ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Распределение изучаемого признака ***в выборке*** называется ***эмпирическим*** или ***выборочным***, а во всей ***генеральной совокупности - теоретическим.***

Для изучения **эмпирического распределения** используется ***вариационный ряд,*** а для изучения **теоретического распределения - *закон (функция) распределения.***

Одна из задач вариационного анализа - *установление вида функции (закона)* теоретического распределения, которому соответствует эмпирическое распределение выборочной совокупности.

Законом или функцией распределения случайной величины называется соотношение, устанавливающее связь между значениями случайной величины и соответствующими им вероятностями.

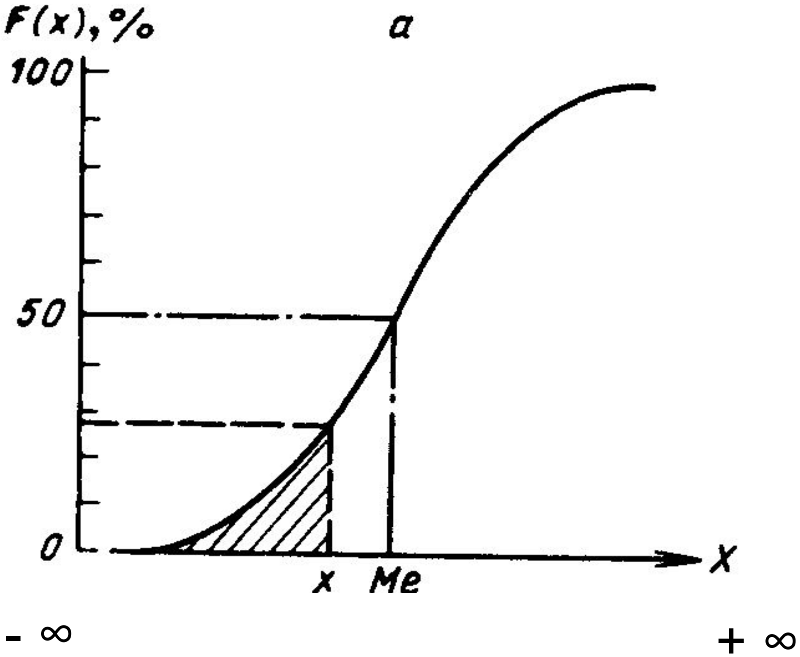
Закон распределения является *наиболее полной и исчерпывающей* характеристикой генеральной совокупности, также как вариационный ряд для выборочной совокупности.

Закон распределения, как и вариационный ряд, может быть задан в трех формах:

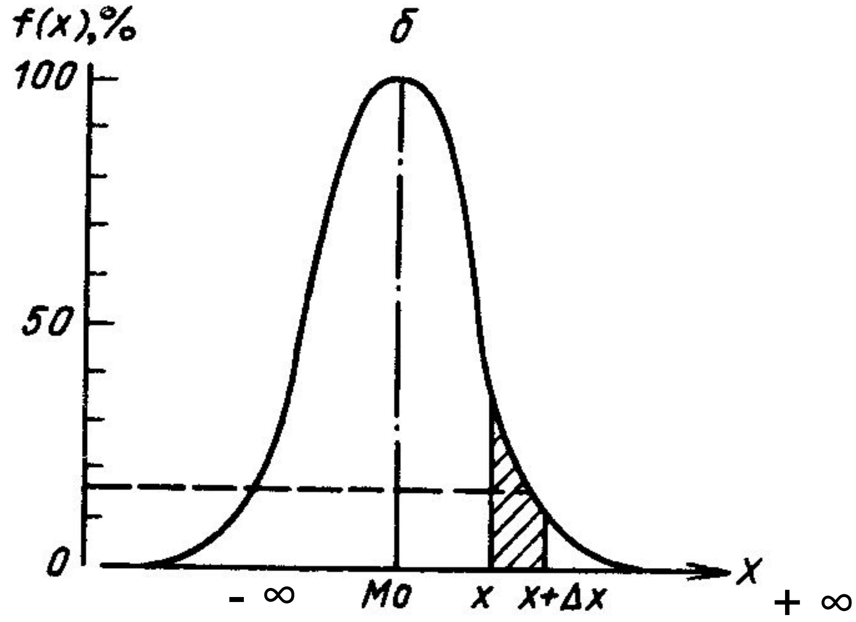
табличной,

графической

аналитической



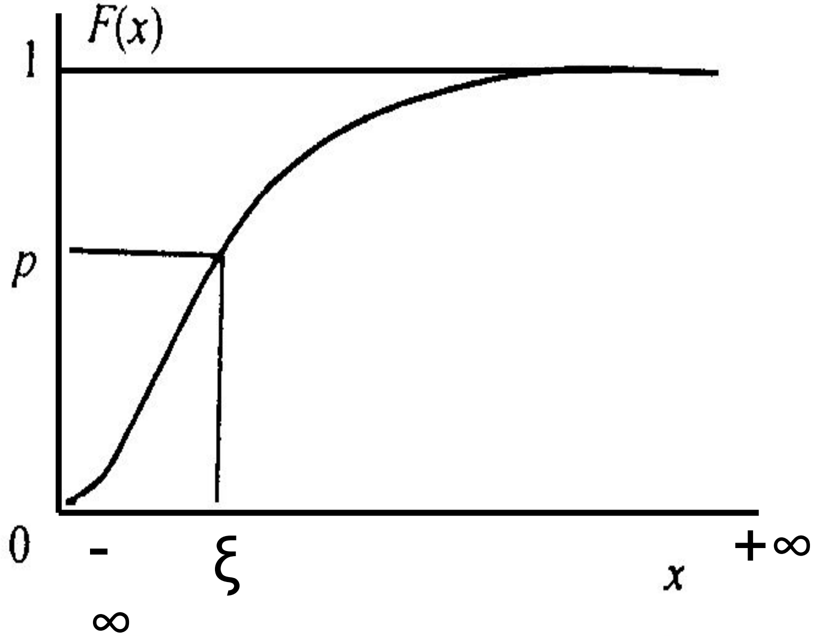
Интегральная функция (закон) распределения



Дифференциальная функция (закон) плотности распределения

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

Интегральная функция распределения -



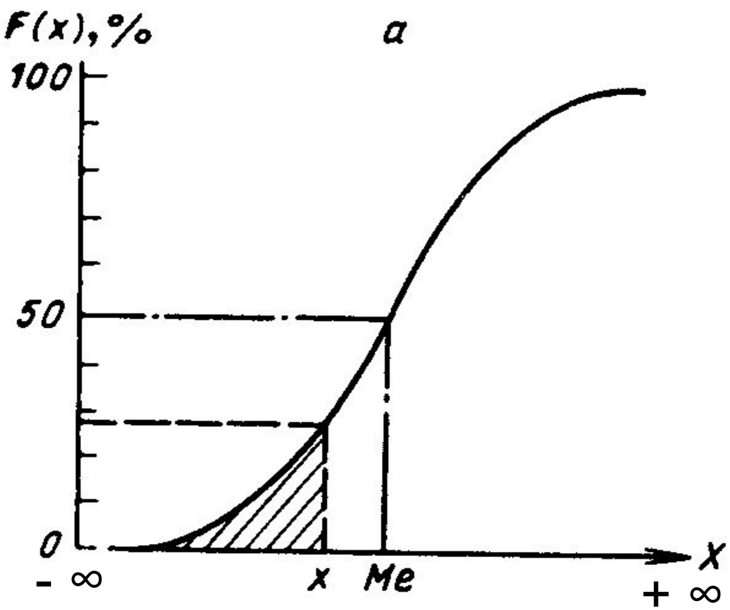
F(x), или просто *__функция распределения__ F(x)*, выражает вероятность того, что случайная величина ξ, окажется меньше или равна некоторому фиксированному значению *x*, т.е $F(x)=P(\xi \le x)$.

