|>3$.

3. Корреляционный момент ССВ, его свойства. Коэффициент корреляции. Корреляционная матрица

Числовые характеристики случайного вектора.
Условные числовые характеристики СВ.

Определение. *Начальным моментом порядка* $k+s$ системы двух случайных величин $(X;Y)$ называется действительное число $\alpha_{k,s}$, определяемое по формуле:

$\alpha_{k,s}=\sum\limits_i\sum\limits_j x_i^k y_j^s p_{ij}$, если $(X;Y)$ – система двух дискретных случайных величин;

$$\alpha_{k,s}=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}x^k y^s f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
, если $(X;Y)$ – система двух непрерывных случайных величин.

Определение. *Центральным моментом порядка* $k+s$ системы двух случайных величин $(X;Y)$ называется действительное число $\mu_{k,s}$, определяемое по формуле:

$$\mu_{k,s}=\sum_i\sum_j(x_i-m_X)^k(y_j-m_Y)^s p_{ij},$$

если $(X;Y)$ – система двух дискретных случайных величин;

$$\mu_{k,s}=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}(x-m_X)^k(y-m_Y)^s f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

если $(X;Y)$ – система двух непрерывных случайных величин.

На практике чаще всего встречаются моменты первого и второго порядков. Очевидно, что начальные моменты первого порядка есть не что иное, как математические ожидания компонент *X* и *Y*:

$\alpha_{1,0}=m_X$, $\alpha_{0,1}=m_Y$.

Точка с координатами $(m_X;m_Y)$ на плоскости *xOy* представляет собой характеристику положения случайной точки $(X;Y)$, а ее рассеивание (разброс) происходит вокруг $(m_X;m_Y)$.

Центральные моменты первого порядка, очевидно, равны нулю, т.е.

$\mu_{1,0}=\mu_{0,1}=0$.

Имеются три начальных момента второго порядка – $\alpha_{2,0}$, $\alpha_{0,2}$ и $\alpha_{1,1}$. Причем первые два из них есть не что иное, как начальные моменты второго порядка компонент *X* и *Y*:

$\alpha_{2,0}=\alpha_2[X]$, $\alpha_{0,2}=\alpha_2[Y]$.

Имеются три центральных момента второго порядка $\mu_{2,0}$, $\mu_{0,2}$ и $\mu_{1,1}$. Первые два из них представляют собой дисперсии компонент *X* и *Y* соответственно:

$\mu_{2,0}=\mathrm{D}[X]$, $\mu_{0,2}=\mathrm{D}[Y]$.

Рассмотрим $\mu_{1,1}$ отдельно.

Определение. Центральный момент второго порядка $\mu_{1,1}$ называется *ковариацией* случайной величины $(X;Y)$.

Для момента $\mu_{1,1}$ используется обозначение $K_{X,Y}=\mathrm{cov}(X;Y)$.

Замечание. По определению ковариации: $K_{X,Y}=K_{Y,X}$.

В механической интерпретации, когда распределение вероятностей на плоскости *xOy* трактуется как распределение единичной массы на этой плоскости, точка $(m_X;m_Y)$ есть не что иное, как *центр масс с* распределения; дисперсии $\mathrm{D}[X]$ и $\mathrm{D}[Y]$ – *моменты инерции* распределения относительно точки $(m_X;m_Y)$ в направлении осей *Ox* и *Oy* соответственно, а ковариация – это *центробежный момент инерции* распределения масс.

Теорема. Если случайные величины *X* и *Y* независимы, то $K_{X,Y}=0$.

Замечание. Как правило, $K_{X,Y}$ удобнее вычислять по формуле

$$K_{X,Y} = \alpha_{1,1} - \alpha_{1,0} \cdot \alpha_{0,1}.$$

Ковариация $K_{X,Y}$ характеризует не только степень зависимости двух случайных величин $(X;Y)$, но также их рассеивание вокруг точки $(m_X; m_Y)$. Однако размерность ковариации $K_{X,Y}$ равна произведению размерностей случайных величин *X* и *Y*. Чтобы получить безразмерную величину, характеризующую только зависимость, а не разброс, ковариацию $K_{X,Y}$ делят на произведение $\sigma_X \sigma_Y$:

$$\rho_{X,Y} = \frac{K_{X,Y}}{\sigma_X \sigma_Y}.$$

Определение. Величина $\rho_{X,Y}$ называется *коэффициентом корреляции* случайных величин *X* и *Y*.

Коэффициент корреляции $\rho_{X,Y}$ характеризует степень зависимости случайных величин *X* и *Y*, причем не любой зависимости, а только *линейной*, проявляющейся в том, что при возрастании одной случайной величины другая проявляет тенденцию также возрастать (или убывать). В первом случае $\rho_{X,Y} > 0$ и говорят, что случайные величины *X* и *Y связаны положительной корреляцией*, во втором случае $\rho_{X,Y} < 0$ и говорят, что случайные величины *X* и *Y связаны отрицательной корреляцией*. Модуль коэффициента корреляции случайных величин *X* и *Y* характеризует *степень тесноты линейной зависимости* между ними. Если линейной зависимости нет, то $\rho_{X,Y} = 0$.

Теорема. Если случайные величины *X* и *Y* связывает линейная зависимость $Y = aX + b$, то $\rho_{X,Y} = +1$ при $a > 0$, $\rho_{X,Y} = -1$ при $a < 0$.

Пример 1. Найти коэффициент корреляции $\rho_{X,Y}$ между случайными величинами: 1) *X* и $Y = 13X - 2$; 2) *X* и $Y = 9 - 7X$.

Решение. Согласно теореме 3.5.2: 1) $\rho_{X,Y} = 1$, т.к. $a = 13$, $a > 0$; 2) $\rho_{X,Y} = -1$, т.к. $a = -7$, $a < 0$.

Ответ: 1) $\rho_{X,Y} = 1$; 2) $\rho_{X,Y} = -1$.

Пример 2. Игральная кость размечена таким образом, что сумма очков на противоположных гранях равна 7 (т.е. 1 и 6, 2 и 5, 3 и 4). Пусть *X* – число очков на верхней грани, *Y* – число очков на нижней грани. Построить совместный закон распределения случайных величин *X* и *Y*, найти коэффициент корреляции между ними.

Решение. По условию задачи $X + Y = 7$. Поэтому $\mathbf{P}\{X + Y \neq 7\} = 0$. Следовательно, для построения таблицы распределения случайного вектора $(X;Y)$ остается вычислить вероятности:

$$\mathrm{P}\{X = 1, Y = 6\} = \mathrm{P}\big(\{X = 1\} \cdot \{Y = 6\}\big) = \mathrm{P}\big(\{X = 1\} \cdot \{7 - X = 6\}\big) = \mathrm{P}\{X = 1\} = 1/6,$$

$$\mathrm{P}\{X = 6, Y = 1\} = \mathrm{P}\big(\{X = 6\} \cdot \{Y = 1\}\big) = \mathrm{P}\big(\{X = 6\} \cdot \{7 - X = 1\}\big) = \mathrm{P}\{X = 6\} = 1/6.$$

Аналогично можно показать, что

$$\mathrm{P}\{X = 2, Y = 5\} = \mathrm{P}\{X = 5, Y = 2\} = 1/6, \ \mathrm{P}\{X = 3, Y = 4\} = \mathrm{P}\{X = 4, Y = 3\} = 1/6.$$

Тогда закон распределения случайного вектора $(X;Y)$ задается следующей таблицей:

Y	1	2	3	4	5	6

X						
1	0	0	0	0	0	1/6
2	0	0	0	0	1/6	0
3	0	0	0	1/6	0	0
4	0	0	1/6	0	0	0
5	0	1/6	0	0	0	0
6	1/6	0	0	0	0	0

Поскольку между случайными величинами *X* и *Y* имеется линейная связь $Y = \mathbf{7} - X$, то $\rho_{X,Y} = -1$.

Ответ: $\rho_{X,Y} = -1$.

Теорема. Для любых случайных величин *X* и *Y*: $|\rho_{X,Y}| \leq \mathbf{1}$.

Определение. Случайные величины *X* и *Y* называются *некоррелированными*, если $\rho_{X,Y} = 0$ (или $K_{X,Y} = 0$), иначе *X* и *Y* называются *коррелированными*.

Замечание. *Из независимости случайных величин следует их некоррелированность. Но из некоррелированности* ($\rho_{X,Y} = 0$) *не вытекает их независимость.* Действительно, если $\rho_{X,Y} = 0$, то это означает только отсутствие линейной связи между случайными величинами, однако любой другой вид связи может при этом присутствовать.

Пример 3. Закон распределения случайного вектора $(X; Y)$ задан таблицей:

Y X	0	2	5
1	0,1	0	0,2
2	0	0,3	0
4	0,1	0,3	0

Выяснить, зависимы или нет случайные величии ны *X* и *Y*. Найти: $m_X, m_Y, D_X, D_Y, \sigma_X, \sigma_Y, K_{X,Y}, \rho_{X,Y}$.

Решение. Найдем законы распределения компонент *X* и *Y*:

Y X	0	2	5	$p_{i\bullet}$
1	0,1	0	0,2	0,3
2	0	0,3	0	0,3
4	0,1	0,3	0	0,4
$p_{\bullet j}$	0,2	0,6	0,2	□

Очевидно, что компоненты *X* и *Y* являются зависимыми, т.к.

$p_{11} = 0{,}1 \neq p_{1\bullet} \cdot p_{\bullet 1} = 0{,}3 \cdot 0{,}2 = 0{,}06$.

$m_X = 1 \cdot 0{,}3 + 2 \cdot 0{,}3 + 4 \cdot 0{,}4 = 2{,}5$; $m_Y = 0 \cdot 0{,}2 + 2 \cdot 0{,}6 + 5 \cdot 0{,}2 = 2{,}2$;

$D_X = \mathbf{1^2 \cdot 0{,}3 + 2^2 \cdot 0{,}3 + 4^2 \cdot 0{,}4 - 2{,}5^2 = 1{,}65}$, $\sigma_X = \sqrt{1{,}65} \approx 1{,}285$;

$D_Y = \mathbf{0^2 \cdot 0{,}2 + 2^2 \cdot 0{,}6 + 5^2 \cdot 0{,}2 - 2{,}2^2 = 2{,}56}$, $\sigma_Y = \sqrt{2{,}56} = 1{,}6$;

$K_{X,Y} = \alpha_{1,1} - \alpha_{1,0} \cdot \alpha_{0,1} = 1 \cdot 0 \cdot 0{,}1 + 1 \cdot 2 \cdot 0 + 1 \cdot 5 \cdot 0{,}2 +$
$+ 2 \cdot 0 \cdot 0 + 2 \cdot 2 \cdot 0{,}3 + 2 \cdot 5 \cdot 0 + 4 \cdot 0 \cdot 0{,}1 + 4 \cdot 2 \cdot 0{,}3 + 4 \cdot 5 \cdot 0 - 2{,}5 \cdot 2{,}2 = -0{,}9$;

$$\rho_{X,Y} \approx \frac{-0{,}9}{1{,}285 \cdot 1{,}6} \approx -0{,}438.$$

Так как $\rho_{X,Y} < 0$, то это показывает, что между случайными величинами X и Y существует отрицательная линейная зависимость, т.е. при увеличении одной из них другая имеет тенденцию уменьшаться.

Пример 4. Закон распределения случайного вектора $(X;Y)$ задан таблицей:

Y / X	1	2
–1	0,15	0,05
0	0,3	0,05
1	0,35	0,1

Выяснить, являются ли случайные величины X и Y: 1) зависимыми; 2) коррелированными.

Решение. Найдем законы распределения компонент X и Y:

Y / X	1	2	$p_{i\bullet}$
–1	0,15	0,05	0,2
0	0,3	0,05	0,35
1	0,35	0,1	0,45
$p_{\bullet j}$	0,8	0,2	□

Очевидно, что компоненты X и Y являются зависимыми, т.к.

$p_{11} = 0{,}15 \neq p_{1\bullet} \cdot p_{\bullet 1} = 0{,}2 \cdot 0{,}8 = 0{,}16$.

$m_X = -1 \cdot 0{,}2 + 0 \cdot 0{,}35 + 1 \cdot 0{,}45 = 0{,}25$; $m_Y = 1 \cdot 0{,}8 + 2 \cdot 0{,}2 = 1{,}2$;

$D_X = (-1)^2 \cdot 0{,}2 + 0^2 \cdot 0{,}35 + 1^2 \cdot 0{,}45 - 0{,}25^2 = 0{,}5875$, $\sigma_X = \sqrt{0{,}5875} \approx 0{,}766$;

$D_Y = 1^2 \cdot 0{,}8 + 2^2 \cdot 0{,}2 - 1{,}2^2 = 0{,}16$, $\sigma_Y = \sqrt{0{,}16} = 0{,}4$;

$K_{X,Y} = \alpha_{1,1} - \alpha_{1,0} \cdot \alpha_{0,1} = -1 \cdot 1 \cdot 0{,}15 + (-1) \cdot 2 \cdot 0{,}05 + 0 \cdot 1 \cdot 0{,}3 +$

$+ 0 \cdot 2 \cdot 0{,}05 + 1 \cdot 1 \cdot 0{,}35 + 1 \cdot 2 \cdot 0{,}1 - 0{,}25 \cdot 1{,}2 = 0$; $\rho_{X,Y} = 0$.

Этот пример показывает, что случайные величины X и Y могут быть некоррелированными, но при этом являться зависимыми.

Пример 5. Двумерный случайный вектор $(X;Y)$ подчинен закону распределения с плотностью

$$f_{X,Y}(x,y) = \begin{cases} axy, & (x;y) \in D; \\ 0, & (x;y) \notin D. \end{cases}$$

Область D – треугольник, ограниченный прямыми $x + y - 1 = 0$, $x = 0$, $y = 0$.

Найти: коэффициент a, $m_X, m_Y, D_X, D_Y, \sigma_X, \sigma_Y, K_{X,Y}, \rho_{X,Y}$. Выяснить, зависимы или нет случайные величины X и Y.

Решение. Коэффициент a находится из уравнения

$$1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = a\int_0^1 \mathrm{d}x \int_0^{1-x} xy\,\mathrm{d}y.$$

Опуская промежуточные выкладки (в этом примере будем делать так и в дальнейшем), получаем $a = 24$. Далее:

$$m_X = \alpha_{1,0} = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x f_{X,Y}(x,y)\,dx\,dy = 24\int\limits_0^1 dx\int\limits_0^{1-x} x^2 y\,dy = 0{,}4,$$

$$D_X = \mu_{2,0} = \alpha_{2,0} - \alpha_{1,0}^2 = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x^2 f_{X,Y}(x,y)\,dx\,dy - 0{,}4^2 = 24\int\limits_0^1 dx\int\limits_0^{1-x} x^3 y\,dy - 0{,}4^2 = 0{,}04.$$

Заметим, что в силу симметрии по переменным *x* и *y*, можно не вычислять математическое ожидание и дисперсию компоненты *Y*, т.е. $m_Y = m_X = 0{,}4$, $D_Y = D_X = 0{,}04$. Тогда $\sigma_X = \sigma_Y = 0{,}2$.

Вычислим ковариацию и коэффициент корреляции:

$$K_{X,Y} = \mu_{1,1} = \alpha_{1,1} - \alpha_{1,0}\cdot\alpha_{0,1} = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xy f_{X,Y}(x,y)\,dx\,dy - 0{,}4\cdot 0{,}4 =$$

$$= 24\int\limits_0^1 dx\int\limits_0^{1-x} x^2 y^2\,dy - 0{,}16 = -\frac{2}{75};\quad \rho_{X,Y} = \frac{-2/75}{0{,}2\cdot 0{,}2} = -\frac{2}{3}.$$

Поскольку компоненты *X* и *Y* коррелированны, следовательно, они зависимы.

Ответ: $a = \mathbf{24}$, $m_Y = m_X = 0{,}4$, $D_Y = D_X = 0{,}04$, $\sigma_X = \sigma_Y = 0{,}2$, $K_{X,Y} = -2/75$, $\rho_{X,Y} = -2/3$. Компоненты *X* и *Y* зависимы.

3.4.3 Результаты и выводы: В результате проведенного занятия студенты:

- освоили основные понятия многомерного статистического анализа;
- усвоили способы задания многомерной СВ, определение общего и частных законов распределения;
- выработали навыки по вычислению числовых характеристик МСВ, условных числовых характеристик МСВ.

3.5 Практическое занятие 7-8 (ПЗ-7-8)

Тема: Статистическое распределение. Оценки параметров распределения

3.5.1 Задание для работы:

1.Первичная обработка статистических данных. Графическое представление статистических рядов. Эмпирическая функция распределения статистических рядов.

2. Числовые характеристики статистического ряда, их свойства. Точечные оценки параметров статистического распределения.

3. Оценки параметров генеральной совокупности. Метод моментов.

4.Метод доверительных интервалов.

3.5.2 Краткое описание проводимого занятия

1.Первичная обработка статистических данных. Графическое представление статистических рядов. Эмпирическая функция распределения статистических рядов.

Пример 1. Имеется распределение 80 предприятий по числу работающих на них (чел.):

	150	250	350	450	550	650	750
m_i	1	3	7	30	19	15	5

Построить полигон распределения частот.

