Конспект лекций по дисциплине Теория вероятностей и математическая статистика

Кондраев Дмитрий

Содержание

1. Случайные события

1. Основные понятия
2. Теорема сложения вероятностей
3. Теорема умножения вероятностей
4. Следствия теорем сложения и умножения
5. Повторение испытаний

1.1. Основные понятия теории вероятностей

- События

- Достоверные события — обязательно произойдут если будет осуществлена определенная совокупность условий.
  * ⊗ Нормальные условия (**НУ**): температура 20°, давление атмосферное.
  * A: «вода жидкая» — достоверное
- Невозможные — заведомо не произойдут при совокупности определенных условий.
  * ⊗ НУ. B: «вода в твердом состоянии» — невозможное
- Случайные — может произойти или не произойти при совокупности определенных условий.
  * ⊗ При бросании монеты. C: «Выпал герб»

Теория вероятности не ставит задачу предсказать, произойдет случайное событие запятая или нет. Однако, рассматривая случайные события, многократно повторяющиеся при одних и тех же условиях, например, много раз бросая монету можно с небольшой погрешностью предсказать количество выпадений герба или решки.

Предмет теории вероятностей — изучение вероятностных закономерностей массовых, однородных, случайных событий.

1.1.1. Испытание и событие

Опустим фразу «осуществляется совокупность условий» заменим ее на «произведено испытание».

Событие — результат испытания.

⊗ Стрелок стреляет по мишени. Выстрел в тире — испытание. Попадание в определенную область мишени — событие.

⊗ В корзине находятся красные и белые шары. Их достают. Извлечение шара — испытание, появление красного или белого шара — событие.

1.1.2. Виды случайных событий

Попарно несовместными называют события появление одного из которых исключает появление другого события в одном и том же испытании.

⊗ Брошена монета. События «появился герб» и «появилась решка» несовместные.

Несколько событий образуют **полную группу**, если в результате испытаний появиться хотя бы одно из них. Если события, образующие полную группу, попарно несовместны, то в результате испытания (одного испытания) появится *только одно* $\exists!$ из этих событий.

⊗ Стрелок произвел выстрел по цели.

Обязательно произойдет либо попадание, либо промах. Эти два несовместных события образуют полную группу.

⊗ 2 билета Русского лото. Может произойти 4 события:

- выиграл — не выиграл,
- не выиграл — выиграл,
- выиграл — выиграл,
- не выиграл — не выиграл.

Образуют полную группу попарно несовместных событий.

Если события, образующие полную группу попарно несовместны, то в результате испытания появится хотя бы одно событие из неё.

События называются **равновозможными**, если есть основания считать, что ни одно из них не является более возможным, чем другое.

⊗ При бросании игральной кости появление любого количество очков равновозможно.

1.1.3. Классическое определение вероятности

Дать им количественную оценку возможности наступления того или иного случайного события.

Рассмотрим пример.

Пусть в урне содержится 6 одинаковых, тщательно перемешанных шаров, причем 2 из них — красные, 3 — синие и 1 — белый. Очевидно, возможность вынуть наудачу из урны цветной шар больше, чем возможность извлечь белый шар. Можно охарактеризовать такую возможность числом, которое будем называть **вероятностью события.** Таким образом, **вероятность** — число, характеризующее степень возможности появления события.

Поставим перед собой задачу дать оценить вероятность $P(A)$ того, что взятый наудачу шар цветной. Появление цветного шара будем рассматривать в качестве события A:

$$A : \text{“Появление цветного шара”}$$

Испытание — извлечение шара из урны.

Его результат — **элементарный исход** ω_i. В нашем примере возможны следующие 6 элементарных исходов:

- ω_1 — появился белый шар,
- ω_2, ω_3 — красный шар,
- $\omega_4, \omega_5, \omega_6$ — синий шар.

Эти исходы образуют *полную группу попарно несовместных событий* (обязательно появится только один шар) и они *равновозможны* (шар вынимают наудачу, шары одинаковы и тщательно перемешаны).

Благоприятные исходы — те элементарные исходы, в которых интересующее нас событие наступает.

В нашем примере благоприятствуют событию A (появлению цветного шара) следующие 5 исходов: $\omega_2, \omega_3, \omega_4, \omega_5, \omega_6$.

Таким образом, событие A наблюдается, если в испытании наступает один, безразлично какой, из элементарных исходов, благоприятствующих A.

Вероятность события $P(A)$ — отношение числа благоприятных для события исходов m к обычному числу всех возможных несовместных элементарных исходов, образующих полную группу n.

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

1.1.4. Свойства вероятностей

1°. Вероятность достоверного события равна единице.

$m = n$, значит, $P(A) = \frac{m}{n} = \frac{n}{n} = 1$.

2°. Вероятность невозможного события равна нулю.

$m = 0$, значит, $P(A) = \frac{m}{n} = \frac{0}{n} = 0$.

3°. Вероятность случайного события есть положительное число, заключенное между нулем и единицей.

$0 < m < n$ значит, $0 < \frac{m}{n} < 1$, следовательно, $0 < P(A) < 1$.

Значит, $0 \leqslant P(A) \leqslant 1$.

1.1.5. Решение задач

⊗ Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 100. Найдите вероятность того что номер 1 на удачу извлеченного жетона не содержит цифры 5.

$$m = 9 \cdot 9 = 81$$

$$P = 0,81$$

⊗ Брошены 2 игральных кости. Найти вероятность того что сумма выпавших очков равна 7.

Пусть выпала $a \in \{1..6\}$ очков на первом кубике. Тогда исход будет благоприятным тогда и только тогда, когда на втором кубике выпадет $7 - a$ очков. Значит благоприятных исходов $m = 6$, а всего исходов $n = 36$. $P = \frac{1}{6}$

⊗ Брошены две игральные кости. Какова вероятность, что сумма очков равна 8, а разность 4?

С суммой 8 таких пар 5: $(2, 6)$, $(3, 5)$, $(4, 4)$ и наоборот. Из них под второе условие подходит только 2: $(2, 6)$ и $(6, 2)$.

$$m = 2$$

$$n = 36$$

$$P = \frac{1}{18}$$

⊗ Куб, все грани которого окрашены, распилен на 1000 кубиков одинакового размера, а тщательно перемешав. Какова вероятность, что наудачу извлеченный кубик будет иметь окрашенных граней:

1. 1,
2. 2,
3. 3?

Не окрашено $8^3 = 512$ кубиков.

Окрашена одна грань у $6 \cdot 8^2$ кубиков.

Окрашено 2 грани у $12 \cdot 8^1$ кубиков.

Окрашено 3 грани у $8 \cdot 8^0$ кубиков.

Ответ: $P(A) = 0,384$, $P(B) = 0,096$, $P(A) = 0,008$.

1.1.6. Формулы комбинаторики

Комбинаторика — изучает количество комбинаций, подчинённых определенным условиям, которые можно составить из элементов заданного конечного множества.

Перестановки — выборки элементов, состоящие из одних и тех же различных элементов и отличающиеся только порядком их расположения.

$$P_n = n!$$

Число возможных перестановок множества из n элементов.

⊗ Даны числа 1, 2, 3. Сколько трехзначных чисел можно составить из них, если каждая цифра входит в число изображение числа 1 раз?

$$P_3 = 3! = 6$$

Вот они: 121, 132, 213, 231, 312, 321.

Размещения *из n различных элементов по m(мест)* — такие выборки, которые имея по m элементов, выбранных из числа данных n элементов, отличаются одна от другой либо составом элементов, либо порядком их расположения.

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

⊗ Сколько можно составить сигналов из 6 флажков различного цвета взятых по 2 штуки?

$$A_6^2 = 30.$$

Сочетания—неупорядоченные выборки из n различных элементов по m.

$$C_n^m = \frac{A_n^m}{P_m} = \frac{n!}{m!\,(n-m)!}$$

$C_n^1 = n, C_n^n = 1$

⊗ Сколькими способами можно 2 детали из ящика содержащего 10 деталей?

$$C_{10}^2 = 45.$$

При решении задач комбинаторики используют следующие правила:

- **Правило суммы**. Если объект A можно выбрать из совокупности объектов m способами, объект $B - n$ способами, то выбрать либо A, либо B можно $m + n$ способами.
- **Правило произведения**. В аналогичной ситуации *пару* объектов (A, B) можно выбрать $m \times n$ способами.

1.1.7. Решение задач

⊗ Имеется 5 одинаковых кубиков. На всех гранях каждого кубика написана одна из букв: П, Р, О, С, Т. Найти вероятность того, что на вынутых по одному и расположенных в одну линию кубиках можно прочитать слово СПОРТ.

$$P(A) = \frac{1}{P_5} = \frac{1}{120}.$$

⊗ На каждой из шести одинаковых карточек написана одна буква А, Т, М, Р, С, О. Карточки перемешаны. Найти вероятность того, что на 4 вынутых по одной и расположенных в одну линию карточек можно прочесть слово ТРОС.

$$P(A) = \frac{1}{A_6^4} = \frac{1}{360}.$$

⊗ В пачке 20 карт, помеченных номерами от 101 до 120. Наудачу извлекли две карты. Найти вероятность того, что извлеченные карты — 105 и 115.

$$P(A) = \frac{1}{C_{20}^2} = \frac{1}{190}.$$

⊗ В ящике 15 деталей. 10 из них окрашены. Наудачу извлекли 3 детали. Найти вероятность того, что все извлеченные детали окажутся окрашенными.

$$P(A) = \frac{C_{10}^3}{C_{15}^5} = \frac{24}{91}.$$

⊗ В ящике 100 деталей, 10 из них бракованы. Наудачу извлекли 4 детали. Найти вероятность, того, что среди извлеченных деталей:

- нет браков,
- нет годных.

$$P(A) = \frac{C_{90}^4}{C_{100}^4} = \frac{15486}{23765} \approx 0,65$$
$$P(B) = \frac{C_{10}^4}{C_{100}^4} = \frac{2}{37345} \approx 0,00005$$

⊗ В коллективе 6 мужчин и 4 женщины. Наудачу отобраны 7 человек. Найти вероятность того, что три из них — женщины.

$$P(A) = \frac{C_4^3 \cdot C_6^4}{C_{10}^7} = 0,5$$

⊗ На складе имеются 15 объектов, 10 из них изготовлены заводом номер 1. Найти вероятность того, что среди 5 взятых наудачу объектов, три из них окажутся из Ну изготовлены заводом 1.

$$P(A) = \frac{C_{10}^3 \cdot C_{15}^2}{C_{15}^5} = 0,4$$

⊗ В коробке 5 деталей, 3 окрашены. Наудачу извлечены 2. Найти вероятность того, что среди 2 деталей окажется:

- 1 окрашенная деталь,
- 2 окрашенные детали.

$$P(A) = \frac{C_2^1 \cdot C_3^1}{C_5^2} = 0,6$$
$$P(A) = \frac{C_3^2}{C_5^2} = 0,3$$

1.2. Теорема сложения вероятностей

1.2.1. Сложение вероятностей несовместных событий

Теорема. Вероятность появления одного из двух несовместных событий, безразлично какого, равна сумме вероятностей этих событий.

$$P(A+B) = P(A) + P(B)$$

Доказательство. Введём обозначения:

n — общее число возможных элементарных исходов испытания,

m_1 — число благоприятных исходов события A,

m_2 — число благоприятных исходов события B.

Тогда по правилу суммы из комбинаторики, число благоприятных исходов для наступления события A или B равняется $m_1 + m_2$. Тогда

$$P(A+B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B).$$

⊗ В корзине 30 шаров: 10 красных, два синих, 18 белых. Найти вероятность появления цветного шара.

$$P(A+B) = P(A) + P(B) = \frac{10}{30} + \frac{2}{30} = 0,6$$

1.2.2. Полная группа событий

Теорема. Сумма вероятностей событий $A_1 ... A_n$, образующих полную группу равна единице.

$$\sum_{i=1}^{n} P\left(A_i\right) = 1$$

Доказательство. Так как появление одного из событий полной группы достоверно, а вероятность достоверного события равна 1, то

$$P\left(\sum_{i=1}^{n} A_i\right) = 1$$

Все события полной группы попарно несовместны. Значит можно применить теорему сложения:

$$\sum_{i=1}^{n} P\left(A_i\right) = P\left(\sum_{i=1}^{n} A_i\right) = 1$$

⊗ Фирма получает пакеты с документами из своих филиалов, расположенных в городах A, B и C. Вероятность получения пакетов из A равна 0,7. вероятность получения пакетов из B равна 0,2, Какова вероятность получения пакетов из C.

События получение пакетов образуют полную группу, значит

$$P(A) + P(B) + P(C) = 1$$

$$0,7 + 0,2 + P(C) = 1$$

$$P(C) = 0,1$$

1.2.3. Противоположные события

Противоположные события — пара единственно возможных событий, образующих полную группу.

$\overline{A}$ — событие, противоположное A.

⊗ Попадание и промах при выстреле по цели противоположные события.

⊗ Извлечение из ящика с деталями стандартной или нестандартной детали события противоположные.

Теорема. Сумма вероятностей противоположных событий равна единице.

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

$$p + q = 1$$

Доказательство. Противоположные события образуют полную группу, а сумма вероятностей событий, образующих полную группу равна единице.

⊗ Вероятность p того, что день дождливый равна 0,7. Какова вероятность, того, что день ясный?

$$q = 1 - p = 0,3$$

⊗ В ящике 10 деталей 8 стандартных, 2 нестандартных. Какова вероятность того, что среди наудачу извлеченных 2 деталей, есть хотя бы одна стандартная.

$$P\left(A\right)=1-P\left(\overline{A}\right)=1-\frac{C_{2}^{2}}{C_{10}^{2}}=\frac{44}{45}$$

$$P(A)=P(A_1)+P(A_2)=\frac{C_{8}^{2}}{C_{10}^{2}}+\frac{C_{1}^{1}C_{2}^{1}}{C_{10}^{2}}=\frac{44}{45}$$

Задачи со словами «хотя бы» легче решаются, если найти вероятность противоположного события. Например, пусть

$$A:\text{“отобрана хотя бы одна стандартная деталь”},$$

тогда

$$\overline{A}:\text{“среди отобранных нет стандартных деталей”},$$

$\otimes$ На стеллаже в случайном порядке расставлены 15 учебников, 5 из них в переплете. наудачу берут 3 из них. Найти вероятность того, что хотя бы один из них в переплете.

$$P\left(A\right)=1-P\left(\overline{A}\right)=1-\frac{C_{10}^{3}}{C_{15}^{3}}=\frac{67}{91}$$

1.3. Теорема умножения вероятностей

1.3.1. Произведение событий. Условная вероятность

Введём понятие условной вероятности. До сих пор мы говорили о безусловной вероятности.

Безусловная вероятность — вероятность события, которое может произойти, или не произойти при совокупности определенных условий. Если кроме этой совокупности условий накладываются дополнительные условия, то **вероятность** события называют **условной**.

$P_A(B)$ — вероятность события B, при условии что событие A уже произошло.

1.3.2. Теорема умножения вероятностей

Теорема. Вероятность совместного выполнения двух событий равна произведению вероятности одного из них на условную вероятность другого, вычисленное в предположении, что первое уже наступило.