__Свойства интегральной функции распределения:__

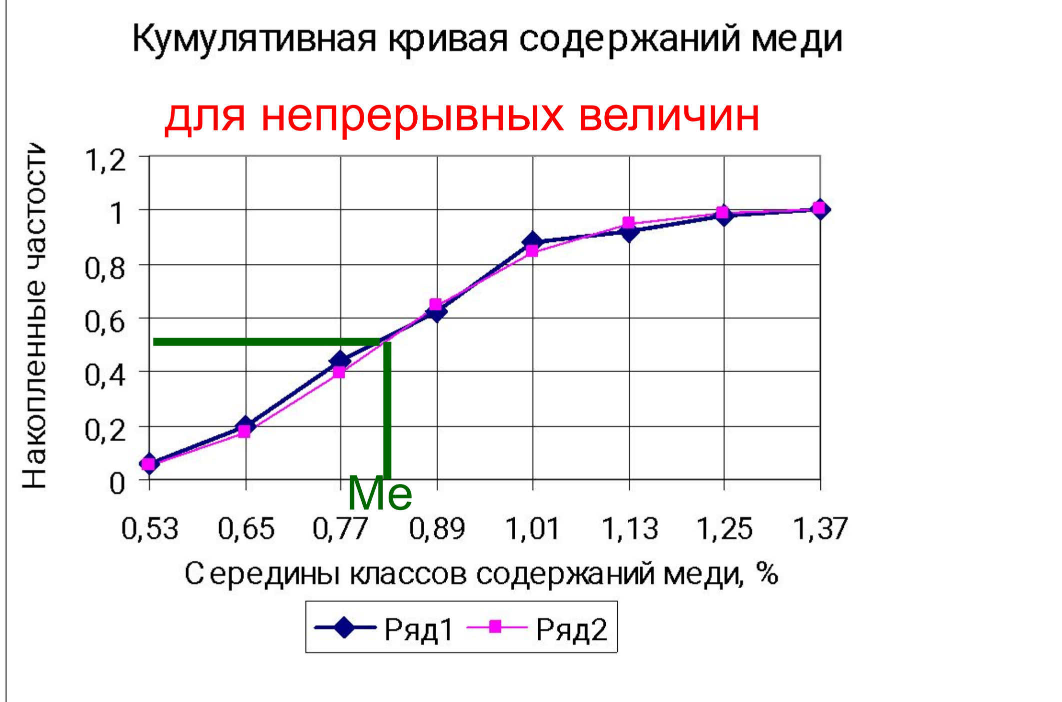
1) функция определена в интервале от -∞ до +∞, является всегда положительной и изменяется, как и любая вероятность, от 0 до 1, т.е. $F(-\infty)=0$ и $F(+\infty)=1$ так как $P(-\infty)=0$ и $P(+\infty)=1$;

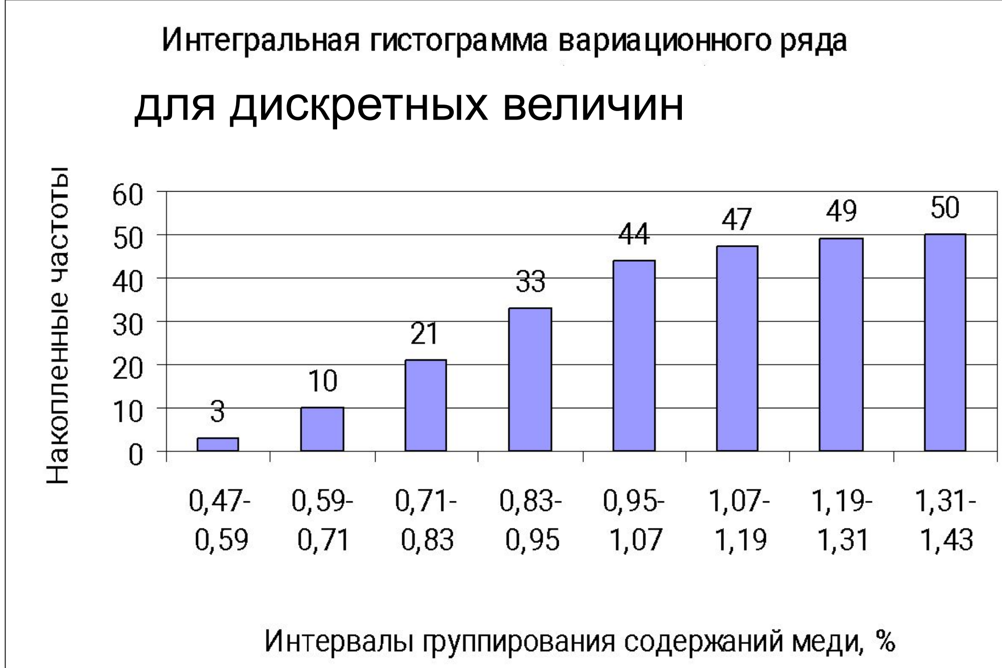
2) функция *F(x)* является монотонно возрастающей (точнее, неубывающей, так как могут существовать участки, на которых она постоянна), т.е. если $X1 \le X2$, то $F(X1) \le F(X2)$.