Решение. Признак *X*- число работающих (чел.) на предприятии. В данной задаче признак *X* является дискретным. Поскольку различных значений признака сравнительно немного -*k*= 7, применять интервальный ряд для представления статистического распределения нецелесообразно (в прикладной статистике в подобных задачах часто используют именно интервальный ряд). Ряд распределения - дискретный. Построим полигон распределения частот (рис. 1).

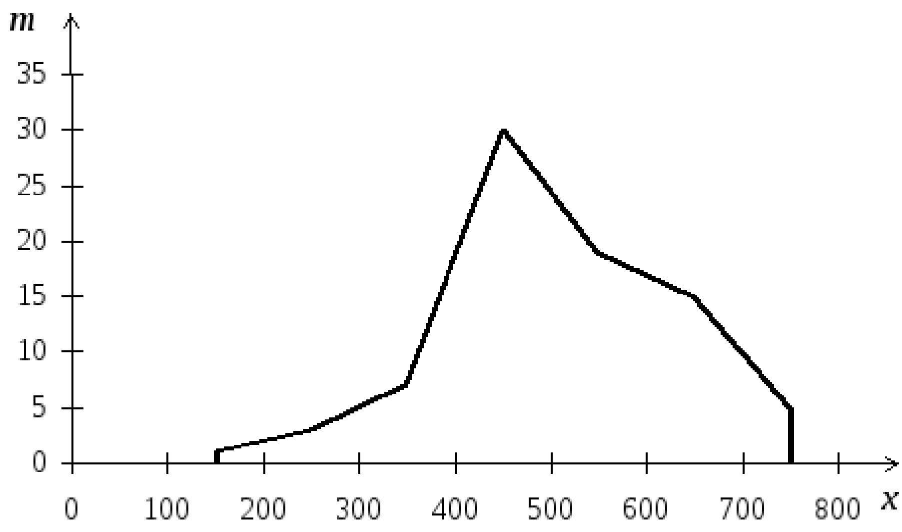


Рис. 1

Пример 2. Дано распределение 100 рабочих по затратам времени на обработку одной детали (мин):

$x_{i\text{-}1}$-x_i	22-24	24-26	26-28	28-30	30-32	32-34	
	2	12	34	40	10	2	.

Построить гистограмму частот.

Решение. Признак *X* - затраты времени на обработку одной детали (мин). Признак *X* -непрерывный, ряд распределения - интервальный. Построим гистограмму частот (рис. 2), предварительно определив $h=(x_k-x_0)/k=(34-22)/6=2$ (*k*= 6) и плотность частоты m_i/h.

$x_{i\text{-}1}$-x_i	22-24	24-26	26-28	28-30	30-32	32-34

m_i/h	1	6	17	20	5	1

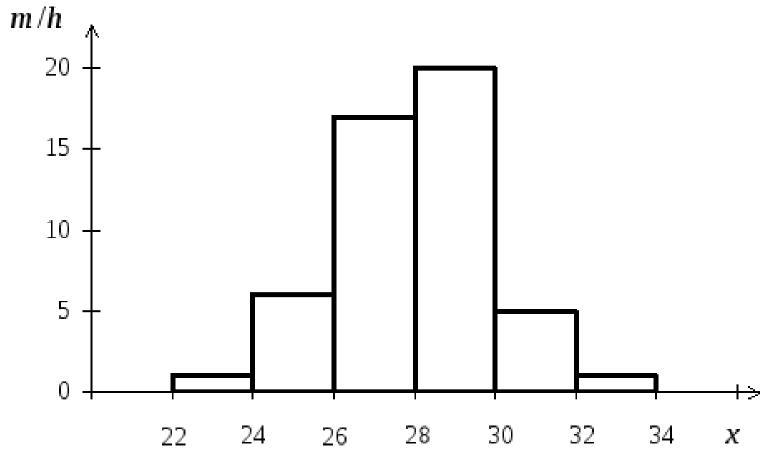


Рис. 2

Построение интервального вариационного ряда рассмотрим на примере.

При измерении диаметра валиков после шлифовки получены следующие результаты:

6,75; 6,77; 6,77; 6,73; 6,76; 6,74; 6,70; 6,75; 6,71; 6,72; 6,77; 6,79; 6,71; 6,78;

6,73; 6,70; 6,73; 6,77; 6,75; 6,74; 6,71; 6,70; 6,78; 6,76; 6,81; 6,69; 6,80; 6,80;

6,77; 6,68; 6,74; 6,70; 6,70; 6,74; 6,77; 6,83; 6,76; 6,76; 6,82; 6,77; 6,71; 6,74;

6,77; 6,75; 6,74; 6,75; 6,77; 6,72; 6,74; 6,80; 6,75; 6,80; 6,72; 6,78; 6,70; 6,75;

6,78; 6,78; 6,76; 6,77; 6,74; 6,74; 6,77; 6,73; 6,74; 6,77; 6.74; 6,75; 6,74; 6,76;

6,76; 6,74; 6,74; 6,74; 6,74; 6,76; 6,74; 6,72; 6,80; 6,76; 6,78; 6,73; 6,70; 6,76;

6,76; 6,77; 6,75; 6,78; 6,72; 6,76; 6,78; 6,68; 6,75; 6,73; 6,82; 6,73; 6,80; 6,81;

6,71; 6,82; 6,77; 6,80; 6,80; 6,70; 6,70; 6,82; 6,72; 6,69; 6,73; 6,76; 6,74; 6,77;

6,72; 6,76; 6,78; 6,78; 6,73; 6,76; 6,80; 6,76; 6,72; 6,76; 6,76; 6,70; 6,73; 6,75;

6,77; 6,77; 6,70; 6,81; 6,74; 6,73; 6,77; 6,74; 6,78; 6,69; 6,74; 6,71; 6,76; 6,76;

6,77; 6,70; 6,81; 6,74; 6,74; 6,77; 6,75; 6,80; 6,74; 6,76; 6,77; 6,77; 6,81; 6,75;

6,78; 6,73; 6,76; 6,76; 6,76; 6,77; 6,76; 6,80; 6,77; 6,74; 6,77; 6,72; 6,75; 6,76;

6,77; 6,81; 6,76; 6,76; 6,76; 6,80; 6,74; 6,80; 6,74; 6,73; 6,75; 6,77; 6,74; 6,76;

6,77; 6,77; 6,75; 6,76; 6,74; 6,82; 6,76; 6,73; 6,74; 6,75; 6,76; 6,72; 6,78; 6,72;

6,76; 6,77; 6,75; 6,78.

Для построения интервального ряда необходимо определить величину частичных интервалов. Считая, что все частичные интервалы имеют одну и ту же длину, для каждого интервала следует установить его верхнюю и нижнюю границы, а затем в соответствии с полученной упорядоченной совокупностью частичных интервалов сгруппировать результаты наблюдении. Длину частичного интервала *h* следует выбрать так, чтобы построенный ряд не был громоздким и в то же время позволял выявить характерные черты изменения значений случайной величины.

Просматривая результаты наблюдений, находим, что наибольшим значением случайной величины $x_{наиб}$ является 6,83, а наименьшим $x_{наим}$ - 6,68. Найдем размах варьирования *R*. :

$$R=6,83-6,68=0,15.$$

Выберем число интервалов. Для того чтобы вариационный ряд не был слишком громоздким, обычно число интервалов берут от 7 до 11. Положим предварительно v=7, тогда длина частичного интервала $h=\frac{R}{v}=\frac{0,15}{7}\approx 0,0214\approx 0,02.$

За начало первого интервала рекомендуется брать величину $x_{нач}=x_{наим}-0,5h$.

В данном случае $x_{нач}=6,67$. Конец последнего интервала должен удовлетворять условию

$$x_{кон}-h\le x_{наиб}<x_{кон}.$$

Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала *h* (в рассматриваемом случае h=0,02).

Теперь, просматривая результаты наблюдений, определяем, сколько значений признака попало в каждый конкретный интервал. При этом в интервал включают значения случайной величины, большие или равные нижней границе и меньшие верхней границы.

В таблице частота m_i, показывает, в скольких наблюдениях случайная величина приняла значения, принадлежащие тому или иному интервалу, причем нижний конец интервала входит в него, а верхний—нет. Такие частоты обычно называют интервальными, а их отношение к общему числу наблюдений—интервальными частостями.

При вычислении интервальных частостей округление результатов следует проводить таким образом, чтобы общая сумма частостей была равна 1:

$$\sum \hat{p}_i=1.$$

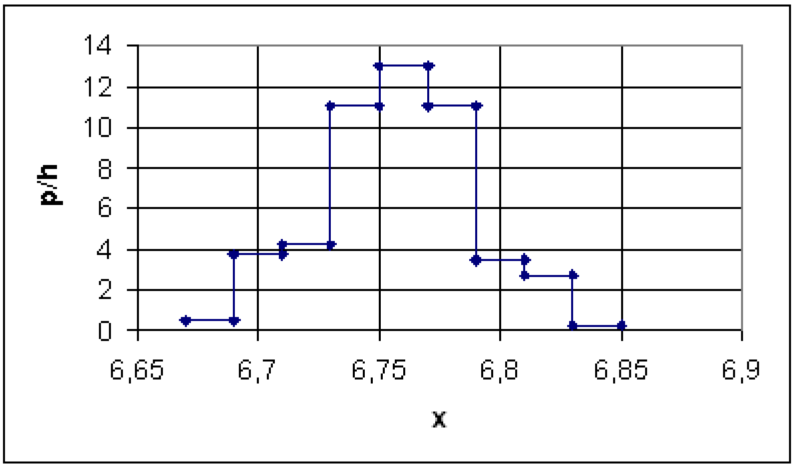
Для данного примера интервальный вариационный ряд имеет вид:

№	x_i - x_{i+1}	m_i	$\hat{p}_i$	m_i/h	$\hat{p}_i/h$
1	6,67-6,69	2	0,01	100	0,5
2	6,69-6,71	15	0,075	750	3,75
3	6,71-6,73	17	0,085	850	4,25
4	6,73-6,75	44	0,22	2200	11
5	6,75-6,77	52	0,26	2600	13
6	6,77-6,79	44	0,22	2200	11
7	6,79-6,81	14	0,07	700	3,5

8	6,81-6,83	11	0,055	550	2,75
9	6,83-6,85	1	0,005	50	0,25
	$\sum$	200	1		

По данным интервального ряда строят гистограмму частот или гистограмму относительных частот:

Ступенчатая фигура, состоящая из прямоугольников, основания которых- частичные интервалы, высоты равны отношению частоты к длине частичного интервала(плотность частоты) (частости к длине частичного интервала (плотность частости)).Гистограмма частостей имеет вид:



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

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

2. Числовые характеристики статистического ряда, их свойства. Точечные оценки параметров статистического распределения.

Каждой числовой характеристике случайной величины X соответствует ее статистическая аналогия.

Для основной характеристики положения математического ожидания случайной величины – такой является среднее арифметическое наблюденных значений случайной величины: $M^*|X| = \frac{\sum_{i=1}^{n} x_i}{n}$, (1) где x_i — случайной величины, наблюденное i-м опыте, n - число опытов.

Эту характеристику мы будем в дальнейшем называть статистическим средним случайной величины.

Согласно закону больших чисел, при ограниченном увеличении числа опытов статистическое среднее приближается (сходится по вероятности) математическому ожиданию.

При ограниченном числе опытов статистическое среднее является случайной величиной, которая, тем не менее, связана с математическим ожиданием и может дать о нем известное представление.

Подобные статистические аналогии существуют для всех числовых характеристик. Условимся в дальнейшем эти статистические аналогии обозначать теми же буквами, что и соответствующие числовые характеристики, но и снабжать их значком *.

Рассмотрим, например, дисперсию случайной величины. Она представляет собой математическое ожидание случайной величины $\overset{\circ}{X}{}^2 = (X - m_x)^2$:

$$D[X] = M\left[\overset{\circ}{X}{}^2\right] = M\left[(X - m_x)^2\right]. \qquad (2)$$

Если в этом выражении заменить математическое ожидание его статистической аналогией – средним арифметическим, мы получим статистическую дисперсию случайной величины X:

$$D^*[X] = \frac{\sum_{i=1}^{n}\left(x_i - m_x^*\right)^2}{n} \qquad (3)$$

где $m_x^* = M^*[X]$ - статистическое среднее.

Аналогично определяются статистические начальные и центральные моменты любых порядков:

$$\alpha_s^*[X] = \frac{\sum_{i=1}^{n} x_i^s}{n} \quad (4), \qquad \mu_s^*[X] = \frac{\sum_{i=1}^{n}\left(x_i - m_x^*\right)^s}{n} \quad (5)$$

Все эти определения полностью аналогичны данным в главе 5 определениям числовых характеристики случайной величины, с той разницей, что в них везде вместо математического ожидания фигурирует среднее арифметическое. При увеличении числа наблюдений, очевидно, все статистические характеристики будут сходиться по вероятности к соответствующим математическим характеристикам и при достаточном n могут быть приняты приближенно равными им.

Нетрудно доказать, что для статистических начальных и центральных моментов справедливы те же свойства, которые были введены в главе 5 для математических моментов. В частности, статистический первый центральный момент всегда равен нулю:

$$\mu_1^* = \frac{\sum_{i=1}^{n}\left(x_i - m_x^*\right)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} - m_x^* = m_x^* - m_x = 0.$$

Соотношения между центральными и начальными моментами также сохраняются:

$$\mu_2^* = D_x^* = \frac{\sum_{i=1}^{n}\left(x_i - m_x^*\right)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - 2m_x^* \frac{\sum_{i=1}^{n} x_i}{n} + \left(m_x^*\right)^2 = \alpha_2^* - \left(m_x^*\right)^2$$

(7.4.6)

и т.д.

При очень большом количестве опытов вычисление характеристик по формулам (1) - (5) становится чрезмерно громоздким и можно применить следующий прием: воспользоваться теми же разрядами, на которые был расклассифицирован статистический материал для построения статистического ряда или гистограммы, и считать приближенно значение случайной величины в каждом разряде постоянным и равным среднему значению, которое выступает в роли «представителя» разряда. Тогда статистические числовые характеристики будут выражаться приближенными формулами:

$$m_x^* = M^*[X] = \sum_{i=1}^{k} \tilde{x}_i p_i^*, \qquad (7)$$

$$D_x^* = D^*[X] = \sum_{i=1}^{k} \left(\tilde{x}_i - m_x^*\right)^2 p_i^*, \qquad (8)$$

$$\alpha_s^*[X] = \sum_{i=1}^{k} \tilde{x}_i^s p_i^*, \qquad (9)$$

$$\mu_s^*[X] = \sum_{i=1}^{k} \left(\tilde{x}_i - m_x^*\right)^s p_i^*, \qquad (10)$$

где $\tilde{x}_i$ — «представитель» i-го разряда, p_i^* - частота i-го разряда, k - число разрядов.

Точечная оценка математического ожидания

Пусть $x_1, x_2, \ldots, x_n$ выборка из генеральной совокупности, соответствующей случайной величине x с неизвестным математическим ожиданием Mx =q и известной дисперсией $D\xi = \sigma^2$.

Рассмотрим оценку неизвестного математического ожидания

$$\hat{\theta}_n = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Оценка несмещённая, поскольку её математическое ожидание равно Mx =q:

$$M\theta_n = M\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) = \frac{1}{n}\left(Mx_1 + Mx_2 + \ldots + Mx_n\right) = \frac{1}{n} \cdot n \cdot \theta = \theta,$$

Оценка состоятельная, поскольку при n- б.б. $D\theta_n = \frac{1}{n}\sigma^2 \to 0$:

$$D\theta_n = D\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) = \frac{1}{n^2}\left(Dx_1 + Dx_2 + \ldots + Dx_n\right) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{1}{n}\sigma^2 \to 0.$$

Итак, для оценки неизвестного математического ожидания случайной величины будем использовать выборочное среднее: $\theta_n = \bar{x} = \frac{1}{n}\left(x_1 + x_2 + \ldots + x_n\right)$.

Точечная оценка дисперсии

Для дисперсии σ^2 случайной величины ξ можно предложить следующую оценку:

$$\overline{DX} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2,$$ где $\bar{x}$ — выборочное среднее.

Доказано, что эта оценка состоятельная, но смещенная.

В качестве состоятельной несмещенной оценки дисперсии используют величину

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

Именно несмещенностью оценки s^2 объясняется ее более частое использование в качестве оценки дисперсии.

3. Оценки параметров генеральной совокупности. Метод моментов. Задача 1. Путем опроса получены следующие данные (N=80):

2 4 2 4 1 1 1 2 0 61 2 1 2 2 4 1 1 5 10 2 4 1 2 2 1 1 1 1
1 1 1 1 2 1 1 4 1 17 4 1 4 2 1 2 1 1 14 1 1 4 5 1 4 2 4 5
1 6 4 1 1 2 4 1 1 10 0 4 6 4 7 4 1 1 5

Выполнить задания:

А) получить дискретный вариационный ряд и статистическое распределение выборки;
Б) построить полигон частот;
В) составить ряд распределения относительных частот;
Г) составить эмпирическую функцию распределения;
Д) построить график эмпирической функции распределения;
Е) найти основные числовые характеристики вариационного ряда (по возможности использовать упрощающие формулы для их нахождения):

1) выборочное среднее $\overline{x}_B$;
2) выборочную дисперсию D(X);
1) выборочное среднее квадратическое отклонение $\sigma(X)$;
4) коэффициент вариации V;
5) интерпретировать полученные результаты.

