Нормальное распределение

- Нормальное распределение было получено К.Ф. Гауссом (1777–1855 гг). Оно является самым распространенным распределением в природе, экономике и т.д. Кроме того, многие другие распределения в некоторых предельных случаях переходят в нормальное распределение.

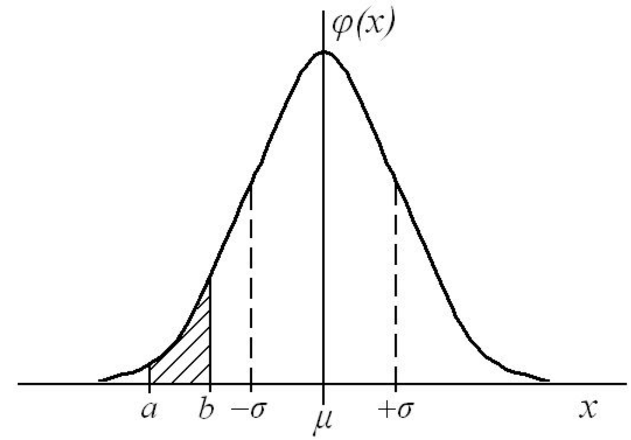
- **Цель** – описать множественный ансамбль, состоящий из однородных элементов таким образом, чтобы появилась возможность количественной оценки свойств отдельных элементов или определенных групп внутри изучаемого ансамбля.

- **Решение** – рассмотрение множественного ансамбля как статистической совокупности, выяснение *характера распределения* случайных величин внутри этой совокупности и последующая *вероятностная оценка*.

- Результаты многократных измерений при наличии случайных погрешностей формируются под влиянием большого числа независимо действующих факторов. На этом основании можно считать, что при отсутствии какого-либо доминирующего влияния результаты прямых многократных измерений подчиняются нормальному распределению.

Закон Гаусса

Случайная величина x с нормальным распределением может принимать любые значения в интервале от $-\infty$ до ∞ и имеет **функцию плотности вероятности** вида

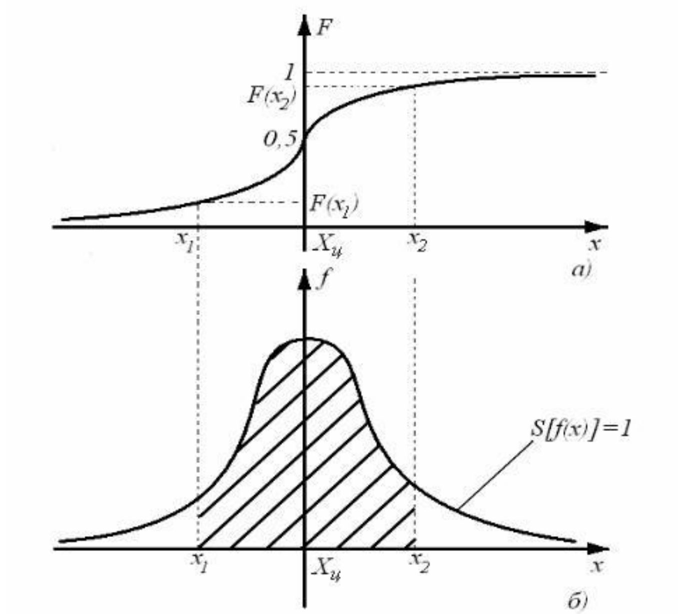


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

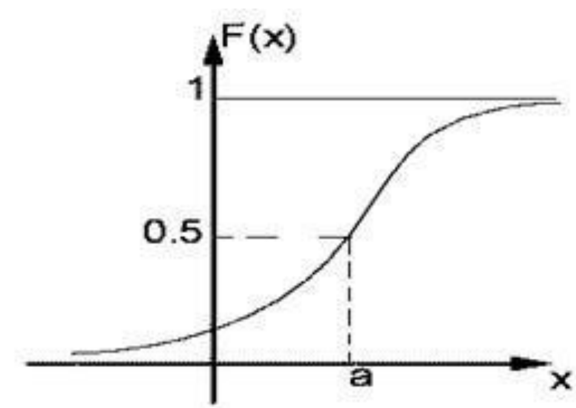
где σ^2 – стандартное отклонение

μ – матожидание

Функция (интегральная) **нормального распределения** случаиной величины



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



Основные свойства нормального распределения

1) Функция плотности вероятности величина положительная $f(x) \geq 0$

2) Площадь, ограниченная функцией $f(x)$ и осью x всегда равна единице, т.е. то, что случайная величина примет любое значение из интервала своего существования, является достоверным событием:

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

3) Распределение симметрично относительно точки $x=\mu$

$$\int_{\mu}^{\mu+a} f(x)dx = \int_{\mu-a}^{\mu} f(x)dx$$

Для отклонений от матожидания, одинаковых по величине, но разных по знаку, плотность вероятности одинакова.

