Лекция 8

1 Метрология: применение математической статистики при измерениях и испытаниях.

2 Проверка статистических гипотез. Ошибки первого и второго рода.

3 Нормальный (гауссовский) закон распределения случайной величины. Вычисление вероятности по нормальному закону распределения.

МЕТРОЛОГИЯ

Применение математической статистики при измерениях и испытаниях

Статистические гипотезы. Проверка гипотез. Односторонний и двухсторонний критерии

Генеральная совокупность случайной величины (СВ) - бесконечное (очень большое) число наблюдений СВ.

Выборочная совокупность – выборка ограниченного объема из генеральной совокупности.

Закон распределения генеральной совокупности. Плотность распределения

$$\left.\begin{aligned} &F(x) = P(X < x) \\ &F(b) - F(a) = P(a \le x \le b) \end{aligned}\right\}$$

- непрерывная функция распределения

X – **случайная величина**

$$f(x) = {dF}/{dx}$$

- плотность непрерывного распределения

$$F(x) = \int f(x)dx$$

- связь между функцией и плотностью распределения

Статистическая гипотеза (СГ) – некоторое предположение относительно <u>вида</u> неизвестного или о <u>параметрах</u> известного распределения генеральной совокупности СВ.

Проверка СГ заключается в сопоставлении неких <u>статистических показателей</u> (Θ - критериев проверки)**, вычисляемых по выборке, со значениями $\Theta_{кр.}$, которые определены при верной проверяемой гипотезе.**

ЭТАПЫ ПРОВЕРКИ СГ

1 Выдвигают <u>основную</u> $(H_0:)$
и одновременно <u>альтернативную</u> $(H_1:)$
гипотезы.

2 Вычисляют по выборке СВ некий статистический критерий $(\Theta_{\text{расч.}})$ **.**

3 Сравнивают рассчитанное значение $\Theta_{\text{расч.}}$ **с критическим** $\Theta_{\text{кр.}}$ **(табличным).**

4 В зависимости от результатов сравнения $\Theta > \Theta_{\text{расч.}} (\Theta < \Theta_{\text{расч.}})$ **принимают** $H_0:$ **или** $H_1:$

$\Theta_{\text{расч.}}$ - это специально подобранная СВ, <u>точное</u> или <u>приблизительное</u> распределение которой <u>известно</u>.

Проверка СГ заключается в сопоставлении $\Theta_{\text{расч.}}$ со значением $\Theta_{\text{кр.}}$, которое определено в <u>предположении</u>, что проверяемая <u>гипотеза</u> <u>верна</u>.

Условия принятия (отбрасывания) гипотез:

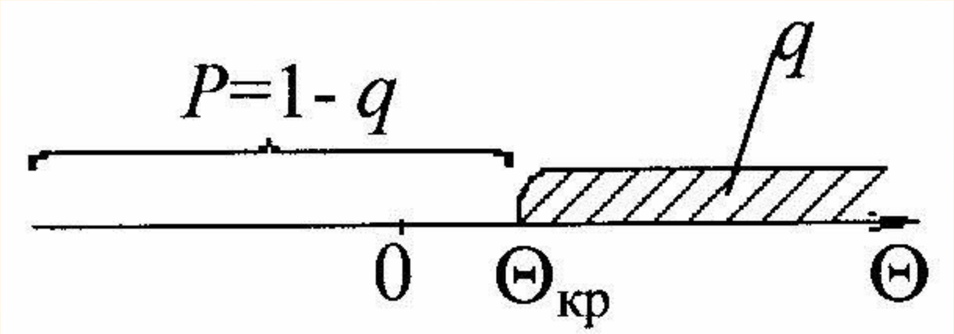
1 H_0: - **принимается, если** $\Theta_{\text{расч.}}$ **<u>не попадает</u> в критическую область, при этом отвергается** H_1:

2 H_1: - **принимается, если** $\Theta_{\text{расч.}}$ **<u>попадает</u> в критическую область, одновременно отвергается** H_0:

Односторонний критерий принятия гипотезы

Для одностороннего (правостороннего) критерия вероятность попасть $\Theta_{\text{расч.}}$ в <u>критическую область</u> (заштрихована) равна $P(\Theta_{\text{расч.}} > \Theta_{\text{кр.}}) = q$

q **- есть уровень значимости**



Для левостороннего критерия вероятность попадания $\Theta_{\text{расч.}}$
в критическую область равна

$$P(\Theta_{\text{расч.}} < \Theta_{\text{кр.}}) = q$$

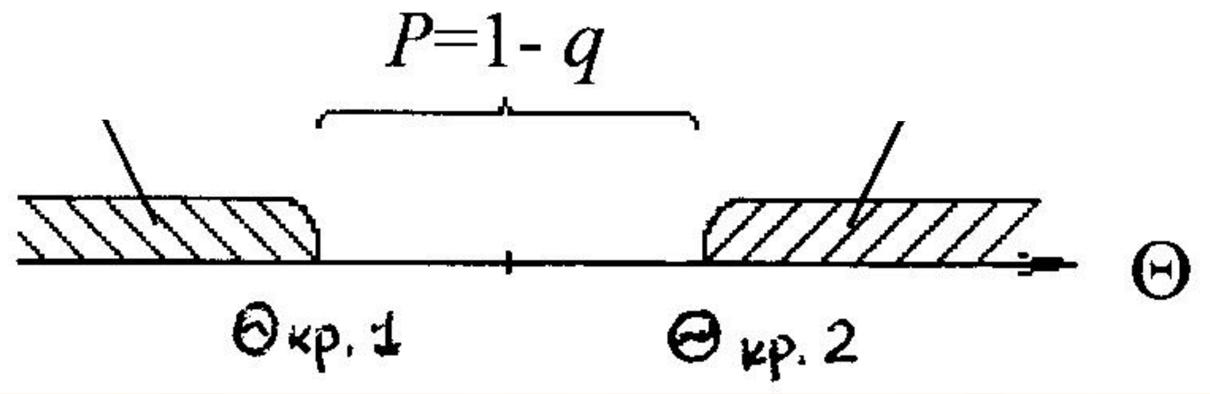
Вероятность попадания $\Theta_{\text{расч.}}$
в <u>область принятия гипотезы</u> H_0
равна $\alpha = 1 - q$

α **- есть доверительная вероятность**

Двухсторонний критерий принятия гипотезы

Вероятность попадания $\Theta_{\text{расч.}}$ в критическую область равна:

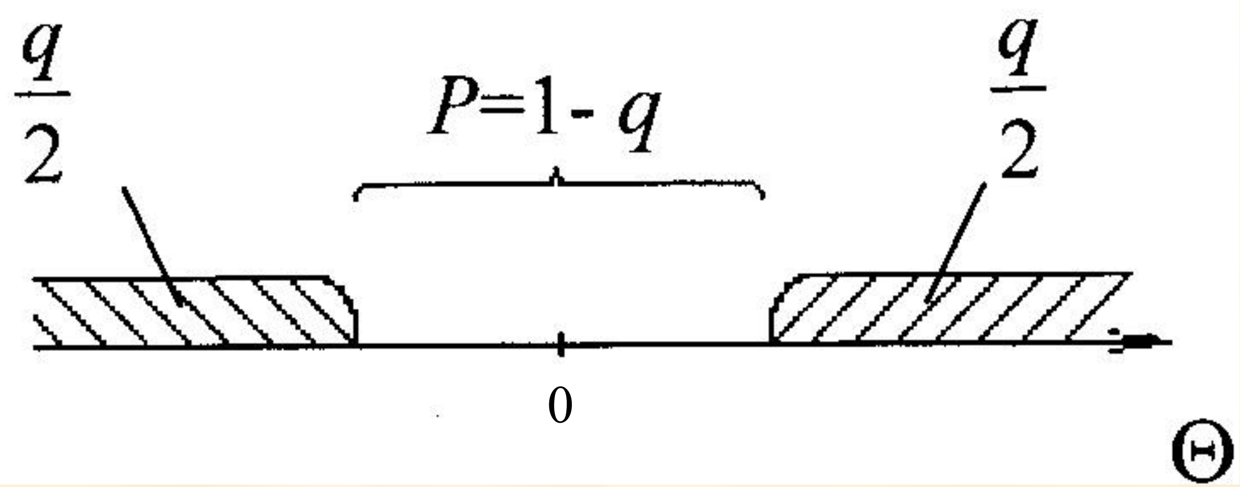
$$P(\Theta_{\text{расч.}} < \Theta_{\text{кр.1}}) + P(\Theta_{\text{расч.}} > \Theta_{\text{кр.2}}) = q$$