Решение.

А) Для составления дискретного вариационного ряда отсортируем данные опроса по величине и расположим их в порядке возрастания:

0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 6 6 6 7 7.

Статистическое распределение выборки представлено в таблице1, в которой первая строка – варианты (наблюдаемые значение), вторая строка – Частоты появления этих вариант).

Таблица1. Варианты и их частоты

Xi	0	1	2	1	4	5	6	7
Ni	4	11	14	24	16	4	1	2

Б) Для построения полигона частот найдем относительные частоты ($w_i = n_i / n$,

где $n = \sum_{i=1}^{m} n_i$, где M – число различных значений признака X ($m \le n$) и в данном примере M=8), которые будем вычислять с одинаковой точностью. Полигон частот – ломаная линия, соединяющая точки с координатами (x_i, w_i) (Рис.1). Расчеты запишем в табл.2.

Таблица 2. Относительные частоты и накопленные частоты

Xi	Ni	Относительные частоты w_i	Накопленные частоты
0	4	0.050	0.050
1	11	0.161	0.211
2	14	0.175	0.188
1	24	0.100	0.688
4	16	0.200	0.888
5	4	0.050	0.918
6	1	0.018	0.975
7	2	0.025	1.000
Сумма	80	1	

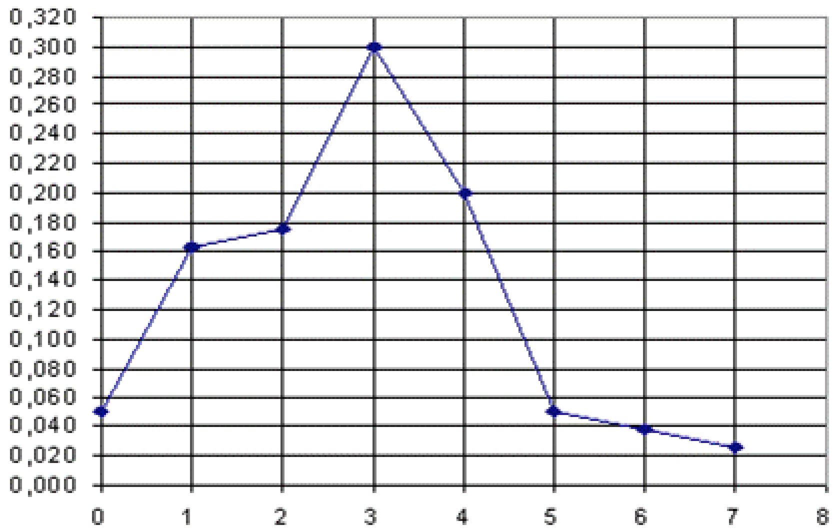


Рис.1. Полигон частот вариационного ряда

В) Запишем ряд распределения (табл. 1) относительных частот в виде таблицы 1, в которой первая строка – варианты (изучаемый признак), вторая строка – относительные частоты (Частости).

Таблица 1. Распределение относительных частот появления признака

Xi	0	1	2	1	4	5	6	7
Ni	0.05	0.161	0.175	0.1	0.2	0.05	0.018	0.025

Г) Эмпирическую функцию распределения найдем, используя накопленные частоты (табл. 1, столбик 4):

$$F^*(x) = \begin{cases} 0, & x \le 0, \\ 0.05, & 0 < x \le 1, \\ 0.213, & 1 < x \le 2, \\ 0.388, & 2 < x \le 3, \\ 0.688, & 3 < x \le 4, \\ 0.888, & 4 < x \le 5, \\ 0.938, & 5 < x \le 6, \\ 0.975, & 6 < x \le 7, \\ 1 & x > 7. \end{cases}$$

Д) Построим график эмпирической функции распределения (рис. 2), используя значения, полученные в пункте г).

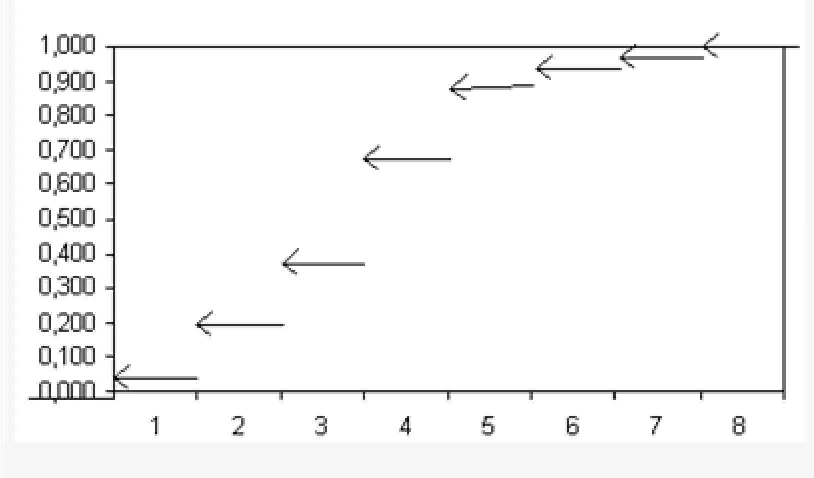


Рис. 2. График эмпирической функции распределения

Е) Для вычисления выборочного среднего $\overline{x}_B$ и выборочной дисперсии D_B с использованием приведенных выше формул, удобно составлять расчетную таблицу 2:

Таблица.2. Расчетная таблица для вычисления выборочных величин

Xi	Ni	Xi×Ni	$\left(x_i - \overline{x}_B\right)^2$	$\left(x_i - \overline{x}_B\right)^2$ ×Ni
0	4	0	8.1796	12.7184
1	11	11	1.4596	44.9748
2	14	28	0.7196	10.1544
1	24	72	0.0196	0.4704
4	16	64	1.2996	20.7916
5	4	20	4.5796	18.1184
6	1	18	9.8596	29.5788
7	2	14	17.1196	14.2792
Сумма	80	229		191.488

Используя суммы, полученные в табл. 2, определим искомые величины.

1) Выборочную среднюю $$\overline{x}_B = \frac{\sum_{i=1}^{m} x_i \cdot n_i}{n} = \frac{229}{80} \approx 2.86.$$

2) Выборочную дисперсию $$D_B = \frac{\sum_{i=1}^{m} \left(x_i - \overline{x}_B\right)^2 \cdot n_i}{n} = \frac{191.488}{80} \approx 2.39.$$

1) Выборочное среднее квадратическое отклонение

$\sigma_B(X) = \sqrt{D_B} = \sqrt{2.39} \approx 1.55.$

4) Коэффициент вариации $$V = \frac{\sigma_B(X)}{\overline{x}_B} \cdot 100\% = \frac{1.55}{2.86} \cdot 100\% \approx 54.2\%.$$

5) Интерпретация полученных результатов:

- величина $\overline{x}_B \approx 2.86$ характеризует среднее значение признака X;

- среднее квадратическое отклонение $\sigma_B(X)$ описывает абсолютный разброс значений показателя X относительно среднего значения и в данном случае составляет $\sigma_B(X) \approx 1.55$;

- коэффициент вариации V характеризует относительную изменчивость показателя X, то есть относительный разброс вокруг его среднего значения $\bar{x}_в$, и в данном случае составляет $V \approx 54{,}2\%$.

Ответ: $\bar{x}_в \approx 2{,}86$; $D_в \approx 2{,}39$; $\sigma_в(X) \approx 1{,}55$; $V \approx 54{,}2\%$.

4. Метод доверительных интервалов.

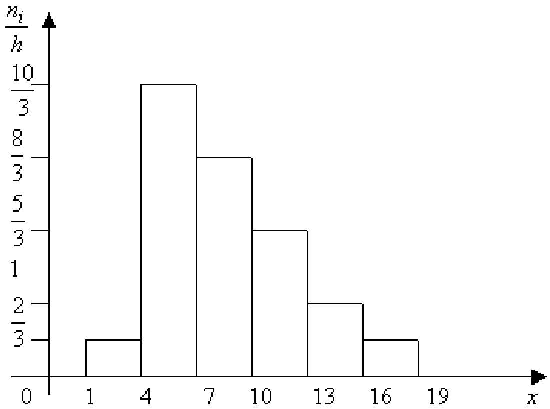
Задача: Выборочно обследование 30 предприятий машиностроительной промышленности по валовой продукции и получены следующие данные, в млн. руб.:

18,0; 12,0; 11,9; 1,9; 5,5; 14,6; 4,8; 5,6; 4,8; 10,9; 9,7; 7,2; 12,4; 7,6;
9,7; 11,2; 4,2; 4,9; 9,6; 3,2; 8,6; 4,6; 6,7; 8,4; 6,8; 6,9; 17,9; 9,6;
14,8; 15,8.

Составить интервальное распределение выборки с началом $x_0 = 1$ и длиной частичного интервала $h = 3$. Построить гистограмму частот. Решение. Для составления интервального распределения составим таблицу, в первой строке которой расположим в порядке возрастания интервалы, длина каждого из которых $h = 3$. Во второй строке запишем количество значений признака в выборке, попавших в этот интервал (т.е. сумму частот вариант, попавших в соответствующий интервал):

$(x_i; x_{i+1})$	1-4	4-7	7-10	10-13	13-16	16-19
n_i	2	10	8	5	3	2

Объем выборки *n* = 2 + 10 + 8 + 5 + 3 + 2 = 30.



Для построения гистограммы частот на оси абсцисс откладываем частичные интервалы, на каждом из них строим прямоугольники высотой $\frac{n_i}{h}$, где n_i - частота *i*-го частичного интервала, *h* – шаг (длина интервала), таким образом, гистограмма примет вид:

Указание. Для построения эмпирической функции распределения и нахождения точечных оценок ряда необходимо преобразовать его к дискретному виду по формуле

$$x_i^* = \frac{x_i + x_{i+1}}{2}.$$

Получим

x_i^*	2,5	5,5	8,5	11,5	14,5	17,5
n_i	2	10	8	5	3	2

Задача: Из большой партии электроламп случайным образом отобрано 100. Средняя продолжительность горения ламп в выборке оказалась равной 1000 ч. Найти с надежностью $\gamma = 0{,}95$ доверительный интервал для средней продолжительности горения ламп во всей партии, если известно, что среднее квадратическое отклонение продолжительности горения лампы $\sigma = 40$ ч и продолжительность горения ламп распределена по нормальному закону. Решение. По условию $\bar{x}_B = 1000,\ \gamma = 0{,}95,\ \sigma = 40.$ Для решения воспользуемся формулой

$$\bar{x}_B - \frac{\sigma t}{\sqrt{n}} < a < \bar{x}_B + \frac{\sigma t}{\sqrt{n}}$$

По приложению 3 находим t из условия: $\Phi(t) = \frac{\gamma}{2} = \frac{0{,}95}{2} = 0{,}475; \Rightarrow t = 1{,}96.$

Тогда доверительный интервал:

$$1000 - \frac{1{,}96 \cdot 40}{\sqrt{100}} < a < 1000 + \frac{1{,}96 \cdot 40}{\sqrt{100}}$$

$$992{,}16 < a < 1007{,}84.$$

Задача: По результатам наблюдений была найдена оценка неизвестного математического ожидания M случайной величины $\xi \sim N(m, \sigma^2)$, если точечная оценка $\overline{m}$ =10.2, а дисперсия оценки σ_X =4. Требуется оценить доверительный интервал для оценки математического ожидания по 36-ти наблюдениям с заданной надежностью ¡=0.99.

Решение. Из (4.1) следует, что $\Phi_{0,1}(t_\gamma) = \frac{0.99}{2} = 0.495$. Отсюда получаем, что $t_\gamma = \frac{\delta\sqrt{n}}{\sigma_X}$ =2.58 и половина искомого интервала $\delta = \frac{2.58 \cdot 4}{\sqrt{36}} = 1.89$. Так как $\overline{m} - \delta < m < \overline{m} + \delta$, то с вероятностью 0.99 доверительный интервал для оценки математического ожидания: $8.31 < m < 12.09$.

Со случаем, когда распределение результатов наблюдений нормально, но их дисперсия неизвестна, можно ознакомится в [3, 4, 6].

3.5.3 Результаты и выводы: В результате проведенного занятия студенты:

- освоили основные понятия математической статистики;
- усвоили алгоритмы первичной обработки статистических данных;
- выработали навыки по вычислению точечных и интервальных оценок параметров генеральной совокупности;
- усвоили основные методы исследования параметров генеральной совокупности.

3.6 Практическое занятие 9-10 (ПЗ-9-10)

Тема: Статистические критерии, их виды

3.6.1 Задание для работы:

1. Статистические гипотезы и их виды. Критерии согласия.
2. Критерии согласия.
3. Оценка параметров неизвестного распределения. Выравнивание рядов.

3.6.2 Краткое описание проводимого занятия

1. Статистические гипотезы и их виды. Критерии согласия.

Пример. Для подготовки к зачету преподаватель сформулировал 100 вопросов (генеральная совокупность) и считает, что студенту можно поставить «зачтено», если тот знает 60 % вопросов (критерий). Преподаватель задает студенту 5 вопросов (выборка из генеральной совокупности) и ставит «зачтено», если правильных ответов не меньше трех. Гипотеза H_0: «студент курс усвоил», а множество $\{3, 4, 5\}$ — область принятия этой гипотезы. Критической областью является множество $\{0, 1, 2\}$ — правильных ответов меньше трех, в этом случае основная гипотеза отвергается в пользу альтернативной H_1 «студент курс не усвоил, знает меньше 60 % вопросов».

Студент А выучил 70 вопросов из 100, но ответил правильно только на два из пяти, предложенных преподавателем, — зачет не сдан. В этом случае преподаватель совершает ошибку первого рода.

Студент Б выучил 50 вопросов из 100, но ему повезло, и он ответил правильно на 3 вопроса — зачет сдан, но совершена ошибка второго рода.

Преподаватель может уменьшить вероятность этих ошибок, увеличив количество задаваемых на зачете вопросов.

Алгоритм проверки статистических гипотез сводится к следующему:

1) сформулировать основную H_0 и альтернативную H_1 гипотезы;

2) выбрать уровень значимости α;

3) в соответствии с видом гипотезы H_0 выбрать статистический критерий для ее проверки, т. е. случайную величину K, распределение которой известно;

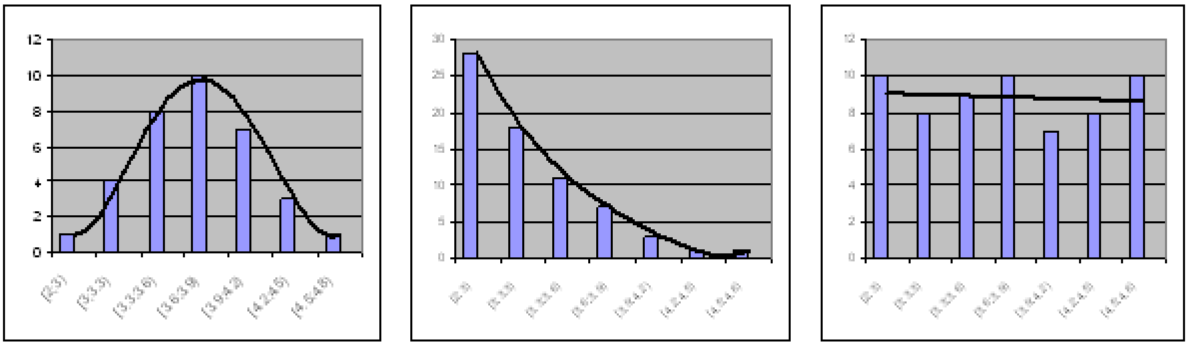
4) по таблицам распределения случайной величины K найти границу критической области $K_{кр}$ (вид критической области определить по виду альтернативной гипотезы H_1);

5) по выборочным данным вычислить наблюдаемое значение критерия $K_{набл}$;

6) принять статистическое решение: если $K_{набл}$ попадает в критическую область — отклонить гипотезу H_0 в пользу альтернативной H_1; если $K_{набл}$ попадает в область допустимых значений, то нет оснований отклонять основную гипотезу.

3. Критерии согласия. Критерии однородности.

Одной из важных задач математической статистики является установление теоретического закона распределения случайной величины, характеризующей изучаемый признак по эмпирическому распределению, представляющему вариационный ряд. Предположение о виде закона распределения можно сделать по гистограмме или полигону (Рис. 1)



A) B) C)

Рис.1. Возможные виды гистограмм:
а) нормального, б) показательного, в) равномерного распределений

Например, по гистограмме (рис. 1, а)) можно сделать предположение о том, что генеральная совокупность распределена по нормальному закону.

Для проверки гипотез о виде распределения служат специальные критерии — *Критерии согласия*. Они отвечают на вопрос: согласуются ли результаты экспериментов с предположением о том, что генеральная совокупность имеет заданное распределение.

Проверим это предположение с помощью ***Критерия согласия Пирсона***. В этом критерии мерой расхождения между гипотетическим (предполагаемым) и эмпирическим распределением служит статистика $K=\sum_{j=1}^{k}\frac{(n_j-np_j)^2}{np_j}$.

Где *N* — объем выборки; *K* — количество интервалов (групп наблюдений); n_j — количество наблюдений, попавших в *J*-й интервал; p_j — вероятность попадания в *J*-й интервал случайной величины, распределенной по гипотетическому закону.

