Определение понятия вероятности

- *Классическое (комбинаторное) определение*
  Если событие может приводить к N равновозможным различным исходам, и в n случаях проявляется признак A, то вероятность A есть

$$\boxed{P(A) = n/N}$$

равновозможно $\equiv$ равновероятно?

- *Частотное определение*
  Пусть событие повторяется большое число раз N и в n случаях обладает признаком A. Если исходы событий в этой последовательности взаимно независимы, то вероятность признака A

$$\boxed{P(A) = \lim_{N\to\infty} n/N}$$

- *Аксиоматическое определение*
  Рассмотрим пространство Ω, элементы которого назовем элементарными событиями. Определим в нем меру, называемую вероятностью, обладающую следующими свойствами:
  (1) Для любого элементарного события или множества событий $A \in \Omega$ $\boxed{P(A) \geq 0}$
  (2) Вероятность, связанная со всем пространством, $\boxed{P(\Omega) = 1}$
  (3) Если множества элементарных событий A и B не имеют общих элементов, вероятность объединения A и B представляет собой сумму двух соответствующих вероятностей:
  $\boxed{P(A \cup B) = P(A) + P(B)}$

  Следствие: Пусть Φ — пустое множество событий. Тогда $\Phi \cup \Omega = \Omega$, и Ω не имеет общих элементов с Φ. Следовательно,
  $\boxed{P(\Omega \cup \Phi) = P(\Omega) + P(\Phi) = P(\Omega)}$
  и $\boxed{P(\Phi) = 0}$

Другие следствия аксиом вероятности:

$P(\overline{A}) = 1 - P(A)$, где $\overline{A}$ комплементарно A

$$P(A \cup \overline{A}) = 1$$

$$0 \leq P(A) \leq 1$$

если $A \subset B$, то $P(A) \leq P(B)$

Сложение вероятностей

Пусть $C = (A \cup B) \equiv A + B$ есть множество элементарных событий, которые встречаются либо в A, либо в B, либо и в том и в другом. Пусть $D = (A \cap B) \equiv AB$ есть множество тех событий, которые встречаются одновременно как в A, так и в B.

$$P(C) = P(A) + P(B) - P(D)$$

Если наборы событий исключают друг друга, т.е. $D = \Phi$, то

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

Умножение вероятностей

Пусть событие E_1 может осуществляться n_1 способами, событие E_2 — n_2 способами, а событие $E_1 \cap E_2$ — n_{12} способами, причем полное число возможных исходов равно N.

$P(E_1) = \frac{n_1}{N}$, $P(E_1 E_2) = \frac{n_{12}}{N}$

$P(E_1 E_2) = \frac{n_1}{N} \cdot \frac{n_{12}}{n_1} = P(E_1)\frac{n_{12}}{n_1}$

$\frac{n_{12}}{n_1}$ — вероятность E_2 при условии, что произошло E_1 или *условная вероятность* события E_2, $P(E_2/E_1)$.

$$\boxed{P(E_1 E_2) = P(E_1) \cdot P(E_2/E_1)}$$

Если тот факт, что E_1 произошло, не влияет на вероятность события E_2, то говорят, что событие E_2 *не зависит* от события E_1. Тогда $P(E_2/E_1) = P(E_2)$ и $P(E_1 E_2) = P(E_1)P(E_2)$.

Дискретные распределения

Событие E с несколькими различными исходами называется *случайным*. Если E — переменная числовая величина, то такая величина называется *случайной*.

Если существует лишь конечное число ν исходов e_i события или бесконечное счетное число исходов, то каждому e_i приписывается вероятность P_i, т.е.

$$P(E = e_i) = P_i.$$

Множество значений вероятностей P_i мы называем *дискретным распределением* вероятности.

Биномиальное распределение (ν=2)

Пусть у события 2 исхода: "благоприятный" с вероятностью p и "неблагоприятный" с вероятностью $1-p$. Найдем вероятность r благоприятных исходов при полном числе исходов N. В этом случае r — случайная величина, пробегающая значения от 0 до N.

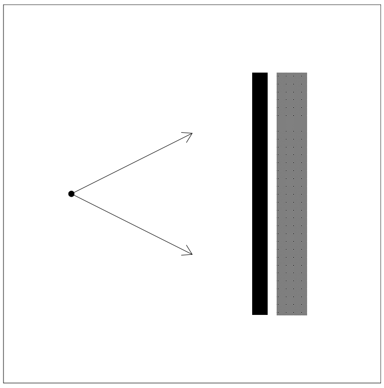
$$P(r) = C_N^r p^r (1-p)^{N-r} = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}$$

Распределение случайной величины r — биномиальное (Бернулли), имеет 2 параметра: p и N.

Заметим, что

$$\sum_{r=0}^{N} P(r) = [p + (1-p)]^N = 1,$$

как и для любого дискретного распределения.



Вероятность поглощения частицы в поглотителе ε. Какова вероятность запуска счетчика (счетчик срабатывает при попадании хотя бы одной частицы)?

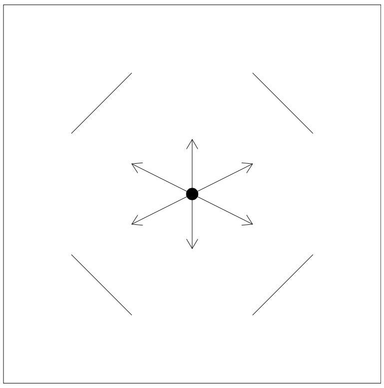
обе поглотились	ε^2
одна поглотилась	$2\varepsilon(1-\varepsilon)$
обе прошли	$(1-\varepsilon)^2$

Ясно, что $\sum P = [\varepsilon + (1-\varepsilon)]^2 = 1$

$\boxed{\text{вероятность запуска } p = 1 - \varepsilon^2}$

Детектор охватывает телесный угол $\alpha \cdot 4\pi$ стерадиан, $\alpha < 1$. Из места встречи пучков вылетает N частиц, распределенных изотропно. Найти аксептанс (вероятность геометрического попадания).

$$\varepsilon_k = C_N^k \alpha^k (1-\alpha)^{N-k}$$



Обычно для запуска детектора требуется не менее 2-х частиц.

$$\varepsilon_{tot} = 1 - \varepsilon_0 - \varepsilon_1 = 1 - (1-\alpha)^N - N\alpha(1-\alpha)^{N-1}$$

Вероятность зарегистрировать все N есть $\varepsilon_N = \alpha^N$

$$\alpha = 0.5$$

N	2	3	4	7	10
ε_{tot}	0.25	0.5	0.69	0.94	0.99
ε_N	0.25	0.125	0.062	0.008	0.001

Распределение Пуассона ($\nu = \infty$)

Допустим нас интересует вероятность того, что за данный промежуток времени произойдет r событий. При этом выполняются следующие предположения:

- Событие в момент t не зависит от истории, т.е. событий до момента t
- Вероятность события за малый интервал времени δt пропорциональна длительности интервала $P_1(t, t + \delta t) = \mu \delta t + o(\delta t)$
- Вероятность двух или большего числа событий за тот же промежуток мала $P_{\geq 2}(t, t + \delta t) = 0 + o(\delta t)$

Вычислим в этих предположениях вероятность того, что за $(0, t)$ произойдет r событий.

За промежуток времени $(0, t + \delta t)$ не произойдет ни одного события, если не будет событий в интервалах $(0, t)$ и $(t, t + \delta t)$, т.е.

$$P_0(t + \delta t) = P_0(t)(1 - \mu\delta t + o(\delta t))$$

$$\frac{P_0(t+\delta t) - P_0(t)}{\delta t} = -\mu P_0(t)$$

при $\delta t \to 0$ получим $\frac{dP_0(t)}{dt} = -\mu P_0(t)$

Решение этого уравнения $P_0(t) = Ae^{-\mu t}$. При $t = 0$ имеем $P_0(0) = 1$; значит, $A = 1$ и $\boxed{P_0(t) = e^{-\mu t}}$

Одно событие: Если $r = 1$, то либо событие произошло в интервале $(0, t)$, либо в $(t, t+\delta t)$. Тогда

$$P_1(t+\delta t) = P_1(t)(1-\mu\delta t + o(\delta t)) + P_0(t)(\mu\delta t + o(\delta t)),$$

$$\frac{P_1(t+\delta t) - P_1(t)}{\delta t} = -\mu P_1(t) + \mu P_0(t)$$

при $\delta t \to 0$ получим $\frac{dP_1(t)}{dt} = -\mu P_1(t) + \mu e^{-\mu t}$.

Решение этого уравнения $P_1(t) = \mu t e^{-\mu t}$.

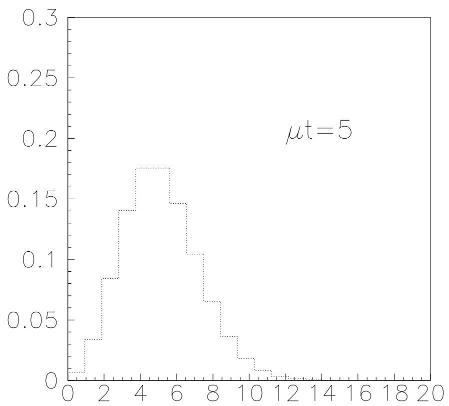
В общем случае: При $r \geq 1$

$$\frac{dP_r(t)}{dt} = -\mu P_r(t) + \mu P_{r-1}(t)$$

Итак, $P_r(t) = \frac{(\mu t)^r}{r!} e^{-\mu t}$.

Найдем полную вероятность (вероятность любого значения r).

$\sum_{r=0}^{\infty} P_r(t) = e^{-\mu t} \sum_{r=0}^{\infty} \frac{(\mu t)^r}{r!} = e^{-\mu t} \cdot e^{\mu t} = 1$, как и следовало ожидать.



$P_r(t)$ называется распределением Пуассона для случайной величины r и содержит единственный параметр μ.

При $r = 0$: $P_0(t) = e^{-\mu t}$

При $r = 1$: $P_1(t) = \mu t e^{-\mu t}$

Связь биномиального распределения с распределением Пуассона

Пусть источник испускает в среднем μt частиц за интервал времени $(0, t)$. Разобьем этот интервал на N интервалов, в каждом из которых есть вероятность p наблюдать одну частицу. Вероятность наблюдать r частиц в N интервалах:

$$P_r(t) = C_N^r p^r (1-p)^{N-r}$$

Среднее число частиц $Np = \mu t$, т.е. $p = \mu t / N$

$$P_r(t) = \frac{N(N-1)...(N-r+1)}{r!}\left(\frac{\mu t}{N}\right)^r \left(1 - \frac{\mu t}{N}\right)^{N-r} = \frac{N(N-1)...(N-r+1)}{N^r}\frac{(\mu t)^r}{r!}\left(1 - \frac{\mu t}{N}\right)^{N-r} = 1\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)...\left(1 - \frac{r-1}{N}\right)\frac{(\mu t)^r}{r!} \cdot \frac{\left(1-\frac{\mu t}{N}\right)^N}{\left(1-\frac{\mu t}{N}\right)^r}.$$

$$\lim_{N\to\infty} P_r(t) = \frac{(\mu t)^r}{r!} e^{-\mu t}$$

Таким образом, биномиальное распределение стремится к распределению Пуассона при $N \to \infty$ и фиксированном Np.

Распределение Пуассона $P_r(\mu) = \frac{\mu^r}{r!} e^{-\mu}$.

Для практических вычислений удобна формула Стирлинга $r! \approx \sqrt{2\pi r}(\frac{r}{e})^r$

Сложение (свертка) двух Пуассоновских распределений. Есть 2 случайные независимые величины, распределенные по Пуассону, n_1 с параметром μ_1 и n_2 с параметром μ_2. Как распределена случайная величина $n = n_1 + n_2$?

$$P_n = \sum_{n_1, n_2} P^{(1)}_{n_1} \cdot P^{(2)}_{n_2} = \sum_{n_1=0}^{n} P^{(1)}_{n_1} \cdot P^{(2)}_{n-n_1} =$$

$$\sum_{n_1=0}^{n} e^{-\mu_1} \frac{\mu_1^{n_1}}{n_1!} \cdot e^{-\mu_2} \frac{\mu_2^{n-n_1}}{(n-n_1)!} = \frac{e^{-(\mu_1+\mu_2)}}{n!}.$$

$$\sum_{n_1=0}^{n} C_n^{n_1} \mu_1^{n_1} \mu_2^{n-n_1} = \frac{e^{-(\mu_1+\mu_2)}}{n!} (\mu_1 + \mu_2)^n$$

Снова получаем распределение Пуассона с параметром $(\mu_1 + \mu_2)$!

Непрерывные распределения

Пусть X — случайная величина, принимающая любые значения в интервале $a \leq X \leq b$.

Если $P(X = x) \neq 0$, то $\sum\limits_x P(X = x) = \infty > 1$.
Если $P(X = x) = 0$, то $\sum\limits_x P(X = x) = 0$?
Суммирование происходит по несчетному множеству точек x!

Будем говорить о вероятности того, что X заключено в малом интервале $P(x \leq X \leq x + dx) = f(x)dx \geq 0$

$$f(x)dx = P(a \leq X \leq x + dx) - P(a \leq X \leq x) = F(x + dx) - F(x)$$

$$f(x) = \frac{dF(x)}{dx} \geq 0; \; F(y) = \int\limits_a^y f(x)dx$$

$f(x)$ — *плотность вероятности* или *функция непрерывного распределения*.