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

$\otimes$ У сборщика 3 конусных и 7 овальных объектов. Какова вероятность того, что первый случайно выбранный объект конусный, а второй — овальный, и наоборот.

$$P(AB)=P(A)\cdot P_A(B)=\frac{3}{10}\cdot\frac{7}{9}=\frac{7}{30}$$

$$P(BA)=P(B)\cdot P_B(A)=\frac{7}{10}\cdot\frac{3}{9}=\frac{7}{30}.$$

1.3.3. Независимые события. Теорема умножения для независимых событий

Событие B называют **независимым** от события A, если появление A не изменяет вероятность события B, то есть $P_A(B)=P(B)$, тогда

$$P(AB)=P(A)P(B)$$

(теорема умножения вероятностей независимых событий)

$\otimes$ Имеется 6 учебников 3 из них в переплете, 3 нет. Наудачу взяты два учебника. Найти вероятность того, что оба они в переплёте.

$$P(AB) = P(A)P(B) = \frac{3}{6} \cdot \frac{2}{5} = \frac{1}{5}$$

1.3.4. Вероятность появления хотя бы одного события

Вероятность появления хотя бы одного из независимых событий $A_1, ...A_n$ имеющих вероятности $p_1...p_n$ равна разности между единицей и произведения вероятностей противоположных им событий,

$P(A) = 1 - q_1 q_2 ... q_n$

В частном случае, если события имеют одинаковые вероятности ($p_1 = ... = p_n = p$), то вероятность наступления хотя бы одного —

$$P(A) = 1 - q^n$$

1.3.5. Решение задач

$\otimes$ Вероятность попадания в цель при стрельбе из трех орудий равны

- $p_1 = 0,8$
- $p_2 = 0,7$
- $p_3 = 0,9$

Найти вероятность хотя бы одного попадания при одном залпе из каждого орудия.

$$P(A) = 1 - q_1 q_2 q_3 = 0,994$$

$\otimes$ Два стрелка стреляли по мишени. Вероятность попадания одного из них $p_1 = 0,7$, а другого — $p_2 = 0,8$. Найти вероятность того, что при одном залпе в мишень попадет только один из стрелков.

$$P(A) = p_1 q_2 + p_2 q_1 = 0,38$$

1.4. Следствия теорем сложения и умножения

1.4.1. Формула полной вероятности

Пусть событие A может наступить при условии появления одного из несовместных событий $B_1, ..., B_n$, образующих полную группу. Пусть также известны вероятности этих событий и условные вероятности события A.

Теорема. Вероятности события A, которое может наступить при условии появления одного из несовместных событий $B_1, ..., B_n$, образующих полную группу равна сумме произведений вероятностей каждого из этих событий на условную вероятность события A.

$$P(A) = \sum_{i=1}^{n} P(B_i A) = \sum_{i=1}^{n} P(B_i) P_{B_i}(A)$$

(формула полной вероятности)

Доказательство. По условию событие A может наступить, если наступит одно из несовместных событий $B_1, ..., B_n$, иными словами, появление A означает появление одного, безразлично какого, из несовместных событий $B_1 A, ..., B_n A$. Пользуясь теоремой сложения

$$P(A) = P(B_1 A) + ... + P(B_n A),$$

Каждое слагаемое здесь можно выписать по теореме умножения вероятностей

$$P(B_iA) = P(B_i)P_{B_i}(A), i \in \overline{1,n}$$

Подставив эти равенства в предыдущую выражение, получим формулу полной вероятности.

$$P(A) = \sum_{i=1}^{n} P(B_i)P_{B_i}(A)$$

1.4.2. Решение задач

⊗ Имеется два набора деталей. Вероятность того, что деталь первого набора стандартна $P_{B_1}(A) = 0,8$. Вероятность того, что деталь второго набора стандартна $P_{B_2}(A) = 0,9$. Найти вероятность того, что взятая наудачу деталь из любого набора стандартна.

Примем обозначения:

A — Извлечена стандартная деталь,

B_1 — Деталь из первого набора,

B_2 — Деталь из второго набора.

$$P(B_1) = P(B_2) = \frac{1}{2}$$

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) = 0,85$$

⊗ В группе спортсменов 20 лыжников, 6 велосипедистов и 4 бегуна. Вероятности выполнить квалификационную норму равны соответственно для лыжника 0,9, для велосипедиста 0,8, для бегуна — 0,75. Найти вероятность того, что спортсмен, выбранный наудачу, выполнит норму.

Пусть:

A — спортсмен выполнил норму,

B_1 — выступал лыжник,

B_2 — выступал велосипедист,

B_3 — выступал бегун.

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) + P(B_3)P_{B_3}(A) = 0,86$$

⊗ В первом ящике из 20 деталей 15 стандартных, во втором ящике из 30 деталей 24 стандартных, в третьем ящике из 10 деталей 6 стандартных. Найти вероятность того, что наудачу извлеченная деталь из любого ящика, стандартна.

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) + P(B_3)P_{B_3}(A) = 0,75$$

⊗ В ателье имеются четыре кинескопа, вероятности того, что объект гарантированно выдержит срок службы, соответственно равны $P_{B_1}(A) = 0,8$, $P_{B_2}(A) = 0,9$, $P_{B_3}(A) = 0,95$, $P_{B_4}(A) = 0,85$. Найти вероятность того, что взятый наудачу кинескоп выдержит гарантийный срок службы.

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) + P(B_3)P_{B_3}(A) + P(B_4)P_{B_4}(A) = 0,875$$

⊗ Студент знает не все экзаменационные билеты. В каком случае вероятность вытащить неизвестный билет будет наименьшей: когда он берёт билет первым или последним?

Вероятности одинаковы. *Почему*?

1.4.3. Формула Байеса. Вероятность гипотез

Пусть событие A может наступить при условии появления одного из несовместных событий $B_1, ..., B_n$, образующих полную группу. Так как заранее неизвестно, какое из этих событий наступит, их называют **гипотезами**. По формуле полной вероятности:

$$P(A) = \sum_{i=1}^{n} P(B_i) P_{B_i}(A)$$

Будем искать условные вероятности $P_A(B_i)$ ($i \in \overline{1, n}$), то есть определим, как изменились в вероятности гипотез, если произведены испытания, в результате которого появились события.

$$P_A(B_i) - ?$$

$$P(AB_i) = P(A)P_A(B_i) = P(B_i)P_{B_i}(A) \Rightarrow P_A(B_i) = \frac{P(B_i)P_{B_i}(A)}{P(A)} = \frac{P(B_i)P_{B_i}(A)}{\sum_{i=1}^{n} P(B_i)P_{B_i}(A)}$$

$$P_A(B_i) = \frac{P(B_i)P_{B_i}(A)}{\sum_{i=1}^{n} P(B_i)P_{B_i}(A)}$$

Формулы Байеса позволяют переоценить вероятности гипотез после того, как становятся известны итог испытания, в результате которого появляется событие A.

$\otimes$ Детали попадают для проверки двум контролерам. Вероятность попадания к первому $P(B_1) = 0,6$, ко второму $P(B_2) = 0,4$. Вероятность того, что деталь признана в стандартной у первого контроллера $P_{B_1}(A) = 0,9$, у второго $P_{B_2}(A) = 0,99$. Известно, что при проверке деталь признана стандартной. Найти вероятность $P_A(B_1)$ того, что ее проверил первый контролер.

$$P_A(B_1) = \frac{P(B_1)P_{B_1}(A)}{P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A)} = \frac{60}{104} \approx 0,58$$

1.5. Повторение испытаний

1.5.1. Формула Бернулли

Пусть проведено n независимых испытаний, в каждом из которых событие A может появиться или не появиться. Пусть в каждом испытании вероятность одинакова и равна $P(A) = p$, а вероятность ненаступления $P\left(\overline{A}\right) = q = 1 - p$.

Задача. Найти вероятность $P_n(k)$ того, что при n испытаниях событие A появилось k раз и не появится $n - k$ раз.

$\otimes$ $P_5(3)$ — вероятность того, что событие появилось в 5 испытаниях ровно 3 раза.

Применим формулу Бернулли:

Пусть n — количество испытаний, событие наступит k раз и не наступит $n - k$ раз. Вероятность одного сложного события A можно найти по теореме умножения вероятностей:

$$P(A) = p^k q^{n-k}$$

Таких сложных событий может быть столько, сколько сочетаний из n по $k - C_n^k$. Так как эти сложные события несовместны, то вероятность наступления одного из них можно найти по теореме о сложении вероятностей, но вероятности события A одинаковы, значит

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

⊗ Вероятность того, что расход электроэнергии в сутки не превосходит нормы — 0,75. Найти вероятность того, что в ближайшие 6 суток расход электроэнергии в течение 4 суток не превысит нормы.

$$P_6(4) \approx 0,3$$

⊗ В семье пять детей. Найти вероятность того, что среди этих детей

1. 2 мальчика,
2. не меньше 2 и не более 4 мальчиков.

Если вероятность рождения мальчика
$$p = 0,51.$$

$$P(A) = P_5(2) \approx 0,3$$

$$P(B) = P_5(2) + P_5(3) + P_5(4) = 0,79$$

1.5.2. Локальная теорема Лапласа

Пользоваться формулой Бернулли при больших n затруднена, так как нужно производить действия над большими числами, например, 400, поэтому в таких ситуациях стоит пользоваться другой формулой, позволяющей найти вероятность появление события k раз в n испытаниях, если число испытаний велико — **формулой Лапласа**.

Теорема. Пусть вероятность появления события в каждом испытании постоянна и отлична от нуля и единицы, тогда

$$P_n(k) = \frac{\varphi(x)}{\sqrt{npq}}$$

$$x = \frac{k - np}{\sqrt{npq}}$$

Значение функции
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}$$

собраны в таблицах. Надо помнить, что

$$\varphi(-x) = \varphi(x)$$

(функция четна)

⊗ Найти вероятность того, что событие наступит ровно 80 раз в 400 испытаниях, если $p = 0,2$.

$$x = \frac{k - np}{\sqrt{npq}} = 0; \varphi(x) = 0,3989$$

$$P_{400}(80) = 0,0499$$

⊗ Вероятность поражения мишени при одном выстреле $p = 0,75$. Найти вероятность 8 поражений, если произведено 10 выстрелов.

$$P_{10}(8) = C_{10}^{8} p^8 q^2 = 0,2815$$

$$P_{10}(8) = \frac{\varphi(2,52)}{npq} = 0,2721$$

Примечание. Расчет по формуле Бернулли ближе к истине, так как формула Лапласа дает хорошие результаты только при достаточно больших n.

1.5.3. Интегральная формула Лапласа

Задача. Пусть проведено n независимых испытаний, в каждом из которых событие A может появиться или не появиться. Пусть в каждом испытании вероятность одинакова и равна $P(A) = p$. Необходимо вычислить вероятность $P_n(k_1;k_2)$ того, что событие A появится не менее k_1 раз и не более k_2 раз.

Теорема. Пусть $P(A) = const = p$ и отлично от нуля и единицы, то вероятность того, что того, что событие A появится не менее k_1 раз и не более k_2 раз равна

$$P_n(k_1;k_2) = \Phi(x'') - \Phi(x'),$$

где

$$x' = \frac{k_1 - np}{\sqrt{npq}},$$

$$x'' = \frac{k_2 - np}{\sqrt{npq}},$$

$$\Phi(x) = \int_0^x \varphi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{x^2}{2}}.$$

$\Phi(x)$ нечетна и при $x \geqslant 5$: $\Phi(x) = 0,5$.

$\otimes$ Вероятность того, что деталь не прошла проверку $p = 0,2$. Найти вероятность того, что среди 400 случайно отобранных деталей окажутся непроверенными от 70 до 100 деталей.

$$x' = -1,25$$

$$x'' = 2,5$$

$$P_{400}(70;100) = \Phi(x'') - \Phi(x') = 0,8882$$

2. Случайные величины

1. Виды СВ. Задание ДСВ
2. МО ДСВ
3. Дисперсия ДСВ
4. Подготовка к КР
5. Функция распределения вероятностей СВ
6. Плотность распределения вероятностей НСВ
7. Нормальное распределение

2.1. Виды случайных величин. Задание дискретной случайной величины

2.1.1. Случайные величины. Дискретные и непрерывные случайные величины

Мы рассмотрим событие, состоящее в появлении какого-либо числа.

Например, появление числа при бросании кости (нельзя наперёд определить число выпавших очков, так как это случайная величина).

Возможные значения случайной величины.

Случайная величина (СВ) — величина, которая в результате испытания примет единственное $\exists!$ значение, зависящее от случайных причин, которые не могут быть учтены заранее.

Обозначаются прописными буквами X, Y, Z, а их возможные значения с соответствующими строчными буквами: x, y, z.

$\otimes$ СВ X имеет три возможных значения $\{x_1, x_2, x_3\}$

Дискретная случайная величина (ДСВ) — величина, принимающая отдельные возможные значения с определенными вероятностями.

$\otimes$ Число родившихся мальчиков из 100 новорождённых — ДСВ с возможными значениями от 0 до 100.

Эти возможные значения отделены друг от друга промежутками, в которых возможных значений нет. Число возможных значений ДСВ может быть конечным или бесконечным.

Непрерывная случайная величина (НСВ) — величина, принимающая все значения из некоторого конечного или бесконечного промежутка.

$\otimes$ Расстояние, которое пролетит снаряд при выстреле — НСВ с возможными значениями, принадлежими некоторыми промежутку $[a; b]$. Здесь нет интервалов между возможными значениями случайной величины. Число возможных значений НСВ бесконечно.

2.1.2. Закон распределения вероятностей ДСВ

На первый взгляд, для задания ДСВ достаточно перечислить все возможные значения. Но СВ могут принимать одинаковые возможные значения, то есть различные вероятности.

Закон распределения ДСВ — соответствие между возможными значениями СВ и их вероятностями.

Его можно задать

- аналитически (формулой),
- графически (многоугольник распределения),
- таблично.

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

В одном испытании ДСВ принимает единственное возможное значение, следовательно события вида $A_i = \{X = x_i\}$ образуют полную группу. То есть

$$\sum p_i = 1$$

$\otimes$ В денежной лотерее 100 билетов. Разыгрывается один приз по 500₽ и 10 призов по 100₽. Найти закон распределения стоимости возможных выигрышей для лотерейного билета.

X	0	100	500
P	$\frac{89}{100}$	$\frac{1}{10}$	$\frac{1}{100}$

2.1.3. Биномиальное распределение

Пусть произведено n независимых испытаний, в каждом из которых событие A может появиться или не появиться. Рассмотрим в качестве ДСВ число появлений события в этих испытаниях.

Найдём закон распределения ДСВ. Для этого нужно найти возможные значения ДСВ и их вероятности. В n испытаниях событие может появиться от 1 до n раз. Вероятности рассчитываются по формуле Бернулли.

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

Биномиальное распределение — распределение, рассчитывающееся по формуле Бернулли.

⊗ Монета брошена два раза. Написать в виде таблицы закон распределения ДСВ — числа выпавших гербов.

$p_0 = P_2(0) = \frac{1}{4}, p_1 = P_2(1) = \frac{1}{2}, p_2 = P_2(2) = \frac{1}{4}.$

X	0	1	2
P	0.25	0.5	0.25

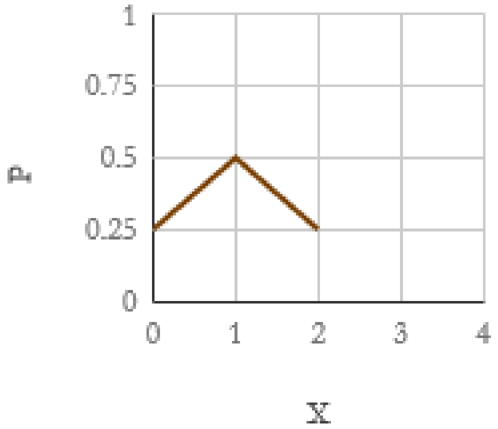


Рис. 1: Многоугольник вероятности: по оси Ox значения СВ, по Oy — соответствующие вероятности

Многоугольник распределения — графическое представление биномиального закона.

⊗ В партии $p = \frac{1}{10}$ нестандартных деталей. Наудачу отобраны 4 детали. Найти биномиальный закон распределения ДСВ числа нестандартных деталей среди 4 отобранных и построить многоугольник этого распределения.

