

# Некоторые законы распределения случайных величин

## Нормальный закон распределения («закон Гаусса»)

Плотность вероятностей

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad X \sim N(a, \sigma)$$

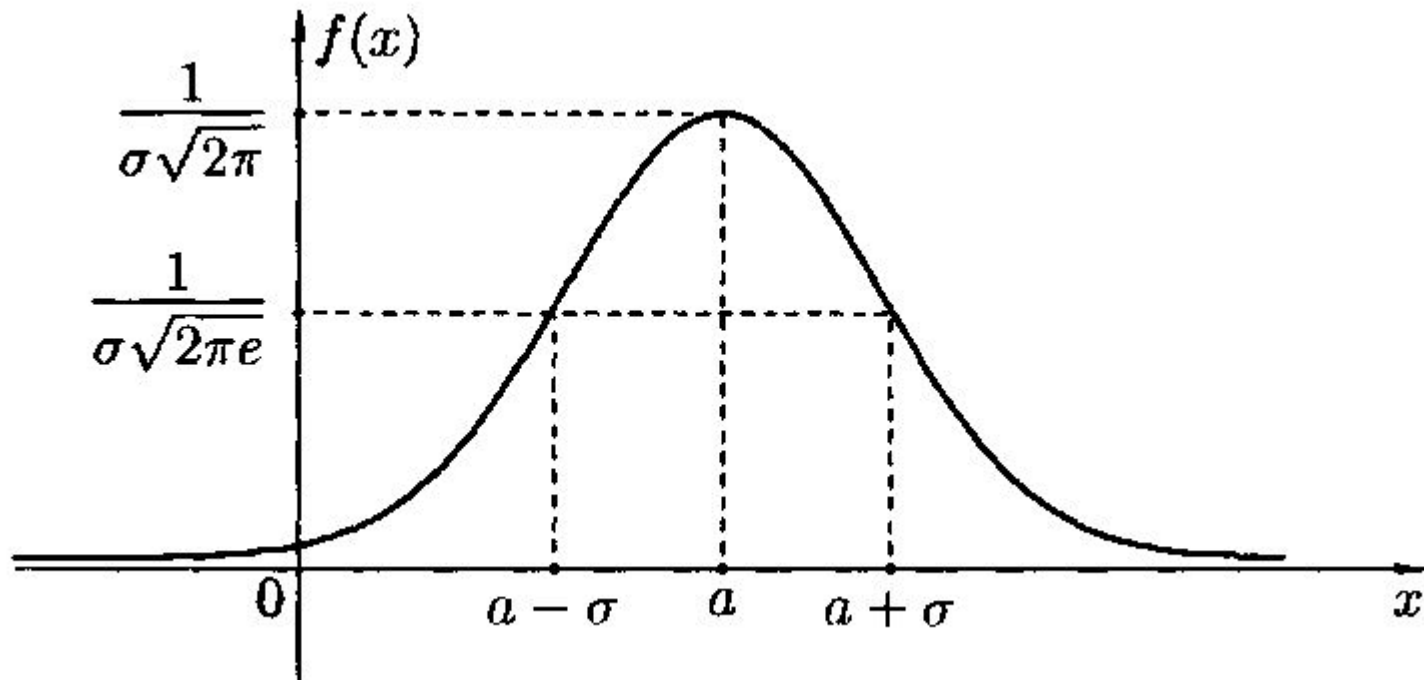
Свойства плотности вероятностей

1.  $f(x) > 0$  при любом  $x \in (-\infty, +\infty)$

2. Ось абсцисс является горизонтальной асимптотой

$$\lim_{x \rightarrow \mp\infty} f(x) = 0$$

## График нормального закона



$$f_{\max} = f(a) = \frac{1}{\sigma\sqrt{2\pi}}$$

Максимальное значение

Точки перегиба

$$M_1 \left( a - \sigma, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}} \right) \quad \text{и} \quad M_2 \left( a + \sigma, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}} \right)$$

# Характеристическая функция гауссовской случайной величины

$$g(v) = e^{iva - \frac{\sigma^2 v^2}{2}}$$

Линейное преобразование гауссовской случайной величины  $Y = cX + b$

$$Y \sim N(ca + b, |c|\sigma)$$

Сумма  $X + Y$  двух независимых гауссовских случайных величин  $X \sim N_X(a_1, \sigma_1)$  и  $Y \sim N_Y(a_2, \sigma_2)$

$$X + Y \sim N_{X+Y}(a_1 + a_2, \sigma_1 + \sigma_2)$$

Центральные моменты гауссовской случайной величины  $\mu_k[X]$

Нечетные моменты  $\mu_k[X] = 0$  ( $A=0$  - коэффициент асимметрии)