В наиболее общем случае $-\infty < X < +\infty$ и тогда

$$F(x) = P(-\infty < X \leq x) = \int\limits_{-\infty}^{x} f(y)dy,$$

причем

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

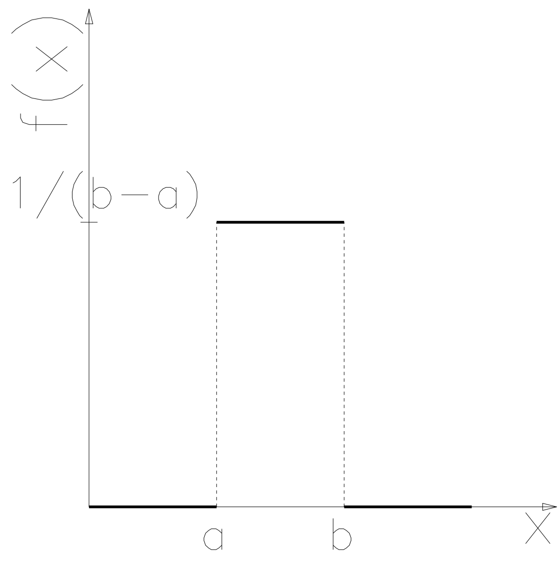
$F(x)$ называется *распределением накопленной вероятности*. Это неубывающая функция с предельными значениями

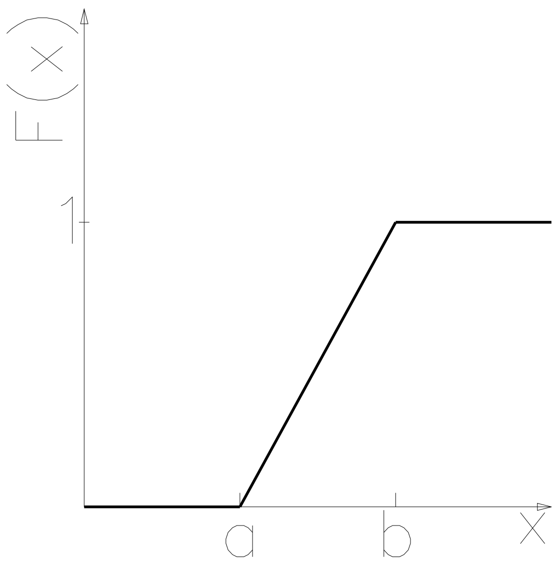
$$F(-\infty) = 0;\ F(+\infty) = 1.$$

Равномерное распределение

Если все значения случайной величины, лежащие в интервале $[a, b]$ равновероятны, то плотность вероятности и распределение накопленной вероятности:

$$f(x) = \begin{cases} 0 \\ \frac{1}{b-a} \\ 0 \end{cases} \quad F(x) = \begin{cases} 0 & \text{при } -\infty < x < a \\ \frac{x-a}{b-a} & \text{при } a \le x \le b \\ 1 & \text{при } b < x < +\infty \end{cases}$$





Задача. Пусть случайные величины x_1, x_2 независимы и распределены равномерно, каждая в интервале [0;1]. Найти распределение величины $x = x_1 + x_2$.

$$P(x) = \int\limits_0^1 P^{(1)}(x_1)P^{(2)}(x - x_1)dx_1$$

- при $0 \leq x \leq 1$ пределы интегрирования по x_1 будут от 0 до x, т.к. при $x_1 > x$ $P^{(2)}(x - x_1) = 0$. В результате интеграл равен

$$P(x) = \int\limits_0^x 1 \cdot dx_1 = x$$

- при $1 \leq x \leq 2$ пределы интегрирования по x_1 будут от $x - 1$ до 1, т.к. при $x - x_1 > 1$ $P^{(2)}(x - x_1) = 0$. В результате интеграл равен

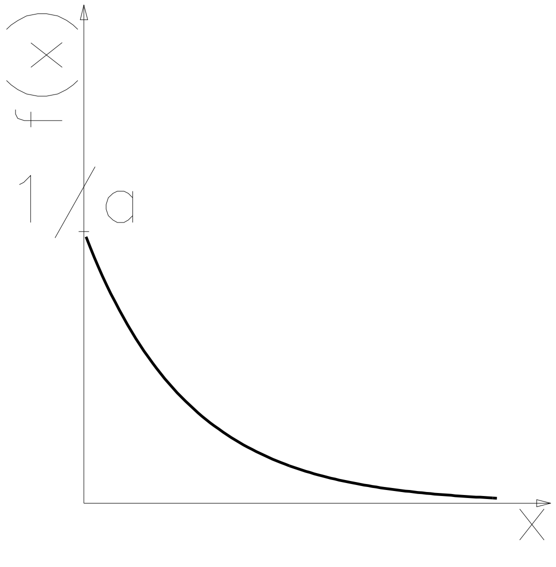
$$P(x) = \int\limits_{x-1}^1 1 \cdot dx_1 = 2 - x$$

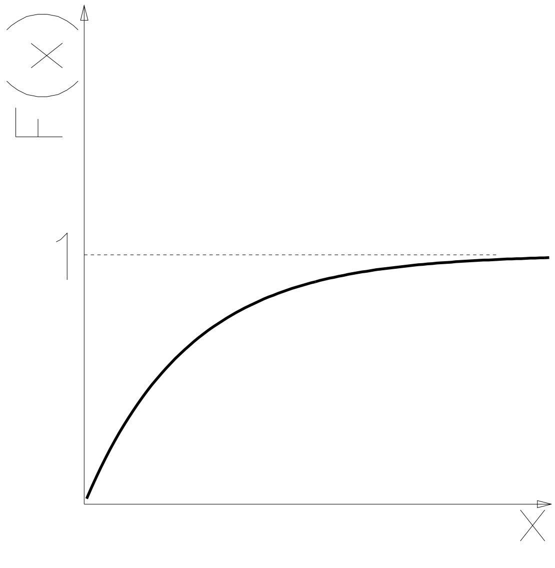
- при $x < 0$ или $x > 2$ $P(x) = 0$

Экспоненциальное распределение

$$f(x) = \begin{cases} \frac{1}{a} e^{-\frac{x}{a}} & \text{при } 0 \le x < \infty \\ 0 & \text{при } x < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{a}} & \text{при } 0 \le x < \infty \\ 0 & \text{при } x < 0 \end{cases}$$

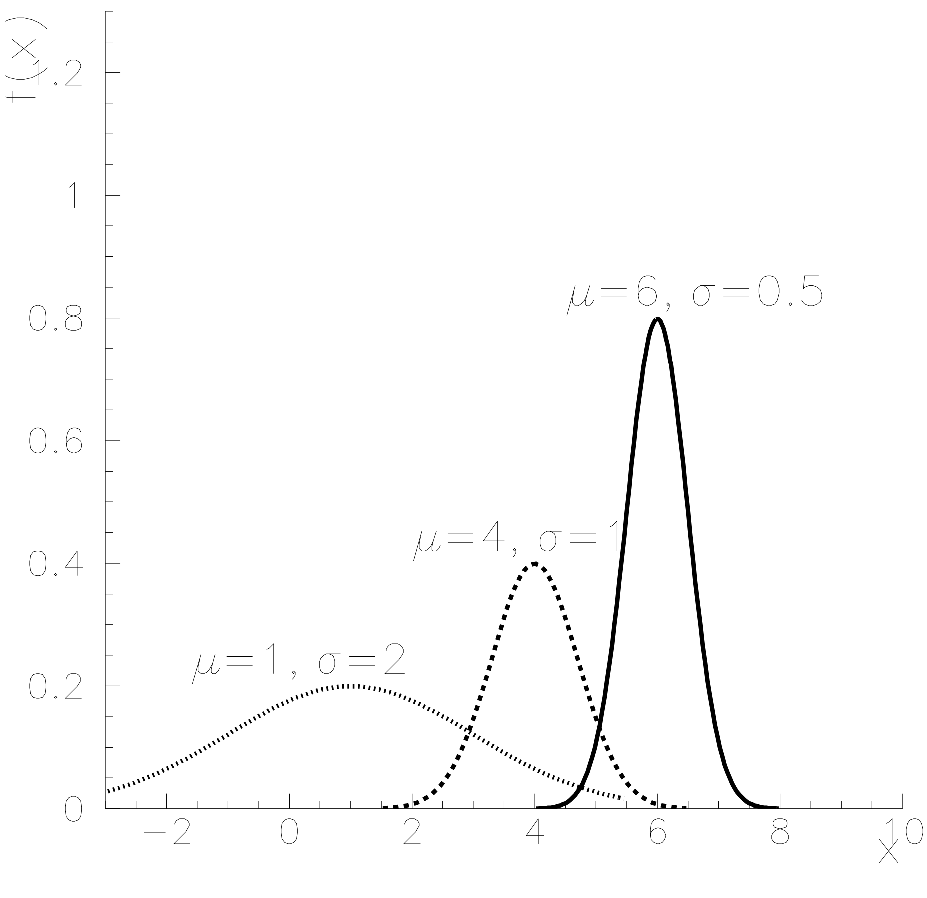




Нормальное (Гауссово) распределение

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

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



Общие характеристики распределения

- *Математическое ожидание* случайной величины x есть среднее значение x с учетом вероятности реализации каждого значения x.

  дискретное: непрерывное:

  $$\mathcal{E}(x) = \sum_r rP(x = r) \quad \mathcal{E}(x) = \int_{-\infty}^{+\infty} xf(x)dx$$

  $\mathcal{E}(x)$ или μ характеризует положение распределения x

- *Математическое ожидание* функции $h(x)$ случайной величины x есть средняя величина $h(x)$ по возможным значениям x.

  $\mathcal{E}[h(x)] = \sum_r h(r)P(x = r)$ (дискретное)

  $\mathcal{E}[h(x)] = \int_{-\infty}^{+\infty} h(x)f(x)dx$ (непрерывное)

- *Дисперсия* случайной величины x равна $\mathcal{E}[(x-\mu)^2]$. Ее обозначают $\mathcal{D}(x)$ или $\sigma^2(x)$. σ называют средним квадратичным или стандартным отклонением x. Это мера разброса x относительно среднего значения μ.

- *Асимметрия* распределения характеризуется параметром $\gamma_1 = \frac{\mathcal{E}[(x-\mu)^3]}{\sigma^3}$.

$\gamma_1 < 0$	$f(x)$ вытянута влево от μ
$\gamma_1 = 0$	$f(x)$ симметрична относительно μ
$\gamma_1 > 0$	$f(x)$ вытянута вправо от μ

Свойства μ и $\mathcal{D}$

$$\mathcal{E}[h_1(x) + h_2(x)] = \mathcal{E}[h_1(x)] + \mathcal{E}[h_2(x)]$$

$$\mathcal{E}[kh(x)] = k\mathcal{E}[h(x)]$$

$$\sigma^2 = \mathcal{D}(x) = \mathcal{E}[(x-\mu)^2] = \mathcal{E}(x^2 - 2\mu x + \mu^2) =$$
$$\mathcal{E}(x^2) - 2\mu\mathcal{E}(x) + \mu^2 = \mathcal{E}(x^2) - 2\mu^2 + \mu^2 =$$
$$\underline{\mathcal{E}(x^2) - \mu^2}$$

$$\mathcal{D}(kx) = \mathcal{E}[(kx - k\mu)^2] = k^2\mathcal{D}(x)$$

Вычисление μ, σ для основных распределений

Биномиальное распределение с параметрами N и p

$P_r = C_N^r p^r q^{N-r}$, где $q = 1 - p$

$$\mu = \sum_{r=0}^{N} \frac{rN!}{(N-r)!r!} p^r q^{N-r} =$$

$$Np \sum_{r=1}^{N} \frac{(N-1)!}{(N-r)!(r-1)!} p^{r-1} q^{N-r} =$$

$$Np \sum_{s=0}^{N-1} \frac{(N-1)!}{(N-s-1)!s!} p^s q^{N-s-1},$$

где $s = r - 1$. Тогда при $n = N - 1$ $\mu = Np \sum\limits_0^n C_n^s p^s q^{n-s} = Np$.

Аналогично получим

$$\mathcal{E}(x^2) = \sum_0^N r^2 P_r = \sum_2^N r(r-1)P_r + \sum_1^N rP_r =$$
$$N(N-1)p^2 + Np$$

Отсюда $\sigma^2 = \mathcal{E}(x^2) - \mu^2 = Np(1-p) = Npq$.

Заметим, что σ растет с N.

Можно аналогично показать, что $\mathcal{E}[(x-\mu)^3] = Np(1-p)(1-2p)$. Тогда асимметрия

$\gamma_1 = \frac{Np(1-p)(1-2p)}{(Np(1-p))^{3/2}} = \frac{1-2p}{\sqrt{Np(1-p)}}$, т.е.

$$\gamma_1 < 0, \text{ если } p > \tfrac{1}{2}$$
$$\gamma_1 = 0, \text{ если } p = \tfrac{1}{2}$$
$$\gamma_1 > 0, \text{ если } p < \tfrac{1}{2}$$

При фиксированном p для любого p $\lim\limits_{N\to\infty} \gamma_1 = 0$.

Пример. Пусть новая случайная величина $h(x) = x/N$. Тогда

$\mathcal{E}(\frac{x}{N}) = \frac{Np}{N} = p$, $\mathcal{D}(\frac{x}{N}) = \frac{1}{N^2}\mathcal{D}(x) = \frac{Npq}{N^2} = \frac{pq}{N}$ и $\sigma(\frac{x}{N}) = \sqrt{\frac{pq}{N}}$.

Типичный пример — вычисление вероятности регистрации: всего попало x событий из полного числа N.

Вероятность регистрации

Пусть детектор охватывает некоторую часть телесного угла. Вероятность регистрации

$$\varepsilon_{det} = \frac{\int\limits_{det} \frac{d\sigma}{d\Omega} d\Omega}{\int\limits_{4\pi} \frac{d\sigma}{d\Omega} d\Omega} = \frac{\sigma_{vis}}{\sigma_{tot}}$$

На практике интеграл по телесному углу удобно вычислять методом Монте-Карло, т.е. имитировать картину случайными числами. Пусть N — полное число разыгранных событий, n из которых попадают в детектор. Тогда $\varepsilon_{det} \approx \frac{n}{N} = \varepsilon_0$. Ясно, что при фиксированном N величина n распределена биномиально. Тогда в качестве оценки $\varepsilon_{det} = \varepsilon_0 \pm \sqrt{N\varepsilon_0(1-\varepsilon_0)}\frac{1}{N}$.

Действительно, $\mathcal{D}(n) = N\varepsilon_0(1-\varepsilon_0)$, $\mathcal{D}(\varepsilon_{det}) = \frac{1}{N^2}\mathcal{D}(n)$.

Распределение Пуассона

$$P_r = \frac{a^r}{r!}e^{-a}$$

$\mu = \mathcal{E}(x) = \sum\limits_{r=0}^{\infty} r\frac{a^r}{r!}e^{-a} = a\sum\limits_{r=1}^{\infty}\frac{a^{r-1}}{(r-1)!}e^{-a} = a\sum\limits_{s=0}^{\infty}\frac{a^s}{s!}e^{-a} = a$, где $s = r-1$.

Итак, $\boxed{\mu = a}$

$\mathcal{E}(x^2) = \sum\limits_{0}^{\infty}[r(r-1)+r]\mu^r\frac{e^{-\mu}}{r!} = \mu^2\sum\limits_{2}^{\infty}\mu^{r-2}e^{-\mu}\frac{1}{(r-2)!} + \mu = \mu^2+\mu.$

Отсюда $\mathcal{D}(x) = \mathcal{E}(x^2) - \mu^2 = \mu^2+\mu-\mu^2 = \mu$, т.е. $\boxed{\sigma = \sqrt{\mu}}$.

Наконец, $\mathcal{E}[(x-\mu)^3] = \mu$, поэтому асимметрия $\gamma_1 = \frac{\mu}{\mu^{3/2}} = \mu^{-1/2}$.

γ_1 всегда положительна и $\lim\limits_{\mu\to\infty}\gamma_1 = 0$, т.е. с ростом μ распределение становится все более симметричным.

Определение сечения реакции

Пусть в эксперименте зарегистрировано n событий изучаемой реакции. Нужно определить полное сечение σ_{tot}:

$n = \mathcal{L}\sigma_{tot}\varepsilon_{det}$, где $\mathcal{L}$ — интегральная светимость, характеризующая интенсивность пучков, их размеры и частоту встречи.

n распределено по Пуассону, т.е. $\mathcal{D}(n) = n$ и $\sigma_n = \sqrt{n}$.

Оценка $\boxed{\sigma_{tot} = \frac{n}{\mathcal{L}\varepsilon_{det}} \pm \frac{\sqrt{n}}{\mathcal{L}\varepsilon_{det}}}$.

Равномерное распределение

$$\mu = \int_a^b \frac{x}{b-a}dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2},$$

т.е. середина интервала, как и следовало ожидать.