Если предположение о виде закона распределения справедливо, то статистика Пирсона распределена по закону «хи-квадрат» с числом степеней свободы $k-r-1$ (*R* — число параметров распределения, оцениваемых по выборке): $K \sim \chi^2_{(k-r-1)}$.

Оцениваются неизвестные параметры с использованием теории точечных оценок, некоторые оценки приведены в табл. 1.

Таблица 1. Оцениваемые параметры и их точечные оценки

Вид распределения	Оцениваемые параметры	Точечные оценки параметров
$f_X(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x}, & x\ge 0\end{cases}$	$\lambda>0$	$\bar{\lambda}=\frac{1}{\bar{x}}$

$f_X(x)=\begin{cases}0, & x<\alpha\\ 1/(\beta-\alpha), & x\in[\alpha,\beta]\\ 0, & x>\beta\end{cases}$	$\alpha,\ \beta$	$\alpha=\overline{x}_в-\sqrt{3}\cdot\sqrt{D_в}$ $\beta=\overline{x}_в+\sqrt{3}\cdot\sqrt{D_в}$
$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-m)^2}{2\sigma^2}}$	$m,\ \sigma^2$	$m=\overline{x}_в,\quad \sigma^2=\sqrt{D_в}$

Здесь $\overline{x}_в=\frac{\sum_{i=1}^{m}x_i\cdot n_i}{n}$, $D_в=\frac{\sum_{i=1}^{m}(x_i-\overline{x}_в)^2\cdot n_i}{n}$

Количество интервалов *K* рекомендуется рассчитывать по формуле Стерджеса $k=1+3.3\cdot\lg n$, где *N* — объем выборки. Длину *I*-го интервала принимают равной $h=\frac{x_{(n)}-x_{(1)}}{k}$, где $x_{(n)}$—наибольшее, а $x_{(1)}$ — наименьшее значение в вариационном ряду.

Пример 1. Для среднего балла среди 30-ти групп (с точностью до сотых долей балла) получили выборку x_i:

3.7, 3.85, 3.7, 3.78, 3.6, 4.45, 4.2, 3.87, 3.33, 3.76, 3.75, 4.03, 3.8, 4.75, 3.25, 4.1, 3.55, 3.35, 3.38, 3.05, 3.56, 4.05, 3.24, 4.08, 3.58, 3.98, 3.4, 3.8, 3.06, 4.38. Проверить гипотезу о нормальном распределении среднего балла на уровне значимости $\alpha=0.025$.

Решение. Сгруппируем эту выборку. Наименьший средний балл равен 3.05, наибольший — 4.75. Интервал [3; 4.8] разобьем на 6 частей длиной $h=0.3$, применяя формулу Стерджеса ($k=5.875\approx 6$). Подсчитаем частоту n_i (относительную частоту $\frac{n_i}{n}$) для каждого интервала и получим сгруппированный статистический ряд (табл. 2).

Таблица 2. Статистический ряд

Интервалы	[3;3.3)	[3.3;3.6)	[3.6;3.9)	[3.9;4.2)	[4.2;4.5)	[4.5;4.8)
Частоты n_i	4	7	10	5	3	1
Относительные частоты $\frac{n_i}{n}$	0.133	0.233	0.3	0.167	0.1	0.033

Сформулируем основную и альтернативную гипотезы.

$H_0: X\sim N(\overline{a},\ \overline{\sigma})$ — случайная величина *X* (средний балл) подчиняется нормальному закону с параметрами $\overline{a},\ \overline{\sigma}$. Так как истинных значений параметров $a,\ \sigma$ мы не знаем, возьмем их оценки, рассчитанные по выборке: $\overline{a}=3.746,\quad \overline{\sigma}=0.399.$

$H_1:$ случайная величина *X* не подчиняется нормальному закону с данными параметрами.

Рассчитаем наблюдаемое значение $K_{набл}$ статистики Пирсона. Эмпирические частоты n_j уже известны (табл. 2), а для вычисления вероятностей p_j (в предположении, что гипотеза H_0 справедлива) применим уже известную формулу (свойство **В**):

И таблицу функции Лапласа. Полученные результаты сведем в таблицу (табл. 3). Наблюдаемое значение статистики Пирсона равно $K_{набл} = 0.978.$

Определим границу критической области. Так как статистика Пирсона измеряет разницу между эмпирическим и теоретическим распределениями, то чем больше ее наблюдаемое значение $K_{набл}$, тем сильнее довод против основной гипотезы. Поэтому критическая область для этой статистики всегда правосторонняя: $[K_{кр}; \ +\infty).$ Её границу $K_{кр} = \chi^2_{(k-r-1;\ \alpha)}$ находим по таблицам распределения «хи-квадрат» и заданным значениям $\alpha = 0.025,\ k = 6$ (число интервалов), $r = 2$ (параметры a и σ оценены по выборке): $K_{кр} = \chi^2(6-2-1;\ 0.025) = \chi^2(3;\ 0.025) = 9.4.$

Наблюдаемое значение статистики Пирсона не попадает в критическую область: $K_{набл} < K_{кр}$, поэтому н*ет оснований отвергать основную гипотезу*.

Вывод: на уровне значимости 0.025 справедливо предположение о том, что средний балл имеет нормальное распределение

Таблица 3. Сравнение наблюдаемых и ожидаемых частот

№ п/п	Интервалы группировки $[a_j; a_{j+1})$	Наблюдаемая частота n_j	Вероятность p_j попадания в *J*-й интервал	Ожидаемая частота $n \cdot p_j$	Слагаемые статистики Пирсона $\frac{(n_j - np_j)^2}{np_j}$
1.	[3; 3.3)	4	0.101	3.032	0.309
2.	[3.3; 3.6)	7	0.225	6.761	0.008
3.	[3.6; 3.9)	10	0.295	8.79	0.166
4.	[3.9; 4.2)	5	0.222	6.665	0.416
5.	[4.2; 4.5)	3	0.098	2.946	0.001
6.	[4.5; 4.8)	1	0.025	0.758	0.077
$\sum$	—	30	0.965	28.95	$K_{набл} = 0.978.$

Пример 2. Проверить с помощью критерия χ2 при уровне значимости 0,05 гипотезу о том, что выборка объема $n = 50$, представленная интервальным вариационным рядом в таблице 4, извлечена из нормальной генеральной совокупности.

Таблица 4

Номер интервала I	Границы интервала	Частота m_i
1	0 – 2	5
2	2 – 4	11
3	4 –6	17
4	6 – 8	10
5	8 – 10	7

Решение.

1. Сформулируем нулевую и альтернативную гипотезы: H_0 – эмпирическое распределение соответствует нормальному; H_1 - эмпирическое распределение не соответствует нормальному.

Для проверки нулевой гипотезы необходимо рассчитать наблюдаемое значение критерия χ2набл по формуле и сравнить его с критическим значением χ2кр.

2. Определим параметры предполагаемого (теоретического) нормального закона распределения.

Найдем середины интервалов $\overline{x_i} = \dfrac{z_{i-1} + z_i}{2}$ и относительные частоты $p_i{}^* = \dfrac{m_i}{n}$.

Получим следующие значения:

$\overline{x_i}$	1	3	5	6	7
$p_i{}^*$	$\frac{5}{50}$	$\frac{11}{50}$	$\frac{17}{50}$	$\frac{10}{50}$	$\frac{7}{50}$

Найдем оценку математического ожидания:

$$m^* = \left(1 \cdot \frac{5}{50} + 3 \cdot \frac{11}{50} + 5 \cdot \frac{17}{50} + 7 \cdot \frac{10}{50} + 9 \cdot \frac{7}{50}\right) =$$

$$= \frac{1}{50}(5 + 33 + 85 + 70 + 63) = \frac{256}{50} = 5{,}12.$$

Вычислим оценки дисперсии и стандартного отклонения по формулам:

$$s^2 = \left((1 - 5{,}12)^2 \cdot 5 + (3 - 5{,}12)^2 \cdot 11 + (5 - 5{,}12)^2 \cdot 17 +\right.$$

$$\left. + (7 - 5{,}12)^2 \cdot 10 + (9 - 5{,}12)^2 \cdot 7\right) \cdot \frac{1}{49} = \frac{1}{49} \cdot 275{,}27 = 5{,}62;$$

$$s = \sqrt{5{,}62} = 2{,}37.$$

3. Выполним расчет теоретических частот $m_i{}^T$.

$$p_1 = \Phi\left(\frac{2 - 5{,}12}{2{,}37}\right) - \Phi(-\infty) = \Phi(-1{,}3) - 0 = 1 - \Phi(1{,}3) = 1 - 0{,}9 = 0{,}1$$

и

$$m_1{}^T = 50 \cdot 0{,}1 = 5;$$

Для интервала $(2,4)$ находим

$$p_2 = \Phi\left(\frac{4 - 5{,}12}{2{,}37}\right) - \Phi\left(\frac{2 - 5{,}12}{2{,}37}\right) = \Phi(-0{,}5) - \Phi(-1{,}3) = 0{,}90 - 0{,}69 = 0{,}21$$

и $m_2{}^T = 50 \cdot 0{,}21 = 10{,}5;$

Для интервала (4,6) соответственно:

$$p_3 = \Phi\left(\frac{6 - 5{,}12}{2{,}37}\right) - \Phi\left(\frac{4 - 5{,}12}{2{,}37}\right) = \Phi(0{,}4) - \Phi(-0{,}5) = 0{,}66 - 0{,}31 = 0{,}35;$$

$$m_3{}^T = 50 \cdot 0{,}35 = 17{,}5;$$

Для интервала (6,8):

$$p_4 = \Phi\left(\frac{8-5{,}12}{2{,}37}\right) - \Phi\left(\frac{6-5{,}12}{2{,}37}\right) = \Phi(1{,}2) - \Phi(0{,}4) = 0{,}88 - 0{,}66 = 0{,}22$$

и $m_4{}^T = 50 \cdot 0{,}22 = 11$;

Для интервала $(8, \infty)$ вычислим

$$p_5 = \Phi(\infty) - \Phi\left(\frac{8-5{,}12}{2{,}37}\right) = \Phi(\infty) - \Phi(1{,}2) = 1 - 0{,}88 = 0{,}12$$;

$m_5{}^T = 50 \cdot 0{,}12 = 6$.

4. По формуле (4.8) найдем значение $\chi^2{}_{набл}$:

$$\chi^2{}_{набл} = \sum_{i=1}^{5} \frac{\left(m_i - m_i{}^T\right)^2}{m_i{}^T} = \frac{(5-5)^2}{5} + \frac{(11-10{,}5)^2}{10{,}5} +$$

$$+ \frac{(17-17{,}5)^2}{17{,}5} + \frac{(10-11)^2}{11} + \frac{(7-6)^2}{6} = 0{,}29$$.

5. По таблице квантилей распределения χ2 с числом степеней свободы $r = k - 3 = 5 - 3 = 2$ находим, что χ2кр = 6,0 для $\alpha = 0{,}05$.

Поскольку $\chi^2{}_{набл} < \chi^2{}_{кр}$ ($0{,}29 < 6{,}0$), то можно считать, что гипотеза о нормальном распределении генеральной совокупности не противоречит опытным данным.

Пример 3: Используя критерий Пирсона, при уровне значимости a = 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n = 200.

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение.

1. Вычислим $\bar{x}_B = \dfrac{\sum_{i=1}^{k} x_i n_i}{n} = 12{,}63$ и выборочное среднее квадратическое отклонение $\sigma_B = \sqrt{\overline{x_B^2} - \left(\bar{x}_B\right)^2} = 4{,}695$.

2. Вычислим теоретические частоты учитывая, что n = 200, h = 2, σ_B = 4,695, по формуле

$$n_i = \frac{200 \cdot 2}{4{,}695} \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right) = 85{,}2 \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right).$$

Составим расчетную таблицу

i	x_i	$u_i = \dfrac{\left(x_i - \bar{x}_B\right)}{\sigma_B}$	$\varphi(u_i)$	$n_i = \dfrac{n \cdot h}{\sigma_B} \cdot \varphi(u_i)$
1	5	-1,62	0,1074	9,1
2	7	-1,20	0,1942	16,5
3	9	-0,77	0,2966	25,3
4	11	-0,35	0,3752	32,0

5	13	0,08	0,3977	33,9
6	15	0,51	0,3503	29,8
7	17	0,93	0,2589	22,0
8	19	1,36	0,1582	13,5
9	21	1,78	0,0818	7,0

3. Сравним эмпирические и теоретические частоты. Составим расчетную таблицу, из которой найдем наблюдаемое значение критерия $\chi^2_{набл} = \sum_{i=1}^{s} \frac{\left(n_i - n_i'\right)^2}{n_i'}$:

i	n_i	n_i'	$\left\|n_i - n_i'\right\|$	$\left(n_i - n_i'\right)^2$	$\frac{\left(n_i - n_i'\right)^2}{n_i'}$
1	15	9,1	5,9	34,81	3,8
2	26	16,5	9,5	90,25	5,5
3	25	25,3	0,3	0,09	0,0
4	30	32,0	2,0	4,0	0,1
5	26	33,9	7,9	62,41	1,8
6	21	29,8	8,8	77,44	2,6
7	24	22,0	2,0	4,0	0,2
8	20	13,5	6,5	42,25	3,1
9	13	7,0	6,0	36,0	5,1
Сумма	200				$\chi^2_{набл}$ =22,2

По таблице критических точек распределения χ^2 (приложение 6), по уровню значимости a = 0,05 и числу степеней свободы k = s – 3 = 9 – 3 = 6 находим критическую точку правосторонней критической области $\chi^2_{кр}$ (0,05; 6) = 12,6.

Так как $\chi^2_{набл}$ =22,2 > $\chi^2_{кр}$ = 12,6, гипотезу о нормальном распределении генеральной совокупности отвергаем. Другими словами, эмпирические и теоретические частоты различаются значимо.

Пример 4: Представлены статистические данные.
Результаты измерений диаметров n = 200 валков после шлифовки обобщены в табл. (мм):
Таблица Частотный вариационный ряд диаметров валков

i	1	2	3	4	5	6	7	8
xi, мм	6,68	6,69	6,7	6,71	6,72	6,73	6,74	6,75
ni	2	3	12	6	11	14	30	25

i	9	10	11	12	13	14	15	16
xi, мм	6,76	6,77	6,78	6,79	6,8	6,81	6,82	6,83
ni	27	31	14	8	5	6	5	1

Требуется:
1) составить дискретный вариационный ряд, при необходимости упорядочив его;
2) определить основные числовые характеристики ряда;

3) дать графическое представление ряда в виде полигона (гистограммы) распределения;
4) построить теоретическую кривую нормального распределения и проверить соответствие эмпирического и теоретического распределений по критерию Пирсона. При проверке статистической гипотезы о виде распределения принять уровень значимости a = 0,05

Решение: Основные числовые характеристики данного вариационного ряда найдем по определению. Средний диаметр валков равен (мм):

$x_{ср} = \frac{1}{n}\sum_{k=1}^{16} n_i \cdot x_k = 6{,}753$; исправленная дисперсия (мм2): $D = \frac{1}{n-1}\sum_{k=1}^{16} n_i \cdot (x_k - x_{cp})^2 = 0{,}0009166$;

исправленное среднее квадратическое (стандартное) отклонение (мм):$s = \sqrt{D} = 0{,}03028$.

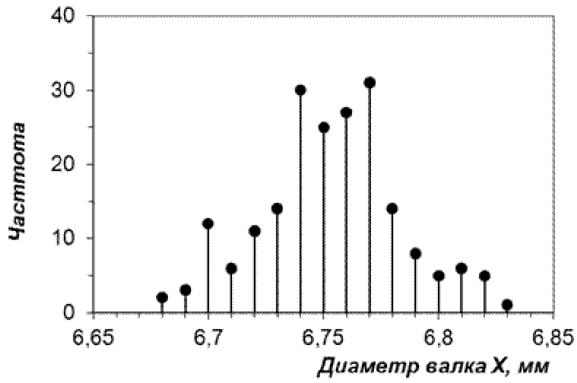


Рис. Частотное распределение диаметров валков

Исходное («сырое») частотное распределение вариационного ряда, т.е. соответствие ni(xi), отличается довольное большим разбросом значений ni относительно некоторой гипотетической «усредняющей» кривой (рис.). В этом случае предпочтительно построить и анализировать интервальный вариационный ряд, объединяя частоты для диаметров, попадающих в соответствующие интервалы.
Число интервальных групп K определим по формуле Стерджесса:
K = 1 + log2n = 1 + 3,322lgn, где n = 200 – объем выборки. В нашем случае
K = 1 + 3,322×lg200 = 1 + 3,322×2,301 = 8,644 » 8.
Ширина интервала равна (6,83 – 6,68)/8 = 0,01875 » 0,02 мм.
Интервальный вариационный ряд представлен в табл.

Таблица Частотный интервальный вариационный ряд диаметров валков.

k	1	2	3	4	5	6	7	8
xk, мм	6,68 – 6,70	6,70 – 6,72	6,72 – 6,74	6,74 – 6,76	6,76 – 6,78	6,78 – 6,80	6,80 – 6,82	6,82 – 6,84
nk	5	18	25	55	58	22	11	6

Интервальный ряд может быть наглядно представлен в виде гистограммы частотного распределения.