Четные моменты  $\mu_k[X] = (k-1)!! \sigma^k$

В частности:  $\mu_2[X] = D(X) = \sigma^2$

$$\mu_4[X] = D(X) = 3\sigma^4 \quad E = \frac{\mu_4[X]}{\sigma^4} - 3 = 0$$

$$\mu_6[X] = 15\sigma^6$$

Вычисление вероятности  $P(\alpha < X < \beta)$

Функция Лапласа  $\Phi(x) = \int_0^x e^{-\frac{t^2}{2}} dt$

$$\Phi(+\infty) = 0,5$$

$$\Phi(-x) = -\Phi(x)$$

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$$

$$P(|X - a| < \delta) = (\delta > 0) = P(a - \delta < X < a + \delta) =$$

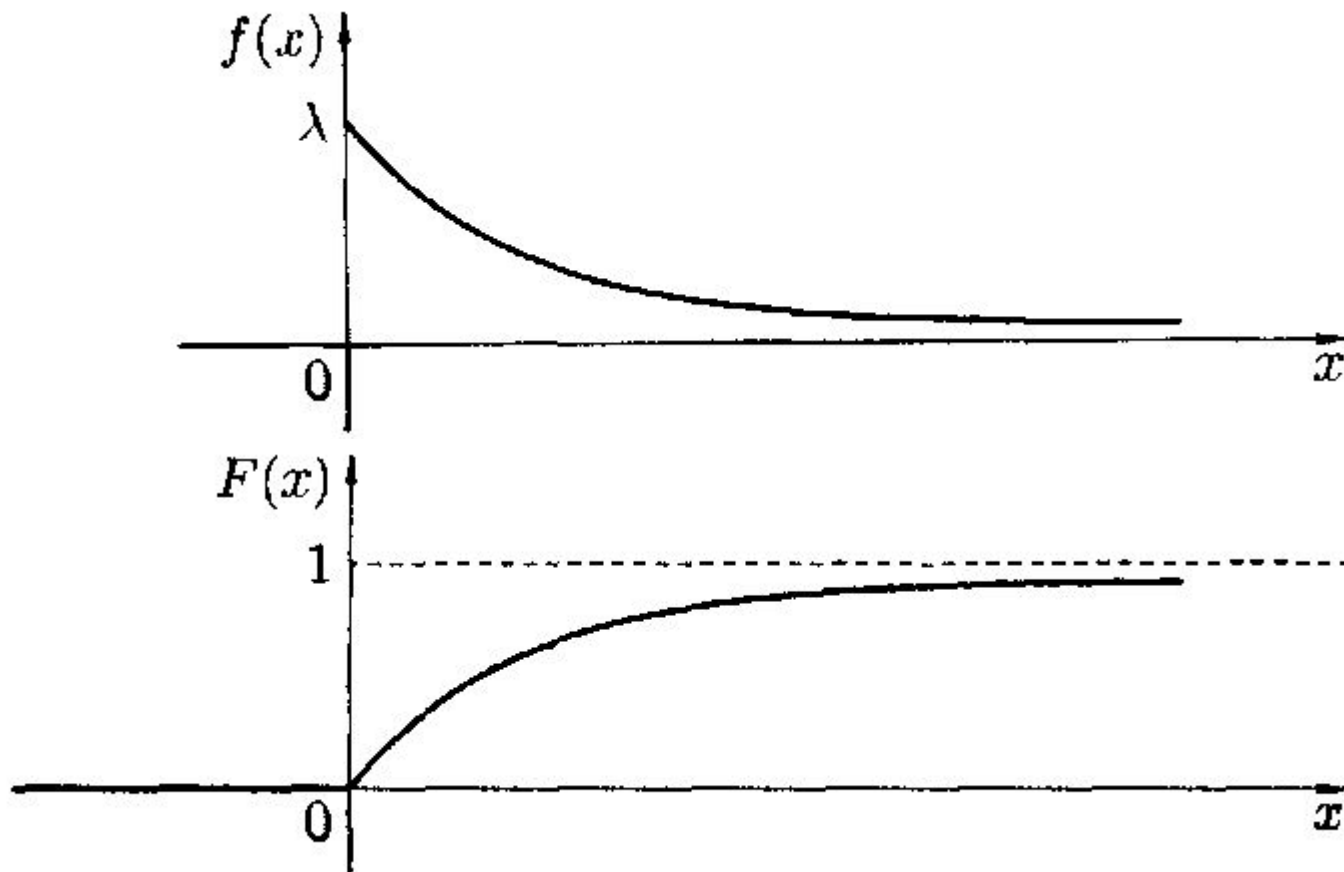
$$= \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right) = 2\Phi\left(\frac{\delta}{\sigma}\right)$$

$$\text{При } \delta = 3\sigma \quad P(|X - a| < 3\sigma) = 0,9973$$

**Пример.** При сортировке случайные значения веса зерна распределены нормально со средним значением 0,15 г среднеквадратическим отклонением 0,03 г. Нормальные всходы дают зерна, вес которых более 0,01 г. Определить процент семян, от которых следует ожидать нормальные всходы.