$\mathcal{E}(x^2) = \int\limits_a^b \frac{x^2}{b-a}dx = \frac{b^2+ab+a^2}{3}$

$\mathcal{D}(x) = \mathcal{E}(x^2) - \mu^2 = \frac{4(b^2+ab+a^2)-3(a+b)^2}{12} = \frac{(b-a)^2}{12}$, $\boxed{\sigma = \frac{b-a}{\sqrt{12}}}$

$\gamma_1 = 0$, т.к. распределение симметрично.

Экспоненциальное распределение

$$\boxed{f(x) = \frac{1}{a}e^{-x/a}}$$

$$\mu = \int_0^\infty \frac{x}{a}e^{-\frac{x}{a}}dx = a\int_0^\infty ye^{-y}dy =$$

$$a[-ye^{-y}|_0^\infty + \int_0^\infty e^{-y}dy] = a$$

$$\mathcal{E}(x^2) = \int_0^\infty \frac{x^2}{a}e^{-\frac{x}{a}}dx = a^2\int_0^\infty y^2e^{-y}dy = a^2[-y^2e^{-y}|_0^\infty +$$
$$2\int_0^\infty ye^{-y}dy] = 2a^2$$

$$\mathcal{D}(x) = \mathcal{E}(x^2) - \mu^2 = 2a^2 - a^2 = a^2;\ \sigma = a$$

$\gamma_1 = 2$, т.е. распределение вытянуто вправо.

Нормальное распределение

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

$$\mu = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2} dx$$

Обозначим $\omega = \frac{x-a}{\sigma}$, $d\omega = \frac{dx}{\sigma}$, и, учитывая, что интеграл от нечетной функции равен 0,

$$\mu = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma\omega + a) e^{-\omega^2/2} d\omega =$$

$$0 + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\omega^2/2} d\omega = a$$

Итак, как и следовало ожидать, $\boxed{\mu = a}$.

Дисперсия $\mathcal{D}(x) = \mathcal{E}[(x-\mu)^2] =$

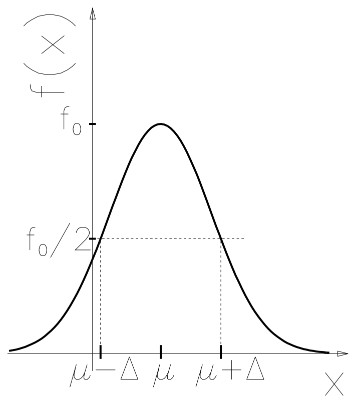
$$\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \int_{-\infty}^{+\infty} \frac{\sigma^2\omega^2}{\sqrt{2\pi}} e^{-\omega^2/2} d\omega =$$

$$\frac{1}{2\pi}[-\sigma^2\omega e^{-\omega^2/2}]|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\omega^2/2} d\omega = \sigma^2,$$

что также следует из принятых обозначений.

И, наконец, асимметрия $\gamma_1 = 0$, т.к. распределение симметрично относительно μ.

Как ширина распределения Гаусса связана с σ? Обычно для характеристики ширины распределения используют ширину на полувысоте FWHM (Full Width at Half Maximum).



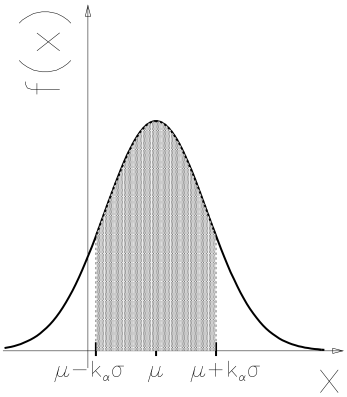
$f_0 = f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$

$\frac{f_0}{2} = \frac{1}{2\sqrt{2\pi}\sigma} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\Delta^2/2\sigma^2}$

Отсюда $\frac{\Delta^2}{\sigma^2} = 2ln2$ и $\Delta = \sqrt{2ln2}\sigma$

$$\boxed{FWHM = 2\Delta \approx 2.3548\sigma}$$

Какова вероятность того, что x лежит в интервале $\pm k_\alpha \sigma$ вокруг среднего?

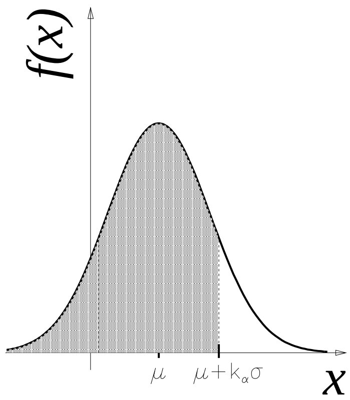


$$\alpha = \mathcal{P}[\mu - k_\alpha \sigma \leq x \leq \mu + k_\alpha \sigma]$$

α	k_α
0.	0.
0.500	0.674
0.683	1.000
0.900	1.645
0.950	1.960
0.990	2.576
0.9975	3.000
0.999	3.261

интервал	вероятность внутри	вероятность снаружи
1σ	0.683	0.317
2σ	0.955	0.045
3σ	0.9975	0.0025

Какова вероятность того, что x меньше какой-то фиксированной величины?



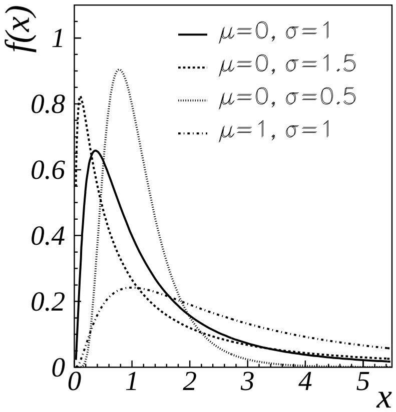
$\mathcal{P}(x \leq \mu + k\sigma) = 1 - \mathcal{P}(|x - \mu| \geq k\sigma)/2$. Тогда верхние пределы на разных уровнях достоверности в единицах σ:

$\mathcal{P}$	k
0.84	1
0.9	1.24
0.95	1.645
0.9775	2

Важность распределения Гаусса (далее Центральная предельная теорема): при воздействии многих случайных факторов результирующее распределение гауссово.

Логарифмическое нормальное распределение

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



В этом случае параметры μ и σ уже не совпадают со средним значением и стандартным отклонением. Обозначив $y = lnx$; $dx = e^y dy$,

$$\mathcal{E}(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(y-\mu)^2}{2\sigma^2}+y}dy =$$

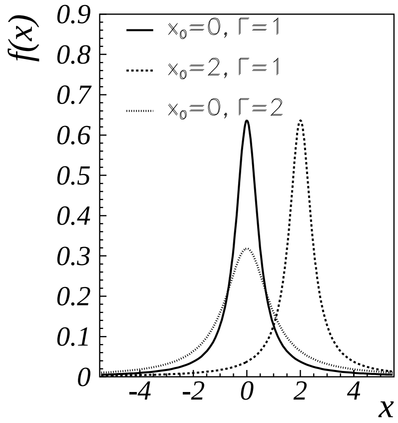
$$e^{\frac{\sigma^2}{2}+\mu} \cdot \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(y-\mu-\sigma^2)^2}{2\sigma^2}}dy = \underline{e^{\frac{\sigma^2}{2}+\mu}}.$$

$$\mathcal{E}(x^2) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(y-\mu)^2}{2\sigma^2}+2y}dy = \underline{e^{2\sigma^2+2\mu}}.$$

$$\mathcal{D}(x) = \mathcal{E}(x^2) - (\mathcal{E}(x))^2 = \underline{e^{\sigma^2+2\mu}(e^{\sigma^2}-1)}.$$

Распределение Коши

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \qquad -\infty < x < \infty$$



Распределение Коши — частный случай важного в физике частиц распределения **Брейта–Вигнера** с параметрами Γ и x_0, описывающего сечение резонансного рождения.

$$f(x) = \frac{1}{\pi} \cdot \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2}$$

Это распределение симметрично относительно x_0, поэтому и среднее значение равно x_0:

$$\mu = \lim_{R\to\infty} \frac{1}{\pi} \int_{-R}^{+R} \frac{x \cdot \Gamma/2}{\Gamma^2/4 + (x - x_0)^2} dx =$$

$$x_0 + \lim_{R\to\infty} \frac{\Gamma}{4\pi} [ln(\Gamma^2/4 + (x - x_0)^2)]_{-R}^{R} =$$

$$x_0 + \frac{\Gamma}{4\pi} \lim_{R\to\infty} ln[\frac{\Gamma^2/4 + R^2}{\Gamma^2/4 + (-R^2)}] = x_0 + 0$$

Дисперсия распределения Брейта-Вигнера:

$$\sigma^2 = \frac{\Gamma/2}{\pi} \int_{-\infty}^{\infty} \frac{(x-x_0)^2}{\Gamma^2/4 + (x-x_0)^2} dx =$$

$$\frac{\Gamma/2}{\pi} \int_{-\infty}^{\infty} [\frac{\Gamma^2}{4} - \frac{1}{\Gamma^2/4 + (x-x_0)^2}] dx = \infty - 1.$$

Т.е. дисперсия бесконечна.

Тем не менее, параметр Г характеризует ширину распределения, представляя собой ширину пика на полувысоте FWHM.

Распределение Ландау

В ядерной физике и физике частиц важным является распределение вероятности $f(\Delta;\beta)$ потерь энергии Δ заряженной частицей при прохождении слоя вещества заданной толщины. Такое распределение впервые было вычислено Ландау.

$$f(\Delta;\beta) = \frac{1}{\xi}\phi(\lambda), \quad 0 \leq \Delta < \infty,$$

$$\xi = \frac{2\pi N_A e^4 z^2 \rho \sum Z}{m_e c^2 \sum A} \cdot \frac{d}{\beta^2}, \ \epsilon' = \frac{I^2 e^{\beta^2}}{2 m_e c^2 \beta^2 \gamma^2},$$

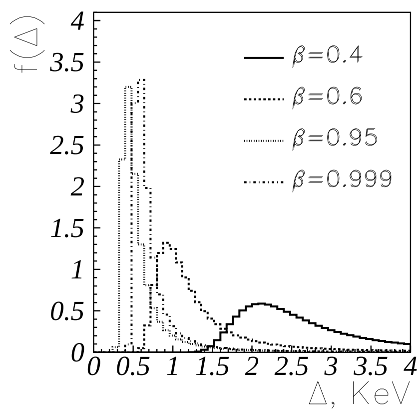
$$\lambda = \frac{1}{\xi}[\Delta - \xi(log\frac{\xi}{\epsilon'} + 1 - \gamma_E)].$$

N_A – число Авогадро, m_e и e – масса и заряд электрона, z – заряд частицы в единицах e, $\sum Z$ и $\sum A$ – сумма атомных номеров и атомных весов молекулы вещества, ρ – плотность, d – толщина слоя, $I = I_0 Z$ при $I_0 \approx 13.5eV$ – потенциал ионизации, $\gamma_E = 0.5772$ – постоянная Эйлера, $\beta = v/c$, $\gamma = 1/\sqrt{1-\beta^2}$.

Распределение от безрамерного параметра λ:

$$\phi(\lambda) = \frac{1}{2\pi i} \int_{\epsilon - i\infty}^{\epsilon + i\infty} e^{-u log u - \lambda u} du =$$

$$\phi(\lambda) = \frac{1}{\pi} \int_{0}^{\infty} e^{-u log u - \lambda u} sin \pi u du.$$



Такой интеграл вычисляется численно. На рисунке приведены распределения потерь энергии Δ для электрона, проходящего слой 4 мм газа аргона при различных скоростях β. Из-за бесконечного хвоста в распределении в сторону больших Δ среднее значение и моменты высоких порядков для распределения Ландау не существуют, т.е. интеграл

$\int\limits_0^\infty \Delta^n f(\Delta) d\Delta$ расходится при $n \geq 1$. Численно можно расчитать *наиболее вероятное значение* в зависимости от β.

$$\Delta_{mp} = \xi[log(\xi/\epsilon') + 0.198]$$

.

Моменты случайной величины

Начальный момент n-го порядка случайной величины x по определению равен $\boxed{\mu'_n = \mathcal{E}(x^n)}$. Заметим, что $\mu'_0 = 1,\ \mu'_1 = \mu$.

Центральный момент n-го порядка случайной величины x по определению равен $\boxed{\mu_n = \mathcal{E}[(x-\mu)^n]}$. Заметим, что $\mu_0 = 1,\ \mu_1 = 0,\ \mu_2 = \sigma^2$.

Полный набор моментов $\{\mu'_n\}$ или $\{\mu_n\}$ полностью определяет распределение вероятности. Часто случается, что проще найти набор моментов неизвестного распределения, чем непосредственно функцию распределения, а затем из набора моментов построить распределение.

Теорема. Если две плотности вероятности $f_1(x)$ и $f_2(x)$ имеют одинаковые моменты $\{\mu'_n\}$ и если можно разложить $f_1(x) - f_2(x)$ в ряд по степеням x, то $f_1(x) \equiv f_2(x)$.

Доказательство. Пусть $f_1(x) - f_2(x) = c_0 + c_1 x + c_2 x^2 + \ldots$. Тогда $\int\limits_{-\infty}^{\infty} [f_1(x) - f_2(x)]^2 dx = \int\limits_{-\infty}^{\infty} [f_1(x) - f_2(x)][c_0 + c_1 x + c_2 x^2 + \ldots] dx = c_0 \int (f_1 - f_2) dx + c_1 \int (x f_1 - x f_2) dx + c_2 \int (x^2 f_1 - x^2 f_2) dx + \ldots = c_0(1-1) + c_1(\mu'_1 - \mu'_1) + c_2(\mu'_2 - \mu'_2) + \ldots = 0$.

Следовательно, $\boxed{f_1(x) \equiv f_2(x)}$

Производящие функции моментов

Введем функцию, зависящую от всех моментов $\{\mu'_n\}$, которая сводит весь набор $\{\mu'_n\}$ к единственному выражению. Для этого введем вспомогательную переменную t.

Производящая функция начальных моментов случайной величины x:

$$M'_x(t) = \mathcal{E}(e^{xt}) = \begin{cases} \int\limits_{-\infty}^{\infty} e^{xt} f(x) dx, \\ \sum\limits_{r=0}^{\infty} e^{rt} P_r. \end{cases}$$

Разлагая экспоненту в ряд, $M'_x(t) = \mathcal{E}[1 + xt + \frac{(xt)^2}{2!} + \frac{(xt)^3}{3!} + \ldots] = 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \mu'_3 \frac{t^3}{3!} + \ldots$

Итак, μ'_n — коэффициент при $\frac{t^n}{n!}$.

Другой способ нахождения μ'_n из $M'_x(t)$ состоит в том, что производящую функцию дифференцируют n раз.

$$\frac{\partial^n M'_x(t)}{\partial t^n} = \int_{-\infty}^{\infty} x^n e^{xt} f(x) dx.$$

Отсюда следует

$$\frac{\partial^n M'_x(0)}{\partial t^n} = \int_{-\infty}^{\infty} x^n f(x) dx = \mu'_n.$$

Производящая функция центральных моментов случайной величины x:

$$M_x(t) = \mathcal{E}[e^{(x-\mu)t}], \quad \mu = \mu_1'.$$

Разлагая экспоненту в ряд, $M_x(t) = \mathcal{E}[1+(x-\mu)t+(x-\mu)^2\frac{t^2}{2!} + (x-\mu)^3\frac{t^3}{3!} + \ldots] = 1 + 0 + \mu_2\frac{t^2}{2!} + \mu_3\frac{t^3}{3!} + \ldots$,

так что μ_n — коэффициент при $\frac{t^n}{n!}$. Заметим, что $M(t) = e^{-\mu t}M'(t)$ и $M'(t) = e^{\mu t}M(t)$.