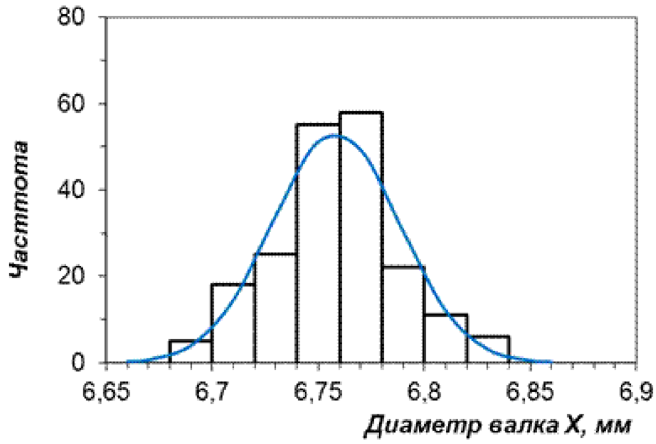


Рис. Частотное распределение диаметров валков. Сплошная линия – сглаживающая нормальная кривая.

Вид гистограммы позволяет сделать предположение о том, что распределение диаметров валков подчиняется нормальному закону, согласно которому теоретические частоты могут быть найдены как

nk, теор = n×N(a; s; x_k)×D_{xk}, где, в свою очередь, сглаживающая гауссова кривая нормального распределения определяется выражением: N(a; s; xk) = $\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x_k-a)^2}{2\sigma^2}\right\}$.

В этих выражениях xk – центры интервалов в частотном интервальном вариационном ряде.

Например, x1 = (6,68 + 6,70)/2 = 6,69. В качестве оценок центра a и параметра s гауссовой кривой можно принять: a = $x_{ср}$.

Из рис. видно, что гауссова кривая нормального распределения в целом соответствует эмпирическому интервальному распределению. Однако следует удостовериться в статистической значимости этого соответствия. Используем для проверки соответствия эмпирического распределения эмпирическому критерий согласия Пирсона. Для этого следует вычислить эмпирическое значение критерия как сумму $\chi^2 = \sum_{k=1}^{8}\frac{(n_k - n_{k,теор})^2}{n_{k,теор}}$,

где n_k и n_k,теор – эмпирические и теоретические (нормальные) частоты, соответственно. Результаты расчетов удобно представить в табличном виде:

Таблица Вычисления критерия Пирсона

[x_k, x_{k+1}), мм	x_k, мм	n_k	n_k,теор	$\frac{(n_k - n_{k,теор})^2}{n_{k,теор}}$
6,68 – 6,70	6,69	5	4,00	0,25
6,70 – 6,72	6,71	18	14,57	0,81
6,72 – 6,74	6,73	25	34,09	2,42
6,74 – 6,76	6,75	55	51,15	0,29
6,76 – 6,78	6,77	58	49,26	1,55
6,78 – 6,80	6,79	22	30,44	2,34
6,80 – 6,82	6,81	11	12,07	0,09
6,82 – 6,84	6,83	6	3,07	2,80
			c2эмп	10,55

Критическое значение критерия найдем по таблице Пирсона [2, 3] для уровня значимости a = 0,05 и числа степеней свободы d.f. = K – 1 – r, где K = 8 – число интервалов интервального вариационного ряда; r = 2 – число параметров теоретического распределения, оцененных на основании данных выборки (в данном случае, – параметры a и s). Таким образом, d.f. = 5. Критическое значение критерия Пирсона есть χ^2 крит(a; d.f.) = 11,1. Так как c2эмп < c2крит, заключаем, что согласие между эмпирическим и теоретическим нормальным распределением является статистическим значимым. Иными словами, теоретическое нормальное распределение удовлетворительно описывает эмпирические данные.

Пример5: Коробки с шоколадом упаковываются автоматически. По схеме собственно-случайной бесповторной выборки взято 130 из 2000 упаковок, содержащихся в партии, и получены следующие данные об их весе:

Вес упаковки (гр.)	Менее 975	975-1000	1000-1025	1025-1050	Более 1050	Всего
Число упаковок	6	38	44	34	8	130

Требуется используя критерий $\chi^2 -$ Пирсона при уровне значимости a=0,05 проверить гипотезу о том, что случайная величина X – вес упаковок – распределена по нормальному закону. Построить на одном графике гистограмму эмпирического распределения и соответствующую нормальную кривую.

Решение

$\bar{x} =$ 1012,5 s^2 = 615,3846

Примечание:

В принципе в качестве дисперсии нормального закона распределения следует взять исправленную выборочную дисперсию. Но т.к. количество наблюдений – 130 достаточно велико, то подойдет и «обычная» s^2.

Таким образом, теоретическое нормальное распределение имеет вид:

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

подставляем a = 1012,5 σ^2 = 615,3846 $\sigma =$ 24,8069

$$f(x) = 0{,}0160819 \cdot e^{-0{,}0008125(x-1012{,}5)^2}$$

Для расчета вероятностей p_i попадания случайной величины в интервал $[x_i ; x_{i+1}]$ используем функцию Лапласа:

$$p_i(x_i \le X \le x_{i+1}) = \frac{1}{2}\left[\Phi\left(\frac{x_{i+1} - a}{\sigma}\right) - \Phi\left(\frac{x_i - a}{\sigma}\right)\right]$$

$$p_i(x_i \le X \le x_{i+1}) \approx \frac{1}{2}\left[\Phi\left(\frac{x_{i+1} - 1012{,}5}{24{,}8069}\right) - \Phi\left(\frac{x_i - 1012{,}5}{24{,}8069}\right)\right]$$

в нашем случае получаем:

$$p_i(950 \le X \le 975) \approx \frac{1}{2}[\Phi(-1{,}51) - \Phi(-2{,}52)] = \frac{1}{2}[0{,}9883 - 0{,}869] = 0{,}0597$$

$$p_i(975 \le X \le 1000) \approx \frac{1}{2}[\Phi(-0{,}5) - \Phi(-1{,}51)] = \frac{1}{2}[0{,}869 - 0{,}3829] = 0{,}2431$$

$$p_i(1000 \le X \le 1025) \approx \frac{1}{2}[\Phi(0{,}5) - \Phi(-0{,}5)] = \frac{1}{2}[0{,}3829 + 0{,}3829] = 0{,}3829$$

$$p_i(1025 \le X \le 1050) \approx \frac{1}{2}[\Phi(1{,}51) - \Phi(0{,}5)] = \frac{1}{2}[0{,}869 - 0{,}3829] = 0{,}2431$$

$$p_i(1050 \le X \le 1075) \approx \frac{1}{2}[\Phi(2{,}52) - \Phi(1{,}51)] = \frac{1}{2}[0{,}9883 - 0{,}869] = 0{,}0597$$

Примечание: Такие симметричные вероятности получились из-за того, что по нашим начальным условиям выборочная средняя попала точно в середину среднего интервала выборки.

Составим таблицу:

i	Интервал [xi ; xi+1]	Эмпирические частоты ni	Вероятности pi	Теоретические частоты npi	(ni-npi)2	$\frac{(n_i - np_i)^2}{np_i}$
1	Менее 975	6	0,0597	7,761	3,101	0,3996
2	975-1000	38	0,2431	31,603	40,922	1,2949
3	1000-1025	44	0,3829	49,777	33,374	0,6705
4	1025-1050	34	0,2431	31,603	5,746	0,1818
5	Более 1050	8	0,0597	7,761	0,057	0,0073
Σ		130	0,9885	128,5		$\chi^2 = 2{,}55$

Итого, значение статистики $\chi^2 = 2{,}55$.

Определим количество степеней свободы по формуле: $k = m - r - 1$.

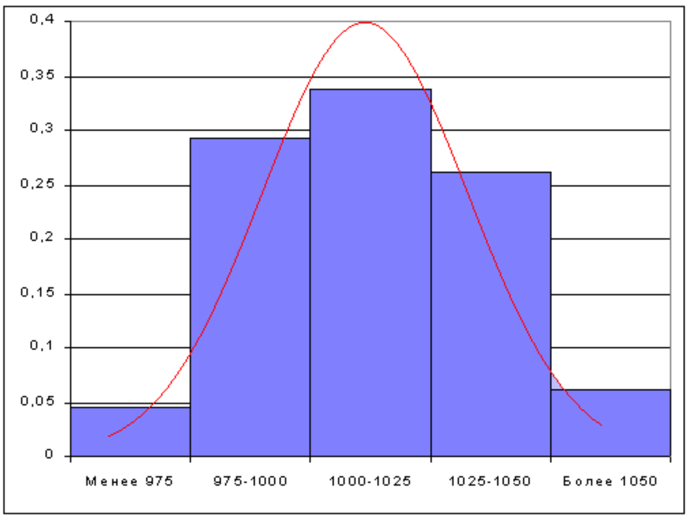
m – число интервалов (m = 5), r– число параметров закона распределения (в нормальном распределении r = 2) Т.е. k = 2.

Соответствующее критическое значение статистики

$\chi^2_{0{,}05;2} = 5{,}99$

Поскольку $\chi^2_{0{,}05;2} > \chi^2$, гипотеза о нормальном распределении с параметрами N(1012,5; 615,3846) согласуется с опытными данными.

Ниже показана гистограмма эмпирического распределения и соответствующая нормальная кривая.



3. Оценка параметров неизвестного распределения. Выравнивание рядов.

Пример. С целью исследования закона распределения ошибки измерения дальности с помощью радиодальномера произведено 400 измерений дальности. Результаты опытов представлены в виде статистического ряда:

I_i (м)	20;30	30;40	40;50	50;60	60;70	70;80	80;90	90;100
m_i	21	72	66	38	51	56	64	32
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,080

Выровнять статистический ряд с помощью закона равномерной плотности.

Решение. Закон равномерной плотности выражается формулой

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \text{при } \alpha < x < \beta, \\ 0 & \text{при } x < \alpha \text{ или } x > \beta\end{cases}$$

и зависит от двух параметров α и β. Эти параметры следует выбрать так, чтобы сохранить первые два момента статистического распределения –

математическое ожидание m_x^* и дисперсию D_x^*. Имеем выражения математического ожидания и дисперсии для закона равномерной плотности:

$$m_x=\frac{\alpha+\beta}{2};$$

$$D_x=\frac{(\beta-\alpha)^2}{12}.$$

Для того, чтобы упростить вычисления, связанные с определением статистических моментов, перенесем начало отсчета в точку $x_0=60$ и примем за представителя его разряда его середину. Ряд распределения имеет вид:

$\tilde{x}_i'$	−35	−25	−15	−5	5	15	25	35
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,080

где $\tilde{x}_i'$ - среднее для разряда значение ошибки радиодальномера X' при новом начале отсчета.

Приближенное значение статистического среднего ошибки X' равно:

$$m_{x'}^*=\sum_{i=1}^{k}\tilde{x}_i' p_i^*=0,26$$

Второй статистический момент величины X' равен: $\alpha_2^*=\sum_{i=1}^{k}(\tilde{x}_i')^2 p_i^*=447,8$,

откуда статистическая дисперсия:

$$D_{x'}^*=\alpha_2^*-\left(m_{x'}^*\right)^2=447,7.$$

Переходя к прежнему началу отсчета, получим новое статистическое среднее:

$m_x^*=m_{x'}^*+60=60,26$ в ту же статистическую дисперсию:

$$D_x^*=D_{x'}^*=447,7.$$

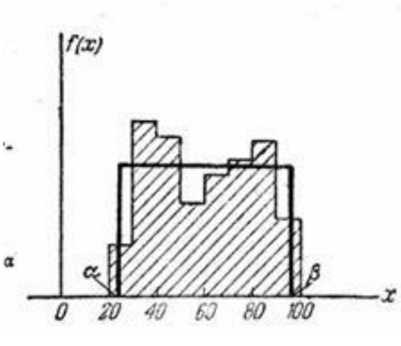


Рис. 1

Параметры закона равномерной плотности определяются уравнениями:

$$\frac{\alpha+\beta}{2}=60,26; \quad \frac{(\beta-\alpha)^2}{12}=447,7.$$

Решая эти уравнения относительно α и β, имеем: $\alpha\approx 23,6; \quad \beta\approx 96,9$, откуда

$$f(x)=\frac{1}{\beta-\alpha}=\frac{1}{73,3}\approx 0,0136.$$

На рис. 1. показаны гистограмма и выравнивающий ее закон равномерной плотности $f(x)$.

3.6.3 Результаты и выводы: В результате проведенного занятия студенты:
- освоили основные понятия теории проверки статистических гипотез;
- усвоили алгоритмы применения статистических критериев;
- выработали навыки по применению критерия Пирсона;
- усвоили основные методы и алгоритмы выравнивания статистических рядов.

3.7 Практическое занятие 11-12 (ПЗ-11-12)

Тема: Стохастическая зависимость межу величинами.

3.7.1 Задание для работы:

1. Виды зависимостей между величинами.
2. Функция регрессии. Корреляционное отношение. Коэффициент детерминации. Значимость выборочных коэффициентов.
3. Линейная парная регрессия.
4. Коэффициент корреляции, его свойства, значимость.

3.7.2 Краткое описание проводимого занятия

1. Виды зависимостей между величинами.

Сущность корреляционно-регрессионного анализа и его задачи. Экономические явления, будучи весьма разнообразными, характеризуются множеством признаков, отражающих определенные свойства этих процессов и явлений и подверженных взаимообусловленным изменениям. В одних случаях зависимость между признаками оказывается очень тесной (например, часовая выработка работника и его заработная плата), а в других случаях такая связь не выражена вовсе или крайне слаба (например, пол студентов и их успеваемость). Чем теснее связь между этими признаками, тем точнее принимаемые решения.

Различают два типа зависимостей между явлениями и их признаками:

• ***функциональная (детерминированная, причинная) зависимость***. Задается в виде формулы, которая каждому значению одной переменной ставит в соответствие строго определенное значение другой переменной (воздействием случайных факторов при этом пренебрегают). Иными словами, ***функциональная зависимость***– это связь, при которой каждому значению независимой переменной x соответствует точно определенное значение зависимой переменной y. В экономике функциональные связи между переменными являются исключениями из общего правила;

• ***статистическая (стохастическая, недетерминированная) зависимость***– это связь переменных, на которую накладывается воздействие случайных факторов, т.е. это связь, при которой каждому значению независимой переменной x соответствует множество значений зависимой переменной y, причем заранее неизвестно, какое именно значение примет y.

Частным случаем статистической зависимости является корреляционная зависимость.

Корреляционная зависимость– это связь, при которой каждому значению независимой переменной x соответствует определенное математическое ожидание (среднее значение) зависимой переменной y.

Корреляционная зависимость является «неполной» зависимостью, которая проявляется не в каждом отдельном случае, а только в средних величинах при достаточно большом числе случаев. Например, известно, что повышение квалификации работника ведет к росту производительности труда. Это утверждение часто подтверждается на практике, но не означает, что у двух и более работников одного разряда / уровня, занятых аналогичным процессом, будет одинаковая производительность труда.

Корреляционная зависимость исследуется с помощью методы корреляционного и регрессионного анализа.

Корреляционно-регрессионный анализ позволяет установить тесноту, направление связи и форму этой связи между переменными, т.е. ее аналитическое выражение.

Основная задача корреляционного анализа состоит в количественном определении тесноты связи между двумя признаками при парной связи и между результативными и несколькими факторными признаками при многофакторной связи и статистической оценке надежности установленной связи.

2. Функция регрессии. Корреляционное отношение. Коэффициент детерминации. Значимость выборочных коэффициентов.