Если распределение Θ симметрично:

$-\Theta_{кр.1} = \Theta_{кр.2} = \Theta_{кр.}$, **то вероятность попадания** $\Theta_{расч.}$ **в любую из критических областей равна:**

$$P(\Theta_{расч.} > \Theta_{кр.}) = q/2$$



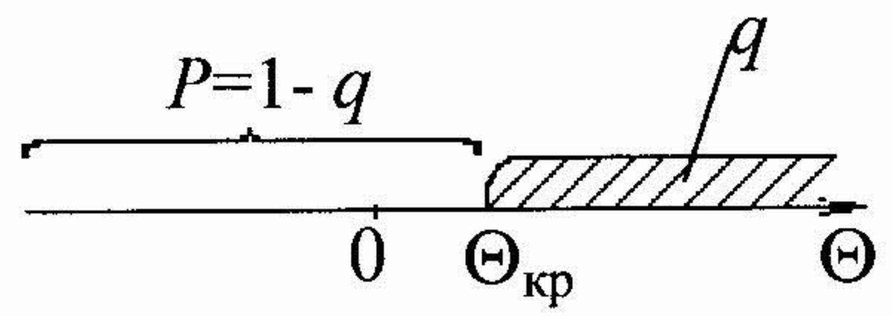
Практическое правило:

Для двухстороннего критерия численное значение $q^*/2 = 0{,}05$ **берут таким же, как для одностороннего критерия** $q = 0{,}05$

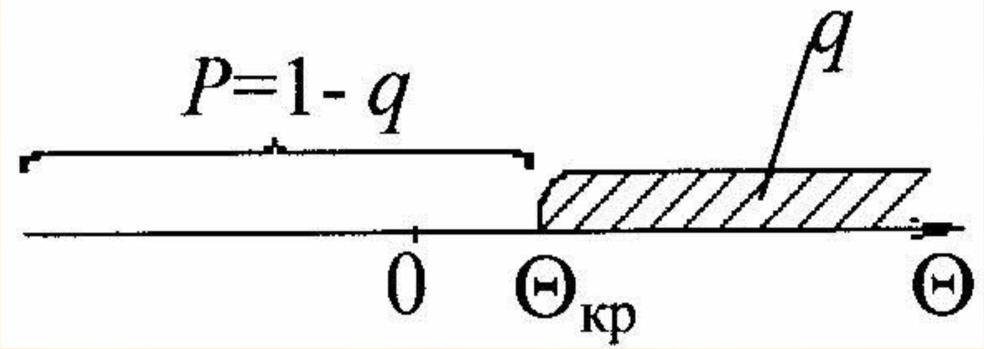
В этом случае вероятность ошибки второго рода будет одинаковой для одностороннего и двухстороннего критерия (для последнего критерия получаем $q^* = 0{,}10$ **)**

Как найти границы критической области ?

Рассмотрим правостороннюю область.
Для нахождения $\Theta_{кр.}$ **задаются** **малой вероятностью** $q\ (0{,}05;\ \ 0{,}025;\ \ 0{,}01)$.
Затем ищут $\Theta_{кр.}$,
исходя из неравенства $P(\Theta > \Theta_{кр.}) = q$



Это означает, что вероятность события $\Theta > \Theta_{кр.}$ мала, и в единичном испытании оно не должно наступить. Если все же событие произошло, то H_0: ложна.



Ошибки первого и второго рода:

Ошибка 1-го рода – отвергнуть верную гипотезу $(H_0:)$.
Причем $H_0:$ **действительно верна.**

Её вероятность составляет не более q.
При $q = 0{,}05$ **ошибка 1-го рода произойдет в** 5 **случаях из** 100.
Ошибку 1-го рода называют рискoм производителя