## Показательное (экспоненциальное) распределение.

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



Характеристическая функция

$$g(v) = \left(1 - \frac{iv}{\lambda}\right)^{-1}$$

Кумулянтная функция

$$\varphi(v) = -\ln\left(1 - \frac{iv}{\lambda}\right)$$

$$m_1[X] = -i\varphi'(0) \qquad m_1[X] = \frac{1}{\lambda}$$

$$D(X) = -\varphi''(0) \qquad D(X) = \frac{1}{\lambda^2}$$



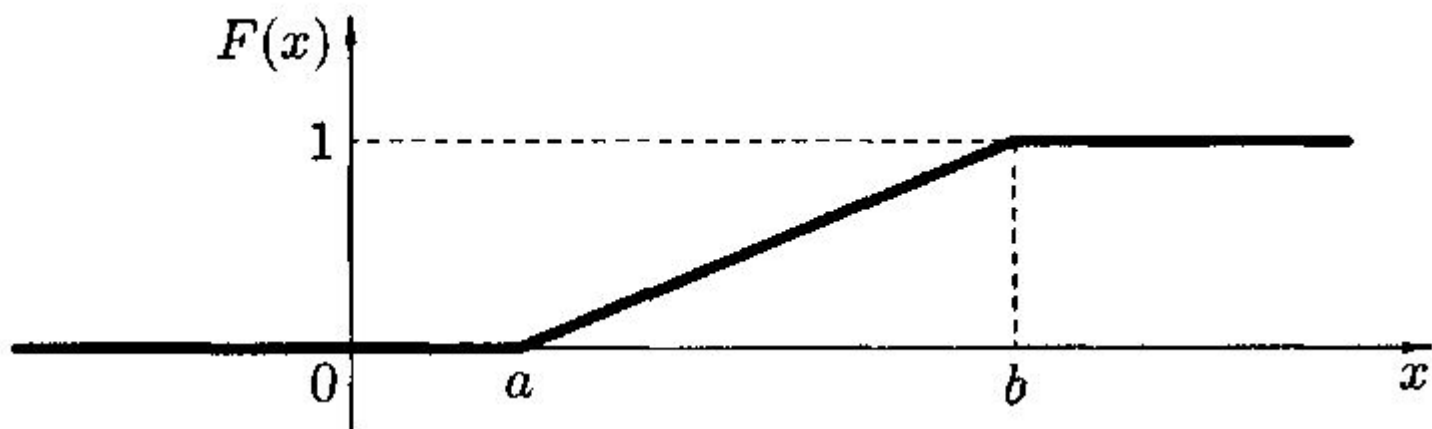
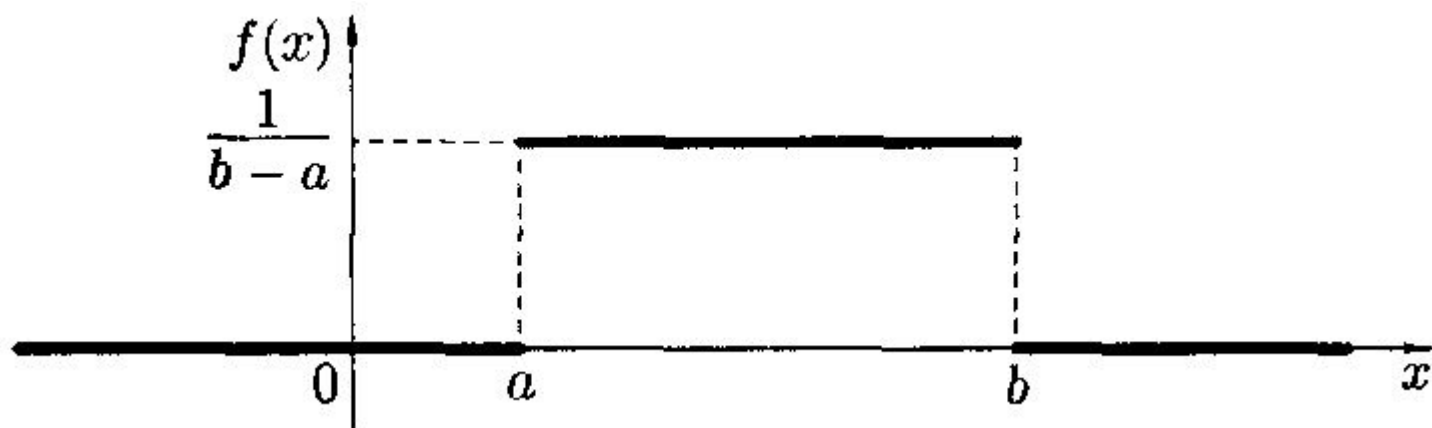
## Равномерное распределение.