$P_4(k) = C_4^k \frac{9^{4-k}}{10^4}$

$P_4(0) = 1 \cdot \frac{9^4}{10^4}$

$P_4(1) = 1 \cdot \frac{9^3}{10^4}$

$P_4(2) = 1 \cdot \frac{9^2}{10^4}$

$P_4(3) = 1 \cdot \frac{9^1}{10^4}$

$P_4(4) = 1 \cdot \frac{9^0}{10^4}$

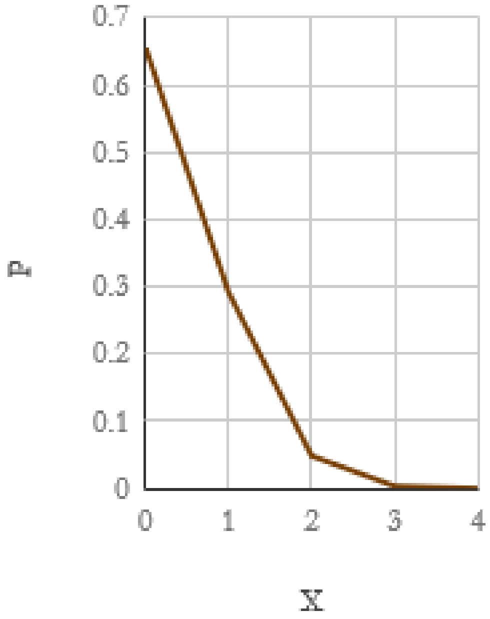


Рис. 2: многоугольник

X	0	1	2	3	4
P	0.6561	0.2916	0.0486	0.0036	0.0001

2.1.4. Распределение Пуассона

Если число испытаний велико, то пользоваться формула Лапласа, можно, однако, она становится непригодной, если вероятность события мала $p \leqslant 0,1$.

Пусть $\lambda = np = \text{const}$. Следовательно среднее число появления события в различных сериях испытаний остается неизменным.

$P_n(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (по формуле Бернулли)

$\otimes$ Отправлено $n = 5000$ деталей. Вероятность повреждения деталей в пути $p = 2 \cdot 10^{-4}$. Найти вероятность того, что на базу прибудут три нестандартные детали.

$$\lambda = 1 \Rightarrow P_5 000(3) = \frac{e^{-1}}{3!} = 0,06$$

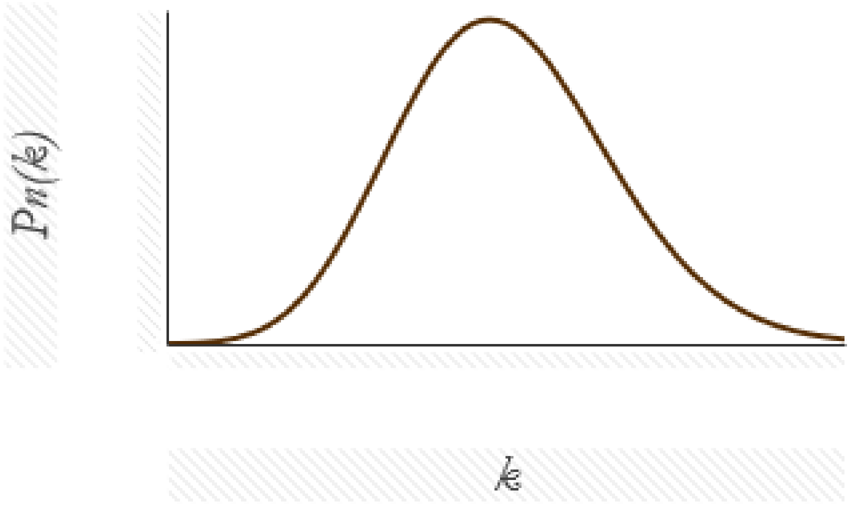


Рис. 3: Максимум при $k = \lambda$

2.1.4.1. График распределения Пуассона

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

то

$$M(X) = \sum x_i p_i$$

$\otimes$ Найти $M(X)$, зная ее закон распределения.

X	2	3	5
P	0,3	0,1	0,6

$$M(X) = 2 \cdot 0,3 + 3 \cdot 0,1 + 5 \cdot 0,6 = 3,9$$

2.2.3. Вероятностный смысл МО

Пусть произведено k испытаний, в которых ДСВ приняла m_i раз значение x_i, $i \in \overline{1,n}$. При этом $\sum m_i = n$. Тогда сумма всех значений, принятых X, будет равна

$$\sum x_i m_i$$

Найдем среднее арифметическое всех значений принимаемых СВ:

$$\overline{X} = \frac{\sum x_i m_i}{k}$$

Относительная частота события — отношение числа испытаний, в которых событие появилось, к общему числу произведенных испытаний. Где число появление события, общее число испытаний.

$$W(A) = \frac{m}{n},$$

где m — число появлений события,

а n — общее число испытаний.

Понятие схоже с вероятностью, но отличается в том, что определение вероятности не требует, чтобы испытание производились в действительности.

Определение относительной частоты предполагает проведение испытаний.

Вернемся к $\overline{X}$.

Пусть $W_i = \frac{m_i}{k}$ — относительная частота появления x_i. Тогда:

$$\overline{X} = \sum x_i W_i$$

Пусть число испытаний k велико, а значит $W_i \approx p_i$:

$$\overline{X} \approx \sum x_i p_i$$

$$\overline{X} \approx M(X)$$

Математическое ожидание примерно равно среднему арифметическому наблюдаемых значений СВ.

2.2.4. Свойства МО

2.2.4.1. 1. МО постоянной

$$M(\text{const}) = \text{const}$$

2.2.4.2. 2. Постоянный множитель можно вынести за знак МО

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

2.2.4.3. 3. МО произведения СВ

Замечание 1. Две СВ называют **независимыми**, если закон распределения одной из них не зависит от того, какие возможные значения приняла другая величинаж; иначе они **зависимы**.

Замечание 2. **Произведение** независимых СВ XY — СВ, возможные значения которой равны произведению всех возможных значений X на каждое возможное значение Y. Вероятности этих возможных значений равны произведению вероятностей сомножителей.

Замечание 3. Сумма независимых СВ $X+Y$ — СВ, возможные значения которой равны суммам всех возможных значений со всеми возможными значениями. Вероятности этих возможных значений равны произведениям вероятностей слогаемых.

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

Доказательство. Пусть X и Y Заданы законами распределения.

X	x_1	x_2	Y	y_1	y_2
P	p_1	p_2	G	g_1	g_2

Составим все возможные значения которые может принимать XY, умножив все значения X на каждое значение Y. Учитывая замечание 2, получаем:

XY	x_1y_1	x_1y_2	x_2y_1	x_2y_2
PG	p_1g_1	p_1g_2	p_2g_1	p_2g_2

Математическое ожидание равно сумме произведений всех возможных значений на их вероятности:

$$M(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_iy_jp_ig_j = \sum_{i=1}^{n} x_ip_i \cdot \sum_{j=1}^{m} y_jg_j = M(X)M(Y)$$

⊗ Независимые случайные величины X и Y заданы следующими законами распределения. Найти $M(XY)$ двумя способами: по свойству 3 и составив общий закон распределения.

X	2	4	5	Y	7	9
P	0,3	0,1	0,6	G	0,8	0,2

$M(X) = 4,4$, $M(Y) = 7,4$. $M(XY) = 32,56$.

XY	$2 \cdot 7$	$4 \cdot 7$	$5 \cdot 7$	$2 \cdot 9$	$4 \cdot 9$	$5 \cdot 9$
pg	0.06	0.02	0.12	0.24	0.08	0.48

$M(XY) = 32,56$

2.2.4.4. 4. МО суммы СВ

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

Доказательство. Аналогично.

⊗ Найти математическое ожидание числа очков, которые могут выпасть при бросании двух кубиков.

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

$$M(X+X)=2M(X)=7$$

2.2.5. МО биномиального распределения

Теорема. Математическое ожидание числа появлений события A в n независимых испытаниях равна произведению числа испытаний на вероятность появления события в каждом испытании.

$$M(X)=np$$

⊗ Вероятность попадания в цель $p=0,6$. Найти математическое ожидание общего числа попаданий, если будет произведено 10 выстрелов.

$$M(X)=np=6$$

2.3. Дисперсия дискретной случайной величины

2.3.1. Отклонение случайной величины от ее МО

Введём понятие отклонение значения от ее математического ожидания.

Отклонение СВ — разность между случайной величиной и ее математическим ожиданием.

$$X-M(X)$$

На практике часто приходится оценивать рассеяния возможных значений случайной величины вокруг ее среднего значения. Например, артиллерия оценивают кучность снарядов вокруг поражаемые цели.

На первый взгляд кажется, что для оценки рассеяния нужно найти все возможные значения отклонений СВ, а потом их среднее значение. Но может получиться так, что среднее значение $M(X-M(X))$ окажется равным нулю.

Поэтому пришли к решению заменить возможные отклонения их квадратами. Дисперсия характеризует степень рассеяния возможных значений оси от своего среднего значения.

Дисперсия СВ — математическое ожидание квадратов отклонений от их среднего значения.

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

Расчетная формула для дисперсии

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

⊗ Найти дисперсию по закону распределения.

X	2	3	5
p	0,1	0,6	0,3

$$M(X) = 3,5$$

$$M(X^2) = 13,3$$

$$D(X) = 13,3 - 12,25 = 1,05$$

2.3.2. Дисперсия биномиального распределения

Теорема. Дисперсия числа появлений события A в n независимых испытаниях равна произведению числа испытаний на вероятности появления и не появления события в одном испытании.

$$D(X) = npq$$

$\otimes$ Произведено 10 независимых испытаний, в каждом из которых $p = 0,6$. Найти дисперсию числа появлений события в этих испытаниях.

$$D(X) = npq = 2,4$$

$\otimes$ СВ X может принимать два значения $x_2 > x_1$. $M(X) = 2,7$, $D(X) = 0,21$. Найти ее значения. ($p_1 = 0,3; p_2 = 0,7$)

$0,3x_1 + 0,7x_2 = 2,7$

$0,3x_1^2 + 0,7x_2^2 - 2,7^2 = 0,21$

$x_{1_1} = 2; x_{2_1} = 3$

$x_{1_2} = 3,4; x_{2_2} = 2,4$

2.4. Функция распределения вероятностей случайной величины

2.4.1. Интегральная функция распределения

ДСВ может быть задана перечнем всех и возможных значений и их вероятности, но такой способ не применим для НСВ.

Пусть X своими возможными значениями сплошь заполняют интервал $[a; b]$, тогда перечень значений составить невозможно. Поэтому для описания такой СВ введем интегральную функцию распределения.

Функция распределения — функция $F(x)$, определяющая для всех $x \in X$ вероятность того, что X примет значение меньше x.

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

Геометрически это равенство можно истолковать так: $F(x)$ есть вероятность того, что случайная величина примет значение, которое изображается на числовой оси точкой, лежащей левее точки x.

2.4.2. Свойства

Свойство 1. Значение функции распределения принадлежат отрезку $[0; 1]$.

Свойство вытекает из определения функции распределения как вероятности

Свойство 2. $F(x)$ — неубывающая функция, то есть $F(x_2) \geqslant F(x_1)$, если $x_2 > x_1$.

Доказательство. Пусть $x_2 > x_1$. Событие, заключающееся в том, что СВ примет значение меньше x_2, можно разделить на два несовместных события по теореме о сложении вероятностей.

$$P(X < x_2) = P(X < x_1) + P(x_1 \leqslant X < x_2)$$

Отсюда

$$P(x_1 \leqslant X < x_2) = F(x_2) - F(x_1).$$

Так как любая вероятность есть число неотрицательное, то

$F(x_2) - F(x_1) \geqslant 0$, или $F(x_2) \geqslant F(x_1)$ (*)

Следствие 1. Вероятность того, что X примет значение, заключенное в интервале (a, b), равна приращению функции распределения на этом интервале:

$$P(a < X < b) = F(b) - F(a)$$

Это следствие вытекает из формулы (*), если положить $x_2 = b$ и $x_1 = a$.

⊗ Случайная величина задана функцией распределения. Найти вероятность того, что в результате испытания случайная величина примет значение в интервале $(0; 2)$.

$$F(x) = \begin{cases} 0, & x \leqslant -1 \\ \frac{x}{4} + \frac{1}{4}, & -1 < x \leqslant 3 \\ 1, & x > 3 \end{cases}$$

$$P(0 < x < 2) = F(2) - F(0) = 0,5$$

Следствие 2. Вероятность того, что X примет ровно одно определённое значение, равна нулю. $P(X = x_0) = 0$

Доказательство.

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x).$$

Так как X — НСВ, то функция $F(x)$ непрерывна. В силу непрерывности $F(x)$

$$\lim_{\Delta x \to 0} F(x + \Delta x) - F(x) = 0$$

Следовательно, $P(X = x_0) = 0$. Используя это положение, легко убедиться в справедливости равенств

$$P(a < X < b) = P(a \leqslant X < b) = P(a < X \leqslant b) = P(a \leqslant X \leqslant b)$$

Таким образом, бессмысленно говорить о том, что НСВ примет одно определенное значение, но имеет смысл рассматривать попадание её в интервал, пусть и сколь угодно малый.

Свойство 3. Если возможно значение случайной величины лежат в интервале $[a; b]$, то

$F(x) = 0$ при $x \leqslant a$,

$F(x) = 1$ при $x \geqslant b$.

2.4.3. График ИФР

1. График расположен в полосе, ограниченной прямыми $y = 0$ и $y = 1$.
2. При возрастающем x на интервале (a, b) график функции идет вверх.
3. При $x \leqslant a$ ординаты графика равны нулю; при $x \geqslant b$ ординаты графика равны единице.

ДСВ также можно представить ИФР. При этом график будет иметь ступенчатый вид.

⊗ ДСВ X задана таблицей распределения. Найти функцию распределения и вычертить ее график.

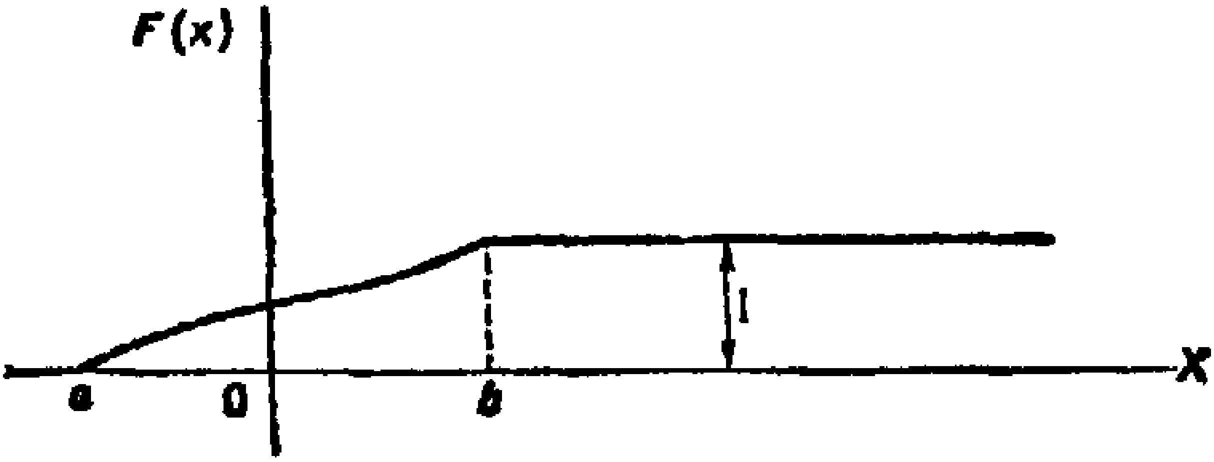
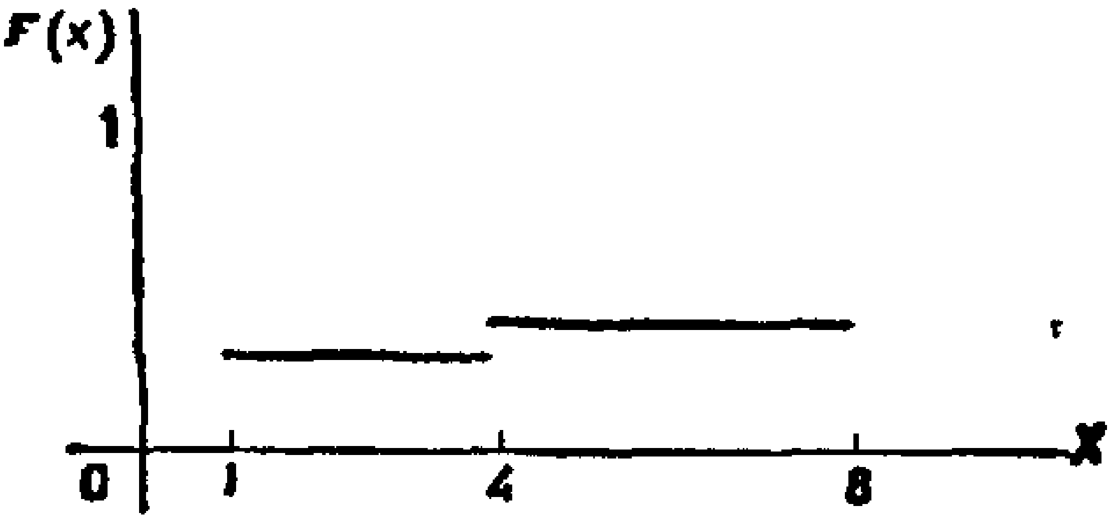


Рис. 4: Пример графика ИФР

X	1	4	8
p	0,3	0,1	0,6

$$F(x) = \begin{cases} 0, & x < 1 \\ 0,3, & 1 \leqslant x < 4 \\ 0,4, & 4 \leqslant x < 8 \\ 1, & x \geqslant 8 \end{cases}$$



$\otimes$ ДСВ X задана таблицей распределения. Найти функцию распределения и вычертить ее график.