Интегральная функция распределения

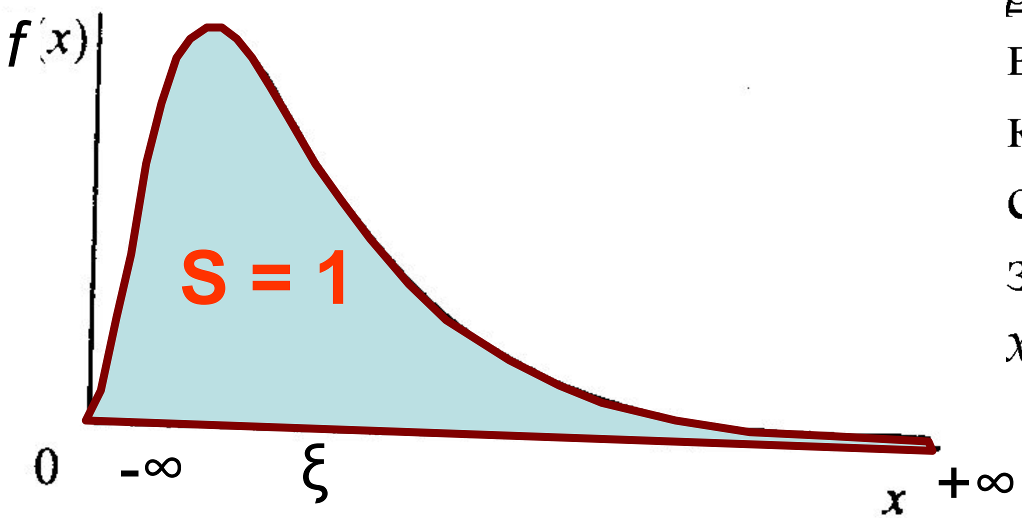


$$F(x) = \int_{-\infty}^{x} f(x)dx$$





Дифференциальная функция распределения -



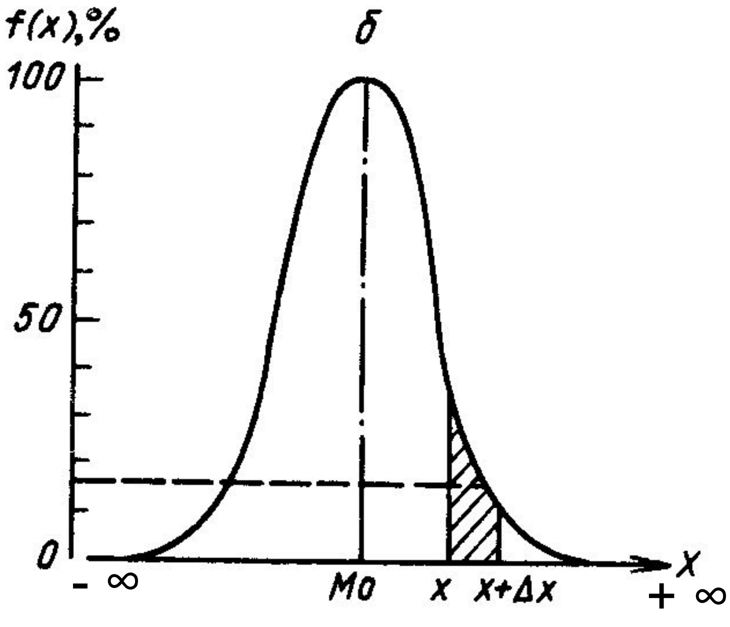
<u>*f(x)*</u> или <u>*функция плотности распределения*</u> характеризует вероятность попадания конкретного значения случайной величины ξ, в заданный интервал от *x* до *x*+Δ*x*, *Т.е.* $f(x)=P(x<\xi<x+\Delta x)$

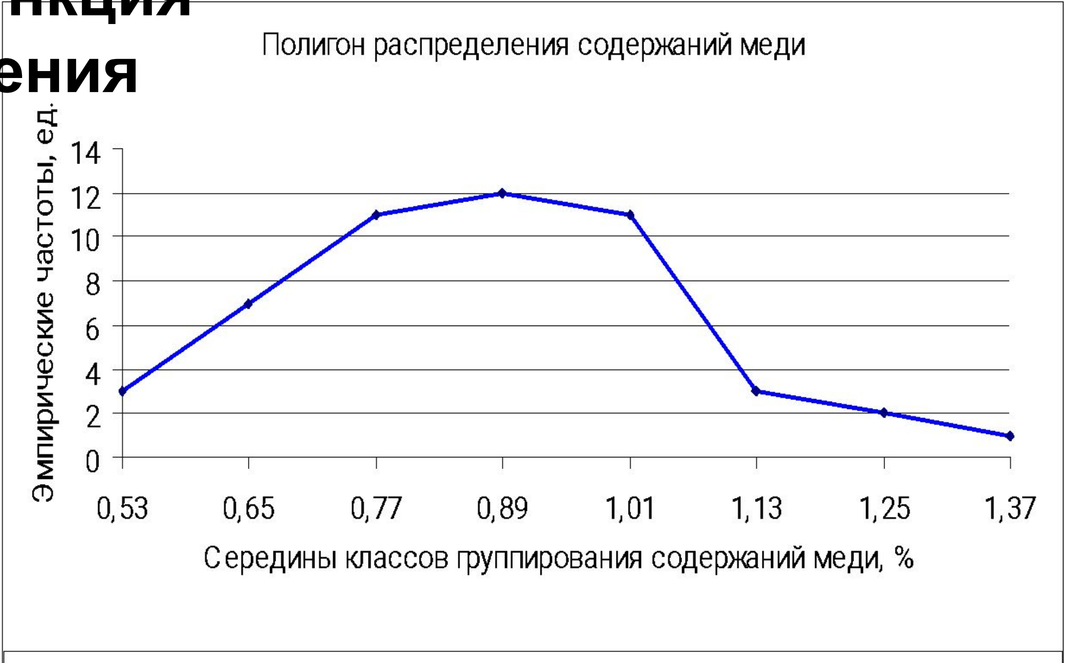
<u>***Свойства дифференциальной функции распределения:***</u>

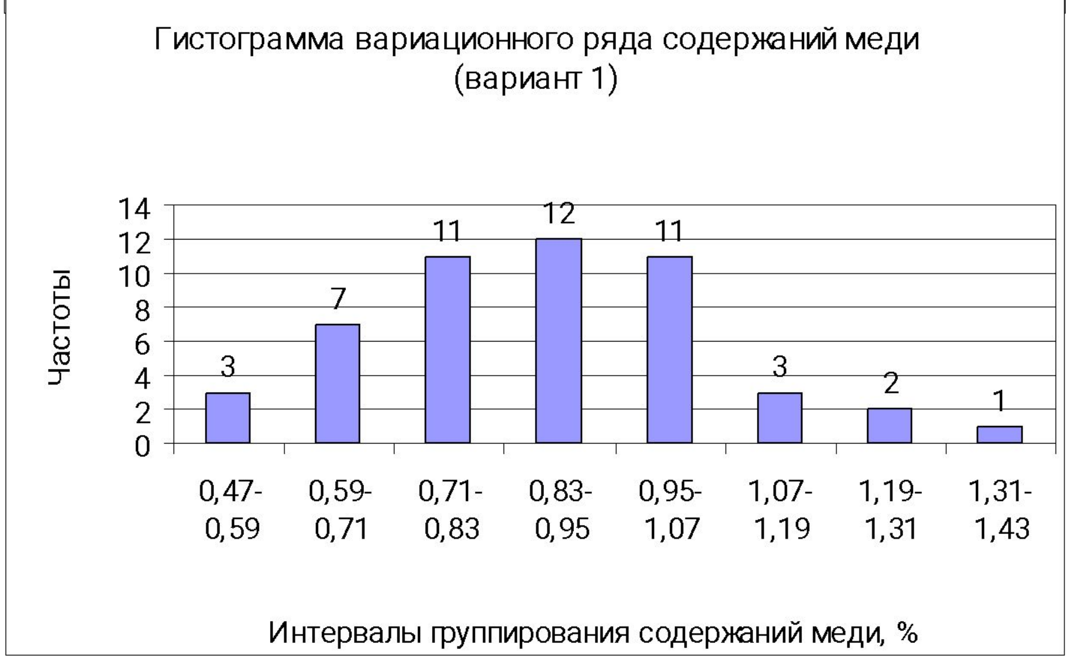
1) функция $f(x)$ определена в интервале от $-\infty$ до $+\infty$ и является всегда положительной, т.е. $f(x)\geq 0$, поэтому кривая плотности распределения располагается выше оси абсцисс;

2) полный интеграл $\int_{-\infty}^{+\infty} f(x)dx=1$, т.к. характеризует полную вероятность; геометрически это означает, что площадь, ограниченная кривой распределения и осью абсцисс, равна 1.