$$f(x) = \begin{cases} \frac{1}{b-a}, x \in (a, b) \\ 0, x \notin (a, b) \end{cases}$$

$$m_1[X] = \frac{a+b}{2}$$

$$D[X] = \frac{(b-a)^2}{12}$$

$$g(v) = -i \frac{e^{iba} - e^{iav}}{v(b-a)}$$



## Распределение Пуассона

$$P(m, \Delta t) = \frac{(\lambda \Delta t)^m}{m!} e^{-\lambda \Delta t}$$

## Биномиальный закон распределения

$$P_n(m) = C_n^m p^m (1-p)^{n-m}$$

## Центральная предельная теорема

**Пример.** Если  $X_1, X_2, \dots, X_n$  независимые случайные величины, имеющие один и тот же закон распределения с нулевым математическим ожиданием и дисперсией  $\sigma^2$ , то при увеличении числа  $n$  закон распределения суммы

$$Y_n = \sum_{i=1}^n X_i$$

неограниченно приближается к нормальному

# Основные понятия математической статистики

Термин статистика происходит от латинского слова «статус» -состояние.

В настоящее время статистика включает в себя следующие разделы:

1. Сбор статистических сведений, характеризующих отдельные составляющие каких-либо массовых совокупностей;
2. Статистическое исследование полученных данных, заключающееся в выяснении тех закономерностей, которые могут быть установлены на основе массовых наблюдений.

3. Разработка приемов статистического наблюдения и анализа статистических данных. Этот раздел составляет основное содержание математической статистики.

На основе полученных статистических данных можно решать следующие задачи:

1. Оценивать значения неизвестной вероятности случайного события.

2. Определить неизвестные функции распределения или моменты случайной величины  $X$ .

В результате  $n$  независимых наблюдений СВ  $X$  получены следующие ее значения  $x_1, x_2, \dots, x_n$ . Требуется определить, хотя бы приближенно, неизвестную функцию распределения случайной величины  $X$  или ее моменты (например, среднее, дисперсия).

### 3. Определение неизвестных параметров распределения.

(Часто исходя из некоторых соображений можно сделать заключение о типе функции распределения интересующей нас СВ. Тогда задача сводится к нахождению неизвестных параметров)

Примеры.

а. Пуассоновский поток  $\lambda$  - ?

б. Гауссовское распределение  $a$  и  $\sigma$  - ?;  $a$  - ?,  $\sigma$  – известно;  $\sigma$  - ?  $a$  – известно.

в. Экспоненциальное распределение -  $\lambda$  ?

### 4. Оценка зависимости

Производится последовательность наблюдений сразу двух СВ  $X$  и  $Y$ . В результате наблюдений получаем пары значений  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

Требуется выяснить наличие функциональной или корреляционной связи между  $X$  и  $Y$ .

## Понятие выборки

**Определение.** Совокупность всех подлежащих изучению результатов всех мыслимых наблюдений, производится над каким-то объектом, называется генеральной совокупностью

**Определение.** Выборка называется репрезентативной (представительной), если она хорошо представляет свойства генеральной совокупности

В результате эксперимента над случайной величиной  $X$  мы получаем множество ее значений  $x_1, x_2, \dots, x_n$  как результатов  $n$  наблюдений.



Множество  $\{x_1, x_2, \dots, x_n\}$  отдельных значений СВ распределенных по неизвестному нам, но одинаковому закону  $F(x)$  называется выборкой объема  $n$  из генеральной совокупности.

Числа  $x_i$  - элементы выборки или варианты

## Форма записи выборки

$x_1' \leq x_2' \leq x_3' \leq \dots \leq x_n'$  - вариационный ряд

$x_n' - x_1'$  - размах выборки

			...		
Номер группы ( $i$ )	1	2	...	k-1	
		2	...	1	
			...		