X	3	4	7	10
p	0,2	0,1	0,4	0,3

$$F(x) = \begin{cases} 0, & x < 3 \\ 0,2, & 3 \leqslant x < 4 \\ 0,3, & 4 \leqslant x < 7 \\ 0,7, & 7 \leqslant x < 10 \\ 1, & x \geqslant 10 \end{cases}$$

$\otimes$ НСВ X задана ИФР. Найти вероятность того, что в результате 4 испытаний X примет значение в интервале $(0,25; 0,75)$ ровно три раза.

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leqslant x < 1 \end{cases}$$

2.5.1. Задачи

⊗ Имеется 5 черных и 3 белых шара. Случайным образом вынимают 2 шара. Найти вероятность то го, что среди них был хотя бы один белый шар.

$$P(A) = 1 - \frac{C_5^2 C_3^0}{C_2^8} = 0,64$$

⊗ вероятность попадания в цель при стрельбе из каждого из трех орудий $p = 0,7$. Найти вероятность хотя бы одного попадания при одном залпе из каждого орудия.

$$P(A) = 1 - q^3 = 0,973$$

⊗ в группе 22 мальчика и восемь девочек. Для девочки вероятность сдать экзамен $P_{B_1}(A) = 0,85$. Для мальчика вероятность сдать экзамен $P_{B_2}(A) = 0,8$. Найти вероятность того, что один случайно взятый(-ая) студент(-ка) сдаст экзамен.

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) = 0,813$$

⊗ В цеху 8 станков. У каждого станка вероятность того, что он работает 0,8. Найти вероятность того, что сейчас включено

1. шесть станков,
2. от 4 до 6 станков включительно.

$$P(A) = C_8^2 p^6 q^2 = 0,293$$

$$P(B) = C_8^2 p^6 q^2 + C_8^3 p^5 q^3 + C_8^4 p^4 q^4 = 0,486$$

⊗ В каждом из 500 независимых испытаний события происходят $p = 0,4$. Найти вероятность того, что событие происходит

1. 220 раз,
2. от 180 до 240 раз.

$$P_{500}(220) = \frac{\varphi(1,826)}{\sqrt{npq}} = 0,0069$$

$$P_{500}(180;240) = \Phi(-1,826) - \Phi(3,652) = 0,96593$$

⊗ Найти математическое ожидание и дисперсию, зная закон распределения.

X	1	2	4	10	20
p	0,05	0,3	0,1	0,2	0,35

$M(X) = 10,05$
$D(X) = 61,8475$

⊗ Дано распределение X. Найти недостающие части, зная, что $M(X) = 5$.

X	8	3
p	p_1	p_2

$$8p_1 + 3p_2 = 5$$

$$p_1 + p_2 = 1$$

$$p_1 = 0,4; p_2 = 0,6$$

2.6. Плотность распределения вероятностей непрерывной случайной величины

2.6.1. Дифференциальная функция распределения вероятностей НСВ

Способ задания на с помощью интегральной функцией не единственный. Ее можно также задать с помощью дифференциальной функцией распределения вероятностей.

Дифференциальная функция $f(x)$ или **плотность вероятностей** — первая производная от интегральной функции распределения.

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

Зная плотность распределения, можно вычислить вероятность того, что она примет значение, принадлежащее заданному интервалу.

Теорема. Вероятность того, что НСВ примет значение, принадлежащее интервалу, равно определенному интегралу плотности распределения.

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

Доказательство. Из свойств интегральной функции

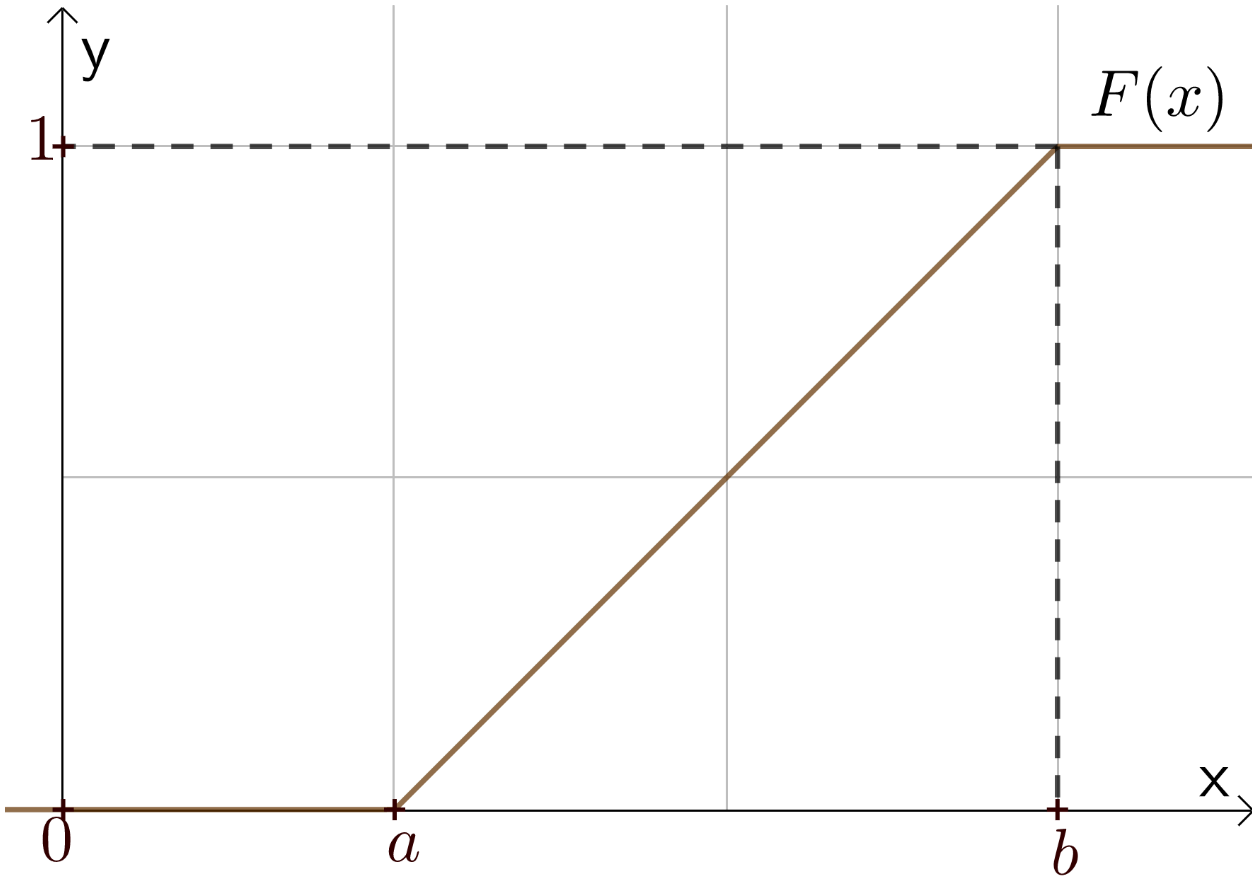
$$P(a < X \leqslant b) = F(b) - F(a)$$

А с другой стороны по формуле Ньютона-Лейбница.

$$F(b) - F(a) = \int_a^b F'(x)dx = \int_a^b f(x)dx$$

Геометрически полученный результат можно истолковать так: вероятность того, что НСВ примет значение, принадлежащее интервалу (a, b), равна площади криволинейной трапеции, ограниченной осью Ox, кривой распределения $f(x)$ и прямыми $x = a$ и $x = b$.

В общем виде, имея плотность распределения $f(x)$, можно найти функцию распределения $F(x)$ по формуле



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

⊗ Найти функцию распределения по данной плотности распределения

$$f(x) = \begin{cases} 0, & x \leqslant a \\ \frac{1}{b-a}, & a < x \leqslant b \\ 0, & x \geqslant b \end{cases}$$

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0dx, & x \leqslant a \\ \int_{-\infty}^{a} 0dx + \int_{a}^{x} \frac{1}{b-a}dx, & a < x \leqslant b \\ \int_{-\infty}^{a} 0dx + \int_{a}^{b} \frac{1}{b-a}dx + \int_{b}^{x} 0dx, & x \geqslant b \end{cases} = \begin{cases} 0, & x \leqslant a \\ \frac{x-a}{b-a}, & a < x \leqslant b \\ 1, & x \geqslant b \end{cases}$$

2.6.2. Свойства дифференциальной функции распределения

2.6.2.1. Плотность распределения — неотрицательная функция

$$f(x) \geqslant 0$$

Доказательство. Функция распределения — неубывающая функция, следовательно, ее производная $F'(x) = f(x)$ — функция неотрицательная.

Геометрически это свойство означает, что точки, принадлежащие графику плотности распределения, расположены либо над осью Ox, либо на этой оси.

2.6.2.2. Несобственный интеграл от плотности распределения в пределах от −∞ до +∞ равен единице

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

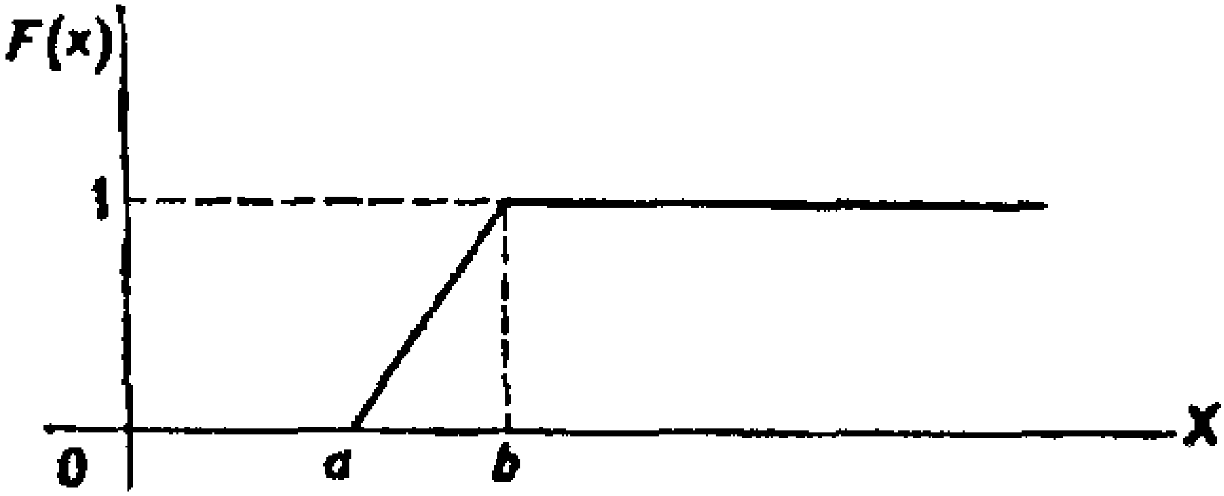


Рис. 5: График $F(x)$

Доказательство. Несобственный интеграл $\int_{-\infty}^{+\infty} f(x)dx$ выражает вероятность события, состоящего в том, что случайная величина примет значение, принадлежащее интервалу $(-\infty; +\infty)$. Очевидно, такое событие достоверно, следовательно, вероятность его равна единице.

Геометрически это означает, что вся площадь криволинейной трапеции, ограниченной осью Ox и кривой распределения, равна единице.

⊗ НСВ задана дифференциальной функцией $f(x) = \frac{2}{3}\sin 3x$ при $0 < x < \frac{\pi}{3}$. Вне этого интервала функция равна нулю. Найти вероятность того, что случайная величина примет значение из интервала $(\frac{\pi}{6}; \frac{\pi}{4})$.

$$P(\frac{\pi}{6} < x < \frac{\pi}{4}) = \int_{\pi/6}^{\pi/4} \frac{2}{3}\sin 3x = \frac{\sqrt{2}}{9}$$

⊗ НСВ задана дифференциальной функцией. Найти функцию распределения.

$$f(x) = \begin{cases} 0, & x \leqslant 0 \\ \cos x, & 0 < x \leqslant \frac{\pi}{2} \\ 0, & x > \frac{\pi}{2} \end{cases}$$

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0dx, & x \leqslant 0 \\ \int_{-\infty}^{0} 0dx + \int_{0}^{x} \cos xdx, & 0 < x \leqslant \frac{\pi}{2} \\ \int_{-\infty}^{0} 0dx + \int_{0}^{\frac{\pi}{2}} \cos xdx + \int_{\frac{\pi}{2}}^{x} 0dx, & x > \frac{\pi}{2} \end{cases} = \begin{cases} 0, & x \leqslant 0 \\ \sin x, & 0 < x \leqslant \frac{\pi}{2} \\ 1, & x > \frac{\pi}{2} \end{cases}$$

⊗ НСВ задана функцией. Найти $F(x)$ и $P(0 < x < \frac{\pi}{4})$.

$$f(x) = \begin{cases} 0, & x \leqslant 0 \\ \frac{\sin x}{2}, & 0 < x \leqslant \pi \\ 0, & x > \pi \end{cases}$$

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ \int_{0}^{x} \frac{1}{2}\sin xdx, & 0 < x \leqslant \pi \\ 1, & x > \pi \end{cases} = \begin{cases} 0, & x \leqslant 0 \\ \frac{1-\cos x}{2}, & 0 < x \leqslant \pi \\ 1, & x > \pi \end{cases}$$

$$P\left(0 < x < \frac{\pi}{4}\right) = F\left(\frac{\pi}{4}\right) = \frac{2-\sqrt{2}}{4}$$

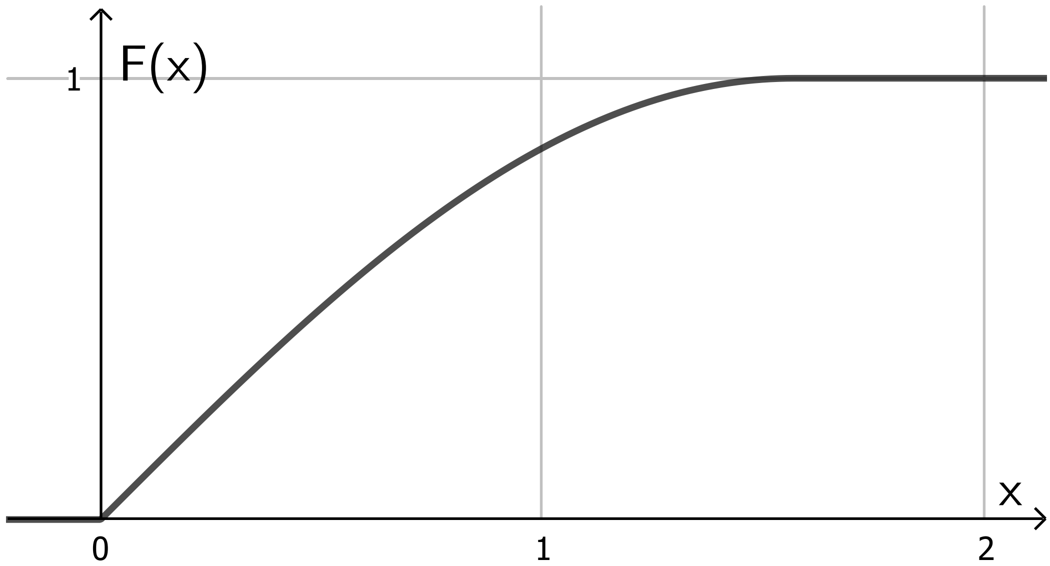


Рис. 6: $F(x)$

2.6.3. Числовые характеристики НСВ

Будем определять их по аналогии с ДСВ.