Дифференциальная функция плотности распределения







Дифференциальная функция является первой *производной* от *интегральной* функции. Обе функции *связаны соотношениями:*

$$f(x) = F'(x) = \frac{dF(x)}{dx}$$

и

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

.

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ТЕОРЕТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Меры положения и меры рассеяния.

М*атем. ожиданием* называется сумма произведений значений случайной величины на соответствующие им вероятности или плотности вероятности. В отличие от среднего рассчитывается не через частоты, а через вероятности по формулам:

$$M_x = \sum_{i=1}^{n} x_i \cdot p_i \quad \text{и} \quad M_x = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

где x_i, - значения случайной величины, p_i, - соответствующие им вероятности, $f(x)$ - плотность вероятности. Обозначается M_X, $M[X]$, a, μ, m_x или МОЖ

Математическое ожидание по своему смыслу эквивалентно среднему значению вариационного ряда. Оно характеризует центр теоретического распределения, около которого группируется подавляющее количество значений случайной величины.

__Дисперсия__ характеризует степень рассеяния отдельных значений или интервалов значений случайной величины относительно ее математического ожидания *Mx*. Обозначается *Dx*, *D[X]*, σ^2_x и рассчитывается по формулам соответственно для дискретной и непрерывной величины:

$$\sigma_x^2 = \sum_{i=1}^{n} (x_i - Mx)^2 \cdot p_i \quad \text{и} \quad \sigma_x^2 = \int_{-\infty}^{+\infty} (x - Mx)^2 \cdot f(x)dx$$

Квадратный корень из дисперсии называется *__средним квадратическим отклонением__* σ_x, а *__среднекаадратическое отклонение,__* нормированное на математическое ожидание, - *__коэффициентом вариации__* *V* :

$$\sigma_x = \sqrt{\sigma_x^2} \quad \text{и} \quad V = \frac{\sigma_x}{Mx} \cdot 100\%$$

Показатели ***<u>асимметрии и эксцесса</u>***:

$$A = \frac{\sum_{i=1}^{n} (x_i - Mx)^3 \cdot p_i}{\sigma_x^3};$$

$$A = \frac{\int_{-\infty}^{+\infty} (x - Mx)^3 \cdot f(x) dx}{\sigma_x^3};$$

$$E = \frac{\sum_{i=1}^{n} (x_i - Mx)^4 \cdot p_i}{\sigma_x^4};$$

$$E = \frac{\int_{-\infty}^{+\infty} (x - Mx)^4 \cdot f(x) dx}{\sigma_x^4}$$

ТЕОРЕТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ В ГЕОЛОГИИ

для непрерывных величин:
нормальный, логнормальный

для дискретных величин:
биномиальный, закон Пуассона

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Нормальным называется *распределение* случайной величины, интегральная $F(x)$ и дифференциальная $f(x)$ функции которого имеют вид:

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

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

где $\boldsymbol{Mx}$ и $\boldsymbol{\sigma^2}$ - параметры распределения (математическое ожидание и дисперсия).

Обозначается N **(X, Mx, σ_x)**,

N означает принадлежность случайной величины X к нормальному распределению с параметрами Mx и σ_x.

СВОЙСТВА ДИФФЕРЕНЦИАЛЬНОЙ ФУНКЦИИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

1) колоколообразная кривая, симметричная относительно *Mx*, при этом *Mx* = *Mo* = *Me*

2) ветви кривой асимптотически приближаются к оси абсцисс, нигде с ней не пересекаясь;

3) функция *f(x)* имеет максимум при $x = Mx$, $f(x)_{max} = \dfrac{1}{\sigma_x \sqrt{2\pi}}$

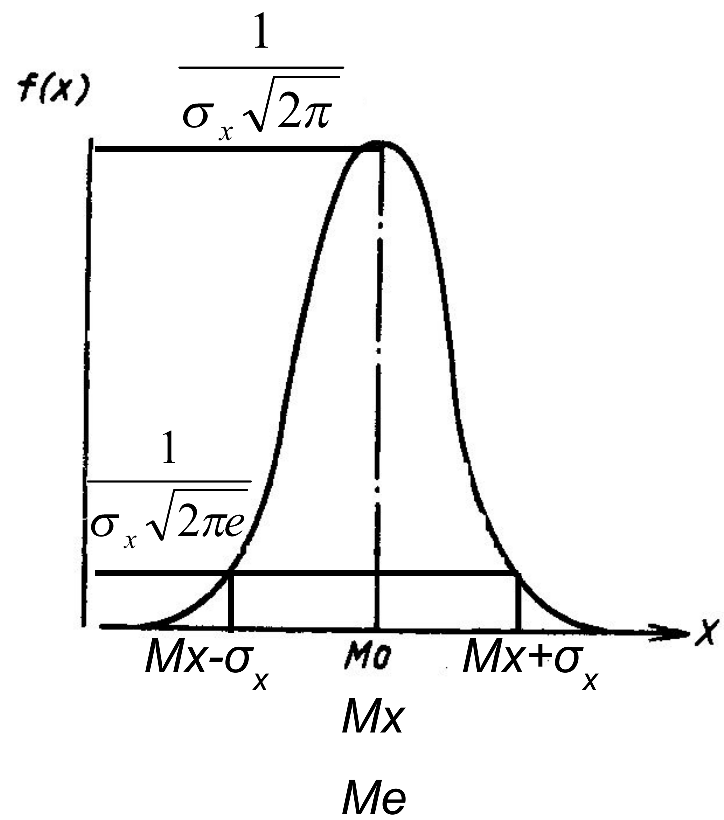
4) точки перегиба кривой *f(x)* имеют координаты: $\left(Mx-\sigma_x;\ \dfrac{1}{\sigma_x\sqrt{2\pi e}}\right)$ и $\left(Mx+\sigma_x;\ \dfrac{1}{\sigma_x\sqrt{2\pi e}}\right)$