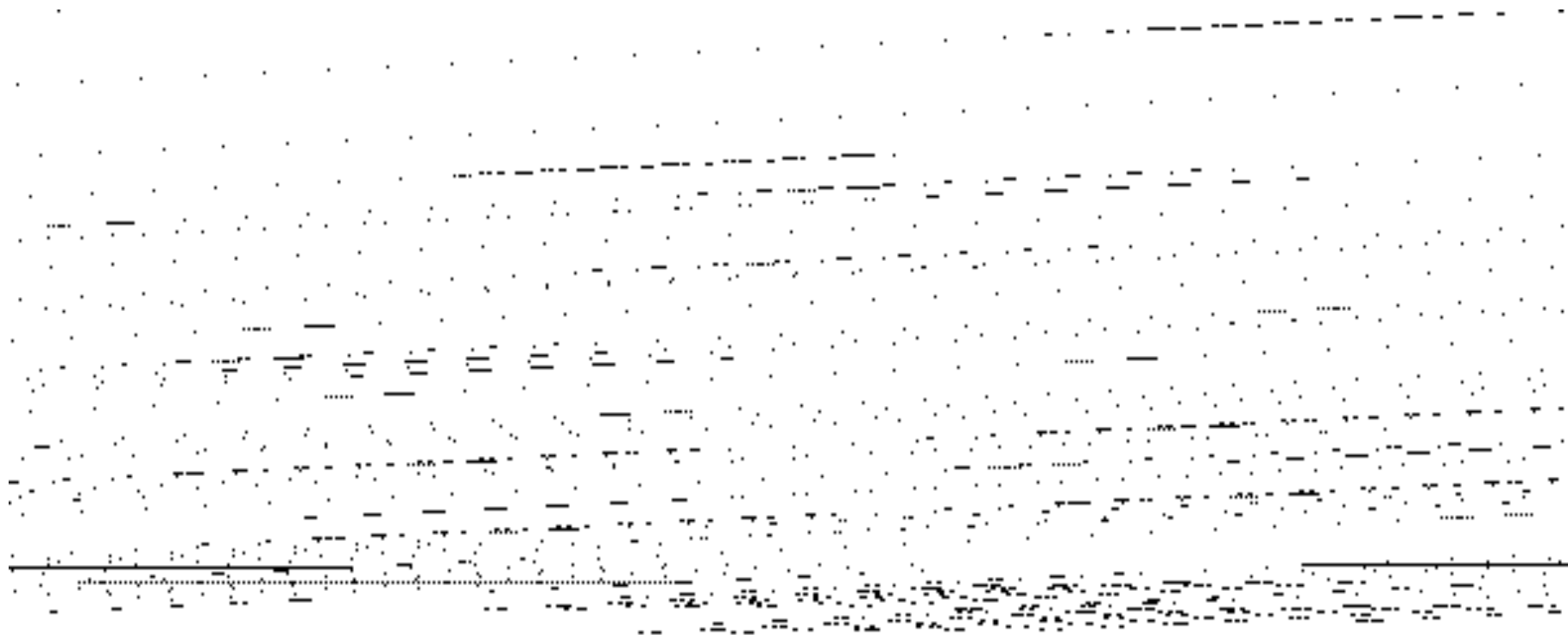
**Пример 1.** При измерениях частоты пульса в однородных группах обследуемых получены следующие результаты: 71, 72, 74, 70, 70, 72, 71, 74, 71, 72, 73, 71, 72, 72, 72, 73, 72, 74, 72, 74. Составить по этим результатам статистический ряд распределения частот и относительных частот.

Объем выборки:  $n = 2 + 4 + 8 + 2 + 4 = 20$

Ряды распределения частот и относительных частот

Пульс			72		
		4	8	2	
			0,4		

# Полигон частот и полигон относительных частот



**Пример 2.** При измерениях роста в однородных группах обследуемых получены следующие результаты:

178, 160, 154, 183, 155, 153, 167, 186, 163, 155,  
157, 175, 170, 166, 159, 173, 182, 167, 171, 169,  
179, 165, 156, 179, 158, 171, 175, 173, 164, 172.

Составить по этим результатам группированный статистический ряд распределения частот и относительных частот. Построить гистограмму и полигон относительных частот