Данные наблюдений над двумерной случайной величиной (X, Y) представлены в корреляционной таблице. Найти выборочное уравнение прямой регрессии Y на X:

$$\overline{y}_x - \overline{y} = r_в \frac{\sigma_y}{\sigma_x}(x - \overline{x}).$$

x \ y	5	6	7	8	9	10	ny
2	-	-	-	-	6	4	10
4	-	-	-	6	6	8	20
6	-	3	4	14	3	-	24
8	2	5	18	2	-	-	27
10	-	7	10	2	-	-	19
nx	2	15	32	24	15	12	100

Решение:

Предварительно вычислим суммы

$$\sum_{i=1}^{100} x_i = 5 \cdot 2 + 6 \cdot 15 + 7 \cdot 32 + 8 \cdot 24 + 9 \cdot 15 + 10 \cdot 12 = 771.$$

$$\sum_{i=1}^{100} y_i = 2\cdot10+4\cdot20+6\cdot24+8\cdot27+10\cdot19=650.$$

$$\sum_{i=1}^{100} x_i^2 = 5^2\cdot2+6^2\cdot15+7^2\cdot32+8^2\cdot24+9^2\cdot15+10^2\cdot12=6109.$$

$$\sum_{i=1}^{100} y_i^2 = 2^2\cdot10+4^2\cdot20+6^2\cdot24+8^2\cdot27+10^2\cdot19=4852.$$

$$\sum_{i=1}^{100} x_i y_i = 2\cdot(9\cdot6+10\cdot4)+4\cdot(8\cdot6+9\cdot6+10\cdot8)+6\cdot(6\cdot3+7\cdot4+8\cdot14+9\cdot3)+$$

$$+8\cdot(5\cdot2+6\cdot5+7\cdot18+8\cdot2)+10\cdot(6\cdot7+7\cdot10+8\cdot2)=4762.$$

Средние арифметические значения

$$\overline{x}=\frac{1}{n}\sum x_i=\frac{771}{100}=7{,}71.\ \overline{y}=\frac{1}{n}\sum y_i=\frac{650}{100}=6{,}5.$$

Дисперсии и средние квадратические отклонения

$$\sigma_x^2=\frac{1}{n}\sum x_i^2-(\overline{x})^2=\frac{1}{100}\cdot6109-7{,}71^2=1{,}6459.$$

$$\sigma_y^2=\frac{1}{n}\sum y_i^2-(\overline{y})^2=\frac{1}{100}\cdot4852-6{,}5^2=6{,}27.$$

$$\sigma_x=\sqrt{1{,}6459}=1{,}28.\ \sigma_y=\sqrt{6{,}27}=2{,}5.$$

Найдем корреляционный момент

$$C_{xy}=\frac{1}{n}\sum x_i y_i-\overline{x}\cdot\overline{y}=\frac{1}{100}\cdot4762-7{,}71\cdot6{,}5=-2{,}5.$$

Находим коэффициент корреляции

$$r_в=\frac{C_{xy}}{\sigma_x\sigma_y}=\frac{-2{,}5}{1{,}28\cdot2{,}5}=-0{,}78$$

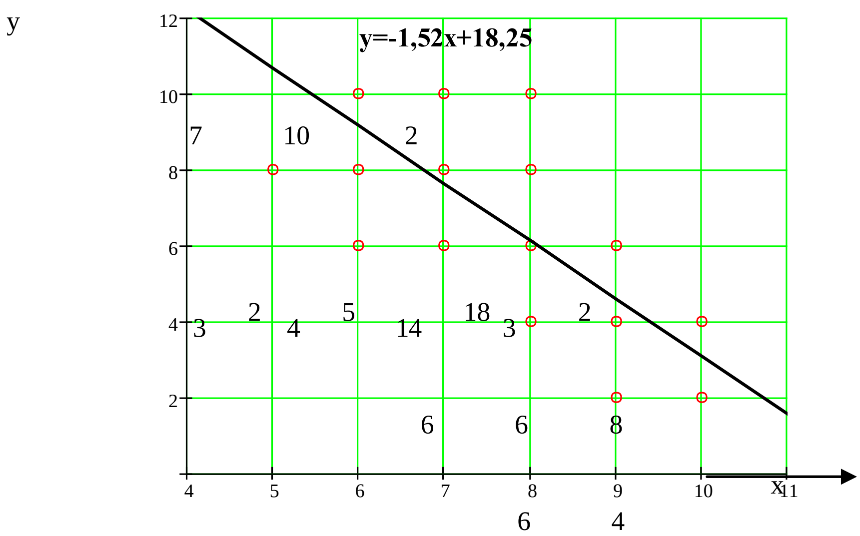
Находим уравнение линии эмпирической регрессии

$$\overline{y}_x-\overline{y}=r_в\frac{\sigma_y}{\sigma_x}(x-\overline{x}).\ \text{Или}$$

$$\overline{y}_x-6{,}5=-0{,}78\frac{2{,}5}{1{,}28}(x-7{,}71).\ \text{Окончательно}$$

$$\overline{y}_x=-1{,}52x+18{,}25.$$

В системе координат x и y, используя корреляционную таблицу, соответствующими точками изображаем корреляционное поле и наносим прямую выборочной регрессии согласно полученного уравнения. На корреляционном поле цифрами показано количество совпадающих точек.



Коэффициент детерминации характеризует долю вариации (дисперсии) результативного признака *y*, объясняемую регрессией, в общей вариации (дисперсии) *y*.

Коэффициент детерминации рассчитывается для оценки качества подбора уравнения регрессии. Для приемлемых моделей предполагается, что коэффициент детерминации должен быть хотя бы не меньше 50%. Модели с коэффициентом детерминации выше 80% можно признать достаточно хорошими. Значение коэффициента детерминации $R^2 = 1$ означает функциональную зависимость между переменными.

3. Линейная парная регрессия.

Парная линейная регрессия $\tilde{y}_x = a + bx$

Предварительные расчеты:

$\bar{x} = \frac{1}{n}\sum x_i$; $\bar{y} = \frac{1}{n}\sum y_i$; $\overline{x^2} = \frac{1}{n}\sum x_i^2$; $\overline{y^2} = \frac{1}{n}\sum y_i^2$; $\overline{xy} = \frac{1}{n}\sum x_i y_i$;

$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2}$; $\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2}$.

Построение таблицы вида

	x	y	xy	x^2	y^2
…………	……….	………	……..	………	………
Среднее значение					

Формулы для расчетов параметров:

$$a = \bar{y} - b\bar{x}, \quad b = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2}.$$

При компьютерном подборе использовать встроенную функцию Линейн

Оценка тесноты связи:

а) коэффициент корреляции $r_{xy} = b\frac{\sigma_x}{\sigma_y}$, или $r_{xy} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x\sigma_y}$;

При компьютерном подборе использовать встроенную функцию Коррел

б) коэффициент эластичности $Э_{xy} = b\frac{\bar{x}}{\bar{y}}$;

в) коэффициент детерминации r_{xy}^2.

Оценка значимости уравнения регрессии в целом:

Предварительные расчеты с построением таблицы вида

	x	y	$\tilde{y}$	$y-\tilde{y}$	$(y-\tilde{y})^2$	$\left\|\frac{y-\tilde{y}}{y}\right\| \cdot 100\%$
…………	……….	………	……..	………	………	

а) F-критерий Фишера при числе степеней свободы $k_1 = 1$ и $k_2 = n-2$ и уровне значимости 0,05 смотреть в таблице. Расчетное значение критерия:

$$F = \frac{r_{xy}^2}{1-r_{xy}^2}(n-2).$$

Если расчетное значение F- критерия больше табличного, нулевая гипотеза об отсутствии значимой связи признаков x и y отклоняется, и делается вывод о существенности этой связи.

б) Средняя ошибка аппроксимации

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \cdot 100\%.$$

Оценка значимости параметров регрессии:

а) Стандартная ошибка параметра a рассчитывается по формуле

$$m_a = S\frac{\sqrt{\sum x^2}}{n\sigma_x}, \text{ где } S^2 = D_{ост} = \frac{\sum (y-\tilde{y})^2}{n-2}.$$

б) Стандартная ошибка коэффициента регрессии b рассчитывается по формуле

$$m_b = \frac{S}{\sigma_x \sqrt{n}}.$$

в) Стандартная ошибка коэффициента корреляции r_{xy} рассчитывается по формуле

$$m_r = \sqrt{\frac{1-r_{xy}^2}{n-2}}.$$

t-критерий Стъюдента при числе степеней свободы $n-2$ и уровне значимости 0,05 смотреть в таблице.

Фактические значения t-статистики: $t_a = \frac{a}{m_a}$, $t_b = \frac{b}{m_b}$, $t_r = \frac{r}{m_r}$.

Если фактическое значение по абсолютной величине превышает табличное, гипотезу о несущественности параметра регрессии можно отклонить, параметр признается значимым.

Связь между F-критерием Фишераи t-критерием Стъюдента выражается равенством

$$t_b^2 = t_r^2 = F.$$

Расчет доверительных интервалов для параметров регрессии:

Доверительный интервал для параметра a определяется как $a \pm t_{табл} \cdot m_a$;

доверительный интервал для коэффициента регрессии определяется как $b \pm t_{табл} \cdot m_b$.

При компьютерном анализе использовать в ExcelСервис/Анализ данных/Регрессия.

Интервальный прогноз на основе линейного уравнения регрессии:

Пусть x_p – прогнозное значение факторного признака; $\tilde{y}_{x_p}$ – точечный прогноз результативного признака. Тогда

а) средняя ошибка прогноза $m_{\tilde{y}}$: $m_{\tilde{y}} = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \overline{x})^2}{n\sigma_x^2}}$;

б) доверительный интервал прогноза $\tilde{y}_{x_p} - t_\alpha \cdot m_{\tilde{y}} \le \tilde{y}_p \le \tilde{y}_{x_p} + t_\alpha \cdot m_{\tilde{y}}$.

Задача:
Имеется связанная выборка из 26 пар значений (x_k,y_k):

k	1	2	3	4	5	6	7	8	9	10
xk	25.20000	26.40000	26.00000	25.80000	24.90000	25.70000	25.70000	25.70000	26.10000	25.80000
yk	30.80000	29.40000	30.20000	30.50000	31.40000	30.30000	30.40000	30.50000	29.90000	30.40000

k	11	12	13	14	15	16	17	18	19	20
xk	25.90000	26.20000	25.60000	25.40000	26.60000	26.20000	26.00000	22.10000	25.90000	25.80000
yk	30.30000	30.50000	30.60000	31.00000	29.60000	30.40000	30.70000	31.60000	30.50000	30.60000

k	21	22	23	24	25	26
xk	25.90000	26.30000	26.10000	26.00000	26.40000	25.80000
yk	30.70000	30.10000	30.60000	30.50000	30.70000	30.80000

Требуется вычислить/построить:

- коэффициент корреляции;
- проверить гипотезу зависимости случайных величин X и Y, при уровне значимости α = 0.05;
- коэффициенты уравнения линейной регрессии;
- диаграмму рассеяния (корреляционное поле) и график линии регрессии;

Коэффициент корреляции — это показатель взаимного вероятностного влияния двух случайных величин. Коэффициент корреляции **R** может принимать значения от **-1** до **+1**. Если абсолютное значение находится ближе к **1**, то это свидетельство сильной связи между величинами, а если ближе к **0** — то, это говорит о слабой связи или ее отсутствии. Если абсолютное значение **R** равно единице, то можно говорить о функциональной связи между величинами, то есть одну величину можно выразить через другую посредством математической функции.

Вычислить коэффициент корреляции можно по следующим формулам:
cov(X,Y) - ковариация случайных величин X и Y

$$R_{x,y} = \frac{cov(X,Y)}{\sigma_x \sigma_y} \qquad (1),$$ где:

$$\sigma_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - M_x)^2, \quad \sigma_y^2 = \frac{1}{n}\sum_{k=1}^{n}(y_k - M_y)^2 \quad (2),$$ - оценки дисперсий случайных величин X и Y соответственно.

$$M_x = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad M_y = \frac{1}{n}\sum_{k=1}^{n} y_k \quad (3),$$ - оценки математического ожидания случайных величин X и Y соответственно.

или по формуле

$$R_{x,y} = \frac{M_{xy} - M_x M_y}{S_x S_y} \quad (4),$$ где:

$$M_x = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad M_y = \frac{1}{n}\sum_{k=1}^{n} y_k, \quad M_{xy} = \frac{1}{n}\sum_{k=1}^{n} x_k y_k \quad (5)$$

$$S_x^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - M_x^2, \quad S_y^2 = \frac{1}{n}\sum_{k=1}^{n} y_k^2 - M_y^2 \quad (6)$$

На практике, для вычисления коэффициента корреляции чаще используется формула (4) т.к. она требует меньше вычислений. Однако если предварительно была вычислена ковариация **cov(X,Y)**, то выгоднее использовать формулу (1), т.к. кроме собственно значения ковариации можно воспользоваться и результатами промежуточных вычислений.

1.1 Вычислим коэффициент корреляции по формуле (4), для этого вычислим значения x_k^2, y_k^2 и $x_k y_k$ и занесем их в таблицу 1.

Таблица 1

k	x_k	y_k	x_k^2	y_k^2	$x_k y_k$
1	2	3	4	5	6
1	25.2	30.8	635.04000	948.64000	**776.16000**
2	26.4	29.4	696.96000	864.36000	**776.16000**
3	26.0	30.2	676.00000	912.04000	**785.20000**
4	25.8	30.5	665.64000	930.25000	**786.90000**
5	24.9	31.4	620.01000	985.96000	**781.86000**
6	25.7	30.3	660.49000	918.09000	**778.71000**
7	25.7	30.4	660.49000	924.16000	**781.28000**
8	25.7	30.5	660.49000	930.25000	**783.85000**
9	26.1	29.9	681.21000	894.01000	**780.39000**
10	25.8	30.4	665.64000	924.16000	**784.32000**
11	25.9	30.3	670.81000	918.09000	**784.77000**
12	26.2	30.5	686.44000	930.25000	**799.10000**
13	25.6	30.6	655.36000	936.36000	**783.36000**
14	25.4	31	645.16000	961.00000	**787.40000**
15	26.6	29.6	707.56000	876.16000	**787.36000**

16	26.2	30.4	686.44000	924.16000	**796.48000**
17	26	30.7	676.00000	942.49000	**798.20000**
18	22.1	31.6	488.41000	998.56000	**698.36000**
19	25.9	30.5	670.81000	930.25000	**789.95000**
20	25.8	30.6	665.64000	936.36000	**789.48000**
21	25.9	30.7	670.81000	942.49000	**795.13000**
22	26.3	30.1	691.69000	906.01000	**791.63000**
23	26.1	30.6	681.21000	936.36000	**798.66000**
24	26	30.5	676.00000	930.25000	**793.00000**
25	26.4	30.7	696.96000	942.49000	**810.48000**
26	25.8	30.8	665.64000	948.64000	**794.64000**

1.2. Вычислим M_x по формуле (5).

1.2.1. Сложим последовательно все элементы x_k

$x_1 + x_2 + \ldots + x_{26} =$ 25.20000 + 26.40000 + ... + 25.80000 = 669.500000

1.2.2. Разделим полученную сумму на число элементов

669.50000 / 26 = 25.75000

M_x = 25.750000

1.3. Аналогичным образом вычислим M_y.

1.3.1. Сложим последовательно все элементы y_k

$y_1 + y_2 + \ldots + y_{26} =$ 30.80000 + 29.40000 + ... + 30.80000 = 793.000000

1.3.2. Разделим полученную сумму на число элементов выборки

793.00000 / 26 = 30.50000

M_y = 30.500000

1.4. Аналогичным образом вычислим M_{xy}.

1.4.1. Сложим последовательно все элементы 6-го столбца таблицы 1

776.16000 + 776.16000 + ... + 794.64000 = 20412.830000

1.4.2. Разделим полученную сумму на число элементов

20412.83000 / 26 = 785.10885

M_{xy} = 785.108846

1.5. Вычислим значение S_x^2 по формуле (1.6.).

1.5.1. Сложим последовательно все элементы 4-го столбца таблицы 1

635.04000 + 696.96000 + ... + 665.64000 = 17256.910000

1.5.2. Разделим полученную сумму на число элементов

17256.91000 / 26 = 663.72731

1.5.3. Вычтем из последнего числа квадрат величины M_x получим значение для S_x^2

S_x^2 = 663.72731 - 25.75000^2 = 663.72731 - 663.06250 = **0.66481**

1.6. Вычислим значение S_y^2 по формуле (1.6.).

1.6.1. Сложим последовательно все элементы 5-го столбца таблицы 1

948.64000 + 864.36000 + ... + 948.64000 = 24191.840000

1.6.2. Разделим полученную сумму на число элементов

24191.84000 / 26 = 930.45538

1.6.3. Вычтем из последнего числа квадрат величины M_y получим значение для S_y^2

S_y^2 = 930.45538 - 30.50000^2 = 930.45538 - 930.25000 = **0.20538**

1.7. Вычислим произведение величин S_x^2 и S_y^2.

$S_x^2 S_y^2$ = 0.66481 • 0.20538 = 0.136541

1.8. Извлечем и последнего числа квадратный корень, получим значение S_xS_y. S_xS_y = 0.36951

1.9. Вычислим значение коэффициента корреляции по формуле (1.4.). R = (785.10885 - 25.75000 • 30.50000) / 0.36951 = (785.10885 - 785.37500) / 0.36951 = -0.72028

ОТВЕТ: $R_{x,y}$ = -0.720279

2. Проверяем значимость коэффициента корреляции (проверяем гипотезу зависимости).

Поскольку оценка коэффициента корреляции вычислена на конечной выборке, и поэтому может отклоняться от своего генерального значения, необходимо проверить значимость коэффициента корреляции. Проверка производится с помощью t-критерия:

Случайная величина **t** следует t-распределению Стьюдента и по таблице t-распределения необходимо найти критическое значение критерия ($t_{кр.\alpha}$) при заданном уровне значимости α. Если вычисленное t по модулю окажется меньше чем $t_{кр.\alpha}$, то зависимости между случайными величинами X и Y нет. В противном случае, экспериментальные данные не противоречат гипотезе о зависимости случайных величин.

2.1. Вычислим значение t-критерия)

2.2. Определим по таблице t-распределения критическое значение параметра $t_{кр.\alpha}$ Искомое значение $t_{кр.\alpha}$ располагается на пересечении строки соответствующей числу степеней свободы и столбца соответствующего заданному уровню значимости α. В нашем случае число степеней свободы есть n - 2 = 26 - 2 = **24** и α = **0.05** , что соответствует критическому значению критерия $t_{кр.\alpha}$ = **2.064**

2.2. Сравним абсолютное значение t-критерия и $t_{кр.\alpha}$

Абсолютное значение t-критерия не меньше критического t = 5.08680, $t_{кр.\alpha}$ = 2.064, следовательно **экспериментальные данные, с вероятностью 0.95** (1 - α), **не противоречат гипотезе** о зависимости случайных величин X и Y.

3. Вычисляем коэффициенты уравнения линейной регрессии.