Если две производящие функции моментов $M'_{x_1}(t)$ и $M'_{x_2}(t)$ одинаковы, то и исходные распределения вероятностей совпадают.

Зачем нужны производящие функции моментов?

- По заданному распределению вероятности легко найти $M(t)$, а, значит, и моменты μ_n. Часто это проще, чем вычислять моменты из самого распределения вероятности.

- При нахождении распределения вероятности нескольких случайных величин бывает легче найти сперва $M_x(t)$, а затем уже само распределение $f(x)$.

Биномиальное распределение

$$\boldsymbol{P_r = C_N^r p^r (1-p)^{N-r}}$$

$$M'(t) = \sum_0^\infty e^{rt} P_r = \sum_0^\infty C_N^r (pe^t)^r (1-p)^{N-r} =$$

$$= (pe^t + 1 - p)^N$$

$$\frac{\partial M'(t)}{\partial t} = Npe^t (pe^t + 1 - p)^{N-1}$$

Следовательно, $\mu = \mu_1' = \frac{\partial M'(0)}{\partial t} = Np$.

$$\frac{\partial^2 M'(t)}{\partial t^2} = Npe^t (pe^t + 1 - p)^{N-1} +$$

$$N(N-1)p^2 e^t (pe^t + 1 - p)^{N-2}$$

$\mu_2' = Np + N(N-1)p^2$; $\sigma^2 = \mu_2' - (\mu_1')^2 = Np + p^2(N-1)N - N^2p^2 = Np(1-p)$

Распределение Пуассона $P_r = \frac{\mu^r}{r!}e^{-\mu}$

$$M'(t) = e^{-\mu} \sum_0^{\infty} \frac{(\mu e^t)^r}{r!} = e^{\mu(e^t-1)},$$

$$\frac{\partial M'(t)}{\partial t} = \mu e^t e^{\mu(e^t-1)}$$

Следовательно, $\mu_1' = \mu$.

Производящая функция центральных моментов

$$M(t) = e^{-\mu t}M'(t) = e^{\mu(e^t-t-1)} =$$

$$e^{\mu(1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\dots-t-1)} = e^{\mu(\frac{t^2}{2!}+\frac{t^3}{3!}+\dots)} =$$

$$1+\mu(\frac{t^2}{2!}+\frac{t^3}{3!}+\dots)+\frac{\mu^2}{2}(\frac{t^2}{2!}+\frac{t^3}{3!}+\dots)^2+\dots$$

Коэффициент при $\frac{t^2}{2!}$ равен μ. Значит, $\mu_2 = \sigma^2 = \mu$.

Нормальное распределение

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

$$M'(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{xt} e^{-(x-\mu)^2/2\sigma^2} dx$$

Выделим полный квадрат в экспоненте:
$xt-(x-\mu)^2/2\sigma^2 = \frac{1}{2\sigma^2}[2\sigma^2 xt - x^2 + 2x\mu - \mu^2] = -\frac{1}{2\sigma^2}[x^2 - 2x(\sigma^2 t + \mu) + \mu^2] = -\frac{1}{2\sigma^2}[(x-\mu-t\sigma^2)^2 - 2\mu t\sigma^2 - t^2\sigma^4] = \frac{1}{2\sigma^2}[-(x-\mu-t\sigma^2)^2] + \mu t + \frac{t^2\sigma^2}{2}$.

$$M'(t) = e^{\mu t + \frac{t^2}{2}\sigma^2} = [1 + \mu t + \mu^2 \frac{t^2}{2} + \ldots][1 + \sigma^2 \frac{t^2}{2!} + \ldots]$$

Коэффициент при t	$\mu_1' = \mu$
Коэффициент при $\frac{t^2}{2!}$	$\mu_2' = \mu^2 + \sigma^2$
Следовательно,	$\mu_2 = \mu_2' - (\mu_1')^2 = \sigma^2$.

Распределение $y = ax + b$

$$M'_y(t) = \mathcal{E}(e^{yt}) = \mathcal{E}(e^{bt} \cdot e^{axt}) = e^{bt}\mathcal{E}(e^{axt})$$

Значит $\boxed{M'_y(t) = e^{bt} M'_x(at)}$.

Пусть x имеет нормальное распределение со средним μ и дисперсией σ^2. Тогда

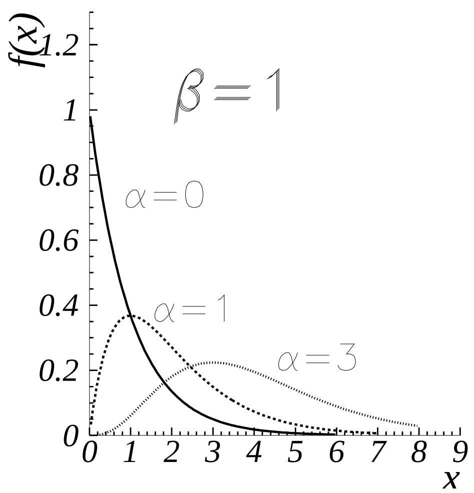
$$M'_x(t) = e^{\mu t + \frac{t^2}{2}\sigma^2}$$

$$M'_y(t) = e^{\mu a t + \frac{t^2}{2}\sigma^2 a^2 + bt} = e^{(a\mu + b)t + \frac{t^2}{2}\sigma^2 a^2}$$

Следовательно, распределение величины y нормальное со средним $a\mu + b$ и дисперсией $a^2\sigma^2$.

Гамма-распределение

$$f(x) = \frac{1}{\alpha!\beta^{\alpha+1}} x^{\alpha} e^{-x/\beta}, \quad 0 \leq x < \infty$$



Распределение с двумя параметрами $\alpha > -1$ и $\beta > 0$. α задает форму распределения, β — масштабный множитель. При $\alpha = 0$ гамма-распределение превращается в экспоненциальное

$$f(x) = \frac{1}{\beta} e^{-x/\beta}.$$

Далее увидим, что распределение χ^2 есть частный случай гамма-распределения.

$$M'(t) = \frac{1}{\alpha!\beta^{\alpha+1}} \int_0^{\infty} x^{\alpha} e^{-x/\beta} e^{xt} dx$$

Подставим $-x/\beta + xt = -\frac{x - x\beta t}{\beta} = -x\frac{1-\beta t}{\beta}$ и $\beta^{\alpha+1} = (\frac{\beta}{1-\beta t})^{\alpha+1}(1-\beta t)^{\alpha+1}$. Тогда $M'(t) =$

$$\frac{1}{(1-\beta t)^{\alpha+1}} \int_0^{\infty} (\frac{1}{\alpha!})(\frac{1-\beta t}{\beta})^{\alpha+1} x^{\alpha} e^{-x(1-\beta t)/\beta} dx =$$

$$\frac{1}{(1-\beta t)^{\alpha+1}} \int_0^{\infty} f(x;\alpha;\frac{\beta}{1-\beta t}) dx = \underline{\frac{1}{(1-\beta t)^{\alpha+1}}}$$

$\frac{\partial M'}{\partial t} = (\alpha+1)(1-\beta t)^{-\alpha-2}\beta;$

$\frac{\partial^2 M'}{\partial t^2} = (\alpha+1)(\alpha+2)(1-\beta t)^{-\alpha-3}\beta^2$

$\mu'_1 = \frac{\partial M'(0)}{\partial t} = \beta(\alpha+1);$ $\boxed{\mu = \beta(\alpha+1)}$
$\mu'_2 = \frac{\partial^2 M'(0)}{\partial t^2} = \beta^2(\alpha+1)(\alpha+2);$

$\mu_2 = \mu'_2 - {\mu'}_1^2 = \beta^2(\alpha+1)[\alpha+2-\alpha-1] = \beta^2(\alpha+1)$; $\boxed{\sigma = \beta\sqrt{\alpha+1}}$

Многомерные случайные величины

Сначала рассмотрим переход от одномерной случайной величины к двумерной.

- *Дискретные* X_1 и X_2:
  $\mathcal{P}[X_1 = r_1$ и $X_2 = r_2] = P_{r_1,r_2}$

$$\sum_{r_1} \sum_{r_2} P_{r_1,r_2} = 1$$

- *Непрерывные* X_1 и X_2:
  $\mathcal{P}[x_1 \le X_1 \le x_1 + dx_1$ и $x2 \le X_2 \le x_2 + dx_2] = f(x_1, x_2)dx_1dx_2$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2)dx_1dx_2 = 1$$

Характеристики распределения (совместной плотности вероятности)

$$\mu_{x_1} = \mathcal{E}(x_1) = \int\int x_1 f(x_1, x_2) dx_1 dx_2$$

$$\mu_{x_2} = \mathcal{E}(x_2) = \int\int x_2 f(x_1, x_2) dx_1 dx_2$$

$$\sigma_{x_1}^2 = \mathcal{E}[(x_1 - \mu_{x_1})^2]; \;\; \sigma_{x_2}^2 = \mathcal{E}[(x_2 - \mu_{x_2})^2]$$

Смешанный второй момент (ковариация)

$$\mathcal{D}(x_1, x_2) = \mathcal{E}[(x_1 - \mu_{x_1})(x_2 - \mu_{x_2})] =$$

$$\mathcal{E}[x_1 x_2 - \mu_{x_1} x_2 - \mu_{x_2} x_1 + \mu_{x_1} \mu_{x_2}] =$$

$$\mathcal{E}(x_1 x_2) - \mathcal{E}(x_1)\mathcal{E}(x_2)$$

Коэффициент корреляции

$$\boldsymbol{\rho(x_1, x_2) = \frac{\mathcal{D}(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}}}$$

- $\boldsymbol{|\rho(x_1, x_2)| \leq 1}$

  Введем $y_i = \frac{x_i - \mu_{x_i}}{\sigma_i}, \ i = 1, 2$

  Очевидно $\mu_{y_i} = 0, \ \mathcal{D}_{y_i} = 1$

  $\mathcal{D}(y_1 \pm y_2) = \mathcal{E}[(y_1 \pm y_2)^2] - (\mu_{y_1} \pm \mu_{y_2})^2 =$

  $\mathcal{D}_{y_1} + \mathcal{D}_{y_2} \pm 2\mathcal{D}(y_1, y_2) = 2(1 \pm \rho(x_1, x_2)) \geq 0.$

  Следовательно, $\boxed{-1 \leq \rho(x_1, x_2) \leq 1}$.

- $|\boldsymbol{\rho(x_1, x_2)}| = \mathbf{1}$ тогда и только тогда, когда x_1 и x_2 связаны линейной зависимостью.

  Пусть $x_2 = ax_1 + b$.

  $\mathcal{D}(x_1, x_2) = \mathcal{E}(x_1 \cdot x_2) - \mathcal{E}(x_1)\mathcal{E}(x_2) = \mathcal{E}(ax_1^2 + bx_1) - \mathcal{E}(x_1)(a\mathcal{E}(x_1) + b) = a[\mathcal{E}(x_1^2) - \mathcal{E}(x_1)^2] + b[\mathcal{E}(x_1) - \mathcal{E}(x_1)] = a\mathcal{D}(x_1)$;
  $\rho(x_1, x_2) = \frac{a\mathcal{D}(x_1)}{\sigma_{x_1} \cdot |a| \sigma_{x_1}} = \frac{a}{|a|} = \pm 1$.

Случайные переменные $x_1, \ldots, x_n$ взаимно независимы, если их совместная плотность вероятности полностью факторизуется:

$$\boxed{f(x_1, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n)}$$

Если X, Y взаимно независимы, то $\mathcal{E}(XY) = \int XY f_x(X) f_y(Y) dX dY = \mathcal{E}(X)\mathcal{E}(Y)$. Тогда $\mathcal{D}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) = 0$ и $\rho(X, Y) = 0$.

Корреляция независимых переменных равна нулю. Обратное, вообще говоря, неверно.

Переменные могут быть некоррелированы ($\rho = 0$), но не обязательно при этом независимые.

Пусть X распределено симметрично относительно 0 с плотностью вероятности $f(x)$, а $Y = X^2$. X и Y функционально связаны, значит, зависимы, а корреляция $\rho(X, Y) = 0$.

Действительно,

$\mathcal{E}(X) = 0$; $\mathcal{E}(Y) = \int X^2 f(X) dX = \mathcal{E}(X^2) = \mathcal{E}(X^2) - \mathcal{E}(X)^2 = \mathcal{D}(X) = \sigma_X^2$.

$\mathcal{D}(X, Y) = \mathcal{E}[(X - \mathcal{E}(X))((Y - \mathcal{E}(Y))] = \mathcal{E}[X(X^2 - \sigma_X^2)] = \mathcal{E}(X^3) - \sigma_X^2 \mathcal{E}(X) = 0$, т.к. $f(x)$ симметрична.

Итак, $\rho(X, Y) = 0$.

Линейные случайные переменные

Пусть $Y = \sum\limits_{i=1}^{n} a_i X_i$

$\mathcal{E}(Y) = \mathcal{E}(\sum a_i X_i) = \sum a_i \mathcal{E}(X_i) = \sum a_i \mu_i$.

$\mathcal{D}(Y) = \mathcal{D}(\sum a_i X_i) = \mathcal{E}[(\sum a_i X_i - \sum a_i \mu_i)^2] = \mathcal{E}[(\sum a_i (X_i - \mu_i))^2] = \sum\limits_{i=1}^{n} a_i^2 \mathcal{E}[(X_i - \mu_i)^2] + \sum\limits_{i}^{(i \neq k)} \sum\limits_{k} a_i a_k \mathcal{E}[(X_i - \mu_i)(X_k - \mu_k)] = \sum\limits_{i=1}^{n} a_i^2 \mathcal{D}(X_i) + 2 \sum\limits_{i=1}^{n-1} \sum\limits_{k=i+1}^{n} a_i a_k \mathcal{D}(X_i, X_k) = \vec{a} \mathcal{D} \vec{a}$ в матричном виде.

$\mathcal{D}(X_i, X_k)$ — матрица вторых моментов или *матрица ошибок* или *ковариационная матрица*.

Если X_i не коррелированы, то $\mathcal{D}(X_i, X_k) = 0$ при $i \neq k$ и $\mathcal{D}(Y) = \sum\limits_{i=1}^{n} a_i^2 \mathcal{D}(X_i)$.