$$x_{min} = 153$$

$$x_{max} = 186$$

Формула Стерджеса:  $h = \frac{x_{max} - x_{min}}{1 + \log_2 n}$

$$\text{В нашем случае: } h = \frac{186 - 153}{1 + \log_2 30} \approx 5,59$$

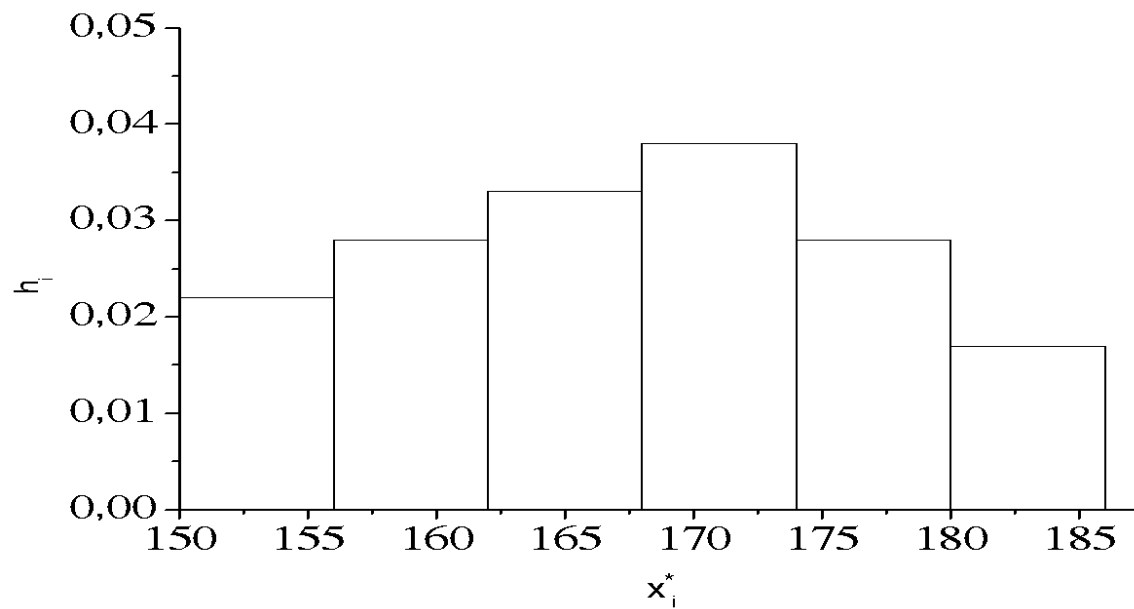
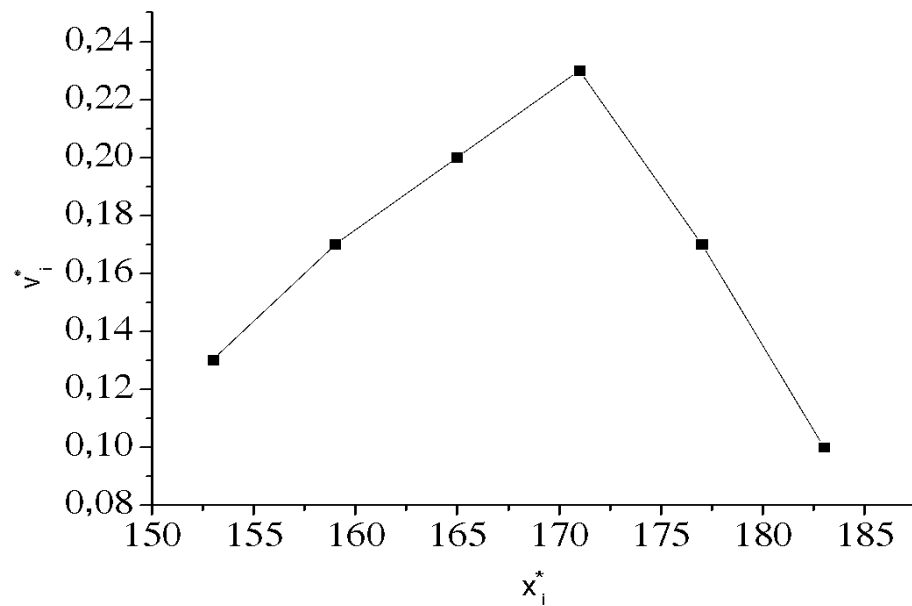
Примем  $h = 6$        $x_{\text{нач}} = 150$

Исходные данные разобьем на 6 интервалов:  $[150,156)$ ,  $[156,162)$ ,  $[162,168)$ ,  $[168,174)$ ,  $[174,180)$ ,  $[180,186]$

Определим число выборочных данных  $n_i$ , попавших в каждый промежуток

	[150,156)	[156,162)	[162,168)	[168,174)	[174,180)	[180,186]
	153	159	165	171	177	183
Частота	4	5	6	7	5	3
	0,13	0,17	0,20	0,23	0,17	0,1
	0,022	0,028	0,033	0,038	0,028	0,017

$H_i$  высота прямоугольника гистограммы



## Эмпирическая функция распределения

**Определение.** Характеристики СВ, найденные на основе выборочных данных называются эмпирическими или выборочными

Пусть есть выборка  $x_1, x_2, \dots, x_n$ ,  $n_x$  – число элементов выборки, меньших  $x$ .

**Определение.** Эмпирической функцией распределения называется

$$F_x^* = \frac{n_x}{n}$$

Свойства  $F_x^*$  :

1.  $0 \leq F_x^* \leq 1$

2.  $F_x^*$  неубывающая функция

3. Если  $x_1$  наименьшая, а  $x_k$  наибольшая варианты, то

$$F_x^* = \begin{cases} 0, & x < x_1 \\ 1 & x > x_k \end{cases}$$

<b>Пультс</b>			72		
		4	8	2	
			0,4		



$$F_x^* = \begin{cases} 0, & x \leq 70 \\ 0,3 & 71 < x \leq 72 \\ 0,7 & 72 < x \leq 73 \\ 0,8 & 73 < x \leq 74 \\ 1, & x > 75 \end{cases}$$