Пусть НСВ X задана плотностью распределения $f(x)$. Пусть все возможные значения НСВ принадлежат интервалу $[a;b]$. Разобьем этот отрезок на n частичных отрезков с длинами $\Delta x_1, .., \Delta x_n$, и выберем в каждом произвольную точку x_i.

Необходимо определить математическое ожидание НСВ. составим сумму произведений возможных значений x_i на вероятности попадания их в интервал Δx_i (напомним, что произведение $f(x)\Delta x$приближенно равно вероятности попадания X в интервал Δx):

$$\sum x_i f(x_i)\Delta x_i$$

Перейдем к пределу:

$$\lim_{\max \Delta x \to 0} \sum x_i f(x_i)\Delta x_i = \int_a^b x f(x) dx$$

Математическое ожидание НСВ X, возможные значения которой принадлежат отрезку $[a;b]$ —

$$M(X) = \int_a^b x f(x) dx$$

Аналогично определяется дисперсия НСВ.

Дисперсия НСВ — математическое ожидание квадрата ее отклонения.

$$D(X) = \int_a^b (x - M(X))^2 f(x) dx$$

Для вычисления используется формула

$$D(X) = \int_a^b x^2 f(x) dx - M(X)^2$$

⊗ Найти математическое ожидание и дисперсию НСВ, заданной функцией распределения

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ x, & 0 < x \leqslant 1 \\ 0, & x > 1 \end{cases}$$

$$f(x) = F'(x) \begin{cases} 0, & x \leqslant 0 \\ 1, & 0 < x \leqslant 1 \\ 0, & x > 1 \end{cases}$$

$$M(X) = \int_0^1 x dx = \frac{1}{2}$$

$$D(X) = \int_0^1 x^2 dx - \frac{1}{4} = \frac{1}{12}$$

2.6.4. Закон равномерного распределения вероятностей НСВ

Плотность распределения НСВ называют **законом распределения**. Чаще всего встречаются законы равномерного и нормального распределения.

Равномерное распределение вероятностей — если на интервале, которому принадлежат все возможные значения случайной величины, плотность распределения сохраняет постоянное значение.

Найдем плотность равномерного распределения $f(x)$, считая, что все возможные значения случайной величины заключены в интервале (a, b), на котором функция $f(x)$ сохраняет постоянные значения:

По условию, X не принимает значений вне интервала (a, b), поэтому $f(x) = 0$ при $x < a$ и $x > b$.

Найдем постоянную C. Так как все возможные значения случайной величины принадлежат интервалу (a, b), то должно выполняться соотношение

$\int_a^b f(x)dx = 1$, или $\int_a^b Cdx = 1$.

Отсюда

$$C = 1/\int_a^b dx = \frac{1}{b-a}$$

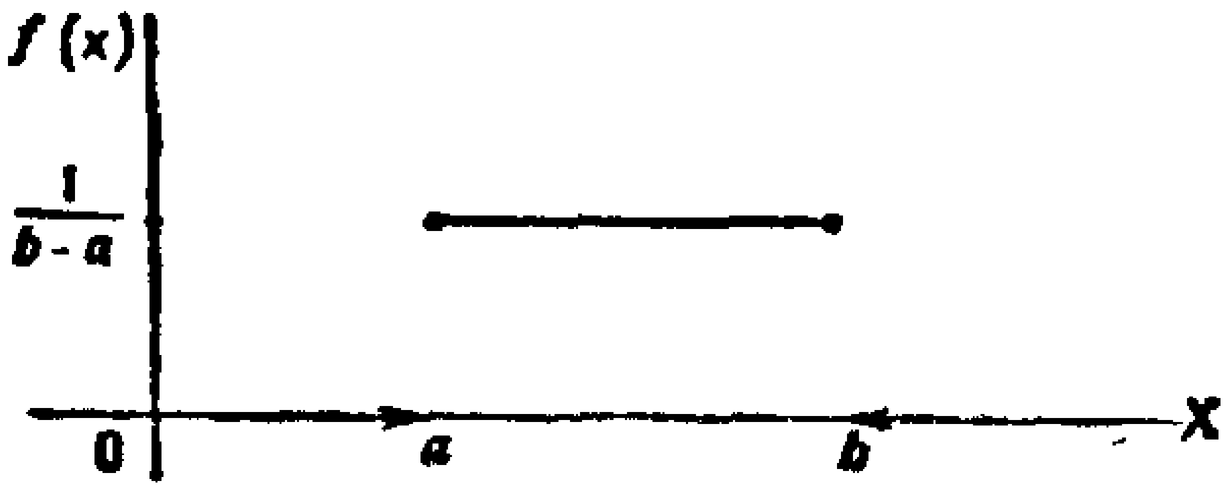


Рис. 7: $f(x)$

Итак, искомая плотность вероятности равномерного распределения

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

$$F(x) = \begin{cases} 0, & x \leqslant a \\ \frac{x-a}{b-a}, & a < x \leqslant b \\ 1, & x > b \end{cases}$$

⊗ Найти математическое ожидание и дисперсию равномерного распределения.

$$M(X) = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{b+a}{2}$$

$$D(X) = \int_a^b x^2 \cdot \frac{1}{b-a} dx - M(X)^2 = \frac{(b-a)^2}{12}$$

2.7. Нормальное распределение

2.7.1. Нормальное распределение

Нормальное распределение — распределение, описываемые плотностью

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

Нормальное распределение описывается двумя параметрами:

- $\mu = M(X)$
- $\sigma = \sqrt{D(X)}$ — среднеквадратичное отклонение

2.7.2. Нормальная кривая

Нормальная кривая (кривая Гаусса) — график плотности нормального распределения.

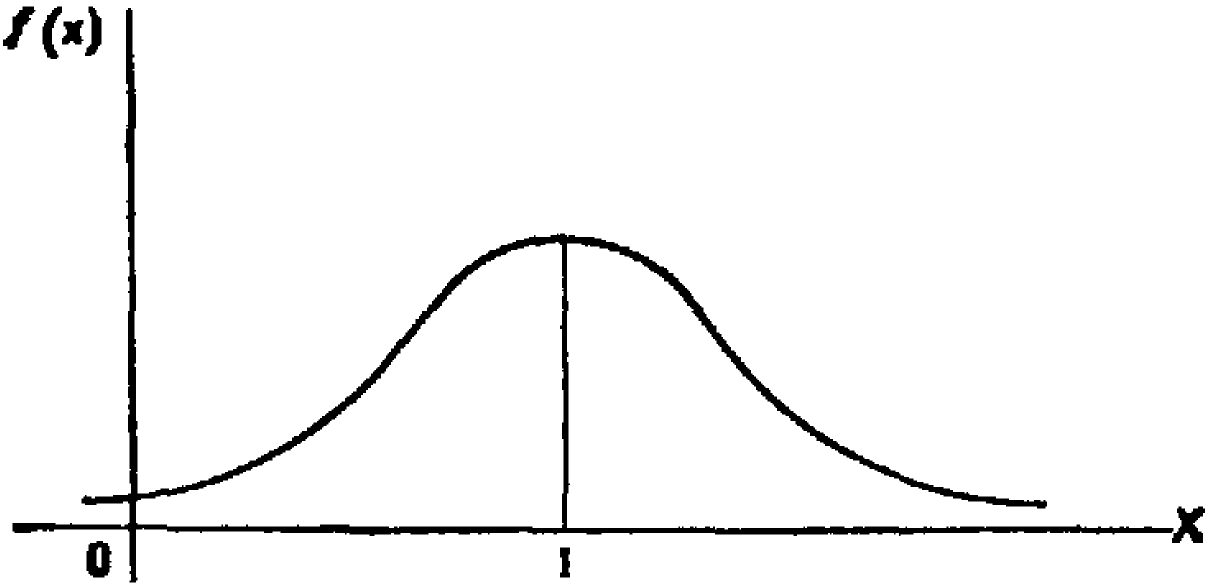


Рис. 8: $\Gamma f(x)$

2.7.2.1. Свойства

1. $D(f) = \mathbb{R}$
2. $f(x) \geqslant 0$
3. Ox — асимптота.
4. Точка максимума $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$
5. С помощью первой производной
6. График симметричен относительно $x = a$
7. Разность $x - a$ содержится в аналитическом выражении функции в квадрате
8. Точки перегиба $\left(\mu \pm \sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$
9. С помощью второй производной
10. Площадь под графиком всегда единица (по свойству плотности распределения)

2.7.3. Влияние параметров нормального распределения на форму нормальной кривой

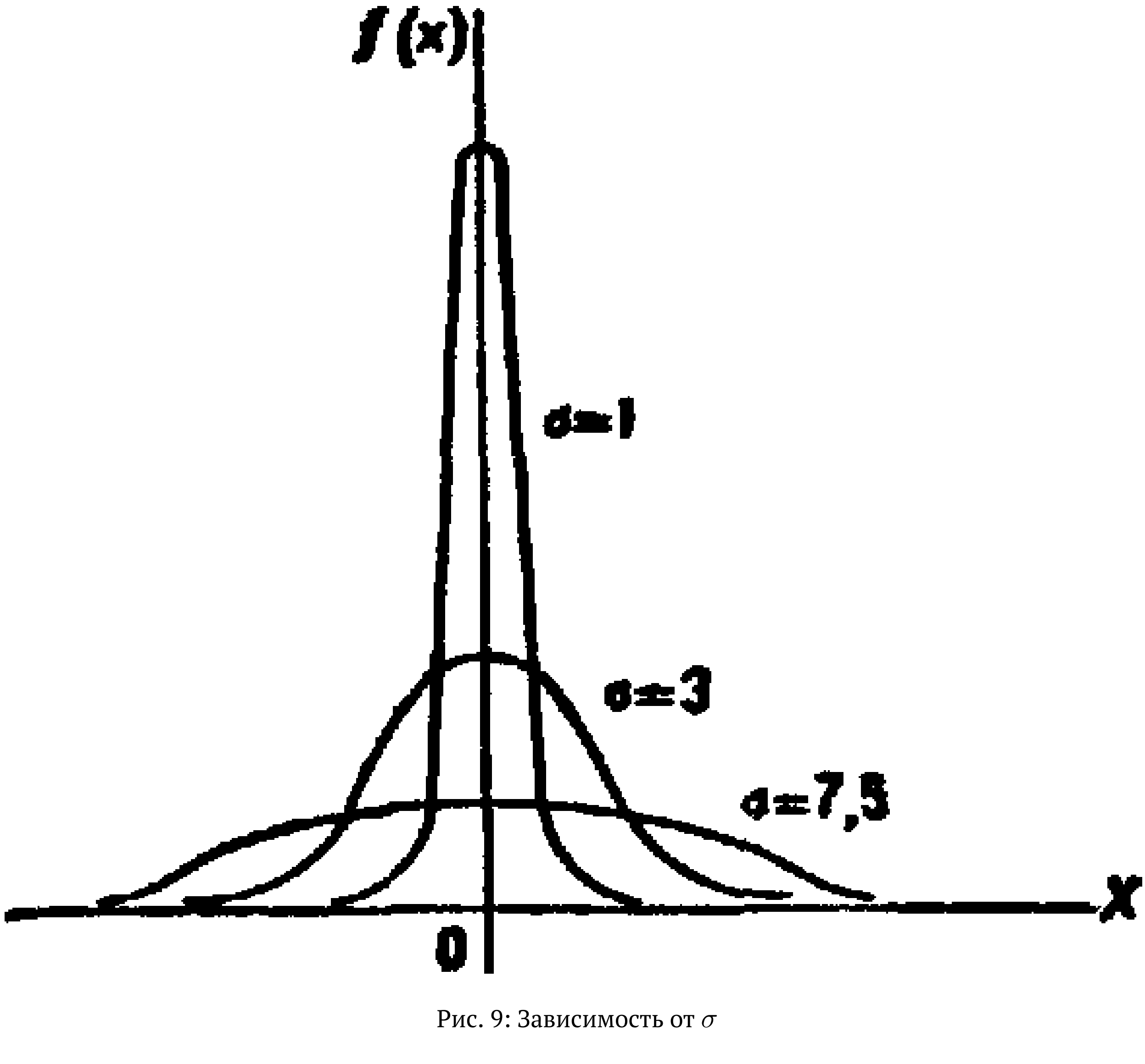


Рис. 9: Зависимость от σ

- μ
  - *увеличивается* $\to$ график движется *вправо* ($\to$)
  - *уменьшается* $\to$ график движется *влево* ($\leftarrow$)
- σ
  - *увеличивается* в n раз $\to$ максимальное значение *уменьшается* в n раз, а сама кривая становится более пологой, т. е. сжимается к оси Ox;
  - и наоборот.

Функция распределения находится по формуле

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

Особое значение среди нормальных распределений имеет **нормированное** нормальное распределение ($\mu = 0, \sigma = 1$).

2.7.4. Вероятность попадания в заданный интервал нормальной случайной величины

Если НСВ X задана плотностью $f(x)$, то вероятность того, что X примет значение, принадлежащее интервалу (a, b), такова

$$P(a < X < b) = \int_{a}^{b} f(x)dx$$

Для нормального распределения такая вероятность находится с помощью формулы Лапласа

$$P(a < X \leqslant b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

2.7.5. Вычисление вероятности заданного отклонения

Часто требуется вычислить вероятность отклонения нормального распределения по абсолютной величине меньше заданного $\varepsilon > 0$.

$$P\left(|x-\mu| \leqslant \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

2.7.5.1. Правило трех сигм

$$P\left(|x-\mu| \leqslant 3\sigma\right) = 2\Phi\left(3\right) = 0,9973.$$

Если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения.

⊗ НСВ задана нормальным законом распределения ($\mu = 30$, $\sigma = 10$). Найти

1. $P(10 < X \leqslant 50) = 2\Phi(2) = 0,9544$
2. $P\left(|x-\mu| \leqslant 3\right) = 2\Phi(0,3) = 0,2358$

⊗ Производится взвешивание некоторого вещества без систематических ошибок. Случайные ошибки взвешивания подчиняются нормальному закону со средним квадратичным отклонением $\sigma = 20$ г. Найти вероятность того, что взвешивание будет производиться с ошибкой, не превосходящей по абсолютной величине $\varepsilon = 10$ г.

$$P\left(|x| \leqslant \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) = 0,385$$

3. Элементы математической статистики

1. Выборочный метод
2. Статистические оценки параметров распределения
3. Методы расчета сводных характеристик выборки

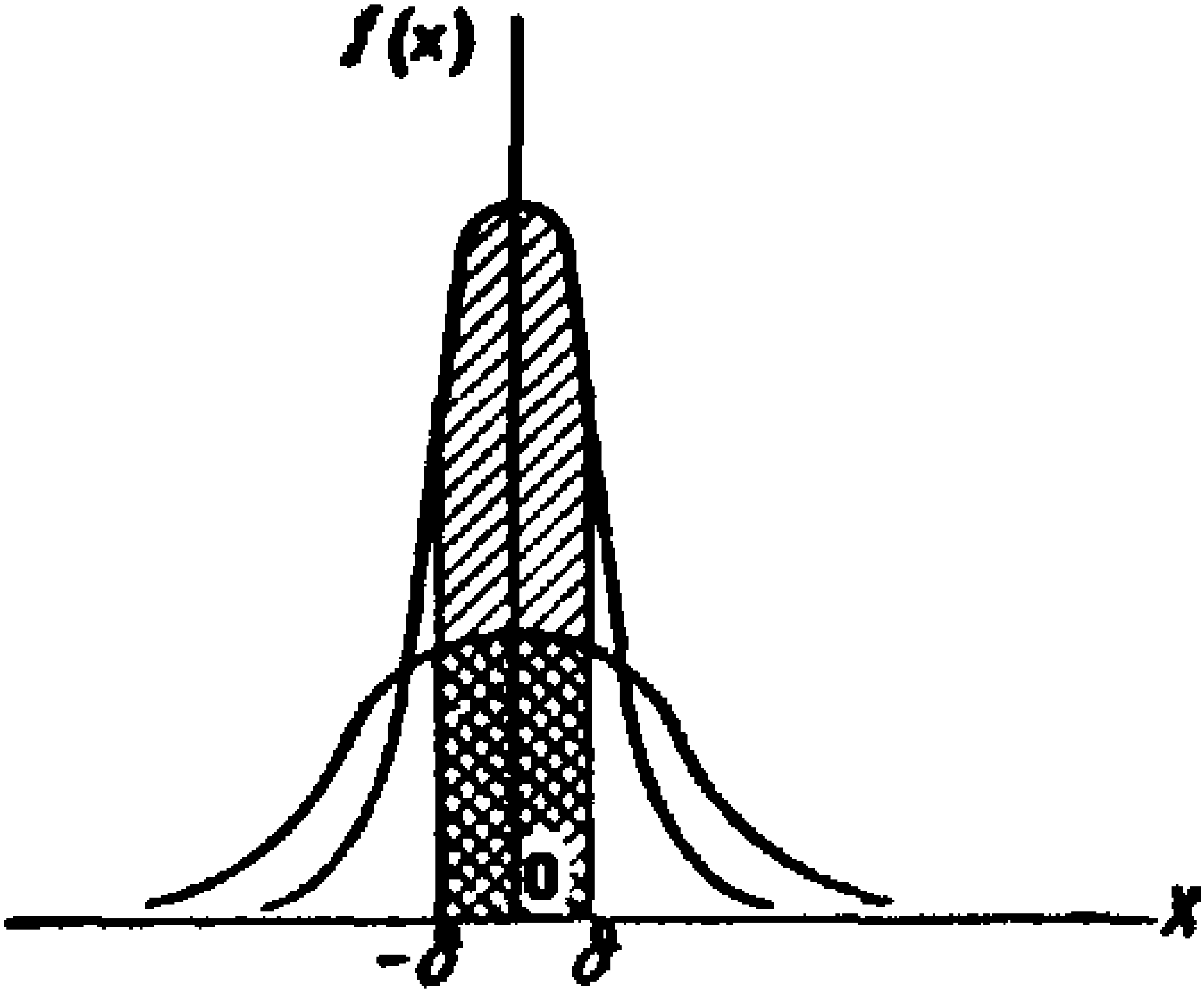


Рис. 10: если две случайные величины нормально распределены и $\mu = 0$ то вероятность принять значение, принадлежащее интервалу $(-\varepsilon; \varepsilon)$, больше у той величины, которая имеет меньшее значение ε.