Уравнение линейной регрессии представляет собой уравнение прямой, аппроксимирующей (приблизительно описывающей) зависимость между случайными величинами X и Y. Если считать, что величина X свободная, а Y зависимая от X, то уравнение регрессии запишется следующим образом

$$Y = a + b \cdot X \qquad (1), \quad \text{где:}$$

$$b = R_{x,y}\frac{\sigma_y}{\sigma_x} = R_{x,y}\frac{S_y}{S_x} \qquad (2),$$

$$a = M_y - b \cdot M_x \qquad (3)$$

Рассчитанный по формуле (2) коэффициент **b** называют коэффициентом линейной регрессии. В некоторых источниках **a** называют постоянным коэффициентом регрессии и **b** соответственно переменным.

Погрешности предсказания Y по заданному значению X вычисляются по формулам :

$$\sigma_{y/x} = \sigma_y\sqrt{1-R^2_{x,y}} = S_y\sqrt{1-R^2_{x,y}} \qquad (4)$$ - абсолютная погрешность,

$$\delta_{y/x} = \frac{\sigma_{y/x}}{M_y}100\% \qquad (5)$$ - относительная погрешность

Величину $\sigma_{y/x}$ (формула 4) еще называют **остаточным средним квадратическим отклонением**, оно характеризует уход величины Y от линии регрессии, описываемой уравнением (1), при фиксированном (заданном) значении X.

3.1. Вычислим отношение $\frac{S_y^2}{}$.

S_x^2

$S_y^2 / S_x^2 = 0.20538 / 0.66481 = 0.30894$

3.2. Вычислим отношение $\dfrac{S_y}{S_x}$.

Извлечем из последнего числа квадратный корень - получим:
$S_y / S_x = 0.55582$

3.3 Вычислим коэффициент b по формуле (2)

b = $-0.72028 \cdot 0.55582 =$ **-0.40035**

3.4 Вычислим коэффициент a по формуле (3)

a = $30.50000 - (-0.40035 \cdot 25.75000) =$ **40.80894**

3.5 Оценим погрешности уравнения регрессии.

3.5.1 Извлечем из S_y^2 квадратный корень получим:

$S_y = \sqrt{0.20538} = 0.45319$;

3.5.2 Возведем в квадрат $R_{x,y}$ получим:

$R^2_{x,y} = -0.72028^2 = 0.51880$

3.5.3 Вычислим абсолютную погрешность (остаточное среднее квадратическое отклонение) по формуле (4)

$\sigma_{y/x} = 0.45319 \sqrt{1 - 0.51880} = 0.31437$

3.5.4 Вычислим относительную погрешность по формуле (5)

$\delta_{y/x} = (0.31437 / 30.50000)100\% = 1.03073\%$

ОТВЕТ: Уравнение линейной регрессии имеет вид: **Y = 40.80894 -0.40035 X** (6)

Погрешности уравнения: $\sigma_{y/x} = 0.31437$; $\delta_{y/x} = 1.03073$**%**

4. Коэффициент корреляции, его свойства, значимость

Выборочным коэффициентом корреляции принято называть отношение выборочного корреляционного момента к произведению выборочных средних квадратических отклонений этих величин:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x_в})(y_i - \overline{y_в})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x_в})}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y_в})}}$$

Коэффициент корреляции показывает тесноту и направление связи.

Свойства выборочного коэффициента корреляции:

1. значения коэффициента корреляции изменяются на множестве $r \in [-1;1]$;

2. чем больше абсолютное значение коэффициента корреляции, тем теснее связь между изученными признаками;

3. если коэффициент корреляции равен 0 (k=0), то между изученными признаками нет линейной корреляционной зависимости,

если |k|=1, то связь полная;

если 0,7<|k|<0,99, то связь сильная;

если 0,3<|k|<0,7, то связь средняя;

если |k|<0,3, то связь слабая.

В случае если r $\in$ [-1;0), то связь обратная;

если r $\in$ (0;1] – зависимость прямая.

Проверка гипотезы для коэффициента корреляции

Пусть r обозначает выборочный коэффициент корреляции, полученный по извлеченным из двумерного нормального распределения пар наблюдений $(x_1, y_1), \dots, (x_n, y_n)$.

Коэффициент корреляции ρ в популяции неизвестен, но может быть оценен по выборке с помощью выборочного коэффициента корреляции r:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \qquad (1)$$

где оценки среднего равны:

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

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

.

Проверим значимость коэффициента корреляции.

Нулевая гипотеза состоит в том, что коэффициент корреляции равен нулю, альтернативная - не равен нулю:

$H_0: \quad \rho = 0$

$H_1: \quad \rho \neq 0$

Очевидно, достаточно большое по *абсолютной* величине значение величины r будет стремиться опровергнуть нулевую гипотезу.

Возникает вопрос.

Насколько большое должно быть абсолютное значение величины r?

Для того чтобы проверить гипотезу, мы должны знать распределение величины r.

Собственное распределение величины r довольно сложное, поэтому мы применим преобразование:

$$t = \frac{r}{\sqrt{1 - r^2}}\sqrt{n-2} \sim t_{n-2} \qquad (2)$$

Итак, выборочное распределение этой статистики есть распределение Стьюдента с n-2 степенями свободы.

При заданном уровне значимости (α) определяем критическое значение $t_{кр}$.

Принимаем решение об отклонении или не отклонении нулевой гипотезы:

$|t| > t_{\text{кр}}$ - отклоняем H_0

$|t| < t_{\text{кр}}$ - не отклоняем H_0

3.7.3 Результаты и выводы: В результате проведенного занятия студенты:

- освоили основные понятия корреляционно-регрессионного анализа;
- усвоили формулы для вычисления коэффициента корреляции, корреляционного отношения, коэффициента детерминации; выработали навыки по интерпретации полученных результатов.
- освоили основные термины и формулы, необходимые для построения парной линейной регрессии;
- усвоили формулы для вычисления коэффициента корреляции, его интерпретацию;

- выработали навыки по проверки значимости выборочных коэффициентов.

3.8 Практическое занятие 13-14 (ПЗ-13-14)

Тема: Математическая обработка опытных данных

3.8.1 Задание для работы:

1. Особенности обработки ограниченного числа опытов. Оценки характеристик выборки.
2. Оценки неизвестных параметров закона распределения. Оценка вероятности по частоте.

3.8.2 Краткое описание проводимого занятия

1. Особенности обработки ограниченного числа опытов. Оценки характеристик выборки - доклады с презентацией

2. Оценки неизвестных параметров закона распределения. Оценка вероятности по частоте - доклады с презентацией

3.8.3 Результаты и выводы: В результате проведенного занятия студенты:
- освоили основные понятия теории математической обработки экспериментальных данных;
-выработали навыки использования алгоритмов проверки значимости параметров генеральной совокупности;
- усвоили основные принципы составления и обработки выборок.

3.9 Практическое занятие 15-16 (ПЗ-15-16)

Тема: Случайный процесс. Линейные преобразования случайных функций.

3.9.1 Задание для работы:

1. Огибающая и фаза нормального случайного процесса. Свойства и вероятностный смысл характеристик.
2. Определение характеристик случайной функции из опыта. Линейные преобразования случайных функций.
3. Методы определения характеристик преобразованных случайных функций по характеристикам исходных случайных функций.

3.9.2 Краткое описание проводимого занятия

1. Огибающая и фаза нормального случайного процесса. Свойства и вероятностный смысл характеристик.

Зная функцию распределения исходного случайного процесса X(t), можно обычными методами теории вероятности найти функцию распределения для всех вновь введенных процессов: $\tilde{X}(t)$, $\dot{X}(t)$, $A(t)$, $\Psi(t)$, $\Phi(t)$, $\Omega(t)$, $A_c(t)$, $A_к(t)$.

В общем случаи эта не простая задача, поэтому рассмотрим частный случай , когда X(t)-центрированный стационарный Гауссовский процесс.

Известно, что процесс, полученный в результате линейного преобразования Гауссовского процесса, также является Гауссовским, т.к. преобразования Гильберта линейно, можно утверждать, что $\tilde{X}(t)$ также Гауссовский процесс.

Процессы Ас(t) и Ак(t) представляют собой линейные комбинации процессов *X(t)* и $\tilde{X}(t)$, поэтому их распределение будет Гауссовским. Поскольку $\tilde{X}(t)$ отличается от X(t) только фиксированным сдвигом фаз всех гармонических составляющих, то

1. $M_{\tilde{X}} = M_X = 0$, т.к. математическое ожидание имеет смысл среднего значения процесса, который при таком преобразовании не меняется.

2. $G_{\tilde{X}}(f) = G_X(f)$, т.к. энергетический спектр не зависит от фазовых соотношений.

3. $B_{\tilde{X}}(\tau) = B_X(\tau)$, вытекает из второго пункта теоремы Винера – Хинчина.

4. $D_{\tilde{X}} = D_X$, т.к. *D=B(0).*

Процессы X(t) и $\tilde{X}(t)$ являются Гауссовскими и в совпадающие моменты времени между собой не коррелированны, следовательно, они не зависимые. Поэтому, совместную плотность вероятностей X(t) и $\tilde{X}(t)$ характеризующую аналитический сигнал $\dot{X}(t) = X(t) + j\tilde{X}(t)$ в одном сечении t можно записать так:

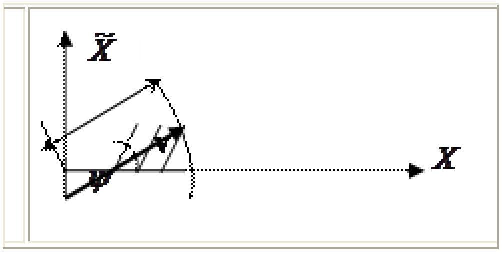
$$W(x,\tilde{x},t) = \frac{1}{\sqrt{2\Pi D_x}}\exp\left\{-\frac{X^2}{2D_x}\right\}\cdot\frac{1}{2\Pi D_x}\left\{-\frac{\tilde{X}^2}{2D_x}\right\} = \frac{1}{2\Pi D_x}\exp\left\{-\frac{X^2+\tilde{X}^2}{2D_x}\right\}$$

Дисперсия аналитического сигнала: $D_{\dot{X}} = M\left\{\left|\dot{X}(t)\right|^2\right\} = M\left\{X^2(t) + \tilde{X}^2(t)\right\} = D_X + D_{\tilde{X}} = 2D_X$

Процессы $A(t)$ и $\Psi(t)$ также стационарны, но поскольку они получены из X(t) и $\tilde{X}(t)$ с помощью нелинейных операций, они не являются Гауссовскими. Найдем совместную интегральную функцию распределения A и Ψ в некотором сечении t, т.е. функцию $F(a,\psi)$.

$$F(a,\psi) = P\{A(t) \le a, \Psi(t) \le \psi\}$$

Геометрическая интерпретация этой задачи представлена на рисунке:



По оси ординат отложены значения $\tilde{X}$, а по оси абсцисс X . искомая функция распределения является вероятностью того, что коней вектора представляющего аналитический сигнал $\dot{X}(t)$ находится внутри заштрихованной зоны. Для нахождения функции распределения $F(a,\psi)$, можно проинтегрировать $W(x,\tilde{x},t)$ по области V. Перейдем к полярной системе координат $\rho(\varphi)$ используя известные формулы:

$$x = \rho\cos\varphi$$

$$\tilde{x} = \rho\sin\varphi$$

$$dx \cdot d\tilde{x} = \rho \cdot d\rho \cdot d\varphi$$

При этом имеем:

$$F(a,\psi) = \iint\limits_V \frac{1}{\sqrt{2\Pi D_X}} \exp\left\{-\frac{1}{2D_X}\left[X^2 + \tilde{X}^2\right]\right\} dx d\tilde{x} = \int\limits_0^a \frac{1}{\sqrt{2\Pi D_X}} \exp\left\{-\frac{1}{2D_X}\rho^2\right\}\rho \cdot d\rho \int\limits_0^\psi d\psi =$$
$$= \left[1 - \exp\left(-\frac{a^2}{2D_X}\right)\right] \cdot \frac{\psi}{2\Pi}$$

Двумерная интегральная функция распределения представляет собой произведение двух одномерных функций:

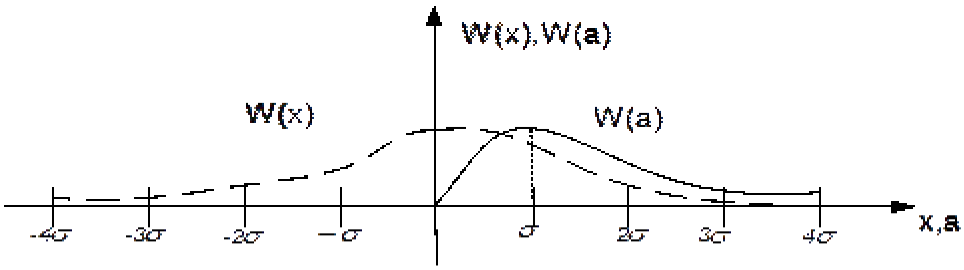
$$F(a) = 1 - \exp\left(\frac{-a^2}{2D_X}\right) \quad a \ge 0$$

$$F(\psi) = \frac{\psi}{2\Pi} \quad 0 < \psi \le 2\Pi$$

Это означает, что огибающая $A(t)$ и $\Psi(t)$ в одном сечении независимы. Плотность распределения огибающей найдем как производную интегральной функции распределения:

$$W(a) = \frac{dF(a)}{da} = \frac{a}{D_X} e^{-\frac{a^2}{2D_X}} \quad a \ge 0$$

Это известное распределение Релея, график которого показан на рисунке совместно с графиком Гауссовского распределения $W(X)$:

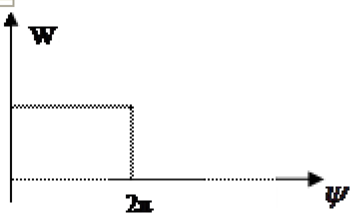


$\sigma = \sqrt{D}$

Наиболее вероятностные значения $A(t) = \sigma$. Математическое ожидание - $M[A(t)] = \sigma \dfrac{\Pi}{2}$. Дисперсия - $D[A(t)] = \sigma^2 \left(2 - \dfrac{\Pi}{2} \right)$. Для плотности распределения фазы имеем:

$$W(\psi) = \frac{dF(\psi)}{d\psi} = \frac{1}{2\Pi} \quad 0 < \psi \le 2\Pi$$

Следовательно, фаза стационарного централизованного Гауссовского процесса распределена равномерно на интервале *(0,2π)*.



Мгновенная начальная фаза $\Phi(t)$ рассмотревшего процесса также равномерно распределена на интервале протяженностью 2π. Действительно, можно отыскать совместную интегральную функцию распределения $F[A(t), \Phi(t)]$ исходя из совместной плотности распределения процесса $W[A_c(t), A_K(t)]$. Процессы Ac(t) и Ак(t) являются центрированными Гауссовскими процессами не зависимыми в совпадающие моменты времени, причем $D_{A_c} = D_{A_K} = D_X$, поэтому совместная плотность вероятности $W[A_c(t), A_K(t)]$ определяется выражением аналогичным $W(x, \tilde{x}, t)$ в котором символы X и $\tilde{X}$ заменены на Ас и Ак. Повторяя только что проведенные рассуждения и выкладки заменой X, $\dot{X}$, Ψ на Ас, Ак, φ найдем плотность распределения для A(t) и равномерную плотность распределения $\Phi(t)$.

2. Определение характеристик случайной функции из опыта. Линейные преобразования случайных функций- доклады с презентацией

3. Методы определения характеристик преобразованных случайных функций по характеристикам исходных случайных функций – доклады с презентацией

3.9.3 Результаты и выводы: В результате проведенного занятия студенты:
- освоить понятие нормального процесса, его фаз, характеристик, вероятностный смысл;
- усвоить методы определения характеристик случайной функции из опыта;
- освоить алгоритм определения характеристик преобразованных случайных функций по характеристикам исходных случайных функций.

3.10 Практическое занятие 17 (ПЗ-17)
Тема: Стационарные случайные процессы

3.10.1 Задание для работы:

1. Линейные преобразования стационарных случайных функций. Методы определения характеристик преобразованных стационарных случайных функций по характеристикам исходных стационарных случайных функций.

2. Применение теории стационарных случайных процессов к решению задач, связанных с анализом и синтезом динамических систем.

3.10.2 Краткое описание проводимого занятия

1. Линейные преобразования стационарных случайных функций. Методы определения характеристик преобразованных стационарных случайных функций по характеристикам исходных стационарных случайных функций - доклады с презентацией

2. Применение теории стационарных случайных процессов к решению задач, связанных с анализом и синтезом динамических систем - доклады с презентацией

3.10.3 Результаты и выводы: В результате проведенного занятия студенты:
- освоить определение линейных, нелинейных преобразований, оператора динамической системы;
- усвоить методы определения характеристик стационарных случайных функций, методы определения характеристик преобразованных стационарных случайных функций по характеристикам исходных стационарных случайных функций;
- выработать навыки применения алгоритмов анализа и синтеза динамических систем.

3.11 Практическое занятие 18-19 (ПЗ-19)
Тема: Преобразования случайных функций

3.11.1 Задание для работы:

1. Определение характеристик эргодической стационарной функции по одной реализации.
2. Представление случайной функции в виде суммы элементарных случайных функций. Каноническое разложение случайной функции. Линейные преобразования случайных функций, заданных каноническими разложениями.

3.11.2 Краткое описание проводимого занятия

1. Определение характеристик эргодической стационарной функции по одной реализации.