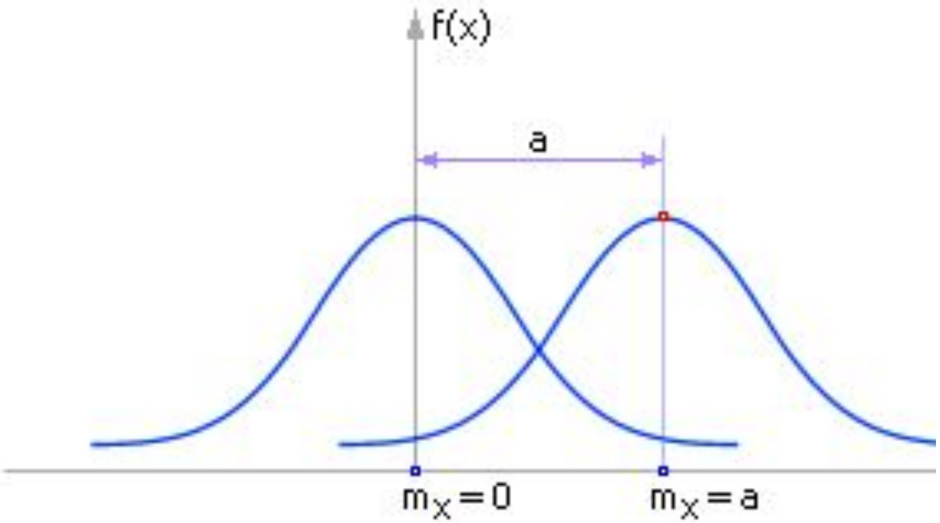
5) изменение *Mx* *__не влияет на форму__* кривой, а вызывает ее смещение вдоль оси *X*;

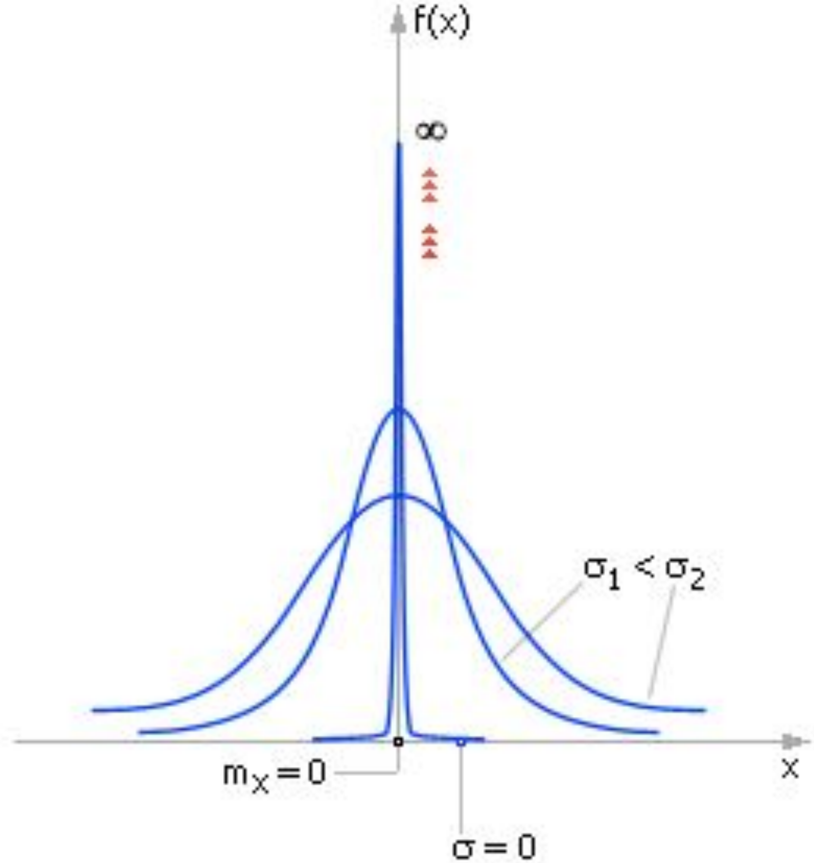
6) σ_x определяет вытянутость или сжатость кривой по оси *X*

7) $A=0$;

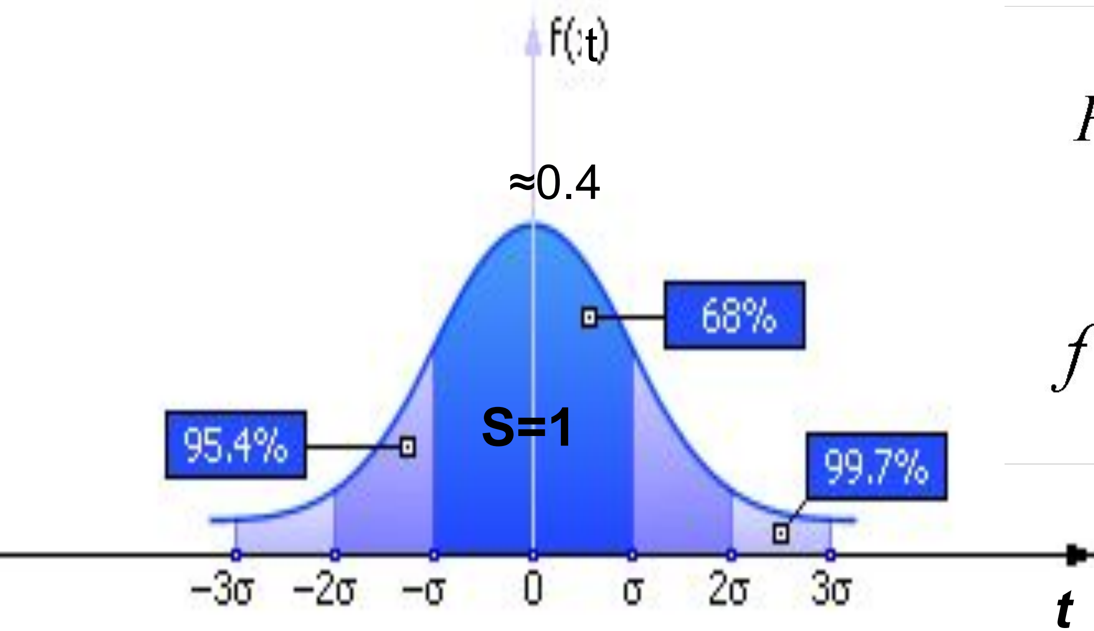
8) поскольку у нормального распределения $\dfrac{\int_{-\infty}^{+\infty}(x-Mx)^4 \cdot f(x)dx}{\sigma_x^4} = 3$, показатель эксцесса E=3-3=0




f(x)
a
$m_X = 0$
$m_X = a$


f(x)
∞
$\sigma_1 < \sigma_2$
$m_X = 0$
x
$\sigma = 0$

СТАНДАРТНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ



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

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

$$t=\frac{x-Mx}{\sigma}$$

Из симметричности следует:

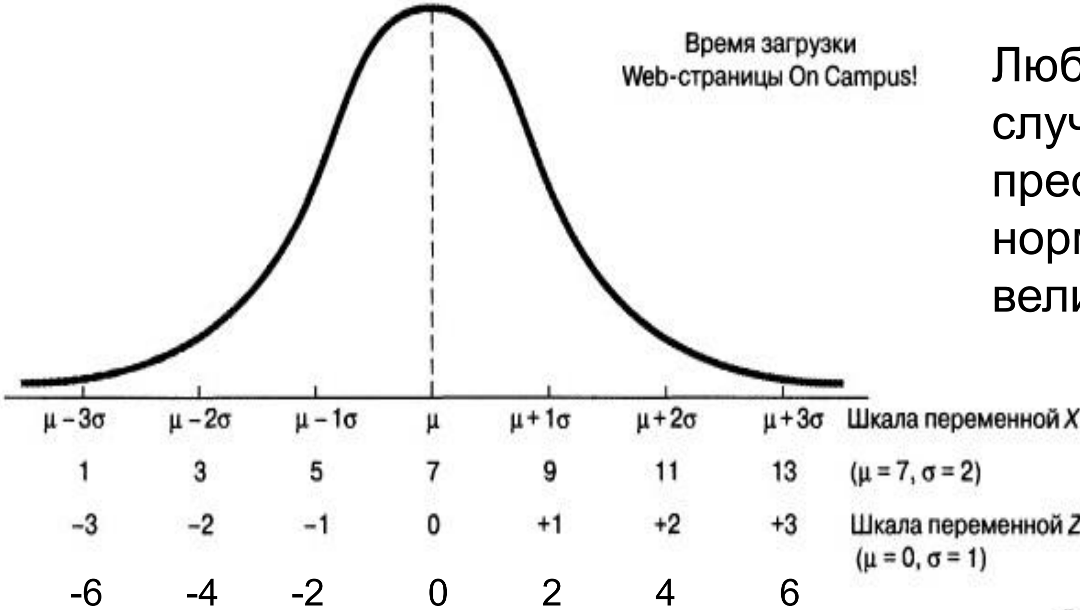
F(-t)=1-F(t)

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

Функция F(t) для t≥0 нормированная функция Лапласа. Обозначается $\Phi_0(t)$ и имеет вид

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

$$\varphi(t)=Z(t)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}$$



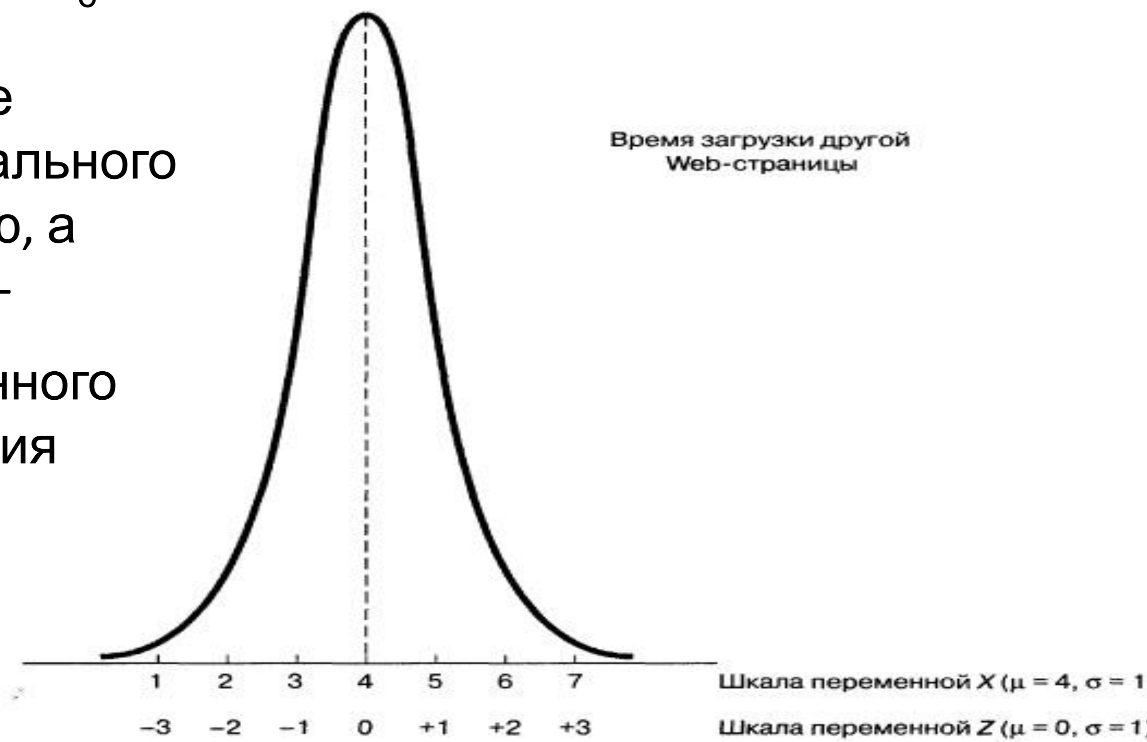
Любую нормально распределенную случайную величину *X* можно преобразовать в нормированную нормально распределенную случайную величину *Z* или *t*.

$$Z = (X - \mu)/\sigma$$

Математическое ожидание стандартизованного нормального распределения равно нулю, а стандартное отклонение — единице.

Плотность стандартизованного нормального распределения

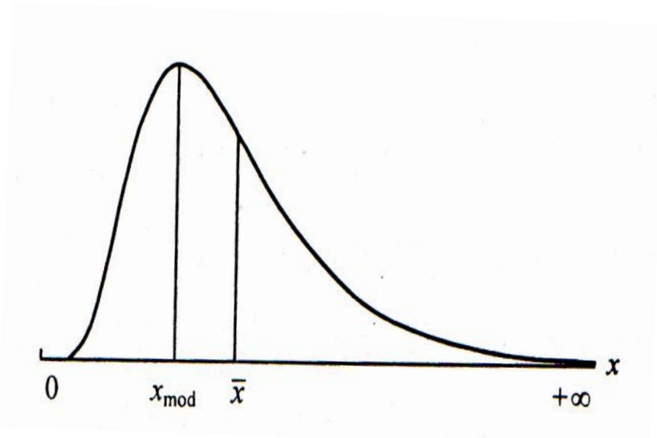
$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} \; e^{\frac{1}{2}Z^2}$$



ЛОГНОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Логарифмически нормальное (логнормальное) распределение, при котором нормально распределены логарифмы значений случайной величины.