**Определение.** Пусть дана последовательность СВ  $\{x_n\}$ .

Говорят, что последовательность  $\{x_n\}$  сходится по вероятности к числу  $a$ , если для любых чисел  $\varepsilon > 0$  и  $\delta > 0$  найдется такое число  $N(\varepsilon, \delta)$  зависящее от  $\varepsilon$ ,  $\delta$ , что для всех  $n > N(\varepsilon, \delta)$  выполняется неравенство

$$P(|x_n - a| < \varepsilon) > 1 - \delta$$

**Теорема.** Если  $X_1, X_2, \dots, X_n$  последовательность попарно независимых СВ с конечными математическими ожиданиями  $m_1, m_2, \dots, m_n$  и дисперсиями  $D_1, D_2, \dots, D_n$  ограниченными одним и тем же числом  $Q$  ( $D_i < Q$   $i = 1, \dots, n$ ) то последовательность СВ

$$Y_n = \frac{\sum_{i=1}^n X_i}{n}$$

сходится по вероятности к среднему арифметическому математических ожиданий

$$\frac{\sum_{i=1}^n m_i}{n}$$

величин  $X_1, X_2, \dots, X_n,$

$L$

что означает

$$P \left( \left| \frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n m_i}{n} \right| < \varepsilon \right) > 1 - \delta$$

**Следствие1.** (Теорема Бернулли). Пусть производится  $n$  независимых опытов, в каждом из которых может появиться некоторое событие  $A$  с постоянной вероятностью  $p$ .

При неограниченном увеличении числа опытов  $n$  относительная частота  $p^*$  появления события  $A$  сходится по вероятности к  $p$ .

**Следствие2.** Пусть  $X_1, X_2, \dots, X_n$  - последовательность независимых СВ распределенных по одному закону, имеющему конечное математическое  $m_X$  ожидание и конечную дисперсию  $D_X$ .

Тогда среднее арифметическое этих величин

$$Y_n = \frac{\sum_{i=1}^n X_i}{n}$$

сходится по вероятности к  $m_X$ .

**Замечание.** Аналогично можно доказать, что любой выборочный момент  $k$ -го порядка

$$\frac{1}{n} \sum X_i^k = \hat{m}_k$$

сходится по вероятности к соответствующему  $k$ -му моменту  $m_k[X]$  исходного распределения, если только существует момент порядка  $2k$  этого распределения.

### **Оценки для неизвестных параметров закона распределения.**

Необходимо отметить, что любое значение искомого параметра, вычисленное на основе ограниченной выборки, будет содержать элементы случайности.

Это приближенное случайное значение мы будем называть оценкой параметра.

Пусть имеется случайная величина  $X$ , закон распределения которой содержит неизвестный параметр  $a$ , т.е.  $f(x; a)$ .

Требуется по выборке  $x_1, x_2, \dots, x_n$ , полученной в результате  $n$  независимых наблюдений, оценить неизвестный параметр  $a$ .

Обозначим  $\hat{a}$  оценку для параметра  $a$ .

$$\hat{a} = \varphi(X_1, X_2, \dots, X_n)$$

$\hat{a}$  является функцией результатов  $n$  наблюдений над случайной величиной  $X$ .

$\hat{a}$  случайная величина

Предъявим к оценке ряд требований, которым она должна удовлетворять, чтобы быть доброкачественной:

1. Оценка должна быть состоятельной. Оценка называется состоятельной, если при неограниченном увеличении объема выборки  $n$  она сходится по вероятности к оцениваемому параметру

$$\hat{a} = \varphi(X_1, X_2, \dots, X_n) \xrightarrow[n \rightarrow \infty]{P} a$$

2. Оценка должна быть несмещенной, т.е. ее математическое ожидание должно совпадать с оцениваемым параметром

$$m_1[\hat{a}] = a$$

$$m_1[\hat{a}] - a = b_n(a) \text{ смещение}$$

Если

$$\lim_{n \rightarrow \infty} b_n(a) = 0$$

то оценка называется асимптотически несмещенной.

3. Оценка  $\hat{a}$  должна быть эффективной, т.е. она должна обладать по сравнению с другими наименьшей дисперсией

$$D(\hat{a}) = V_{min}$$

$V_{min}$ - минимально возможная величина дисперсии, определяемая из неравенства Крамера Рао и называемая потенциальной или предельной точностью.

Для характеристики точности оценки на практике обычно используют величину

$$\delta(\hat{a}) = \sqrt{D(\hat{a})} / a$$

относительная среднеквадратическая погрешность

Отношение  $V_{min}/D(\hat{a}) \leq 1$  называется эффективностью



# Методы нахождения точечных оценок

Наиболее распространенные методы построения точечных оценок:

метод моментов,

метод максимального правдоподобия

метод максимума апостериорной плотности вероятности оцениваемого параметра и т.д.

## Метод моментов

Выборочный момент  $k$ -го порядка определяется по формуле

$$m_k^* = \frac{1}{n} \sum_{i=1}^n x_i^k$$

$x_1, x_2, \dots, x_n$  выборочные данные.

$m_1^*$  — выборочное среднее

Аналогично определим центральные выборочные моменты

$$\mu_k^* = \frac{1}{n} \sum_{i=1}^n (x_i - m_1^*)^k$$

$\mu_2^*$  — центральный выборочный момент второго порядка или выборочная дисперсия  $D_x^*$ . Характеризует меру рассеяния выборочных значений относительно выборочного среднего.