Пусть X_i — n различных результатов одного эксперимента. Введем *выборочное среднее*

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

Будем считать, что для всех i $\mathcal{E}(X_i) = \mu$, $\mathcal{D}(X_i) = \sigma^2$. Вообще говоря, μ не совпадает с $\bar{x}$. Очевидно, что $\bar{x}$ — случайная величина.

$$\mathcal{E}(\bar{x}) = \mathcal{E}(\tfrac{1}{n}\sum X_i) = \tfrac{1}{n}\sum \mathcal{E}(X_i) = \tfrac{1}{n}\cdot n\mu = \mu$$

$$\mathcal{D}(\bar{x}) = \mathcal{D}(\tfrac{1}{n}\sum X_i) = \tfrac{1}{n^2}\sum \mathcal{D}(X_i) + \tfrac{2}{n^2}\sum^{(i\neq k)}\sum \mathcal{D}(X_i, X_k$$
$$\tfrac{\sigma^2}{n} + \tfrac{2}{n^2}\sum^{(i\neq k)}\sum \mathcal{D}(X_i, X_k).$$

Для независимых испытаний $\mathcal{D}(X_i, X_k) = 0$ и $\sigma(\bar{x}) = \sqrt{\mathcal{D}(\bar{x})} = \frac{\sigma}{\sqrt{n}}$.

При наличии корреляции все определяется величиной $\rho \neq 0$. Для двух испытаний $n = 2$ $\sigma^2(\bar{x}) =$

$\frac{1}{2}\sigma^2 + \frac{2}{4}\sigma^2\rho = \frac{1}{2}\sigma^2(1+\rho)$, т.к. $\mathcal{D}(X_1, X_2) = \sigma^2\rho$.
Поскольку $-1 \le \rho \le 1,\ 0 \le \sigma^2(\bar{x}) \le \sigma^2$.

при $\rho = +1$ $\quad \bar{\mathtt{x}} = \mathtt{X}_1 = \mathtt{X}_2 \quad \sigma^2(\bar{\mathtt{x}}) = \sigma^2$
при $\rho = -1$ $\quad \mathtt{X}_2 - \mu = -(\mathtt{X}_1 - \mu),\ \bar{\mathtt{x}} = \mu \quad \sigma^2(\bar{\mathtt{x}}) = 0$

Закон больших чисел

Мы убедились в том, что выборочное среднее $\bar{x}$ является хорошей оценкой среднего μ генеральной совокупности. Верно и более сильное утверждение: с ростом объема выборки выборочное среднее (с очень большой вероятностью) стремится к μ как к своему пределу.

$$\bar{x} \stackrel{n\to\infty}{\longrightarrow} \mu$$

Сначала докажем для произвольной случайной величины x *неравенство Чебышева*:

$$\mathcal{P}[|x-\mu| \geq k\sigma] \leq \frac{1}{k^2}$$

Это неравенство справедливо только для распределений, у которых среднее и дисперсия конечны.

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Разбивая область интегрирования на три части,

$$\sigma^2 = \Big[\int_{-\infty}^{\mu-k\sigma} + \int_{\mu-k\sigma}^{\mu+k\sigma} + \int_{\mu+k\sigma}^{\infty}\Big](x-\mu)^2 f(x)dx$$

$$\sigma^2 \geq \Big[\int_{-\infty}^{\mu-k\sigma} + \int_{\mu+k\sigma}^{\infty}\Big](x-\mu)^2 f(x)dx \geq$$

$$k^2\sigma^2\Big[\int_{-\infty}^{\mu-k\sigma} + \int_{\mu+k\sigma}^{\infty}\Big] f(x)dx = k^2\sigma^2\mathcal{P}[|x-\mu| \geq k\sigma],$$

откуда и следует $\mathcal{P}[|x-\mu| \geq k\sigma] \leq \dfrac{1}{k^2}$

Это неравенство, вообще говоря, весьма слабое. В случае нормального распределения:

k	$\mathcal{P}[\lvert x-\mu\rvert \geq k\sigma]$	неравенство Чебышева
1	0.32	$\leq$ 1.000
2	0.045	$\leq$ 0.250
3	0.003	$\leq$ 0.111

Теперь применим неравенство Чебышева к величине выборочного среднего $\bar{x}$ со среднеквадратичным отклонением $\sigma/\sqrt{n}$:

$$\mathcal{P}[|\bar{x} - \mu| \geq \frac{k\sigma}{\sqrt{n}}] \leq \frac{1}{k^2}.$$

Введя обозначение $\varepsilon = k\sigma/\sqrt{n},\ k = \varepsilon\sqrt{n}/\sigma$, перепишем неравенство:

$$\mathcal{P}[|\bar{x} - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Вероятность противоположного:

$$\mathcal{P}[|\bar{x} - \mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Отсюда видно, что при $n \to \infty$ $\bar{x} \to \mu$.

Центральная предельная теорема

Теорема. Пусть случайная величина x имеет среднее значение μ и дисперсию σ^2. Если σ^2 конечно, то при объеме выборки $n \to \infty$ распределение выборочного среднего $\bar{x}$ будет стремиться к нормальному со средним μ и дисперсией σ^2/n.

Доказательство. Производящая функция центральных моментов

$$M_x(t) = \mathcal{E}[e^{(x-\mu)t}] = 1 + 0 + \sigma^2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \dots$$

Пусть $\omega = (x-\mu)\frac{1}{\sigma\sqrt{n}}$, $\mathcal{E}(\omega) = 0$.

$$M_\omega(t) = \mathcal{E}\Big[e^{\frac{(x-\mu)t}{\sigma\sqrt{n}}}\Big] = M_x(\frac{t}{\sigma\sqrt{n}}).$$

$$M_\omega(t) = 1 + \sigma^2 \frac{t^2}{2!\sigma^2 n} + \mu_3 \frac{t^3}{3!\sigma^3 n\sqrt{n}} + \dots =$$

$$1 + \frac{t^2}{2n} + \mu_3 \frac{t^3}{6\sigma^3 n\sqrt{n}} + \dots.$$

Введем переменную $z = \frac{\bar{x}-\mu}{\sigma(\bar{x})} = \frac{(\bar{x}-\mu)\sqrt{n}}{\sigma} = \frac{1}{n}\sum_{i=1}^{n} \frac{(x_i-\mu)\sqrt{n}}{\sigma} = \sum_{i=1}^{n} \omega_i$. Мы уже показывали ранее, что производящая функция моментов для суммы независимых случайных величин есть произведение их ПФМ. Тогда

$$M_z(t) = [M_\omega(t)]^n = (1+\frac{t^2}{2n}+\mu_3\frac{t^3}{6\sigma^3 n\sqrt{n}}+\ldots)^n.$$

$$M_z(t) \xrightarrow{n\to\infty} e^{t^2/2}.$$

Но $e^{\frac{t^2}{2}}$ представляет собой производящую функцию моментов для нормального распределения со средним значением 0 и дисперсией 1.

$$z = \frac{(\bar{x}-\mu)\sqrt{n}}{\sigma} \Rightarrow \bar{x} = \mu + z\frac{\sigma}{\sqrt{n}}.$$

Следовательно, величина $\bar{x}$ в пределе распределена по нормальному закону со средним μ и дисперсией σ^2/n.

Распределение вероятности для функции случайной величины

Пусть функцию $f(x_1, \ldots, x_n)$, где x_i — случайные переменные, можно разложить в ряд Тейлора вблизи точек $\mathcal{E}(x_1) = \mu_1, \ldots, \mathcal{E}(x_n) = \mu_n$. Опуская члены второго порядка и выше по $x_i - \mu_i$,

$$f(x_1, \ldots, x_n) = f(\mu_1, \ldots, \mu_n) + \sum_{i=1}^{n} (x_i - \mu_i) \frac{\partial f}{\partial x_i}\Big|_{\mu_i}$$

$\mathcal{E}[f(x_1, \ldots, x_n)] \approx f(\mu_1, \ldots, \mu_n)$. Найдем дисперсию (**перенос ошибок**).

$$\mathcal{D}(f) = \mathcal{E}[(f - \mathcal{E}(f))^2] \approx \mathcal{E}([\sum (x_i - \mu_i) \frac{\partial f}{\partial x_i}\Big|_{\mu_i}]^2) =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f}{\partial x_i}\Big|_{\mu_i} \frac{\partial f}{\partial x_j}\Big|_{\mu_j} \mathcal{E}[(x_i - \mu_i)(x_j - \mu_j)] = ADA^T,$$

где $A = (\frac{\partial f}{\partial x_1}\Big|_{\mu_1}, \ldots, \frac{\partial f}{\partial x_n}\Big|_{\mu_n})$, D — ковариационная матрица. Если x_i независимы, то

$$\mathcal{D} \approx \sum_{i=1}^{n} (\frac{\partial f}{\partial x_i}\Big|_{\mu_i})^2 \sigma_i^2$$

Примеры использования формул переноса ошибок:

- Найти ошибку отношения двух случайных величин $\mathcal{D}(\frac{X}{Y})$, если $|\mu_X| \gg \sigma_X,\ |\mu_Y| \gg \sigma_Y,\ Y \neq 0$.

  $\mathcal{D}(\frac{X}{Y}) \approx \frac{\sigma_X^2}{\mu_Y^2} + \frac{\sigma_Y^2}{\mu_Y^4}\mu_X^2 - \frac{2\rho_{XY}\sigma_X\sigma_Y\mu_X}{\mu_Y^3} = (\frac{\mu_X}{\mu_Y})^2[\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2} - \frac{2\rho_{XY}\sigma_X\sigma_Y}{\mu_X\mu_Y}]$.

- В эксперименте найдена величина $m_e^2 = 0.2611 \pm 0.0121$ МэВ^2. Определить m_e.

  $y = x^2 \Rightarrow \mathcal{D}(y) = (2x)^2\mathcal{D}(x) \Rightarrow \sigma_x = \frac{\sigma_y}{2x}$.

$$m_e = 0.5110 \pm 0.0118 \text{МэВ}.$$

Ранее мы получали для линейной функции $f = \sum a_i x_i$ дисперсию $\mathcal{D}(f) = ADA^T$, где $A = (a_1, \ldots, a_n)$. Этот же результат дает и формула переноса ошибок.

Формула переноса ошибок точна для линейных функций. Для остальных это лишь приближение для малых отклонений $x_i - \mu_i$. Что дает учет нелинейности?

$$f(x) = f(\mu) + (x-\mu)\frac{\partial f}{\partial \mu} + \frac{(x-\mu)^2}{2}\frac{\partial^2 f}{\partial \mu^2} + \ldots$$

$$\mathcal{E}(f) = f(\mu) + 0 + \frac{\mathcal{D}(x)}{2}\frac{\partial^2 f}{\partial \mu^2} = f(\mu) + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial \mu^2}.$$

$$\mathcal{D}(f) = \mathcal{E}[(f - \mathcal{E}(f))^2] = \mathcal{E}[((x-\mu)\frac{\partial f}{\partial \mu} + \frac{(x-\mu)^2}{2}\frac{\partial^2 f}{\partial \mu^2} - \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial \mu^2})^2] = \mathcal{E}[(x-\mu)^2 {f'}^2_\mu + \frac{(x-\mu)^4}{4}({f''}_\mu)^2 + \frac{\sigma^4}{4}{f''}^2_\mu + \frac{2(x-\mu)^3}{2}f'_\mu {f''}_\mu - \sigma^2(x-\mu)f'_\mu {f''}_\mu - \frac{(x-\mu)^2}{2}{f''}^2_\mu \sigma^2] =$$

$$\sigma^2 {f'_\mu}^2 + \frac{\mu_4}{4}{f''_\mu}^2 + \frac{\sigma^4}{4}{f''_\mu}^2 + \mu_3 f'_\mu {f''_\mu}^2 - \frac{\sigma^4}{2}f''_\mu =$$

$$\underline{\sigma^2 {f'_\mu}^2 + \frac{1}{4}(\mu_4 - \sigma^4){f''_\mu}^2 + \mu_3 f'_\mu f''_\mu.}$$

Для нормального распределения $\mu_3 = 0,\ \mu_4 = 3\sigma^4$ и

$$\mathcal{D}(f) = {f'_\mu}^2 \sigma^2 + \frac{1}{2} {f''_\mu}^2 \sigma^4$$

Пусть $f = ax + bx^2$. $f' = a + 2bx,\ f'' = 2b$.

$$\mathcal{D}(f) = (a + 2b\mu)^2 \sigma^2 + 2b^2 \sigma^4$$

При $a = 0$ $\mathcal{D}(f) = 4b^2\mu^2\sigma^2 + 2b^2\sigma^4$, т.е. добавка за счет нелинейности есть $2b^2\sigma^4 / 4b^2\mu^2\sigma^2 = \frac{\sigma^2}{2\mu^2}$. Как и должно быть, ее вклад мал при $\mu \gg \sigma$.

Перейдем к нахождению функции распределения функции $h(x)$ случайной величины.

- **Величина x дискретна.**
  a) Если существует лишь одно значение $x = x_0$, для которого $h(x) = h(x_0)$, то $\mathcal{P}[h(x) = h_0] = \mathcal{P}[h(x) = h(x_0)] = \mathcal{P}[x = x_0]$.
  b) Если существует несколько значений $x = x_0, x_1, \ldots, x_m$, для которых $h(x) = h(x_0)$, то $\mathcal{P}[h(x) = h_0] = \mathcal{P}[x = x_0$ или $x = x_1 \ldots$ или $x = x_m] = \sum\limits_{i=0}^{m} \mathcal{P}[x = x_i]$.

- **Величина x непрерывна.** Дана $f(x)dx = \mathcal{P}[x \leq X \leq x + dx]$. Требуется найти $g(h)dh = \mathcal{P}[h \leq H \leq h + dh]$, где $h = h(x)$.
  a) $h(x)$ — однозначная функция x. Нужно найти малый интервал значений h, соответствующий заданному малому интервалу значений x.

Воспользовавшись обратной функцией $x = x(h)$, получим $f(x) = f[x(h)]$. Плотность вероятности неотрицательна, поэтому

$$dx = |\frac{dx(h)}{dh}|dh; \quad f(x)dx = f[x(h)] \cdot |\frac{dx(h)}{dh}|dh.$$

Следовательно, $g(h) = f[x(h)] \cdot |\frac{dx(h)}{dh}|$.

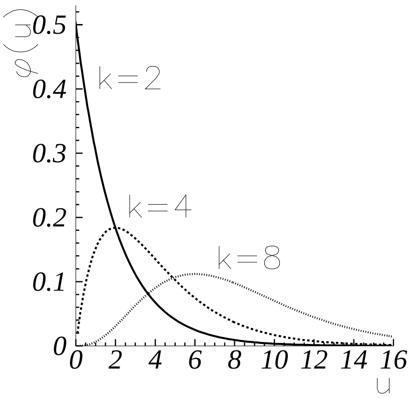
b) $h(x)$ — неоднозначная функция x : $h(x_1) = h(x_2) = \ldots = h(x_m)$. $\mathcal{P}[h \leq H \leq h + dh] = \mathcal{P}[x_1 \leq X \leq x_1 + dx$ или $x_2 \leq X \leq x_2 + dx$ или $\ldots$ или $x_m \leq X \leq x_m + dx] = \sum\limits_{i=1}^{m} \mathcal{P}[x = x_i]$. Тогда

$$g(h)dh = \underset{h(x_i)=h}{} f[x_i(h)] \cdot |\frac{dx_i(h)}{dh}|dh.$$

Пример. x распределено в соответствии с нормальным распределением со средним значением 0 и единичной дисперсией. Найти функцию распределения $h(x) = x^2$.