Интегральная и дифференциальная функции *логнормального* распределения:



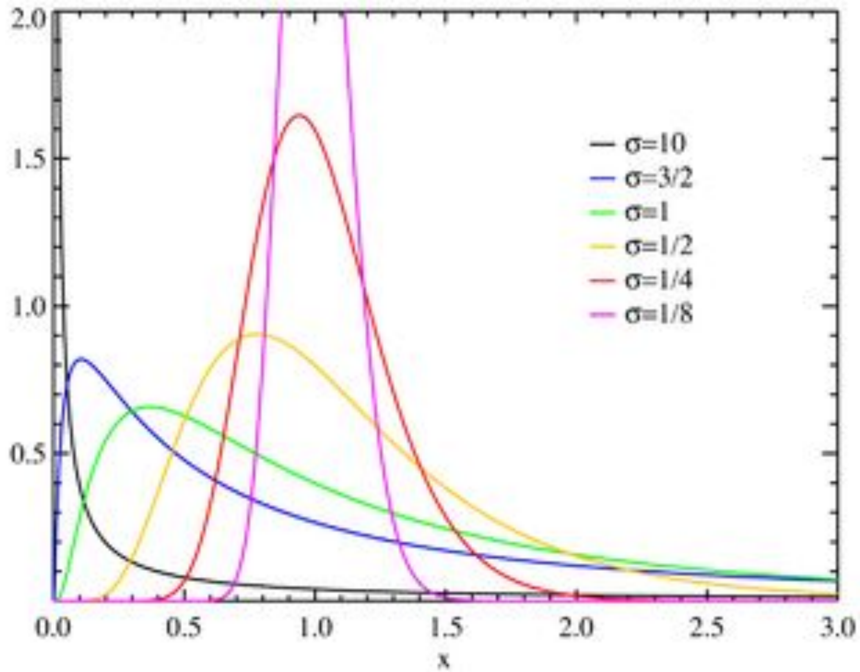
$$F(\ln x) = \frac{1}{\sigma_{\ln x}\sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(\ln x - M\ln x)^2}{2\sigma^2_{\ln x}}} d\ln x$$

$$f(\ln x) = \frac{1}{\sigma_{\ln x}\sqrt{2\pi}} \cdot e^{-\frac{(\ln x - M\ln x)^2}{2\sigma_{\ln x}{}^2}}$$

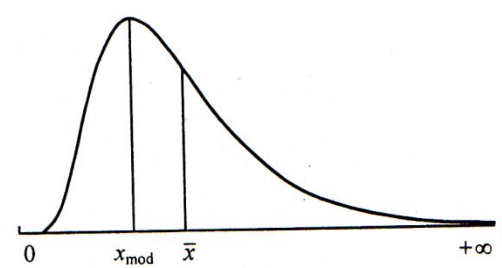
имеет *положительные коэффициенты асимметрии и эксцесса*, т.е. является правоасимметричным (A>0) и крутовершинным (E>0).

Mo<Me<Mx. Определяются из соотношений

$$Mx = e^{M\ln x + 0.5\cdot\sigma^2_{\ln x}} \qquad Me = e^{M\ln x} \qquad Mo = e^{M\ln x - \sigma^2_{\ln x}}$$



2.0
1.5
1.0
0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x
σ=10
σ=3/2
σ=1
σ=1/2
σ=1/4
σ=1/8



0
x_{mod}
$\bar{x}$
$+\infty$
x

ПРОВЕРКА ГИПОТЕЗЫ О ЗАКОНЕ РАСПРЕДЕЛЕНИЯ

Графический

Экспрессный метод графической проверки

Аналитический:

1) χ^2 -критерий Пирсона (хи-квадрат)
2) критерий Колмогорова (критерий λ)

Сравнение эмпирических частот $\boldsymbol{n_{эj}}$ по классам группирования с теоретическими частотами $\boldsymbol{n_{Tj}}$, при этом параметры нормального распределения $\boldsymbol{Mx}$ и $\boldsymbol{\sigma}$ приравниваются к их выборочными оценками $\overline{U}$ и $\boldsymbol{S}$.

Отличие:

в случае *критерия Пирсона* сравниваются *дифференциальные кривые* эмпирического и теоретического распределения (т.е. частотные ряды распределения),

в случае критерия Колмогорова - *интегральные кривые* (т.е. ряды накопленных частот).

Критерий Пирсона χ^2

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_{iЭ} - n_{iT})^2}{n_{iT}}$$

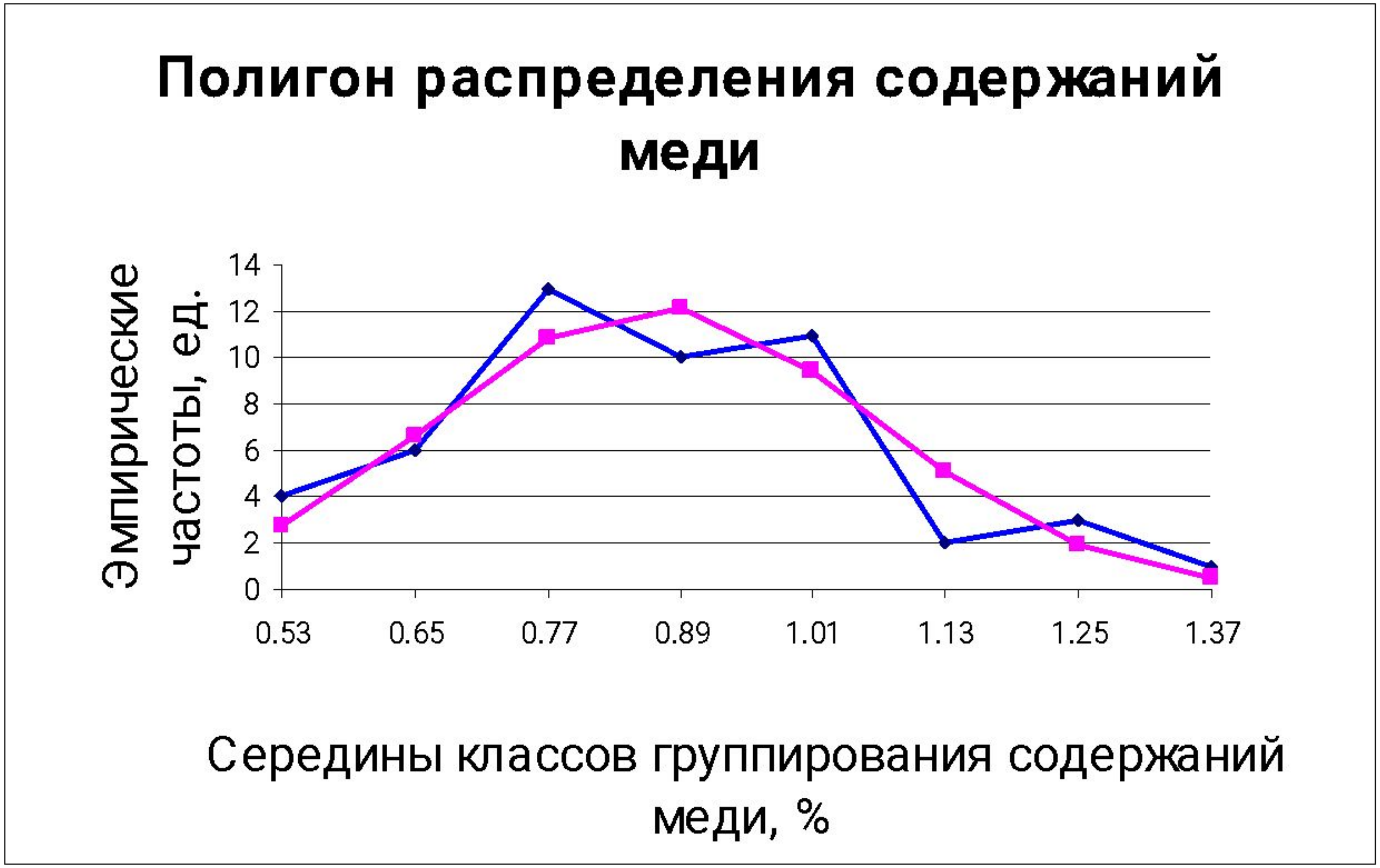
$n_{iэ}$ и n_{iT} - эмпирические и теоретические частоты для проверяемого закона распределения выборочных данных, сгруппированных в k интервалов.