3.1. Выборочный метод

3.1.1. Задачи МС

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

- Оценка неизвестной вероятности события

- оценка неизвестной функции распределения
- оценка параметров распределения
- оценка зависимости случайной величины от одной или нескольких случайных величин

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

Современная математическая статистика разрабатывает способы определения числа необходимых испытаний до начала исследования (планирование эксперимента) и в ходе исследования (последовательный анализ).

Задача математической статистики — создание *методов сбора и обработки* статистических *данных для получения* научных и практических *результатов*.

3.1.2. Генеральная и выборочная совокупность

Выборочная совокупность или **выборка** — совокупность случайно отобранных объектов.

Генеральная совокупность — совокупность объектов, из которых производится выборка.

Объем совокупности — число объектов этой совокупности.

⊗ Пусть из 1000 деталей отобрано 100 деталей для обследования, тогда объем генеральной совокупности 1000, а объем выборки 100.

3.1.3. Повторная и бесповторная выборки. Репрезентативная выборка

- Выборка
  - **Повторная** — такая, при которой отобранный объект возвращается в генеральную совокупность.
  - **Бесповторная** — такая, при которой отобранный объект генеральную совокупность не возвращается, уменьшая объём.

Что по данной выборке можно было достаточно уверенно судить об интересующем нас признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли, то есть выборка должна правильно представлять **пропорции** генеральной совокупности.

3.1.4. Статистическое распределение выборки

Пусть из генеральной совокупности извлечена выборка объема n, при этом

значение x_1наблюдаются n_1 раз,

…

значение x_k наблюдаются n_k раз, при этом $\sum n_i = n$

Варианта — наблюдаемое значение x_i.

Вариационный ряд — последовательность вариант, записанных в порядке возрастания.

Частота (вырианты) — число наблюдений n_i.

Относительная частота (выриanты) — отношение частоты к объему выборки

$$w_i = \frac{n_i}{n}$$

Статистическое распределение выборки — перечень вариант, записанных в порядке возрастания и соответствующих им частот или относительных частот.

⊗ Задано распределение частот выборки, объемом $n = 20$. Найти распределение относительных частот.

x_i	n_i	w_i
2	3	3/20
6	10	10/20
12	7	7/20
$\sum$	20	1

Обычно при этом результат наблюдений фиксируют в протоколе наблюдений в порядке их наблюдения

⊗ 2, 4, 2, 4, 3, 3, 3, 0, 2, 4, 3, 2, 2, 4, 5, 4, 3, 4, 3, 3, 5.

В выборке одно значение варианты может встречаться несколько раз, Следовательно целесообразно составить **таблицу** статистического распределения. Для составления таблицы необходимо

1. Найти минимальное x_{min} и максимальное x_{max} значение выборки
2. В первый столбец записать x_i, начиная от минимального до максимального.
3. Просмотреть все элементы выборки и отметить каждое значение во втором столбце таблицы.
4. Если объем выборки велик, то существует метод подсчёта меток. Каждый 4 метки перечеркиваются пятой.
5. Посчитать количество меток и записать соответствующее число n_i в третий столбец.
6. Подсчитать объем выборки $n = \sum_{i=1}^{k} n_i$, где k — число различных вариант
7. В четвёртом столбце записать относительные частоты w_i

⊗

x_i		n_i	w_i
0	$\vert$	1	1/21
1		0	0
2	$\vert\vert\vert\vert\diagdown$	5	5/21
3	$\vert\vert\vert\vert\diagdown$		
4	$\vert\vert\vert\vert\diagdown$		6
5	$\vert\vert$	21	21/21
$\sum$		21	1

Если k слишком велико, или близко к n, то составляют вариационный ряд по интервалам значений генеральной совокупности.

Вариационный ряд по интервалам можно получить из приведённого выше алгоритма: во 2 пункте заполнить первый столбец интервалами значений генеральной совокупности, все интервалы выбирают одинаковой длины h, так, чтобы минимальное значение вошло в 1 интервал, максимальное — в последний.

Обычно начало интервала учитывается при подсчете меток, а его окончание — нет ($[a, b)$).

⊗

Δx_i	n_i	$x_{i\text{ср}}$
20-25	3	22,5
25-30	10	27,5
30-35	11	32,5

3.1.5. Полигон и гистограмма

- *Полигон частот* это ломаная, отрезки которой соединяют точки с координатами $(x_i; n_i)$

- *Полигон относительных частот* — ломаная, отрезки которой соединяют точки с координатами $(x_i; w_i)$.

- *Гистограмма*. Над каждым значением строят прямоугольник, высота которого пропорциональна n_i.

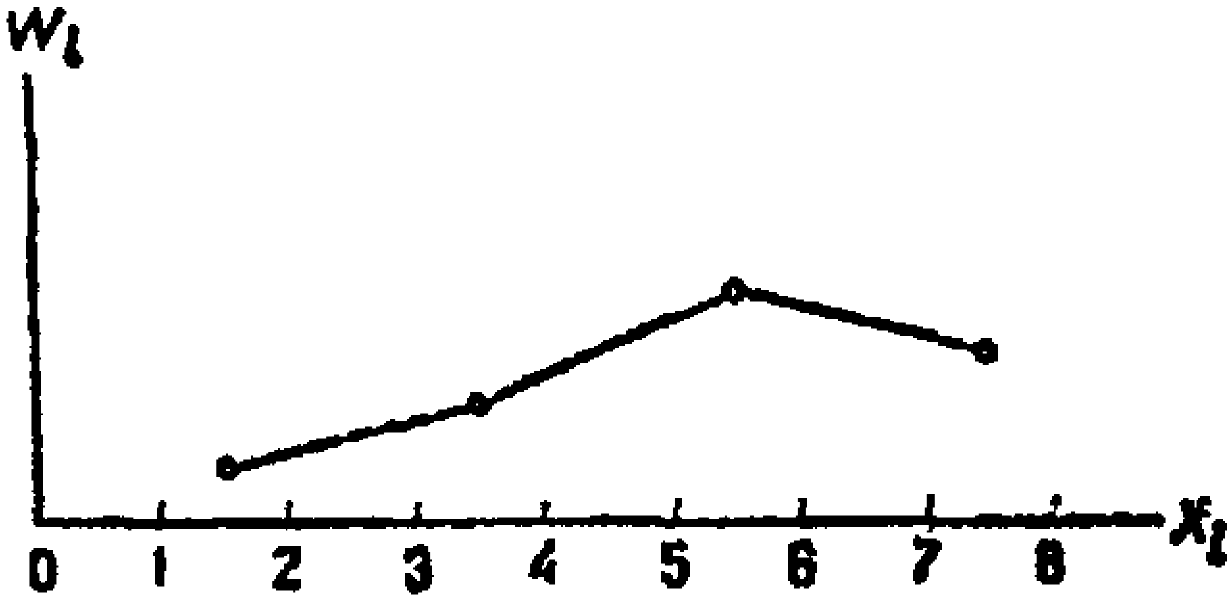


Рис. 11: Полигон относительных частот

3.1.6. Эмпирическая функция распределения

Пусть известно статистическое распределение частот количественного признака X, например контролируемый размер детали, объём выборки n, число наблюдений n_x, при которых наблюдаемое значение признака $X < x$.

Относительная частота события "$X < x$" равна $\frac{n_x}{n}$.

Если изменяется x, то изменяется и относительная частота $\frac{n_x}{n}$, то есть она является функцией от x. Так как эта функция находится опытным путем, то ее называют **эмпирической**.

Эмпирическая функция распределения выборки или просто **функция распределения** — функция, определяющая для всех значений случайной величины x относительную частоту события "$X < x$"

$$F^*(x) = \frac{n_x}{n}$$

,

где n_x — число вариант, для которых "$X < x$".

Теоретическая функция распределения — функция $F(x)$ распределения генеральной совокупности.

$F(x)$ определяет *вероятность* события "$X < x$",

$F^*(x)$ определяет *относительную частоту* этого же события.

Поэтому целесообразно использовать $F^*(x)$ для приближенного представления теоретической (интегральной) функции распределения генеральной совокупности.

⊗ Построить ЭФР по распределению выборки.

x_i	n_i	w_i	n_x
2	12	12/60	0

x_i	n_i	w_i	n_x
6	10	18/60	12
10	30	30/60	30
$\sum$	60	1	

$$F^*(x) = \begin{cases} 0, & x < 2 \\ 12/60, & 2 \leqslant x < 6 \\ 30/60, & 6 \leqslant x < 10 \\ 1, & x \geqslant 10 \end{cases}$$

$\otimes$ Построить ЭФР по распределению выборки.

x_i	n_i	w_i	n_x
2	1	0	0
5	3	1	0,1
7	2	4	0,4
8	4	6	0,6
$\sum$	10		1

$$F^*(x) = \begin{cases} 0, & x < 2 \\ 0,1, & 2 \leqslant x < 5 \\ 0,4, & 5 \leqslant x < 7 \\ 0,6, & 7 \leqslant x < 8 \\ 1, & x \geqslant 8 \end{cases}$$

$\otimes$ То же задание для примера из предыдущего пункта

3.2. Числовые характеристики выборки

3.2.1. Выборочная средняя

Пусть для изучения генеральной совокупности относительно количественного признака X извлечена выборка объема n. Выборочной средней $\overline{x_\text{в}}$ называют среднее значение признака выборочной совокупности. Если все значения $x_1, x_2, \dots x_n$ признака выборки объема n различны, то

$$\overline{x_\text{в}} = \frac{\sum x_i}{n}.$$

Если же значения признака $x_1, x_2, \dots x_k$ имеют соответственно частоты $n_1, n_2, \dots n_k$, причем $n_1 + n_2 + \dots + n_k = n$, то

$$\overline{x_\text{в}} = \frac{\sum x_i n_i}{n}.$$

3.2.2. Выборочная дисперсия

Для того чтобы охарактеризовать рассеяние наблюдаемых значений количественного признака выборки вокруг своего среднего значения $\overline{x_\text{в}}$, вводят сводную характеристику — выборочную дисперсию.

Выборочной дисперсией $D_\text{в}$ называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\overline{x_\text{в}}$.