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

$h(x) = x^2$ — неоднозначная функция, $x_1(h) = \sqrt{h}$, $x_2(h) = -\sqrt{h}$.
$\frac{dx_{1,2}}{dh} = \pm\frac{1}{2\sqrt{h}}$. Значит, $g(h)dh = \frac{2}{\sqrt{2\pi}} e^{-\frac{h}{2}} \frac{dh}{2\sqrt{h}} = \frac{1}{(-\frac{1}{2})!2^{-\frac{1}{2}+1}} h^{-1/2} e^{-h/2} dh = \frac{1}{\alpha!\beta^{\alpha+1}} h^{\alpha} e^{-h/\beta} dh$, где
$0 \leq h < \infty, \alpha = -\frac{1}{2}, \beta = 2$.
Итак, $g(h)$ представляет собой гамма-распределение.
Ранее было показано, что производящая функция моментов для гамма-распределения $M_h'(t) = (1 - \beta t)^{-\alpha-1}$, или в данном случае $M_h'(t) = (1 - 2t)^{-1/2}$.
Пусть есть k независимых x_i, распределенных нормально со средним 0 и дисперсией 1. Как распределена величина $u = \sum\limits_{i=1}^{k} x_i^2$?
$M_u'(t) = (M_h'(t))^k = (1 - 2t)^{-k/2}$. Это снова гамма-распределение с параметрами $\beta = 2$, $\alpha = k/2 - 1$.



Величину $u = \sum_{i=1}^{k} x_i^2$ будем называть χ^2 с k степенями свободы. Функция распределения χ_k^2:

$$\varphi(u) = \frac{1}{(\frac{k}{2}-1)!2^{k/2}} u^{k/2-1} e^{-u/2}$$

Распределение достигает максимума при $u = k-2$, а математическое ожидание величины u равно k. Найдем математическое ожидание и дисперсию из производящей функции моментов.

$$M'(t) = (1-2t)^{-k/2}$$

$$\frac{\partial M'(t)}{\partial t} = k(1-2t)^{-k/2-1}$$

$$\frac{\partial^2 M'(t)}{\partial t^2} = k(k+2)(1-2t)^{-k/2-2}$$

$$\mu_1' = \mathcal{E}(\chi_k^2) = \frac{\partial M'(0)}{\partial t} = k$$

$$\mu_2' = \frac{\partial^2 M'(0)}{\partial t^2} = k(k+2); \quad \mathcal{D}(\chi_k^2) = k^2 + 2k - k^2 = 2k$$

Легко обобщить преобразование распределения вероятности на случай нескольких случайных величин. Пусть $x_1, \ldots, x_n$ нужно преобразовать в $h_1, \ldots, h_n$, где $h_i = h_i(x_1, \ldots, x_n)$ для каждого i. Плотность вероятности для h_i:
$g(h_1, \ldots, h_n) \quad = \quad f[x_1(\vec{h}), \ldots, x_n(\vec{h})] \left|\frac{\partial(x_1,\ldots,x_n)}{\partial(h_1,\ldots,h_n)}\right|$,
где $f(x_1, \ldots, x_n)$ — исходная совместная плотность вероятности, $\left|\frac{\partial(x_1,\ldots,x_n)}{\partial(h_1,\ldots,h_n)}\right|$ — модуль Якобиана преобразования.

Экспериментальное разрешение

Идеальные теоретические распределения всегда искажаются при измерении из-за неточности эксперимента. Для события с истинным значением параметра X будет измерено значение X'. Распределение измеряемых X' определяется некоторой плотностью вероятности — характеристикой измерительного прибора и называется функцией разрешения $r(X, X')$.

Если истинная плотность вероятности распределения X есть $f(X)$, то плотность распределения измеряемых величин

$$g(X') = \int r(X, X') f(X) dX,$$

где после интегрирования по истинной переменной X осталась лишь измеряемая переменная X'. Это может привести к тому, что X' принимает значения, для которых $f(X) = 0$.

Очень часто разрешение описывается нормальным распределением, где σ характеризует точность измерений:

$$r(X, X') = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(X-X')^2}{2\sigma^2}}$$

- Пусть измеряемая величина X — константа, т.е. $X = a$ или истинная плотность вероятности $f(X) = \delta(X - a)$. Тогда

$$g(X') = \frac{1}{\sqrt{2\pi\sigma}} \int\limits_{-\infty}^{\infty} \delta(X-a) e^{-\frac{(X-X')^2}{2\sigma^2}} dX =$$

$$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(X'-a)^2}{2\sigma^2}},$$

т.е. пик δ функции размывается в колокол с центром в точке a и дисперсией σ^2.

- Пусть измеряемая величина X распределена нормально со средним μ и дисперсией τ^2. Свертка двух Гауссов даст Гаусс с дисперсией, равной сумме дисперсий. Поэтому

$$g(X') = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + \tau^2}} e^{-\frac{(X'-\mu)^2}{2(\sigma^2+\tau^2)}}$$

  В предельных случаях:

  - $\sigma \ll \tau$ (прибор очень точный): искажения малы $g(X') \approx f(X')$

  - $\sigma \gg \tau$ (прибор весьма неточный): искажения велики, видимая ширина полностью определяется экспериментальным разрешением $g(X') \approx N(\mu, \sigma^2)$

  - $\sigma \sim \tau$: наблюдаемая ширина $\sim \sigma\sqrt{2}$

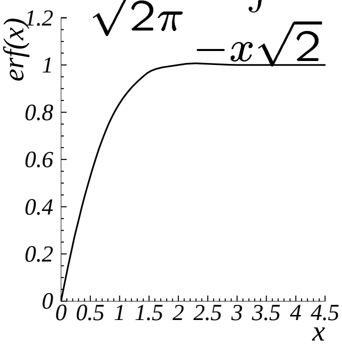
- Плотность вероятности равномерна, т.е.

$$f(x) = \begin{cases} 1 \quad \text{при } |x| \le \frac{1}{2} \\ 0 \quad \text{при } |x| > \frac{1}{2} \end{cases}$$

$$\boldsymbol{g(X') = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-1/2}^{1/2} e^{-\frac{(X-X')^2}{2\sigma^2}} dX =}$$

$$\boldsymbol{\frac{1}{\sqrt{2\pi}} \int\limits_{(-1/2-X')/\sigma}^{(1/2-X')/\sigma} e^{\frac{-y^2}{2}} dy,}$$

при замене $y = \frac{(X-X')}{\sigma}$. Часто используется интеграл вероятности $erf(x) = \frac{2}{\sqrt{\pi}} \int\limits_0^x e^{-t^2} dt$. Замена $z = t\sqrt{2}$ приводит к $erf(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-x\sqrt{2}}^{x\sqrt{2}} e^{-z^2/2} dz$.



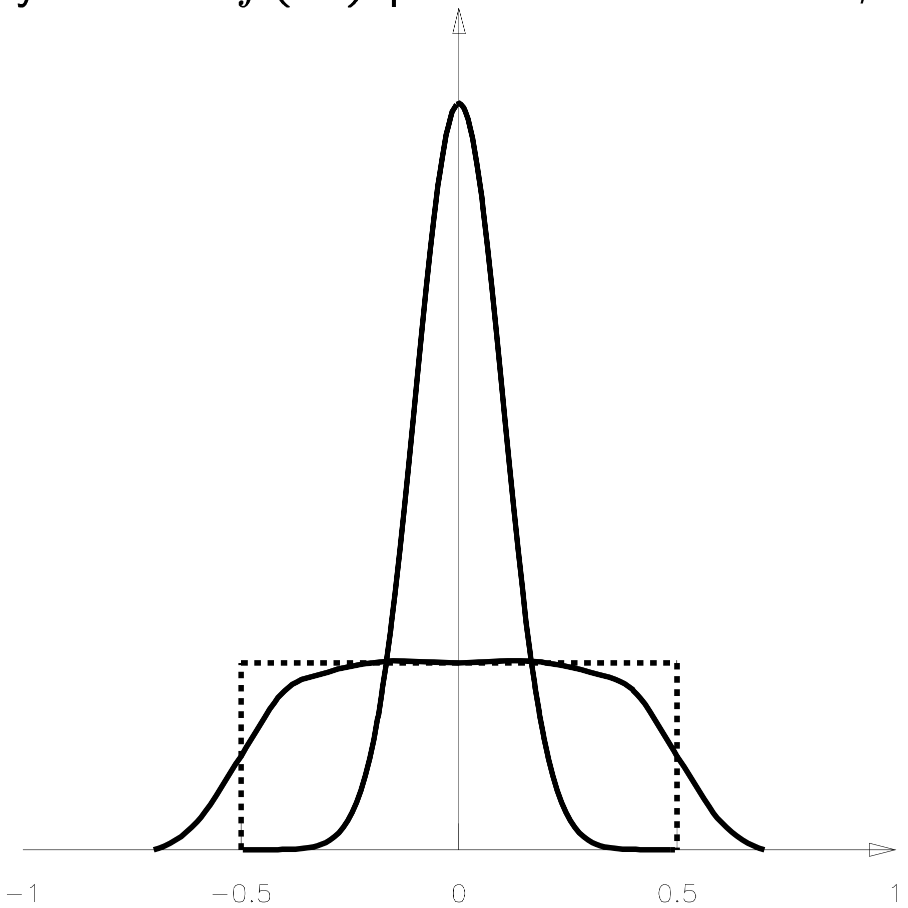
$$erf(\frac{1}{\sqrt{2}}) = 0.688$$
$$erf(\frac{2}{\sqrt{2}}) = 0.954$$

$$g(X') = \frac{1}{\sqrt{2\pi}} \int\limits_{(-1/2-X')/\sigma}^{(1/2-X')/\sigma} e^{\frac{-y^2}{2}} dy =$$

$$\frac{1}{2}[erf(\frac{0.5 - X'}{\sigma\sqrt{2}}) - erf(\frac{-0.5 - X'}{\sigma\sqrt{2}})]$$

Пусть экспериментальное разрешение $\sigma = 0.1$, т.е. много меньше ширины истинного распределения, равной 1. При $X' \approx 0.3$ аргумент первого erf есть $\frac{2}{\sqrt{2}}$, т.е. первый член ≈ 0.954, а второй еще ближе к 1, значит, $g(X') \approx 1$. С дальнейшим ростом X' 1-ый член медленно падает и становится заметно отличным от 1. При $X' = 0.5$ 1-ый член равен 0, второй 1, т.е. $g(X') = 0.5$. При $X' > 0.5$ $g(X')$ продолжает падать, но отлична от 0. Исходное распределение размазывается, хотя в достаточно большой области $|X'| \leq 0.3$ оно не меняется.

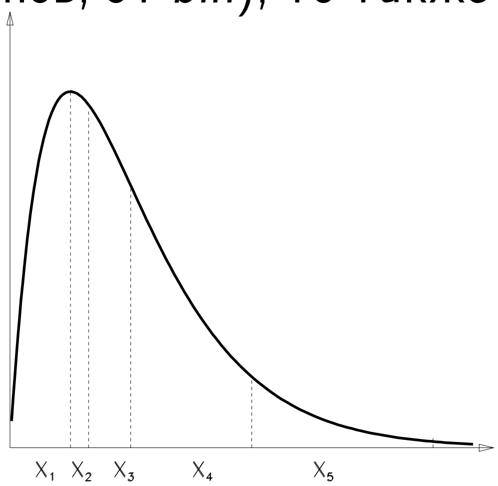
С ухудшением точности измерения (ростом σ) исходный прямоугольник $f(X)$ расползается совсем, и при



$\sigma = 1$ уже почти не отличается от Гаусса с $\sigma = 1$. Это значит, что при плохой точности мы видим не истинное распределение, а просто функцию разрешения (аппаратурную).

Критерий согласия χ^2

Рассмотрим случайную величину x, которая может принимать k значений $x_1, \ldots, x_k$ в случае дискретного распределения. Если x распределена непрерывно, то разобьем область изменения x на k интервалов (или бинов, от *bin*), то также получим k возможных исходов.



Пусть n — полное число повторений опыта, т.е. разных x, а r_i — частота появления значения x_i.

$$\sum_{i=1}^{k} r_i = n$$

Распределение величины r_i — полиномиальное или мультиномиальное, это обобщение биномиального распределения с $k = 2$:

$$\mathcal{P}(r_1, \ldots, r_k) = \frac{n!}{r_1! r_2! \ldots r_k!} P_1^{r_1} P_2^{r_2} \ldots P_k^{r_k},$$

где $\mathcal{P}(x = x_i) = P_i$.

Каждое из r_i распределено по биномиальному закону: 2 исхода — положительный с вероятностью P_i и отрицательный с вероятностью $1 - P_i$. Очевидно, что $\mathcal{E}(r_i) = nP_i$. Ранее показывали, что биномиальное распределение сходится к распределению Пуассона со средним значением и дисперсией $\mu = nP_i$. Покажем, что распределение Пуассона, в свою очередь, с ростом μ сходится к нормальному.

$P_r = \frac{e^{-\mu}}{r!}\mu^r$, $r! \approx \sqrt{2\pi r}(\frac{r}{e})^r$, $P_r \approx \frac{1}{\sqrt{2\pi r}}e^{-\mu+r}(\frac{\mu}{r})^r = \frac{1}{\sqrt{2\pi r}}e^{r-\mu+rln\frac{\mu}{r}}$.

Разлагая по отклонению Δ/μ, где $\Delta = r - \mu$, $r = \Delta + \mu$, $P_r \approx \frac{1}{\sqrt{2\pi\mu}}e^{\Delta+(\Delta+\mu)ln\frac{\mu}{\Delta+\mu}}$.

Показатель экспоненты: $\Delta + (\Delta + \mu)ln\frac{1}{1+\frac{\Delta}{\mu}} = \Delta - (\Delta + \mu)ln(1 + \frac{\Delta}{\mu})$.

Т.к. $ln(1+x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots$, $\Delta - (\Delta+\mu)ln(1+\frac{\Delta}{\mu}) \approx \Delta - (\Delta+\mu)(\frac{\Delta}{\mu} - \frac{\Delta^2}{2\mu^2}) \approx \Delta - \frac{\Delta^2}{\mu} - \Delta + \frac{\Delta^2}{2\mu} = -\frac{\Delta^2}{2\mu}$, т.е. $\boxed{P_r \approx \frac{1}{\sqrt{2\pi\mu}}e^{-\frac{\Delta^2}{2\mu}}}$.

Возвращаемся к нашему рассмотрению случайной величины r_i. При достаточно больших nP_i (≥ 5) r_i распределена примерно по Гауссу со средним значением nP_i и дисперсией nP_i. Значит, распределение величины $z_i = \frac{r_i - nP_i}{\sqrt{nP_i}}$ приблизительно нормальное со средним значением 0 и единичной дисперсией. Тогда величина $u = \sum\limits_{i=1}^{k} z_i^2 = \sum\limits_{i=1}^{k} \frac{(r_i - nP_i)^2}{nP_i}$ распределена примерно по закону χ^2. Число степеней свободы $k-1$, т.к. не все r_i независимы, есть линейное соотношение между r_i: $\sum\limits_{i=1}^{k} r_i = n$. В общем случае число степеней свободы $\nu = k - t$, где t — число соотношений.