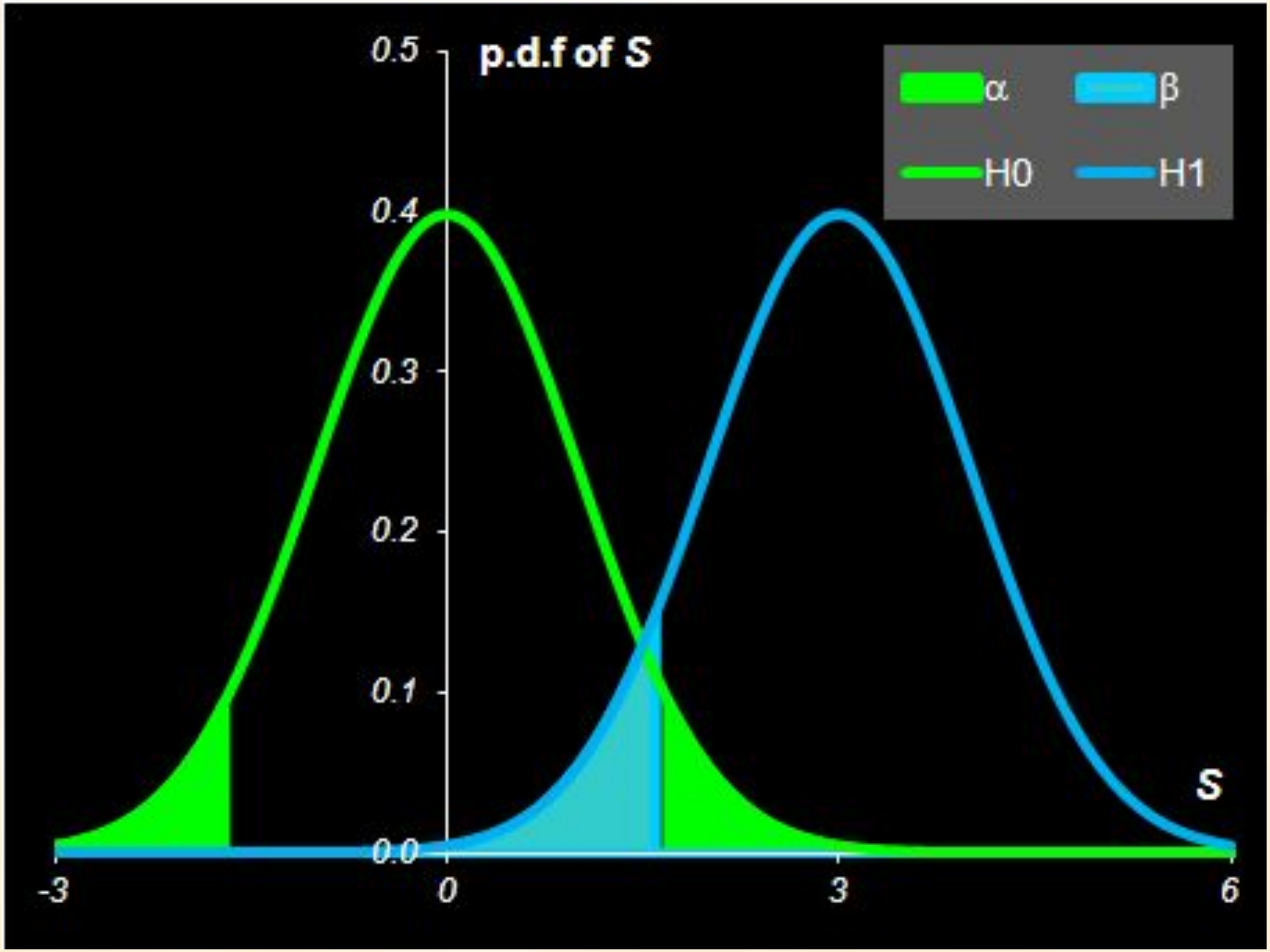
Ошибка 2-го рода – принять ложную гипотезу $(H_0 :)$ за верную. Причем $H_0 :$ действительно ложная. Оценить вероятность её очень сложно. При увеличении q, увеличивается число отвергаемых гипотез! Ошибку 2-го рода называют **риском потребителя**

Если вероятность ошибки 2-го **рода принять равным** β, **то** (1- β) **называют** **<u>мощностью критерия</u>** **-** **это вероятность отклонения** H_0:, **когда она ложная. Мощность критерия характеризует вероятность** **<u>ошибочного применения</u>** **ложной гипотезы** H_0**.**

1-β **–** **<u>вероятность</u>** **<u>не совершить</u>** **ошибку второго рода.**

H_0: $m = 0$;   если H_0 : верна, то q зеленая

H_1: $m = 3$;   если H_0 : ложна, то β синяя



Мощность критерия должна быть <u>максимальной</u>, это обеспечивает <u>минимальность ошибки 2-го рода</u>.