Графически это суммарная площадь между эмпирическим и теоретическим полигонами распределения.

Теоретические частоты рассчитываются:

$$n_{iT} = \Delta X \cdot N \cdot f_j(\hat{X}_j, \overline{X}, S_x^2) = \frac{\Delta X \cdot N}{S_x \cdot \sqrt{2\pi}} \exp\left[-\frac{(\hat{X}_j - \overline{X})^2}{2S_x^2}\right]$$

где $f_j(\hat{X}_j, \overline{X}, S_x^2)$ - значения функции плотности нормального распределения для середины j-го интервала $\hat{X}_j$ при заданном выборочном среднем $\overline{U}$ и дисперсии S_x.



Обычно **χ2** применяется, когда N>60

Если $\chi_p^2 < \chi_T^2$, то с вероятностью ***p=1-α*** ***принимается нулевая гипотеза,*** т.е. гипотеза о том, что эмпирическое распределение ***не противоречит*** **нормальному.**

Если $\chi_p^2 > \chi_T^2$, то нулевая гипотеза отклоняется и принимается альтернативная гипотеза о том, что эмпирическое распределение не соответствует нормальному.

Критерий Колмогорова λ

Численно **критерий λ** равен максимальному отклонению эмпирических накопленных частот от теоретических и рассчитывается по формуле:

$$\lambda = \frac{D_{\max}}{\sqrt{N}} = \frac{\left| n_{s э} - n_{sT} \right|}{\sqrt{N}},$$

где D_{max} - максимальное значение абсолютной разности между эмпирическими накопленными частотами $(n_{sэ})$ и теоретическими (n_{sT}).

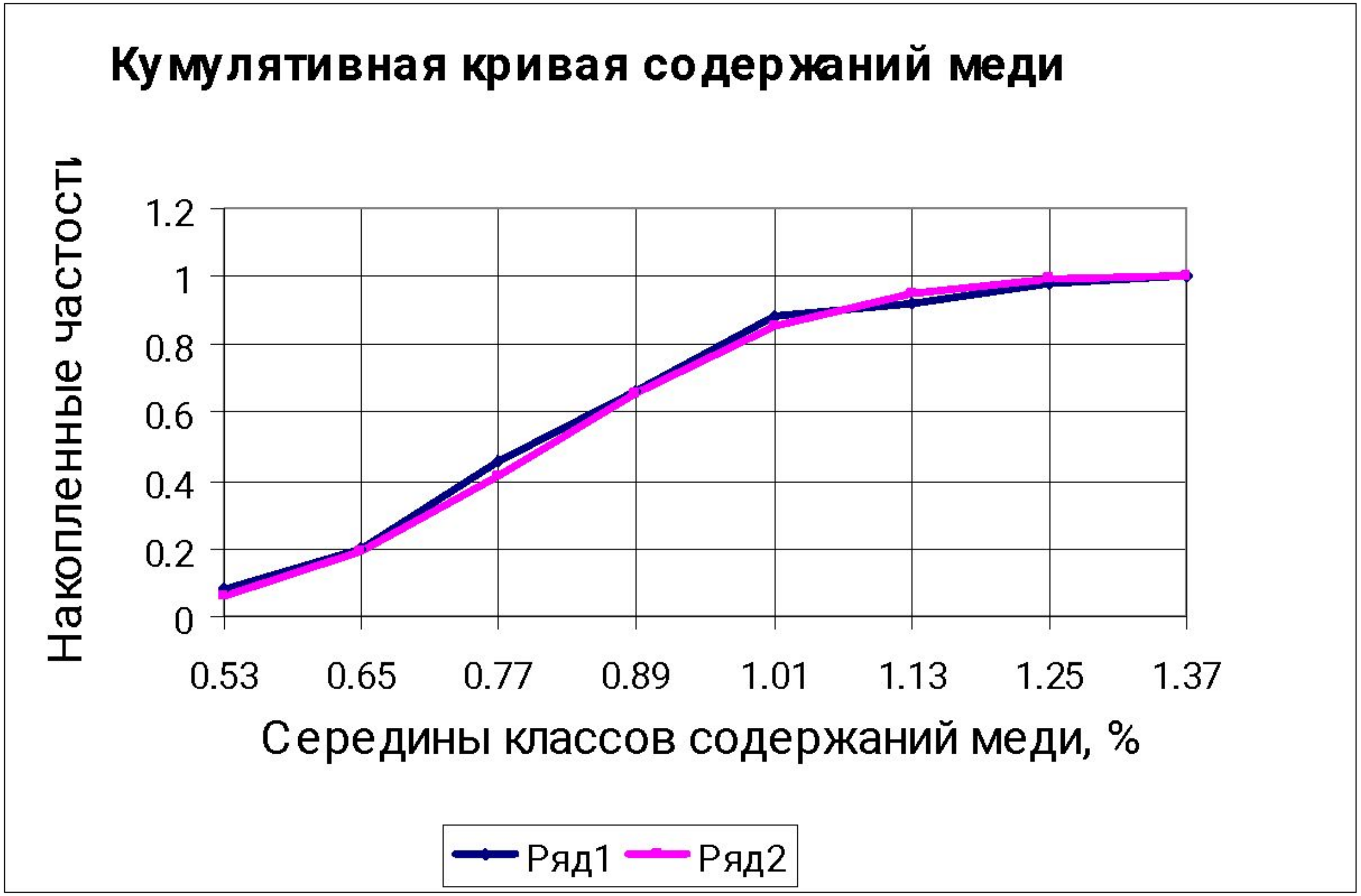
Графически критерий **λ** соответствует максимальному расстоянию между эмпирической и теоретической кумулятивными кривыми.

сравнивают значение λ_p с λ_T для принятого уровня значимости α.

Теоретическое значение λ_T не зависит от N и числа степеней свободы, а определяется только уровнем значимости:

при **α=0,05 , λ_T = 1,36,** при **α=0,01 λ_T =1,63**.

Если $\lambda_p < \lambda_T$, то с вероятностью $p=1-\alpha$ нулевая гипотеза о том, что эмпирическое распределение не противоречит нормальному, принимается.



Кумулятивная кривая содержаний меди
Накопленные частости
1.2
1
0.8
0.6
0.4
0.2
0
0.53
0.65
0.77
0.89
1.01
1.13
1.25
1.37
Середины классов содержаний меди, %
Ряд1
Ряд2