Пусть мы проверяем гипотезу, состоящую в том, что $\mathcal{E}(r_i) = nP_i$ $(i = 1, 2, \ldots, k)$. Для этого вычислим величину $u = \sum\limits_{i=1}^{k} \frac{(r_i - nP_i)^2}{nP_i}$ и посмотрим, настолько ли хорошо согласие между r_i и nP_i, что u можно рассматривать как приемлемую выборку из распределения χ^2. Критерием проверки гипотезы служит сопоставление величины u с табличным значением χ^2_{k-1}.

x	0	1	2	3	4	≥ 5	n
частота r_x	109	65	22	3	1	0	200

Выборочное среднее $\bar{x} = \frac{\sum x r_x}{\sum r_x} = \frac{122}{200} = 0.61$.
Гипотеза. Данные описываются распределением Пуассона, т.е. $P_x = \frac{\mu^x e^{-\mu}}{x!}$. В качестве оценки параметра μ возьмем выборочное среднее $\widehat{\mu} = 0.61$. Тогда можно вычислить $nP_i = 200(0.61)^i \cdot e^{-0.61} \cdot (i!)^{-1}$:

x	0	1	2	3	4	≥ 5	n
nP_x	108.7	66.3	20.2	4.1	0.6	0.1	200

Для проверки гипотезы по критерию χ^2 необходимо, чтобы в каждом интервале ожидаемое число событий было не менее ~ 5. Поэтому последние три интервала просуммируем:

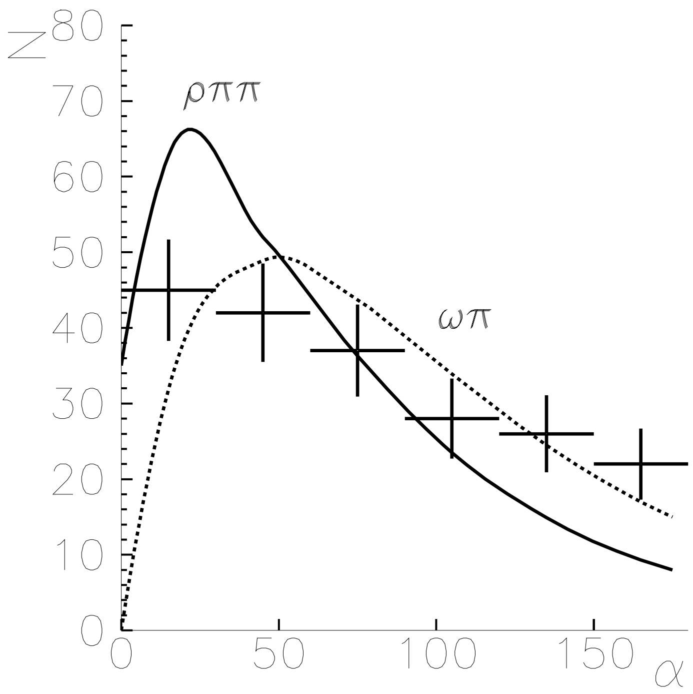
x	0	1	2	≥ 3
r_x	109	65	22	4
nP_x	108.7	66.3	20.2	4.8

Теперь вычислим $u = \sum \frac{(r_x - nP_x)^2}{nP_x} = \frac{(0.3)^2}{108.7} + \frac{(1.3)^2}{66.3} + \frac{(1.8)^2}{20.2} + \frac{(0.8)^2}{4.8} = 0.32$.

Сколько линейных соотношений? 1)$\sum r_x = 200$; 2)$\sum xr_x$ $= 122$ (оценивали μ по данной выборке). Таким образом, число степеней свободы ν=4-2=2. Следовательно, величина u должна быть распределена как χ_2^2. Из таблиц $\mathcal{P}(\chi_2^2 \geq 0.32) \approx 0.85$, т.е. с вероятностью 85% u может быть больше 0.32.

Вывод: гипотеза разумная. Вероятность p того, что $\chi_2^2 > \chi_p^2$:

$\boldsymbol{p}$	0.99	0.95	0.90	0.80	0.70	0.30	0.10	0.01
$\boldsymbol{\chi_p^2}$	0.020	0.10	0.21	0.45	0.71	2.41	4.61	9.21



$e^+e^- \to \rho\pi^0\pi^0 \to \pi^+\pi^-\pi^0\pi^0$

$e^+e^- \to \omega\pi^0 \to \pi^+\pi^-\pi^0\pi^0$

α — угол расколлинеарности $\pi^+\pi^-$

α	N_{exp}	$N_{\omega\pi}$	$\Delta\chi^2_{\omega\pi}$	$N_{\rho\pi\pi}$	$\Delta\chi^2_{\rho\pi\pi}$
$0 \div 30$	45	31.9	**5.34**	62.8	**5.05**
$30 \div 60$	42	48.9	**0.97**	53.2	**2.46**
$60 \div 90$	37	43.7	**1.04**	36.3	**0.02**
$90 \div 120$	28	33.9	**1.03**	23.5	**0.86**
$120 \div 150$	26	24.5	**0.09**	14.9	**8.31**
$150 \div 180$	22	17.0	**1.45**	9.3	**17.34**
$\sum$	200	200	**9.91**	200	**34.04**
$\mathcal{P}(\chi^2_5)$			**0.078**		$\mathbf{2 \cdot 10^{-6}}$

Интервальное оценивание. Распределение Стьюдента.

Пусть необходимо *оценить* некоторый параметр. Есть две возможности:

- точечное оценивание $x = a$ без указания меры уверенности
- интервальное оценивание $x_{min} < x < x_{max}$ с указанием меры уверенности. Чем больше $x_{max} - x_{min}$, тем вероятнее интервал содержит значение величины, т.е. выше мера уверенности.

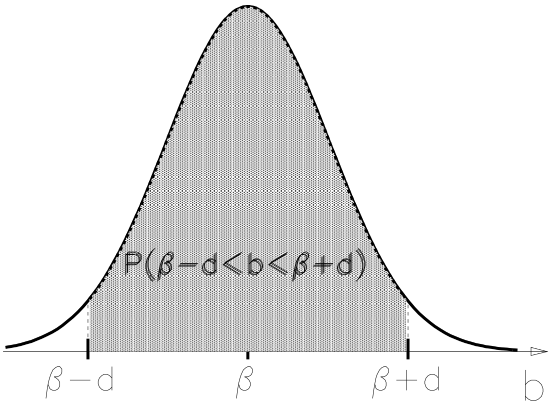
Как находить интервалы и меры уверенности?

Пусть нужно найти доверительный интервал для параметра β совокупности. Величина самого β неизвестна, она оценивается по выборке, вычисляя

$$b = f(x_1, \ldots, x_n).$$

Величина b имеет некоторое распределение со средним $\mathcal{E}(b) = \beta$ и дисперсией $\mathcal{D}(b) = \sigma_b^2$.

По центральной предельной теореме большинство выборочных оценок имеют распределение, близкое к нормальному. Тогда с помощью вероятностей для нормального распределения можно найти интервал $(\beta - d, \beta + d)$, в котором лежит значение β, вычисленное по выборке. Решается и обратная задача — по данной величине вероятности найти d.



Например, для вероятности 0.95 из таблиц нормального распределения получаем $\boldsymbol{d = 1.96\sigma_b}$ и $\boldsymbol{\mathcal{P}[\beta - 1.96\sigma_b \leq b \leq \beta + 1.96\sigma_b] = 0.95}$, что эквивалентно $\boldsymbol{\mathcal{P}[b - 1.96\sigma_b \leq \beta \leq b + 1.96\sigma_b] = 0.95}$.

В общем случае доверительный интервал есть $[b - z\sigma_b, b + z\sigma_b]$, где z получается из выборочного распределения в зависимости от желаемого доверия. σ_b может быть известно из априорной информации о совокупности, в противном случае в больших выборках вместо σ_b можно взять s_b, вычисленное по выборке.

Как интерпретировать $\mathcal{P}[a < \beta < c] = p$*?*

Неверно думать, что β является случайной переменной, лежащей с вероятностью $\boldsymbol{p}$ между $\boldsymbol{a}$ и $\boldsymbol{c}$. β — константа, вероятностные утверждения относительно нее недопустимы. Случайными переменными являются $\boldsymbol{a}$ и $\boldsymbol{c}$, они зависят от конкретной выборки, т.е. доверительный интервал имеет свое выборочное распределение. Утверждение $p = \mathcal{P}[a < \beta < c]$ означает, что *для данного интервала, выбранного из совокупности интервалов, вероятность содержать значение β равна 100p%*. Если рассмотреть большое число выборок и вычислить для них $[a, c]$, то приблизительно 100p% из них содержат β.

Пусть, например, случайная выборка из $n = 100$ измерений параметра x дала выборочное среднее $\bar{x} = 1.00$, выборочную дисперсию $s^2 = 0.04$. Найдем доверительные границы 95% и 99%.

В качестве оценки истинного среднего можно взять $\bar{x}$. Оно имеет распределение, близкое к нормальному, со средним $\mathcal{E}(\bar{x}) = \mu$ и стандартным отклонением $\sigma(\bar{x}) = \sigma/\sqrt{n}$, где μ, σ — среднее значение и стандартное отклонение совокупности.

95% доверительный интервал для μ:

$$\boldsymbol{[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}]}$$

При $n \geq 30$ замена σ на s не меняет уровня достоверности. $\frac{s}{\sqrt{n}} = 0.02 \ \Rightarrow$ 95% доверительный интервал $[0.961, 1.039]$.

При оценивании с большей уверенностью (99%) интервал более широкий. Соответствующий коэффициент 2.58, $2.58\frac{s}{\sqrt{n}} = 0.052$. 99% доверительный интервал $[0.948, 1.052]$.

При небольшом объеме выборки доверительный коэффициент зависит от объема. Это — следствие того, что мы не знаем истинную дисперсию σ^2. Если бы истинная дисперсия была известна, то вводя величину $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, распределенную по закону $N(0,1)$, можно было бы воспользоваться таблицами и т.п.

Если же мы оцениваем σ^2 по выборке (s^2), то величина $\boldsymbol{t = \frac{\bar{x}-\mu}{s/\sqrt{n}}}$ уже не распределена по закону $N(0,1)$.

Перепишем $\boldsymbol{t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{s^2}{\sigma^2}} = \frac{\mathcal{N}}{\sqrt{\mathcal{U}}}}$.

Числитель этой дроби $\mathcal{N}$ распределен по нормальному закону, а величина $\mathcal{U}$ распределена по $\chi^2_{n-1}/(n-1)$. Таким образом, t есть функция двух случайных величин $\mathcal{N}$ и $\mathcal{U}$ с известными распределениями $\Rightarrow$ можно вычислить распределение t.

Запишем двумерную плотность распределения по $\mathcal{N}$ и $\mathcal{U}$:

$$\mathcal{P}(\mathcal{N},\mathcal{U}) = \frac{e^{-\mathcal{N}^2/2}}{\sqrt{2\pi}} \cdot \frac{\mathcal{U}^{\frac{n-1}{2}-1} \cdot e^{-\mathcal{U}/2}}{\left(\frac{n-1}{2}-1\right)! \cdot 2^{(n-1)/2}}$$

при $-\infty \leq \mathcal{N} \leq \infty$ и $0 \leq \mathcal{U} \leq \infty$. Сделаем замену:

$$h_1(\mathcal{N},\mathcal{U}) = t = \frac{\mathcal{N}}{\sqrt{\mathcal{U}}},\ h_2(\mathcal{N},\mathcal{U}) = \mathcal{U}.$$

Обратное преобразование:

$$\mathcal{N} = h_1 \quad \overline{h_2},\ \mathcal{U} = h_2.$$

Якобиан $\begin{vmatrix} \sqrt{h_2} & \frac{h_1}{2\sqrt{h_2}} \\ 0 & 1 \end{vmatrix} = \sqrt{h_2}.$

Двумерная плотность по переменным $g(h_1, h_2) =$

$$\mathcal{P}(\mathcal{N},\mathcal{U}) \quad \overline{h_2} = \frac{e^{-h_1^2 h_2/2} e^{-h_2/2}}{\sqrt{2\pi}\left(\frac{n-1}{2}-1\right)! \cdot 2^{(n-1)/2}} \cdot h_2^{\frac{n-1}{2}-1+\frac{1}{2}}$$

Плотность вероятности по t (т.е. по h_1) получается интегрированием по h_2:

$$f(t) = \int_0^{\infty} g(h_1, h_2) dh_2 = \frac{C}{[1 + \frac{t^2}{n-1}]^{n/2}},$$

где t имеет $(n-1)$ степеней свободы, как и исходная величина χ^2_{n-1}. Постоянная C равна

$$C = \frac{\left(\frac{n-2}{2}\right)!}{\sqrt{\pi(n-1)} \cdot \left(\frac{n-3}{2}\right)!}.$$

$f(t)$ — распределение Стьюдента с $n-1$ степенями свободы.

При $n = 2$ распределение $f(t)$ переходит в $\frac{1}{\pi(1+t^2)}$, т.е. распределение Коши.

Распределение Стьюдента всегда симметрично и при любом конечном n убывает медленнее нормального, имеет более длинные хвосты. Поэтому *доверительный интервал, оцениваемый по t-распределению, больше доверительного интервала по нормальному распределению.* При $n \to \infty$ t-распределение стремится к нормальному. Действительно,

$$\Big(\frac{1}{1+\frac{t^2}{n-1}}\Big)^{\frac{n}{2}} \stackrel{n\to\infty}{\longrightarrow} \Big((1+\frac{t^2}{n})^{-n}\Big)^{1/2} \longrightarrow e^{-t^2/2},$$

$C \stackrel{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}$.

Математическое ожидание $\mathcal{E}(t) = 0$, $\mathcal{D}(t) = \frac{n}{n-2}$ (при $n > 2$).

Существуют лишь те моменты μ_r, для которых $r < n-1$!

Как находить доверительные интервалы по t-распределению? Существуют таблицы значений t_p и вероятности того, что $|t| \geq t_p$ при n степенях свободы.

n	p=0.10	p=0.05	p=0.01
1	6.314	12.706	63.657
2	2.920	4.303	9.925
5	2.015	2.571	4.032
10	1.812	2.228	3.169
30	1.697	2.042	2.750
∞	1.645	1.960	2.576

$$p = 2 \int_{t_p}^{\infty} f_n(t)dt$$

Последняя строка таблицы соответствует нормальному распределению. Видно, как сильно отличается доверительный интервал в большую сторону для малых n.

Пример использования t-распределения. Пусть мы хотим проверить, действительно ли средний рост группы людей равен 167 см. Дана выборка из 10 человек:

$$160,\ 160,\ 167,\ 170,\ 173,\ 176,\ 178,\ 178,\ 181,\ 181.$$

Гипотеза: величина роста распределена нормально со средним значением $\mu = 167$. Вычислим выборочное среднее $\bar{x} = 172.4$. Случайно ли то, что $\bar{x} > \mu$ или надо забраковать гипотезу?

$$\text{Вычислим } s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 62.9,$$

$$\sigma(\bar{x}) = \frac{s}{\sqrt{n}} = 2.51, \quad t_9 = \frac{172.4 - 167.0}{2.51} = 2.15$$

Из таблицы найдем, что при $n = 9$ с вероятностью 10% $|t_9| \geq 1.836$ или с вероятностью 95% $t_9 \leq 1.836$. У нас получилось t_9 больше, т.е. вероятность менее 5%. Таким образом, наблюдаемое отличие от среднего достоверно при 5% уровне значимости. Вряд ли значение 167 может служить надежной оценкой среднего роста группы людей.