Если уменьшать q **, то** β **будет возрастать при** $n = \text{const}$ **.**

Единственный способ <u>одновременного уменьшения</u>

q и β **- это увеличение <u>объема выборки</u>** (n) **!**

Нормальный (гауссовский) закон распределения случайной величины

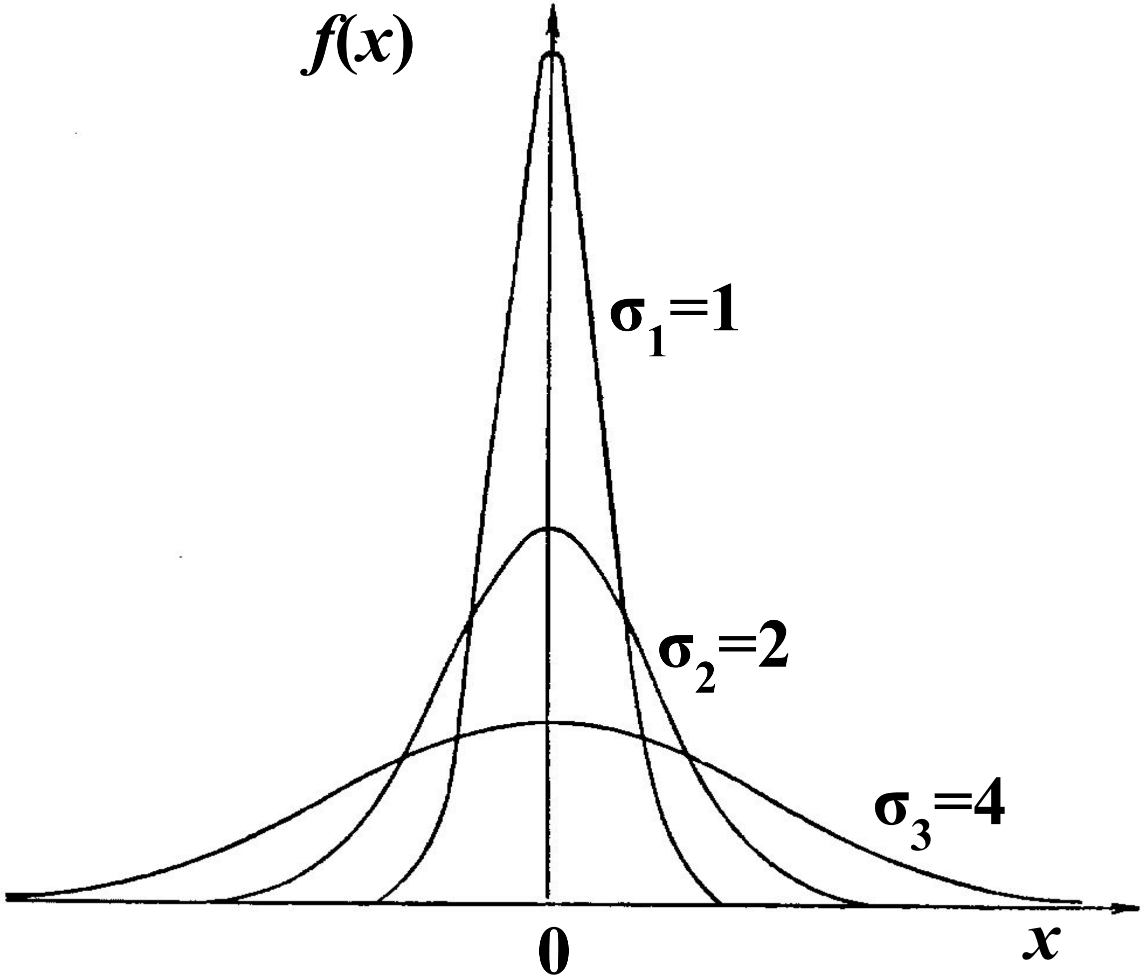
Карл Фридрих Гаусс
30.04.1777 - 23.02.1855
великий немецкий математик, астроном и физик. Считается одним из величайших математиков всех времён
«Король математиков»

СВ распределена по нормальному закону, если плотность её распределения описывается выражением:

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

$\bar{x}$ – **генеральное мат. ожидание СВ**

σ^2 – **генеральная дисперсия СВ**



Семейство кривых нормального распределения
с параметрами
$\mu = 0, \quad \sigma_1 = 1, \quad \sigma_2 = 2, \quad \sigma_3 = 4$