Выборочный коэффициент асимметрии

$$A^* = \frac{\mu_3^*}{(D_x^*)^{3/2}}$$

Выборочный коэффициент эксцесса

$$E^* = \frac{\mu_4^*}{(D_x^*)^2} - 3$$

Рассмотрим свойства выборочных среднего и дисперсии

$$m_1^* = \frac{1}{n} \sum_{i=1}^n x_i$$

$$D_x^* = \frac{1}{n} \sum_{i=1}^n (x_i - m_1^*)^2$$

По следствию из теореме Чебышева  $m_1^*$  сходится по вероятности к  $m_x$ , т.е. является состоятельной

Оценка  $m_1^*$  также является несмещенной

Дисперсия оценки  $m_1^*$

$$D(m_1^*) = \frac{1}{n} D_x$$

Оценка дисперсии  $D_x^* = \frac{1}{n} \sum_{i=1}^n (x_i - m_1^*)^2$

Предложенная оценка является состоятельной

Можно показать, что  $m_1(D_x^*) = \frac{n-1}{n} D_x$

Оценка дисперсии  $D_x^*$  является смещенной

Определим исправленную выборочную дисперсию

$$S^2 = \frac{n}{n-1} D_x^* = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_1^*)^2$$

**Пример.** При измерениях частоты пульса в однородных группах обследуемых получены следующие результаты:  
71, 72, 74, 70, 70, 72, 71, 74, 71, 72, 73, 71, 72, 72, 72, 73, 72,  
74, 72, 74 . Найти выборочные среднее и дисперсию.

$$m_1^* = \frac{1}{n} \sum_{i=1}^n x_i = 72,1$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_1^*)^2 = 1,21$$

Смещенная оценка  $D_x^* = \frac{1}{n} \sum_{i=1}^n (x_i - m_1^*)^2 = 1,15$

# Понятие интервального оценивания параметров. Доверительный интервал.

Назначим некоторую достаточно большую вероятность  $\beta=0.9, 0.95$  или  $0.99$  с которой будет выполняться

событие  $|\hat{a} - a| < \varepsilon$

$$P(|\hat{a} - a| < \varepsilon) = \beta$$

Диапазон максимальных возможных значений ошибки, возникающих при замене  $a$  на  $\hat{a}$  будет равен  $\mp\varepsilon$ .

Интервал  $(\hat{a} - \varepsilon, \hat{a} + \varepsilon)$  называется доверительным интервалом

Нахождение границ доверительного интервала

В качестве примера рассмотрим задачу о доверительном интервале для оценки среднего

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n X_i$$

Найдем вероятность  $P(|\hat{m}_1 - m_x| < \varepsilon) = \beta$

$$\beta = 2\Phi\left(\frac{\varepsilon}{\sigma(\hat{m}_1)}\right) \quad \sigma(\hat{m}_1) = \sqrt{\frac{D_x}{n}}$$

Если дисперсия  $D_x$  не известна, то можно использовать ее оценку

$$D^*_x = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m}_1)^2$$

**Пример.** Дана гауссовская случайная величина  $X$  со среднеквадратическим отклонением  $\sigma=2$ . Найти доверительный интервал для математического ожидания  $a$ , если  $\hat{a} = 20,09, n = 16, \beta = 0,9$ .

$$\chi^2_{\text{набл}}$$



## Критерий согласия Пирсона

На основании выборки  $x_1, x_2, \dots, x_n$  с помощью критерия Пирсона можно проверить гипотезу  $H$  о том, что случайная величина  $X$  имеет некоторый закон распределения  $F(x)$

Введем некоторую величину  $\chi^2_{\text{набл}}$ , характеризующую степень расхождения теоретического и эмпирического законов распределения.

Будем сравнивать  $\chi^2_{\text{набл}}$  с некоторым порогом  $P_\alpha$ :

Если  $\chi^2_{\text{набл}} < P_\alpha$  то гипотезу  $H$  будем принимать

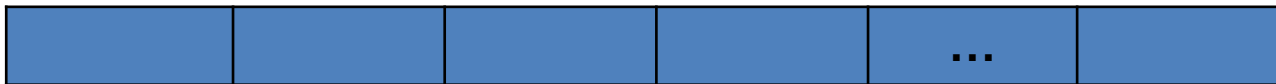
Если  $\chi^2_{\text{набл}} > P_\alpha$  то гипотезу  $H$  будем отклонять

# Группированный статистический ряд

				...	
				...	

$f(x)$  плотность вероятностей  
гипотетического закона  
распределения

$$p_i = \int_{\Delta_i} f(x) dx$$



Мера расхождения между гистограммой и  
 $f(x)$

$$\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

$$n = \sum n_i$$

объем выборки