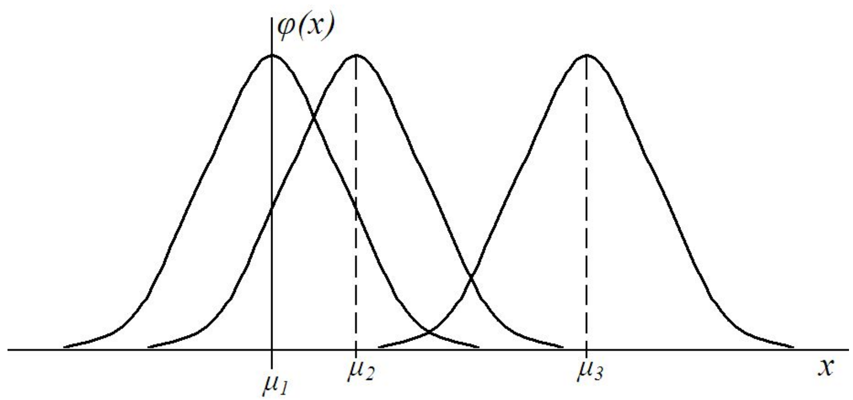
4) Максимум кривой распределения приходится на $x=\mu$.

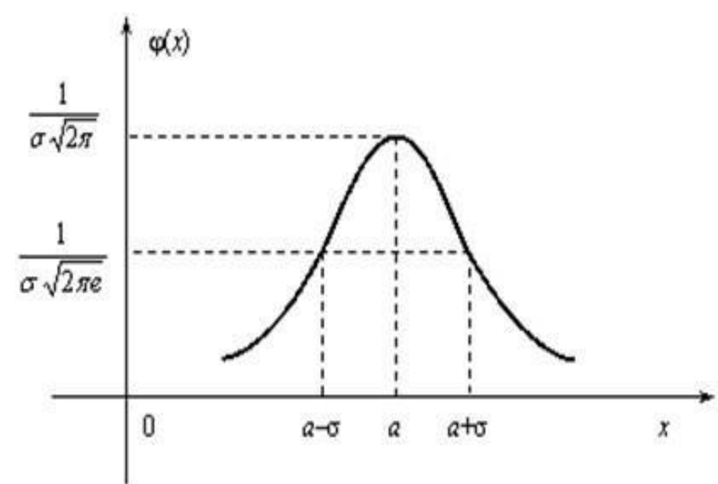
$$df(x)/_{dx} = 0$$

Функция $f(x)$ имеет максимум для результатов анализа, равных среднему арифметическому. Вероятность случайных ошибок тем меньше, чем больше их абсолютное значение. Изменение параметра μ не меняет форму кривой, а приводит к ее смещению по оси x.

5) Кривая $f(x)$ имеет две точки перегиба, симметричные относительно $x=\mu$ на расстояниях от центра рассеивания $\pm\sigma$.

$$d^2 f(x)/_{dx^2} = 0$$

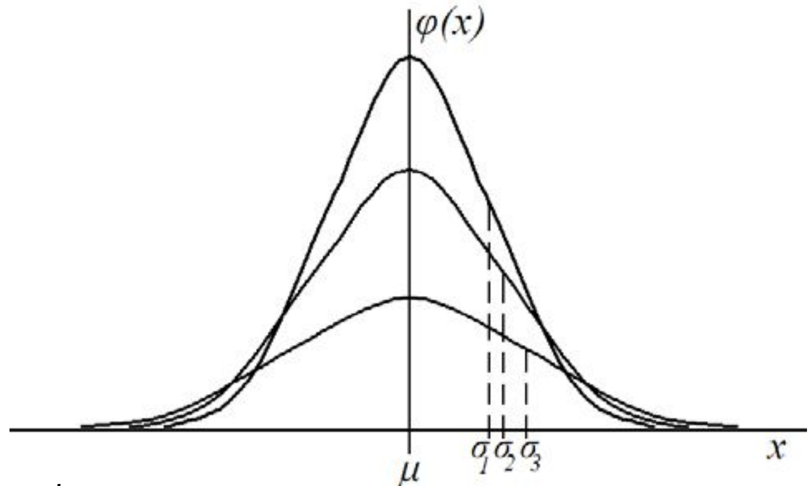




6) Значение функции плотности вероятности в максимуме определяется только параметром σ.

$$f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

Значение параметра σ (среднеквадратического отклонения) определяет степень «размытости» кривой, т. е. степень рассеяния случайной величины относительно ее математического ожидания.