Характеристики кривой НР:

1 Симметрия относительно центра распределения: $F(\bar{x}) = 1/2$
Медиана равна мат. ожиданию $x = \bar{x}$

2 Мода – в центре распределения $f(\bar{x}) = \dfrac{1}{\sigma\sqrt{2\pi}}$

3 $f(x) > 0$

4 Точки перегиба – при $x = \bar{x} \pm \sigma$

Вычисление вероятности по нормальному закону распределения

Пусть случайная величина X распределена по нормальному закону, тогда вероятность найти $X < x$ равна:

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

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

Перейдем к интегралу вероятности

$$F(\bar{x}) = 1/2$$ $\rightarrow$ **из свойства** $F(x)$

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

$\bar{x} = 0; \quad \sigma = 1; \quad x = t$ **- центрирование, нормировка, замена переменной**

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

$$\boxed{\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt}$$

- функция Лапласа

Свойства функции Лапласа:

$$\Phi(0)=0;$$

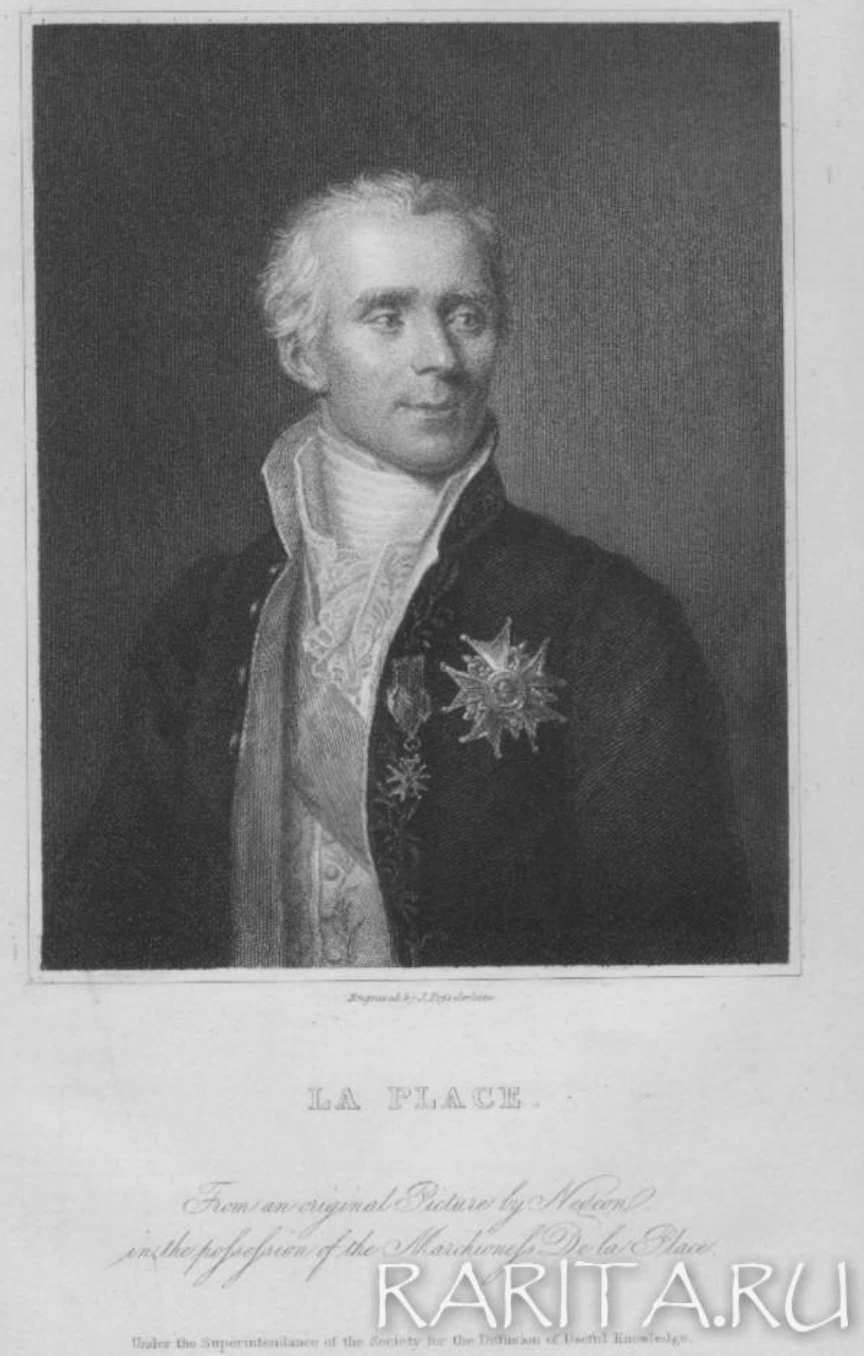
$$\Phi(-\infty)=-0,5;$$

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

$$\Phi(-t)=-\Phi(t)$$

нечетная функция

$$\boxed{F(x)=\tfrac{1}{2}+\Phi(x)}$$



Пьер-Симо́н, маркиз де Лапла́с (*Pierre-Simon de Laplace*) 23.03.1749-05.03.1827.
Французский
математик, механик, физик
и астроном. Известен работами
в области небесной механики,
диф.уравнений, один из
создателей теории вероятности.
Заслуги Лапласа в области
теоретической
и прикладной математики
и, особенно в астрономии,
громадны: он усовершенствовал
почти все отделы этих наук.
Был членом Французского
географического общества.

Таблица значений функции $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

x	$\Phi(x)$	x	$\Phi(x)$
0.00	**0.0000**	**1.00**	**0.3413**
0.10	**0.0398**	**2.00**	**0.4772**
0.30	**0.1179**	**3.00**	**0.49865**
0.50	**0.1915**	**5.00**	**0.499997**