Если все значения $x_1, x_2, \dots x_n$ признака выборки объема n различны, то

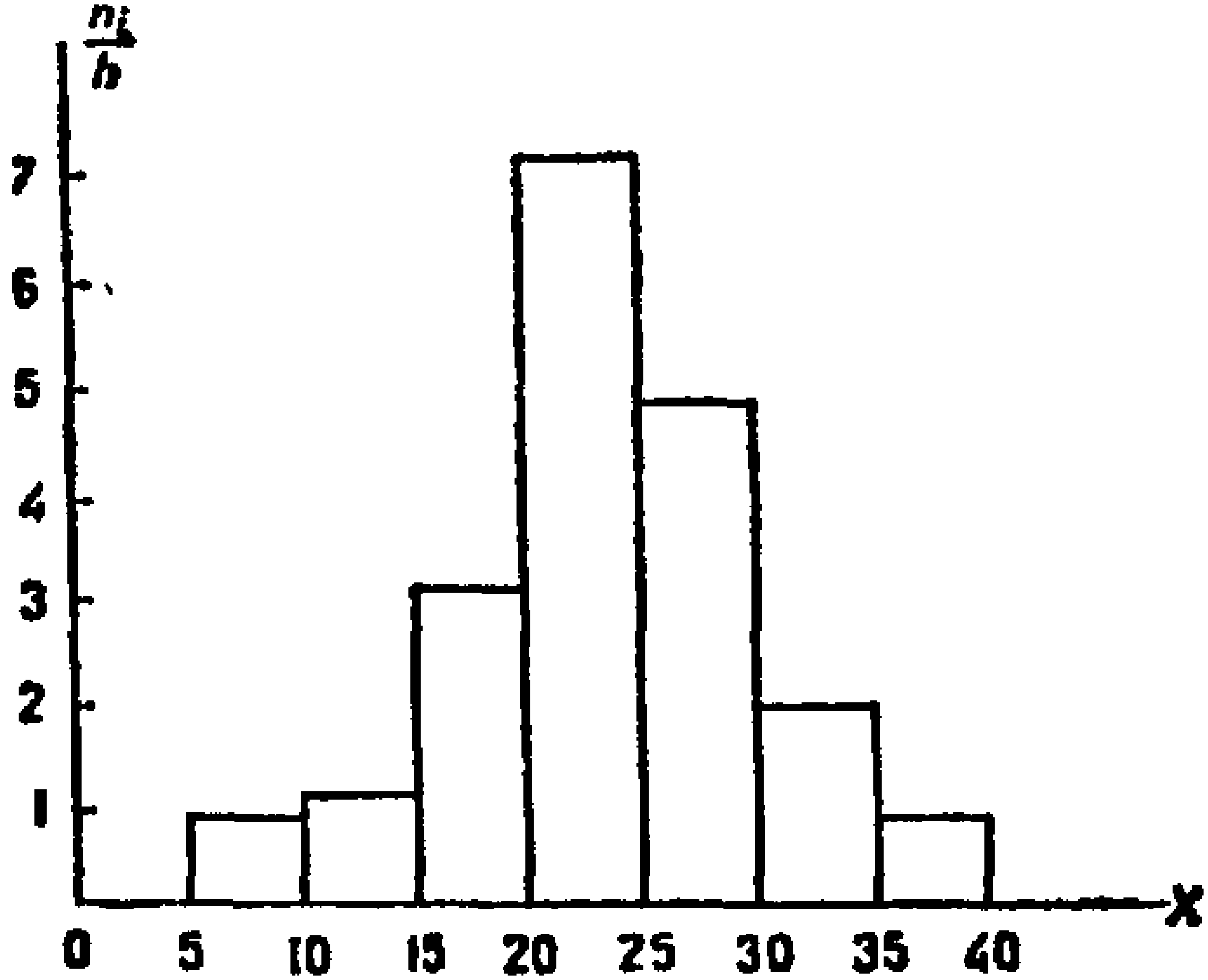


Рис. 12: Гистограмма частот

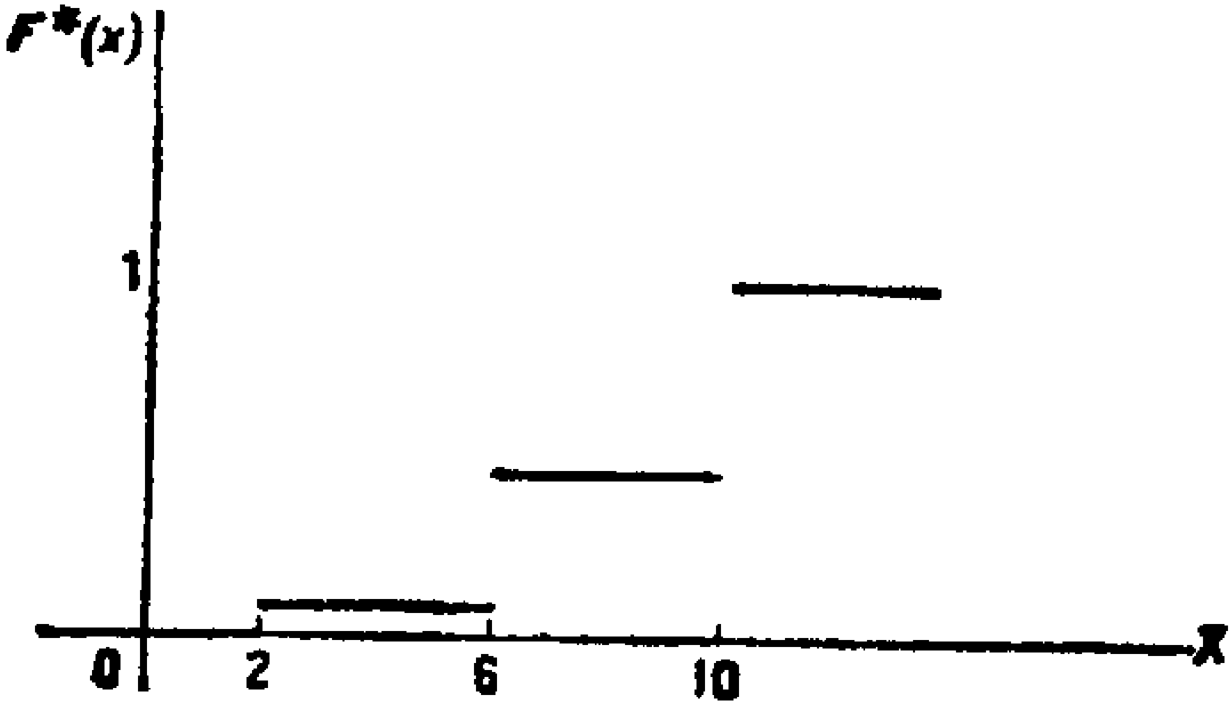


Рис. 13: График ЭФР

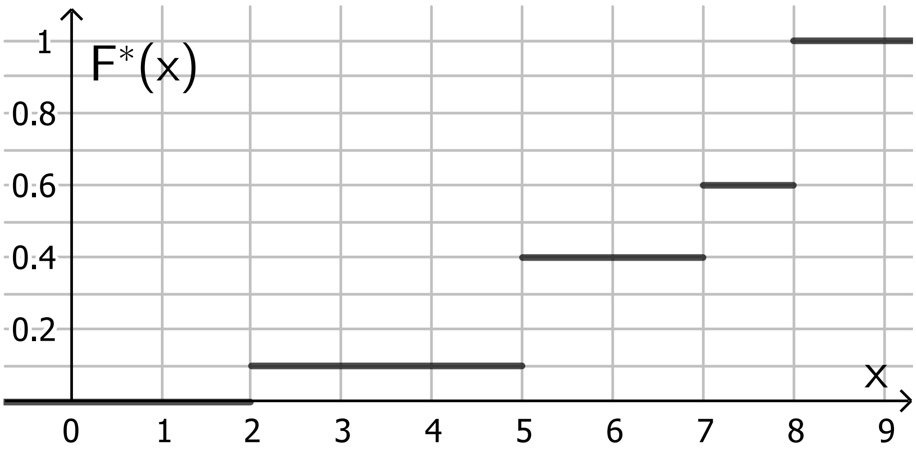


Рис. 14: $F^*(x)$

$$D_{\text{в}} = \frac{\sum (x_i - \overline{x_{\text{в}}})^2}{n}.$$

Если же значения признака $x_1, x_2, \ldots x_k$ имеют соответственно частоты $n_1, n_2, \ldots n_k$, причем $n_1 + n_2 + \ldots + n_k = n$, то

$$D_{\text{в}} = \frac{\sum n_i (x_i - \overline{x_{\text{в}}})^2}{n}.$$

⊗ Выборочная совокупность задана таблицей распределения. Найти $D_{\text{в}}$.

x_i	n_i
1	20
2	15
3	10
4	5
$\sum$	50

$$\overline{x_{\text{в}}} = \frac{\sum x_i n_i}{n} = \frac{1 \cdot 20 + 2 \cdot 15 + 3 \cdot 10 + 4 \cdot 5}{50} = 2$$

$$D_{\text{в}} = \frac{\sum n_i (x_i - \overline{x_{\text{в}}})^2}{n} = \frac{20 \cdot (1-2)^2 + 15 \cdot (2-2)^2 + 10 \cdot (3-2)^2 + 5 \cdot (4-2)^2}{50} = 1$$

3.2.3. Выборочное среднеквадратическое отклонение

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}$$

⊗ Пусть за один час четверо рабочих сделали 25, 21, 30, 28 деталей соответственно. Вычислить среднюю производительность труда на участке и среднеквадратическое отклонение выборки.

$$\overline{x_{\text{в}}} == \frac{25 + 21 + 30 + 27}{4} = 26$$

$$D_{\text{в}} = \frac{(-2)^2 + (-5)^2 + 4^2 + 2^2}{4} = 11,5$$

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = 3,4$$

3.2.4. Мода

3.2.5. Медиана

Медиана Me — варианта, делящая вариационный ряд на 2 части, равные по длине.

- Если число вариант нечетно: $\otimes$ 1, 2, 5, **7**, 10, 13, 18, то $Me = 7$.
- Если число вариант четно: $\otimes$ 1, 2, **5**, **7**, 10, 13, то $Me = \frac{1}{2} \cdot (5 + 7) = 6$.

3.3. Методы расчета сводных характеристик выборки

3.3.1. Условные варианты

Упрощенные методы расчета сводных характеристик выборки основаны на замене первоначальных вариант условными.

Условными называют варианты, определяемые равенством

$$u_i = \frac{x_i - c}{h}$$

где

- c — ложный нуль (новое начало отсчета);
- h — шаг, т. е. разность между любыми двумя соседними первоначальными вариантами (новая единица масштаба).

В качестве ложного нуля обычно применяют моду Mo.

$\otimes$ Найти условные варианты распределения.

x_i	n_i	u_i
23,6	5	-2
28,6	20	-1
33,6	50	0
38,6	15	1
43,6	10	2

$h = 5, c = 33,6.$

$\cancel{u_2 = {}^{28,6-33,6} 5 = 1}$

Если верно выбран ложный ноль, то условные моменты считать не надо.

3.3.2. Обычные, и условные эмпирические моменты

Для вычисления сводных характеристик выборки удобно пользоваться эмпирическими моментами.

Обычным эмпирическим моментом порядка k называют среднее значение k-х степеней разностей $x_i - c$: где x, — наблюдаемая варианта, n_i — частота варианты, $n = \sum n_i$ — объем выборки, c— произвольное постоянное число (ложный нуль).

$$M_k = \frac{n_i(x_i - c)^k}{n}$$

В частном случае

- $c = 0$: $M_1 = \overline{x_в}$
- $c = \overline{x_в}$: $M_2 = D_в$

3.3.2.1. Условные эмпирические моменты Вычисление центральных моментов требует довольно громоздких вычислений. Чтобы упростить расчеты, заменяют первоначальные варианты условными.

Условным эмпирическим моментом порядка k называют начальный момент порядка k, вычисленный для условных вариант:

$$M_k^* = \frac{n_i(\frac{x_i - c}{h})^k}{n}$$

В частности,

$$\overline{x_{\text{в}}} = M_1^* \cdot h + c$$

$$D_{\text{в}} = (M_2^* - M_1^{*^2}) \cdot h^2$$

3.3.3. Метод произведений для вычисления выборочных средней и дисперсии

Пользуются расчетной таблицей, которая составляется так:

1. в первый столбец таблицы записывают выборочные (первоначальные) варианты x_i в возрастающем порядке;
2. во второй столбец записывают частоты вариант n_i; складывают все частоты и их сумму (объем выборки n) помещают в нижнюю клетку столбца;
3. в третий столбец записывают условные варианты $u_i = \frac{x_i - c}{h}$, причем в качестве ложного нуля c выбирают варианту, которая расположена примерно в середине вариационного ряда, и полагают h равным разности между любыми двумя соседними вариантами; практически же третий столбец заполняется так: в клетке строки, содержащей выбранный ложный нуль, пишут 0; в клетках над нулем пишут последовательно -1, -2, -3 и т.д., а под нулем — $1, 2, 3$ и т.д.;
4. умножают частоты на условные варианты и записывают их произведения $n_i u_i$ в четвертый столбец; сложив все полученные числа, их сумму $\sum n_i u_i$ помещают в нижнюю клетку столбца;
5. умножают частоты на квадраты условных вариант и записывают их произведения $n_i u_i^2$ в пятый столбец; сложив все полученные числа, их сумму $\sum n_i u_i^2$ помещают в нижнюю клетку столбца;
6. умножают частоты на квадраты условных вариант, увеличенных каждая на единицу, и записывают произведения $n_i(u_i + 1)^2$ в шестой контрольный столбец; сложив все полученные числа, их сумму $\sum n_i(u_i + 1)^2$ помещают в нижнюю клетку столбца.

Предусмотрен контроль: если сумма $\sum n_i(u_i + 1)^2$ окажется равной сумме $n + 2\sum n_i u_i + \sum n_i u_i^2$ то вычисления проведены правильно.

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i(u_i+1)^2$
...	...	...	...	...	...
Σ	...	Σ	...	...	...

$\otimes$

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i(u_i+1)^2$
23,6	5	-2	-10	20	5
28,6	20	-1	-20	20	0
33,6	50	0	0	0	50
38,6	15	1	15	15	60
43,6	10	2	20	40	90
Σ	100	Σ	5	95	205

Проверка: $95 + 5 \cdot 2 + 100 = 205$.

- $M_1^* = \frac{5}{100} = 0,05$, $\overline{x_{\text{в}}} = M_1^* \cdot 5 + 33,6 = 33,85$
- $M_2^* = \frac{95}{100} = 0,95$, $D_{\text{в}} = (M_2^* - M_1^{*^2}) \cdot h^2 = 23,7$, $\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = 4,87$

3.3.4. Эмпирические и выравнивающие (теоретические) частоты. Построение нормальной кривой по опытным данным

1. находят $\overline{x_{\text{в}}}$ и $\sigma_{\text{в}}$, например, по методу произведений;

2. находят ординаты, выравнивающие частоты теоретической кривой по формуле

$$n_i' = \frac{nh}{\sigma_\text{в}} \cdot \varphi(v_i),$$

где n — сумма наблюдаемых частот, $\sigma_\text{в}$ — разность между двумя соседними вариантами: $v_i = \frac{x_i - \overline{x_\text{в}}}{\sigma_\text{в}}$ и $\varphi(x)$ — функция Лапласа.

Выравнивающие частоты находят, чтобы проверить, согласуется ли выбранный закон распределения с данными наблюдений.

Результаты пишут в таблицу:

x_i	n_i	$x_i - \overline{x_\text{в}}$	v_i	$\varphi(v_i)$	n_i'
23,6	5	-10.25	-2.1060	0.0434	4.4616
28,6	20	-5.25	-1.0787	0.2230	22.9061
33,6	50	-0.25	-0.0514	0.3984	40.9305
38,6	15	4.75	0.9760	0.2478	25.4558
43,6	10	9.75	2.0033	0.0536	5.5102

(23.6, 4.4616) (28.6, 22.9061) (33.6, 40.9305) (38.6, 25.4558) (43.6, 5.5102)

В одной координатной системе строят кривые: опытную (n_i) и теоретическую (n_i')

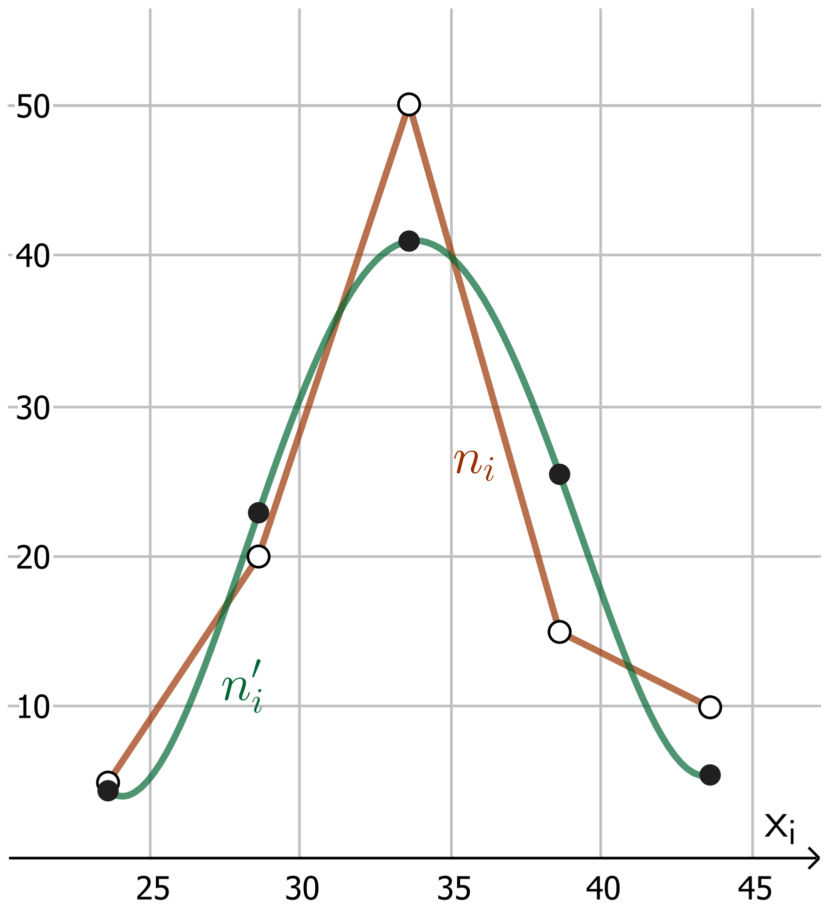


Рис. 15: график рис 22 с 250

3.3.5. Оценка отклонения эмпирического распределения от нормального. Асимметрия и эксцесс

Для оценки отклонения эмпирического распределения от нормального используют характеристики: асимметрию и эксцесс.

Ассиметрия характеризует симметричность нормальной кривой относительно ее вершины. Она положительна, если кривая сдвинута влево; отрицательна, если — вправо (см. рис.).