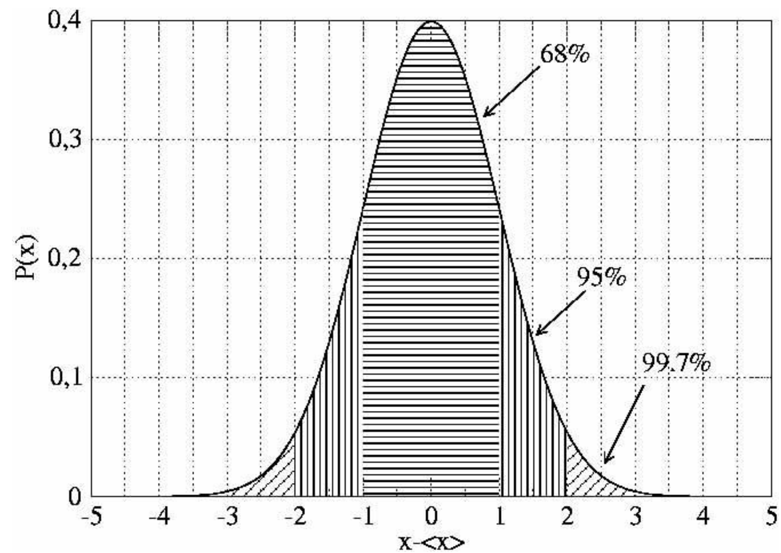
Функции плотности вероятности нормального распределения при постоянной μ, $\sigma^1 < \sigma^2 < \sigma^3$.

7) Правило трех сигм

Площадь фигуры с границами μ±σ составляет 0,68.

Площадь фигуры с границами μ±2σ составляет 0,95

Площадь фигуры с границами μ±3σ составляет 0,99



Это означает, что вероятность того, что случайная погрешность отдельного анализа не привысит по абсолютному значению σ (или 2σ, 3σ), составляет 0,68 (или 0,95 , 0,99).

Нормированное стандартное распределение. Функция Лапласа

- Применение закона нормального распределения Гаусса для оценки результатов химического анализа не является удобным, т. к. табулирование функции *f(x)* предполагает создание отдельных таблиц для каждой пары значений μ и σ. Для решения данной проблемы вводится коэффициент Лапласа *u*:

$$u = \frac{(x-\mu)}{\sigma}$$

- Случайная величина *u* есть мера рассеяния случайной величины x относительно μ в единицах измерения σ.

- Матожидание μ_u=0, т.к. $M(u)=M((x-\mu)/\sigma)=1/\sigma(M(x)-\mu)= 1/\sigma(\mu - \mu)=0$

- Дисперсия σ_u=1, т.к. $D(u)=D(x-\mu)/\sigma)= 1/\sigma^2 D(x- \mu)=\sigma^2 /\sigma^2=1$

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$ - функция (плотности вероятности) Гауса-Лапласа

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du$$ - функция (интегральная) Лапласа - ТАБУЛИРУЕТСЯ!!!

- Результаты многократного химического анализа и сопутствующих им случайные погрешности принято характеризовать с помощью двух статистических критериев:

- ширины доверительного интервала, внутри которого могут лежать результаты отдельных анализов

Доверительным называют ***интервал***, который покрывает неизвестный параметр с заданной надёжностью (вероятностью).

-доверительной вероятности α того, что они не выпадут из этого интервала.

Доверительная вероятность – вероятность того, что значение параметра генеральной совокупности находится в построенном для него доверительном интервале.

Расчет функции Лапласа

$\alpha = P(a \le x \le в) = F(в) - F(a) = \int_a^в f(x)dx$ - в общем случае

Для функции Лапласа

$\alpha = P(a \le x \le в) = \Phi(в) - \Phi(a) = \Phi((в-\mu)/\sigma) - \Phi((a-\mu)/\sigma)$

$\Phi(u_i) = F(u_i) - F(0) = P(0 \le u \le u_i) = P(\mu \le x \le (\mu + u_i\sigma)) = \int_0^{u_i} f(u)du = \frac{1}{\sqrt{2\pi}} \int_0^{u_i} e^{-u^2/2} du$

Т.к. интеграл обладает симметрией, то в таблице задается доверительная вероятность α и функция Лапласа для положительной u.

$$\frac{1}{\sqrt{2\pi}} \int_0^{u} e^{-u^2/2} du = \frac{1}{\sqrt{2\pi}} \int_u^{0} e^{-u^2/2} du$$

Тогда, для нахождения доверительной вероятности того, что случайная погрешность отдельного анализа не превышает величины $\pm u\sigma$ табличные значения вероятности α следует увеличить вдвое.

$$\boxed{P((x - u\sigma) \le x_i \le (x + u\sigma)) = 2\alpha}$$

Статистика малых выборок. Распределение Стьюдента.