Вычисление вероятности нахождения СВ, распределенной по НЗ, в интервале от a до b по функции (интегралу) Лапласа

Заменим СВ $x \rightarrow$ **по НЗ на** $x = \dfrac{x - \overline{x}}{\sigma}$

получим $F(x) = {}^{1}\!/_{2} + \Phi\left(\dfrac{x - \overline{x}}{\sigma}\right)$

Если генеральные МО и дисперсия известны, то вероятность найти "x" **в интервале** (a, b) :

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

$$= \Phi\left(\frac{b - \overline{x}}{\sigma}\right) - \Phi\left(\frac{a - \overline{x}}{\sigma}\right)$$

Перейдем от "x" к отклонению от мат.ожидания $U = x - \bar{x}$**,**

и переформулируем задачу:

$$P(A < U < B) = \Phi\left(\frac{B}{\sigma}\right) - \Phi\left(\frac{A}{\sigma}\right),$$

где $: A = a - \bar{x};$

$B = b - \bar{x}\ .$

Если интервал (A, B) **симметричен:**

$A = -B$ **и** $B > 0$ **, т.к.** $\Phi(x)$ **- нечетная:**

$$\Phi\left(\frac{B}{\sigma}\right) = -\Phi\left(\frac{A}{\sigma}\right)$$

Тогда вероятность найти:

$$P\left(\left|U\right| < B\right) = 2\Phi\left(\frac{B}{\sigma}\right) \qquad (1)$$

Пример 1: вычислим по (1) какова вероятность, что $|U| < \sigma$ т.к. $B \equiv \sigma$,

то $\dfrac{B}{\sigma} = 1$ **есть аргумент**

$$P\left(|U| < \sigma\right) = 2\Phi(1)$$

$$\Phi(1) = 0{,}3413$$

$$P\left(|U| < \sigma\right) = 0{,}6826 \approx 2/3$$

При $n \to \infty$ в 2/3 наблюдений $U = x - \bar{x}$ не превышает σ !!!

Пример 2: **вычислим какова вероятность, что** $|U| < 2\sigma$, **т.е.** $\dfrac{B}{\sigma} = 2$

$$P\left(|U| < 2\sigma\right) = 2\Phi(2) \approx 0{,}95$$

Пример 3: **вычислим вероятность, что** $|U| < 3\sigma$, **т.е.** $\dfrac{B}{\sigma} = 3$

$$P\left(|U| < 3\sigma\right) = 2\Phi(3) = 0{,}9973$$

Это правило 1, 2 **и** 3 σ **!!!**

Таблица значений функции $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

x	$\Phi(x)$	x	$\Phi(x)$
0.00	**0.0000**	**1.00**	**0.3413**
0.10	**0.0398**	**2.00**	**0.4772**
0.30	**0.1179**	**3.00**	**0.49865**
0.50	**0.1915**	**5.00**	**0.499997**