Рассмотрим стационарную случайную функцию $X(t)$, обладающую эргодическим свойством, и предположим, что в нашем распоряжении имеется всего одна реализация этой случайной функции, но зато на достаточно большом участке времени T. Для эргодической стационарной случайной функции одна реализация достаточно большой продолжительности практически эквивалентна (в смысле объема сведений о случайной функции) множеству реализаций той же общей продолжительности; характеристики случайной функции могут быть приближенно определены не как средние по множеству наблюдений, а как средние по времени t. В частности, при достаточно большом T математическое ожидание m_x может быть приближенно вычислено по формуле

$$m_x \approx \frac{1}{T}\int_0^T x(t)dt \quad . \qquad (1)$$

Аналогично может быть приближенно найдена корреляционная функция $k_x(\tau)$ при любом τ. Действительно, корреляционная функция, по определению, представляет собой не что иное, как математическое ожидание случайной функции $\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)$:

$$k_x(\tau) = M[\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)] \qquad (2)$$

Это математическое ожидание также, очевидно, может быть приближенно вычислено как среднее по времени.

Фиксируем некоторое значение τ и вычислим указанным способом корреляционную функцию $k_x(\tau)$. Для этого удобно предварительно «центрировать» данную реализацию $x(t)$, т. е. вычесть из нее математическое ожидание (1):

$$\overset{\circ}{x}(t) = x(t) - m_x . \qquad (3)$$

Вычислим при заданном τ математическое ожидание случайной функции $\overset{\circ}{X}(t)\,\overset{\circ}{X}(t+\tau)$ как среднее по времени. При этом, очевидно, нам придется учитывать не весь участок времени от 0 до T, а несколько меньший, так как второй сомножитель $\overset{\circ}{X}(t+\tau)$ известен нам не для всех t, а только для тех, для которых $t+\tau \le T$.

Вычисляя среднее по времени указанным выше способом, получим:

$$k_x(\tau) \approx \frac{1}{T-\tau}\int_0^{T-\tau} \overset{\circ}{x}(t)\,\overset{\circ}{x}(t+\tau)dt \quad . \qquad (4)$$

Вычислив интеграл (4) для ряда значений τ, можно приближенно воспроизвести по точкам весь ход корреляционной функции.

На практике обычно интегралы (1) и (4) заменяют конечными суммами. Покажем, как это делается. Разобьем интервал записи случайной функции на n равных частей длиной Δt и обозначим середины полученных участков $t_1, t_2, \dots, t_n$ (рис. 1).

Рис. 1.

Предоставим интеграл (1) как сумму интегралов по элементарным участкам Δt и на каждом из них вынесем функцию $x(t)$ из-под знака интеграла средним значением, соответствующим центру интервала $x(t_i)$. Получим приближенно:

$$m_x = \frac{1}{T}\frac{T}{n}\sum_{i=1}^{n} x(t_i),$$

или

$$m_x = \frac{1}{n}\sum_{i=1}^{n} x(t_i). \qquad (5)$$

Аналогично можно вычислить корреляционную функцию для значений τ, равных $0, \Delta t, 2\Delta t, \dots$. Придадим, например, величине τ значение

$$\tau = m\Delta t = \frac{mT}{n}$$

вычислим интеграл (4), деля интервал интегрирования

$$T - \tau = T - \frac{mT}{n} = \frac{n-m}{n}T$$

на $n-m$ равных участков длиной Δt и вынося на каждом из них функцию $\overset{\circ}{x}(t)\,\overset{\circ}{x}(t+\tau)$ за знак интеграла средним значением. Получим:

$$k_x\left(\frac{mT}{n}\right) = \frac{n}{(n-m)T}\frac{T}{n}\sum_{i=1}^{n-m} \overset{\circ}{x}(t_i)\,\overset{\circ}{x}(t_{i+m}),$$

или окончательно

$$k_x\left(\frac{mT}{n}\right) = \frac{1}{n-m}\sum_{i=1}^{n-m} \overset{\circ}{x}(t_i)\,\overset{\circ}{x}(t_{i+m}). \qquad (6)$$

Вычисление корреляционной функции по формуле (6) производят для $m = 0, 1, 2, \dots$ последовательно вплоть до таких значений m, при которых корреляционная функция становится практически равной нулю или начинает совершать небольшие нерегулярные колебания около нуля. Общий ход функции $k_x(\tau)$ воспроизводится по отдельным точкам (рис. 2).

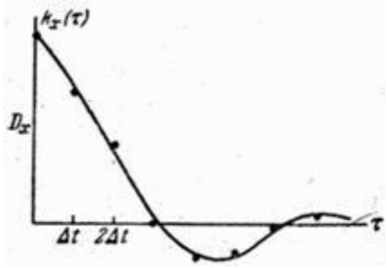


Рис. 2.

Для того чтобы математическое ожидание m_x и корреляционная функция $k_x(\tau)$ были определены с удовлетворительной точностью, нужно, чтобы число точек n было достаточно велико (порядка сотни, а в некоторых случаях даже нескольких сотен). Выбор длины элементарного участка Δt определяется характером изменения случайной функции. Если случайная функция изменяется сравнительно плавно, участок Δt можно выбирать большим, чем когда она совершает резкие и частые колебания. Чем более высокочастотный состав имеют колебания, образующие случайную функцию, тем чаще нужно располагать опорные точки при обработке. Ориентировочно можно рекомендовать выбирать элементарный участок Δt так, чтобы на полный период самой высокочастотной гармоники в составе случайной функции приходилось порядка 5-10 опорных точек.

Часто выбор опорных точек вообще не зависит от обрабатывающего, а диктуется темпом работы записывающей аппаратуры. В этом случае следует вести обработку непосредственно полученного из опыта материала, не пытаясь вставить между наблюденными значениями промежуточные, так как эхо не может повысить точности результата, а излишне осложнит обработку.

Пример. В условиях горизонтального полета самолета произведена запись вертикальной перегрузки, действующей на самолет. Перегрузка регистрировалась на участке времени 200 сек с интервалом 2 сек. Результаты приведены в таблице 1.

Таблица 1

t (сек)	Перегрузка $N(t)$	t (сек)	Перегрузка $N(t)$	t (сек)	Перегрузка $N(t)$	t (сек)	Перегрузка $N(t)$
0	1,0	50	1,0	100	1,2	150	0,8
2	1,3	52	1,1	102	1,4	152	0,6
4	1,1	54	1,5	104	0,8	154	0,9
6	0,7	56	1,0	106	0,9	156	1,2
8	0,7	58	0,8	108	1,0	158	1,3
10	1,1	60	1,1	110	0,8	160	0,9
12	1,3	62	1,1	112	0,8	162	1,3
14	0,8	64	1,2	114	1,4	164	1,5
16	0,8	66	1,0	116	1,6	166	1,2
18	0,4	68	0,8	118	1,7	168	1,4
20	0,3	70	0,8	120	1,3	170	1,4
22	0,3	72	1,2	122	1,6	172	0,8
24	0,6	74	0,7	124	0,8	174	0,8
26	0,3	76	0,7	126	1,2	176	1,3
28	0,5	78	1,1	128	0,6	178	1,0
30	0,5	80	1,2	130	1,0	180	0,7
32	0,7	82	1,0	132	0,3	182	1,1
34	0,8	84	0,6	134	0,8	184	0,9
36	0,6	86	0,9	136	0,7	186	0,9

38	1,0	88	0,8	138	0,9	188	1,1
40	0,5	90	0,8	140	1,3	190	1,2
42	1,0	92	0,9	142	1,5	192	1,3
44	0,9	94	0,9	144	1,1	194	1,3
46	1,4	96	0,6	146	0,7	196	1,6
48	1,4	98	0,4	148	1,0	198	1,5

Считая процесс изменения перегрузки стационарным, определить приближен но математическое ожидание перегрузки m_N, дисперсию D_N и нормированную корреляционную функцию $\rho_N(\tau)$. Аппроксимировать $\rho_N(\tau)$ какой-либо аналитической функцией, найти и построить спектральную плотность случайного процесса.

Решение. По формуле (5) имеем: $m_N = \dfrac{\sum\limits_{i=1}^{100} N(t_i)}{100} \approx 0,98$.

Центрируем случайную функцию (табл. 2).

Таблица 2

t (сек)	$\overset{\circ}{N}(t)$	t (сек)	$\overset{\circ}{N}(t)$	t (сек)	$\overset{\circ}{N}(t)$	t (сек)	$\overset{\circ}{N}(t)$
0	0,02	50	0,02	100	0,22	150	-0,18
2	0,32	52	0,12	102	0,42	152	-0,38
4	0,12	54	0,52	104	-0,18	154	-0,08
6	-0,28	56	0,02	106	-0,08	156	0,22
8	-0,28	58	-0,18	108	0,02	158	0,32
10	0,12	60	0,12	110	-0,18	160	-0,08
12	0,32	62	0,12	112	-0,18	162	0,32
14	-0,18	64	0,22	114	0,42	164	0,52
16	-0,18	66	0,02	116	0,62	166	0,22
18	-0,58	68	-0,18	118	0,72	168	0,42
20	-0,68	70	-0,18	120	0,32	170	0,42
22	-0,68	72	0,22	122	0,62	172	-0,18
24	-0,38	74	-0,28	124	-0,18	174	-0,18
26	-0,68	76	-0,28	126	0,22	176	0,32
28	-0,48	78	0,12	128	-0,38	178	0,02
30	-0,48	80	0,52	130	0,02	180	-0,28
32	-0,28	82	0,02	132	-0,38	182	0,12
34	-0,18	84	-0,38	134	-0,18	184	-0,08
36	-0,38	86	-0,08	136	-0,28	186	-0,08
38	0,02	88	-0,18	138	-0,08	188	0,12
40	-0,48	90	-0,18	140	0,32	190	0,22
42	0,02	92	-0,08	142	0,52	192	0,32
44	-0,08	94	-0,08	144	0,12	194	0,32
46	0,42	96	-0,38	146	-0,28	196	0,62
48	0,42	98	-0,58	148	0,02	198	0,52

Возводя в квадрат все значения $\overset{\circ}{N}(t)$ и деля сумму на $n=100$ получим приближенно дисперсию случайной функции $N(t)$: $D_N = \dfrac{\sum_{i=1}^{100}\left[\overset{\circ}{N}(t_i)\right]^2}{100} \approx 0{,}1045$

и среднее квадратическое отклонение: $\sigma_N \approx 0{,}323$.

Перемножая значения $\overset{\circ}{N}(t)$, разделенные интервалом $\tau = 2,4,6,\ldots$, и деля сумму произведений соответственно на $n-1=99$; $n-2=98$; $n-3=97$; …, получим значения корреляционной функции $k_N(\tau)$. Нормируя корреляционную функцию делением на $D_N = 0{,}1045$, получим таблицу значений функции $\rho_N(\tau)$ (табл. 3).

Таблица 3

τ	$\rho_N(\tau)$	$\rho_N^*(\tau) = e^{-a\lvert\tau\rvert}$
0	1,000	1,000
2	0,505	0,598
4	0,276	0,358
6	0,277	0,214
8	0,231	0,128
10	-0,015	0,077
12	0,014	0,046
14	0,071	0,027

График функции $\rho_N(\tau)$ представлен на рис. 3 в виде точек, соединенных пунктиром.

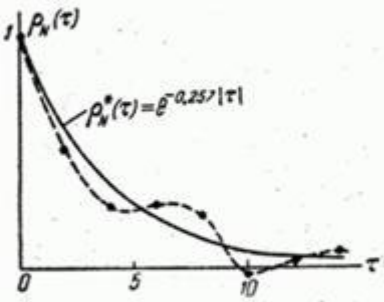


Рис. 3.

Не вполне гладкий ход корреляционной функции может быть объяснен недостаточным объемом экспериментальных данных (недостаточной продолжительностью опыта), в связи с чем случайные неровности в ходе функции не успевают сгладиться. Вычисление $\rho_N(\tau)$ продолжено до таких значений при которых фактически корреляционная связь пропадает.

Для того чтобы сгладить явно незакономерные колебания экспериментально найденной функции $\rho_N(\tau)$, заменим ее приближенно функцией вида: $\rho_N^*(\tau) = e^{-a|\tau|}$,

где параметр a подберем методом наименьших квадратов.

Применяя этот метод, находим $a = 0,257$. Вычисляя значения функции $\rho_N^*(\tau)$ при $\tau = 0, 2, 4, \ldots$, построим график сглаживающей кривой. На рис. 3 он проведен сплошной линией. В последнем столбце таблицы 3 приведены значения функции $\rho_N^*(\tau)$.

Пользуясь приближенным выражением корреляционной функции (6), получим нормированную спектральную плотность случайного процесса в виде:

$$S_x^*(\omega) = \frac{a}{\pi(a^2 + \omega^2)} = \frac{0,257}{\pi(0,257^2 + \omega^2)}.$$

График нормированной спектральной плотности представлен на рис. 4.

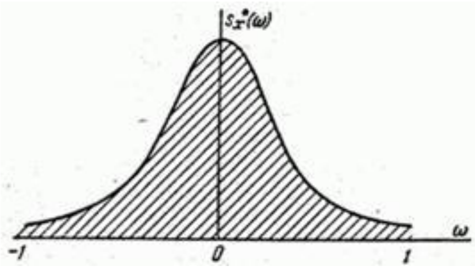


Рис. 4.

2. Представление случайной функции в виде суммы элементарных случайных функций. Каноническое разложение случайной функции. Линейные преобразования случайных функций, заданных каноническими разложениями.

Пусть случайный процесс **X (t)** представлен в виде: $$X(t) = m_x(t) + \sum_{i=1}^{m} V_i \phi_i(t) \quad (1)$$

где ***mx(t)*** – математическое ожидание случайного процесса ***X (t)***; ***φi (t)*** –неслучайные функции времени; ***Vi*** – случайные величины, причем:

M [Vi] = 0; M [Vi*V j] = 0, если $i \neq j$ M [Vi2] = Di.

Здесь **Di** – дисперсия случайной величины **Vi**, **m** – количество неслучайных функций в каноническом разложении.

Соотношение (1) называется каноническим разложением случайного процесса **X (t)**.

Соотношению (1) соответствует корреляционная функция вида:

$$Kx(t_1, t_2) = \sum_{i=1}^{m} \phi_i(t_1)\phi_i(t_2).D_i \quad (2)$$

Соотношение (2) называется каноническим разложением корреляционной функции **K x (t_1, t_2).** Из (2) определим дисперсию **Dx (t)** случайного процесса **X (t)**. Имеем:

$$Dx(t) = Kx(t,t) = \sum_{i=1}^{m} [\phi_i(t)]^2 .D_i \quad (3)$$

Задача 1. Пусть случайная функция X (t) задана каноническим разложением:

X (t) = 3t + X_1 · cos ωt + X_2 · sin ωt + X_3 · cos 2ωt + X_4 · sin 2ωt.

Случайные величины X 1 , X 2 , X 3 , X 4 имеют следующие математические ожидания и дисперсии: m x_1= m x_2 = m x_3 = m x_4 = 0; D x_1 = D x_2 = 1; D x_3 = D x_4 = 3.

Определить m x (t), K x (t_1, t_2), Dx (t).

Решение Прежде всего найдем $m_x(t)$, получим:
$M_x(t) = 3t$.
Определим $K_x(t_1, t_2)$.
Будем иметь:
$K_x(t_1, t_2) = \cos \omega t_1 \cdot \cos \omega t_2 + \sin \omega t_1 \cdot \sin \omega t_2 ++ 3 \cos 2\omega t_1 \cdot \cos 2\omega t_2 + 3 \sin 2\omega t_1 \sin 2\omega t_2 = \cos \omega(t_1 - t_2) + 3 \cos 2\omega(t_1 - t_2)$.
Определим $D_x(t)$, получим: $D_x(t) = K_x(t, t) = 4$.

Задача 2. Пусть случайная функция $X(t)$ задана каноническим разложением вида:
$X(t) = 2t + X_1 \cdot \sin t + X_2 \cdot \cos t$.
Случайные величины X_1, X_2 имеют следующие математические ожидания
и дисперсии: $m_{x_1} = m_{x_2} = 0$; $D_{x_1} = D_{x_2} = 3$.
Найти каноническое разложение случайной функции $Y(t)$ вида:
$Y(t) = t \cdot X(t) - t_2$.
Определить $m_y(t)$, $K_y(t_1, t_2)$, $D_y(t)$.

Решение. Найдем каноническое разложение $Y(t)$. Будем иметь:
$Y(t) = t_2 + X_1 t \cdot \sin t + X_2 t \cdot \cos t$.
Определим $m_y(t)$. Получим: $m_y(t) = t_2$. Найдем $K_y(t_1, t_2)$, получим:

$$K_y(t_1, t_2) = 3t_1 t_2 \sin t_1 \cdot \sin t_2 + 3t_1 t_2 \cos t_1 \cdot \cos t_2 = = 3t_1 t_2 \cos(t_1 - t_2).$$

Определим $D_y(t)$, будем иметь: $D_y(t) = K_y(t, t) = 3t_2$.

3.11.3 Результаты и выводы: В результате проведенного занятия студенты:

- освоить понятие эргодической стационарной функции, ее характеристик;

- усвоить методы определения характеристик эргодической стационарной функции по одной реализации;
- освоить алгоритм представления случайной функции в виде суммы элементарных случайных функций; задание случайной функции каноническим разложением;
- усвоить методы исследования линейных преобразований случайных функций, заданных каноническими разложениями.