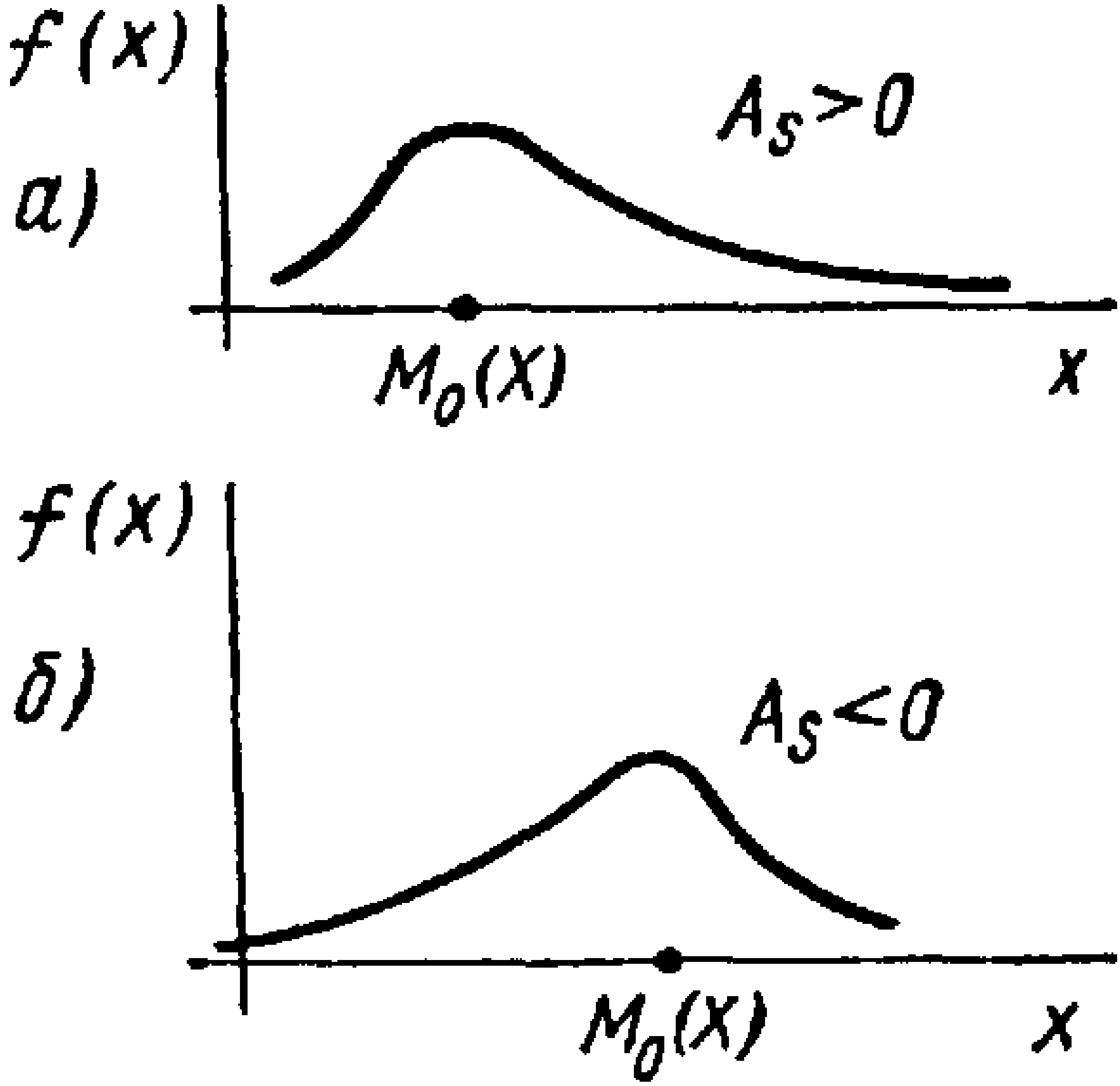


Рис. 16: Ассиметрия

Асимметрия эмпирического распределения определяется равенством

$$a_s = \frac{m_3}{\sigma_{\text{в}}^3}$$

где m_3— центральный эмпирический момент третьего порядка.

$$m_3 = (M_3^* - 3M_1^* M_2^* + 2M_1^{*^3})h^3$$

Эксцесс характеризует расплывание или сужение нормальной кривой. Положителен, если вершина выше и острее; отрицателен, если — ниже и положе (см. рис.).

Эксцесс эмпирического распределения определяется равенством

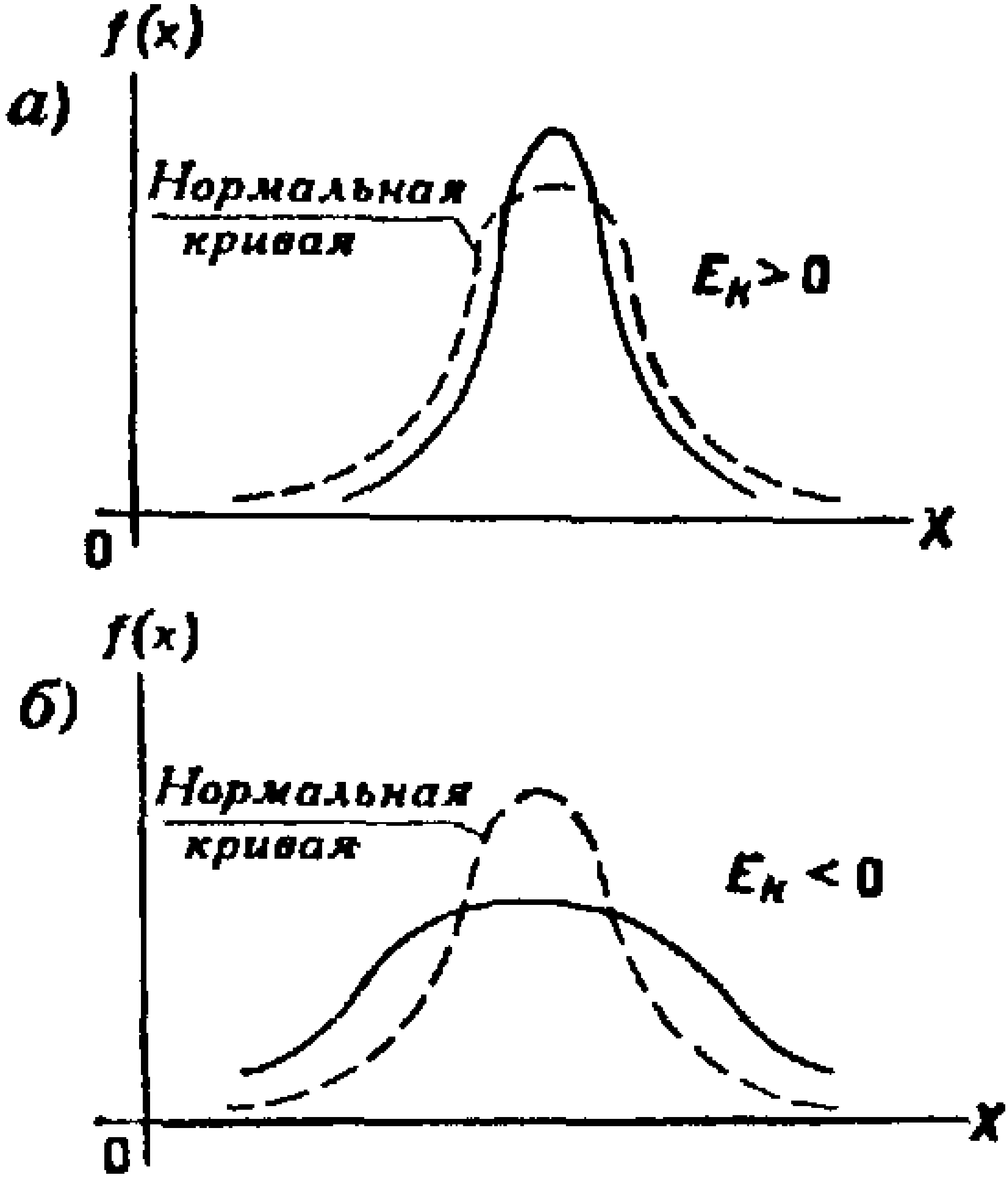


Рис. 17: Эксцесс

$$e_k = \frac{m_4}{\sigma_в^4} - 3$$

, где m_4—центральный эмпирический момент четвертого порядка.

$$m_4 = (M_4^* - 4N1M_3^* + 6M_1^{*^2} - 3M_1^{*^4})h^4$$

3.4. Статистические оценки параметров распределения

3.4.1. Понятие оценки

Генеральная совокупность характеризуется некоторыми постоянными числовыми характеристиками распределения. По выборкам можно найти оценки этих характеристик. Так как выборки случайны, то значение оценок одной числовой характеристики, вычисленной по разным выборкам одной и той же генеральной совокупности могут быть различны.

Обозначим неизвестный параметр распределения, то есть числовую характеристику генеральной совокупности через Θ, а оценку этого параметра T_n. Оценка T_n — функция от выборки. Оценки неизвестного параметра можно находить по выборкам различными способами.

⊗ Если нужно оценить среднее значение $\Theta = \mu$ нормального распределения, то можно воспользоваться следующими оценками:

1. x_1. На практике так и поступают: измеряют величину один раз и этот результат используют как среднее значение этой величины.
2. $\frac{x_{\min}+x_{\max}}{2}$
3. Mo. Для нормального распределения $\mu = Mo$.
4. Me. Для нормального распределения $\mu = Me$.
5. $\overline{x_в}$

Для того, чтобы установить, какая из оценок лучше, надо знать основные свойства и виды оценок.

3.4.2. Несмещенные и эффективные оценки

Для того чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определенным требованиям:

Несмещенной называют статистическую оценку T_n, математическое ожидание которой равно оцениваемому параметру Θ при любом объеме выборки, т. е.

$$M(T_n) = \Theta$$

Иначе оценку называют смещенной и вычисляют смещение:

$$M(T_n) - \Theta$$

⊗ Для НР несмещенной оценкой μ является $\overline{x_в}$.

Однако было бы ошибочным считать, что несмещенная оценка всегда дает хорошее приближение оцениваемого параметра. Действительно, возможные значения T_n могут быть сильно рассеяны вокруг своего среднего значения, т е. дисперсия $D_в$ может быть значительной. Поэтому к статистической оценке предъявляется требование эффективности.

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

⊗ Рассмотренная выше выборочная дисперсия является смещенной оценкой генеральной дисперсии, то есть

$$M(D_в) = \frac{n-1}{n} D_г$$

тогда смещение равно

$$\frac{1}{n}D_{\text{г}}$$

Видно, что при $n \to \infty$ смещение стремится к 0. Значит, при достаточно большом объеме выборки n $D_{\text{в}}$ может приблизительно принимать за несмещенную оценку генеральной дисперсии.

$$D_{\text{г}} \approx D_{\text{в}}$$

Для оценки дисперсии при малом объеме выборки используют исправленную выборочную дисперсию S^2.

$$S^2 = \frac{n}{n-1}D_{\text{в}}$$

S^2 является несмещенной оценкой $D_{\text{г}}$

$$M(S^2) = D_{\text{г}}$$

А исправленное среднеквадратическое отклонение S

$$S = \sqrt{\frac{n}{n-1}D_{\text{в}}}$$

3.4.3. Доверительная вероятность (надежность). Доверительный интервал

Точечной называют оценку, которая определяется одним числом. Все оценки, рассмотренные выше,— точечные.

Все точечные оценки параметров распределения вычисляются по выборкам, но из-за их случайности оценки T_n являются СВ, отличающиеся от истинного значения Θ.

Обозначим точность оценки $\delta > 0$, тогда

$$|\Theta - T_n| \leqslant \delta$$

Чем меньше δ, тем точнее оценка.

Надежностью (доверительной вероятностью) оценки Θ по T_n называют вероятность γ, с которой осуществляется неравенство $|\Theta - T_n| \leqslant \delta$.

$$P(|\Theta - T_n| \leqslant \delta) = \gamma$$
$$P(T_n - \delta \leqslant \Theta \leqslant T_n + \delta) = \gamma$$

Обычно надежность оценки задается наперед, причем в качестве у берут число, близкое к единице. Наиболее часто задают надежность, равную 0,95; 0,99 и 0,999.

3.4.4. Доверительные интервалы для оценки математического ожидания нормального распределения при известном σ

Пусть количественный признак X генеральной совокупности распределен нормально, причем σ этого распределения известно. Требуется оценить неизвестное μ по выборочной средней $\overline{x_{\text{в}}}$. Поставим своей задачей найти доверительный интервал, покрывающий параметр μ с надежностью γ.

Все элементы выборки, имеющие то же распределение, что и ГС

$$x_i \equiv N(\mu i \sigma)$$

$$\overline{x_в} \equiv N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

(глава 8 п 9)

Решаем

$$P(|\overline{x_в} - \mu| \leqslant \delta) = \gamma$$

Из ТВ мы знаем, как найти вероятность того, что СВ, распределенная нормально, принимает значения, отличающиеся от μ не более, чем на δ

$$P(|\overline{x_в} - \mu| \leqslant \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$$

Заменив σ на $\frac{\sigma}{\sqrt{n}}$ можем сказать

$$P(|\overline{x_в} - \mu| \leqslant \delta) = 2\Phi\left(t\right),$$

где $t = \frac{\delta\sqrt{n}}{\sigma}$.

Тогда

$$P(\overline{x_в} - \delta \leqslant \mu \leqslant \overline{x_в} + \delta) = \gamma = 2\Phi\left(t\right),$$

где $\delta = \frac{\sigma t}{\sqrt{n}}$.

То есть с надежностью γ можно утверждать, что доверительный интервал

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

заключает в себя неизвестный параметр μ, где t — решение уравнения $\gamma = 2\Phi\left(t\right)$

$\otimes$ X имеет НР с $\sigma = 3$. Найти доверительный интервал для оценки μ по $\overline{x_в} = 22,6$, если $n = 36$, $\gamma = 0,95$.

$2\Phi(t) = \gamma$

$\Phi(t) = 0,475$

$t = 1,96$

$\delta = \frac{1,96 \cdot 3}{6} = 0,98$

$22,6 - 0,98 \leqslant \mu \leqslant 22,6 + 0,98$

$21,62 \leqslant \mu \leqslant 23,58$

3.4.5. Доверительные интервалы для оценки математического ожидания нормального распределения при неизвестном σ

Пусть количественный признак X генеральной совокупности распределен нормально, причем σ этого распределения известно. Требуется оценить неизвестное μ с помощью доверительных интервалов.

Оказывается, что по данным выборки можно построить случайную величину T,ее возможные значения будем обозначать через t:

$$T = \frac{\overline{x_в} - \mu}{S/\sqrt{n}},$$

которая имеет распределение Стьюдента; здесь

$\overline{x_в}$ — выборочная средняя, S — «исправленное» среднее квадратическое отклонение, n — объем выборки.

Плотность этого распределения описывается формулой, зависящей только от объема выборки.

Пользуясь этим распределением, определим доверительный интервал для μ.

$$\left(\overline{x_в} - \frac{st_\gamma}{\sqrt{n}}; x_в + \frac{st_\gamma}{\sqrt{n}}\right)$$

$t_\gamma = t(n; \gamma)$.

$\otimes$ Количественный признак распределен нормально. По выборке с $n = 16$ найдено $\overline{x_в} = 20,2$ и $S = 0,8$. Оценить μ с помощью довериьтельного интервала с $\gamma = 0,95$.

$t(16; 0,95) = 2,13$

$\left(20,2 - \frac{2,13 \cdot 0,8}{4}; 20,2 + \frac{2,13 \cdot 0,8}{4}\right)$

$19,77 \leqslant \mu \leqslant 20,53$

3.4.6. Доверительные интервалы для оценки среднего квадратического отклонения σ нормального распределения

Пусть количественный признак X генеральной совокупности распределен нормально. совокупности распределен нормально. Требуется оценить неизвестное генеральное среднее квадратическое отклонение σ по «исправленному» выборочному среднему квадратическому отклонению s. Поставим перед собой задачу найти доверительные интервалы, покрывающие параметр σ с заданной надежностью γ.

Потребуем

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

$$s(1 - \delta/s) < \sigma < s(1 + \delta/s)$$

Положим $\delta/s = q$, получим

$$s(1 - q) < \sigma < s(1 + q)$$

q находим по таблице распределения Пирсона

$$q = q(n, \gamma)$$

$\otimes$ Количественный признак X распределен нормально. По выборке с $n = 25$ найдено $S = 0,8$. Найти доверительный интервал, включающий σ с $\gamma = 0,95$.

$q(25; 0,95) = 0,32\ 0,8 \cdot 0,68 \leqslant \sigma \leqslant 0,8 \cdot 1,32$

$0,544 \leqslant \sigma \leqslant 1,056$

Замечание. Если $q > 1$, то доверительный интервал односторонний:

$$0 \leqslant \sigma \leqslant s(1 + q)$$

То есть

$$\max(0; s(1 - q)) \leqslant \sigma \leqslant s(1 + q)$$