- $2 \leq n \leq 20$

- Если полученные результаты заведомо подчиняются нормальному закону, то для их статистической обработки используют *t-распределение Стьюдента*.

- Связывает три основных характеристики выборочной совокупности:

- ширину доверительного интервала
- соответствующую ему доверительную вероятность
- объем выборки (число степеней свободы $f=n-1$)

Коэффициент нормировки Стьюдента t для выборочной совокупности с дисперсией S^2 имеет вид

$$t = (x - \mu)/S$$

Случайная величина t есть мера рассеяния случайной величины x относительно μ в единицах измерения S.

Эта дробь имеет распределение Стьюдента

Аналитический вид зависимости сложен (его не приводят) и используют таблицы.

Табулированными являются коэффициенты Стьюдента $t_{p,f}$ при заданных значениях доверительной вероятности p и числе степеней свободы $f = n - 1$.

Закон Стьюдента – это закон распределения ошибок измерений нормальных (Гауссовских) случайных величин.

Сравнение распределений Гауса-Лапласа и Стьюдента

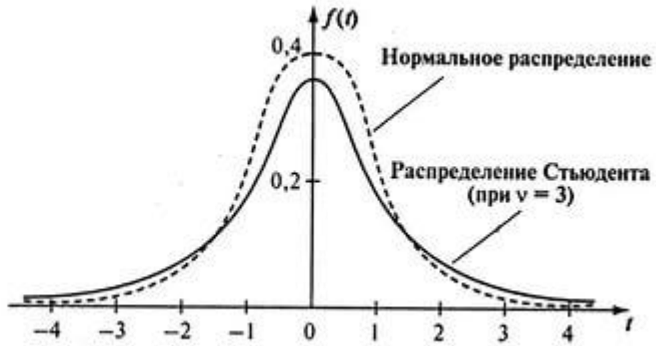
Распределение

Гауса-Лапласа

n→∞, n≥20

$$u = \frac{(x-\mu)}{\sigma}$$

$$\sigma = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n}}$$



Распределение

Стьюдента

2≤n≤20

$$t = (x-\mu)/S$$

$$s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$$

u≠t

σ≠S

t→u

S→σ

при n→∞

Доверительная вероятность, оцененная по Стьюденту меньше, доверительной вероятности по нормальному распределению Гауса-Лапласа (т.е. учитывает неполноту выборки).

При одинаковой ширине интервала доверительная вероятность по Стьюденту всегда меньше доверительной вероятности распределения Гаусса – Лапласа.

Доверительная оценка результатов эксперимента

- Для **единичного результата** анализа $t = (x - \mu)/S$

$$P(\mu - ts \le x_i \le \mu + ts)$$

Однако при анализе результатов эксперимента, как правило, необходима доверительная оценка не единичного, а среднего результата $\bar{x}$, которая с учетом того, что $S_{\bar{x}} = S/\sqrt{n}$ имеет вид

$$\boxed{\mu = \bar{x} \pm ts/\sqrt{n}}$$

$$\mu = \bar{x} \pm \Delta x$$

- Т.о. для **среднего результата** анализа

$$P(\bar{x} - tS/\sqrt{n} \le \mu \le \bar{x} + tS/\sqrt{n})$$

$$\Delta x = \pm ts/\sqrt{n}$$

$$P(\bar{x} - \Delta x \le \mu \le \bar{x} + \Delta x)$$

Для оценки случайной погрешности прямых многократных измерений некоторой физической величины *x* необходимо выполнить следующие расчеты:

- Оценить среднее арифметическое значение результатов измерений

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Вычислить среднеквадратичное отклонен

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{(n-1)}}$$

- Определить число степеней свободы $f = n-1$
- Выбрать доверительную вероятность $\alpha = 0{,}95$ (для большинства работ в курсе химии).
- Зная f и α по таблице определить коэффициент Стьюдента $t_{\alpha f}$.
- Определить доверительный интервал (погрешность серии многократных измерений)

$$\Delta x = \pm ts/\sqrt{n}$$

- Записать результат в виде: $\underline{x} = x \pm \Delta x$
- Результат читается как: *С заданной (0,95) доверительной вероятностью (надежностью) искомая величина заключена в доверительном интервале от….. до ….*