Различие между двумя выборочными средними — другой пример использования t-распределения. Пусть имеется выборка из n_1 значений нормально распределенной величины x и n_2 значений нормально распределенной величины y. Предположим, что

$$x \sim N(\mu_1, \sigma^2) \text{ и } y \sim N(\mu_2, \sigma^2),$$

где величина σ^2 неизвестна, и мы хотим проверить гипотезу $\boxed{\mu_1 = \mu_2}$.

Вычислим выборочные средние $\bar{x}$ и $\bar{y}$, их дисперсии σ^2/n_1 и σ^2/n_2. Если бы σ^2 была известна, то можно было бы вычислить

$$\omega = \frac{\bar{x} - \bar{y}}{\sigma(\bar{x} - \bar{y})} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

и проверить, значимо ли отлична от 0 нормально распределенная величина ω.

Когда σ^2 неизвестна, воспользуемся объединенной оценкой σ^2 из обеих выборок:

$$s^2 = \frac{\sum\limits_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum\limits_{i=1}^{n_2}(y_i - \bar{y})^2}{n_1 + n_2 - 2}.$$

Вместо ω используем статистику t с числом степеней свободы $n_1 + n_2 - 2$:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{\bar{x} - \bar{y}}{\widehat{\sigma}(\bar{x} - \bar{y})}.$$

Необходимо обосновать использование t-распределения. В числителе стоит разность двух нормально распределенных величин, т.е. тоже описывается нормальным распределением. Утверждение $\mu_1 = \mu_2$ эквивалентно тому, что среднее значение числителя равно 0. Покажем, что объединенная оценка s^2 следует распределению χ^2 и не зависит от числителя $\bar{x} - \bar{y}$. Действительно, $\bar{x}$ и $\bar{y}$ не зависят от $\sum(x_i - \bar{x})^2$, $\sum(y_i - \bar{y})^2$ соответственно. Значит, s^2 и $(\bar{x} - \bar{y})$ независимы.

$$\mathop{}\limits_{1}^{n_1}(x_i-\bar{x})^2 \sim \sigma^2\chi^2_{n_1-1} \text{ и } \mathop{}\limits_{1}^{n_1}(y_i-\bar{y})^2 \sim \sigma^2\chi^2_{n_2-1},$$

т.е. по существу надо доказать, что $\chi^2_{n_1-1}+\chi^2_{n_2-1} \sim \chi^2_{n_1+n_2-2}$. Производящие функции моментов для $\chi^2_{n_1-1}$ и $\chi^2_{n_2-1}$ соответственно равны $(1-2s)^{-(n_1-1)/2}$ и $(1-2s)^{-(n_2-1)/2}$. ПФМ суммы независимых случайных величин — произведение соответствующих ПФМ. Следовательно,

$$\text{ПФМ}(\chi^2_{n_1-1}+\chi^2_{n_2-1}) = (1-2s)^{-(n_1+n_2-2)/2},$$

что и есть ПФМ величины χ^2 с n_1+n_2-2 степенями свободы.

Пример. 13 измерений параметра: 7 измерений при одних условиях, 6 — при других.

$$x: \ 15.7, \ 10.3, \ 12.6, \ 14.5, \ 12.6, \ 13.8, \ 11.9.$$

$$y: \ 12.3, \ 13.7, \ 10.4, \ 11.4, \ 14.9, \ 12.6.$$

$$\bar{y} = 12.55, \quad \bar{x} = 13.06.$$

Случайно ли полученное различие $\bar{x} - \bar{y} = 0.51$?

$$\sum_{1}^{7}(x_i - \bar{x})^2 = 18.98; \qquad \sum_{1}^{6}(y_i - \bar{y})^2 = 12.86$$

$$\text{Отсюда } s^2 = \frac{12.86 + 18.98}{11} = 2.89,$$

$$\hat{\sigma}^2(\bar{x}-\bar{y}) = 2.89 \cdot \left(\frac{1}{7} + \frac{1}{6}\right) = 0.895, \quad \hat{\sigma}(\bar{x}-\bar{y}) = 0.95.$$

$$\text{Следовательно, } t_{11} = \frac{13.06 - 12.55}{0.95} = 0.54$$

Из таблиц при $t = 0.54$ и $n = 11$ p=0.60. Т.е. если между $\bar{x}$ и $\bar{y}$ нет различия, такие большие t будут получаться в 60% случаях. Различие незначимо.

Доверительный интервал.

Как можно представить результат предыдущего примера с помощью понятия доверительного интервала?

Из таблиц распределения t с числом степеней свободы $n = 11$ получим p(2.2) = 0.05. Это означает, что случайная величина t_{11} с вероятностью 95% заключена в интервале от -2.2 до 2.2. Пусть действительное различие между средними значениями при разных условиях есть некоторая постоянная $\delta = \mu_1 - \mu_2$. Тогда

$$\frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = t_{11} \text{ и}$$

$$-2.2 \leq \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \leq 2.2$$

с вероятностью 95%. Случайными величинами являются $\bar{x}$, $\bar{y}$ и s^2, а не δ!

$$(\bar{x}-\bar{y})-2.2s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq \delta \leq (\bar{x}-\bar{y})+2.2s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

с вероятностью 95%. Снова подчеркнем, что именно границы интервала представляют собой случайные величины. Подставив наши числа, получим

$$-1.57 \leq \delta \leq 2.59$$

Таким образом, интервал $(-1.57;\ 2.59)$ с вероятностью 95% содержит истинное значение δ. Поскольку этот интервал включает 0, мы говорим, что найденное значение $(\bar{x}-\bar{y})$ незначимо отлично от нуля. Этот интервал называется *95%-ным доверительным интервалом* для величины δ.

Физики обычно приводят результат $\pm$ среднеквадратичное отклонение. В соответствии с этим

$$\delta = (\bar{x}-\bar{y}) \pm \hat{\sigma}(\bar{x}-\bar{y}) = 0.51 \pm 0.95.$$

В этой записи имеется указание доверительного интервала $(-0.44;\ 1.46)$. Из таблиц t-распределения следует, что доверительный интервал шириной в одно среднеквадратичное отклонение содержит истинный результат с вероятностью всего лишь 66%.

Принцип максимального правдоподобия.

Метод максимального правдоподобия представляет собой технику оценки величин параметров на основе конечного набора данных. Дальнейшее применение метода позволяет найти приближенные ошибки полученных оценок.

Пусть измерение случайной величины x было повторено n раз и дало величины $x_1, \ldots, x_n$. x может представлять собой вектор для (многомерная случайная величина). Пусть x распределено в соответствии с известной плотностью вероятности $f(x; \Theta)$, но величина хотя бы одного параметра Θ неизвестна.

В рамках гипотезы $f(x; \Theta)$ вероятность того, что первое измерение даст результат в интервале $[x_1; x_1 + dx_1]$, есть $f(x_1; \Theta)dx_1$.

Т.к. все измерения предполагаются независимыми, вероятность того, что каждое из n измерений соответствует интервалу $[x_i; x_i + dx_i]$, дается произведением $\prod\limits_{i=1}^{n} f(x_i; \Theta) dx_i$.

Введем *функцию максимального правдоподобия*

$$L(\Theta) = \prod_{i=1}^{n} f(x_i; \Theta).$$

Заметим, что это, в сущности, объединенная плотность вероятности для x_i, рассматриваемая как функция параметра Θ. С другой стороны, величины x_i уже рассматриваются как фиксированные (т.к. эксперимент завершен).

Широко используется также *логарифмическая функция правдоподобия*, представляющая $\mathcal{L}(\Theta) = lnL(\Theta)$.

Чтобы найти величину Θ, которая с наибольшей вероятностью приводит к наблюдаемым значениям x_i, необходимо определить, при каком Θ функция правдоподобия принимает максимальное значение.

Пример с неизвестным средним. Пусть случайная величина распределена по нормальному закону с единичной дисперсией и неизвестным средним значением.

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

Тогда $L(\mu) = e^{-\sum (x_i-\mu)^2/2}$ (множитель $(2\pi)^{-n/2}$ можно опустить, на максимум не влияет), а

$$\mathcal{L}(\mu) = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2}.$$

Функция правдоподобия $\mathcal{L}(\mu)$ представляет собой параболу по переменной μ. Максимум $\mathcal{L}$ достигается при

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

Пример с неизвестной дисперсией. Если неизвестна и дисперсия распределения, то задача будет двухпараметрическая.

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

Следовательно, $L(\mu,\sigma^2) = (\sigma^2)^{-n/2} e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}}$,

а $\mathcal{L}(\mu,\sigma^2) = -\frac{n}{2} ln\sigma^2 - \quad \frac{(x_i-\mu)^2}{2\sigma^2}$.

Чтобы найти пару значений $(\widehat{\mu}, \widehat{\sigma^2})$, которые наиболее вероятно описывали наблюдаемые результаты, нужно решить уравнения:

$$\frac{\partial\mathcal{L}}{\partial\mu} = 0 \text{ и } \frac{\partial\mathcal{L}}{\partial\sigma^2} = 0.$$

$$\frac{\partial\mathcal{L}}{\partial\mu} = \frac{1}{\sigma^2} \quad (x_i-\mu) = 0 \;\Rightarrow\; \widehat{\mu} = \bar{x}.$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \quad (x_i - \mu)^2 = 0$$

$$\text{Следовательно, } \widehat{\sigma^2} = \frac{1}{n} \quad (x_i - \bar{x})^2.$$

Однако, математическое ожидание для $\widehat{\sigma^2}$ $\mathcal{E}(\widehat{\sigma^2}) = \frac{n-1}{n}\sigma^2$. Как видно, $\widehat{\sigma^2}$ оказывается несколько смещенной оценкой дисперсии! Это смещение пропадает в пределе больших n.

Вспомним, что выборочное среднее является несмещенной оценкой дисперсии

$$s^2 = \frac{1}{n-1} {}_{i=1}^{n} (x_i - \widehat{\mu})^2.$$

Но это не максимально правдоподобная оценка.

Равномерное распределение. $f(x; a, b) = \frac{1}{b-a}$, $a \leq x \leq b$.

Имеется выборка. Упорядочим элементы в порядке возрастания:

$$x_{(1)} < x_{(2)} < \ldots < x_{(n-1)} < x_{(n)}.$$

Функция правдоподобия от двух параметров a и b $L(a, b) = \frac{1}{(b-a)^n}$. Дифференцированием максимум найти не удается. Максимум достигается при наименьшем из возможных значений знаменателя, т.е. при $\widehat{b} = x_{(n)}$, $\widehat{a} = x_{(1)}$. Оставшаяся часть выборки не добавляет никакой информации. Элементы выборки $x_{(1)}$ и $x_{(n)}$ являются достаточными для оценки a и b.

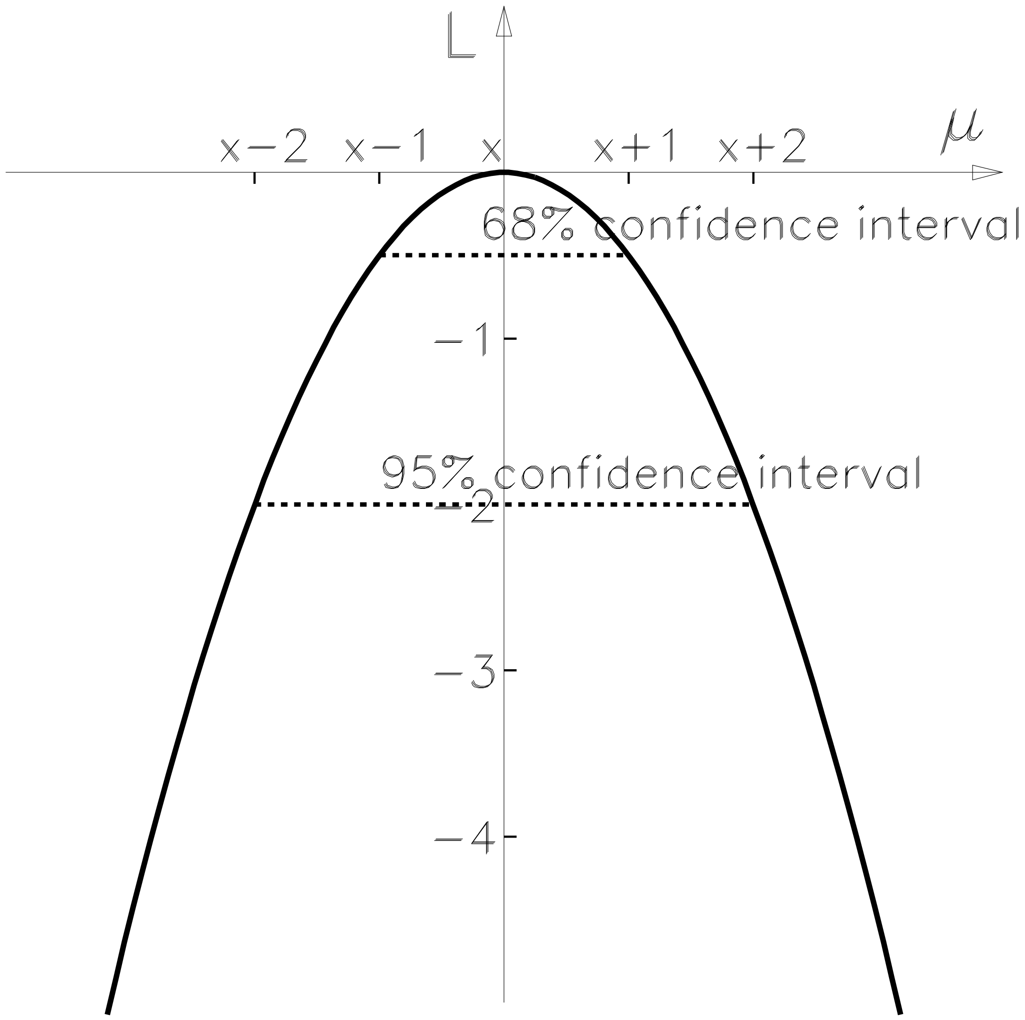
Графический анализ функции правдоподобия.

Одно из достоинств использования функции правдоподобия при анализе экспериментальных данных состоит в том, что можно найти графический вид этой функции и наглядно представить результат. Используя график, можно указать и доверительный интервал или ошибку полученной оценки.

При большом количестве параметров (три или более), оценки которых не являются независимыми, этот метод неудобен. В случае двух параметров функцию максимального правдоподобия изображают контурами.

Рассмотрим нормальное распределение с единичной дисперсией и неизвестным параметром μ. Логарифмическая функция максимального правдоподобия $\mathcal{L}(\mu) = -\frac{1}{2}\sum(x_i - \mu)^2$. К выражению можно добавить любую произвольную постоянную. Если в качестве нее взять $\frac{1}{2}[\sum x_i^2 - n\bar{x}^2]$, то получим функцию $\mathcal{L}(\mu) = -\frac{n}{2}(\mu - \bar{x})^2$. Это парабола, ширина которой растет пропорционально $\sqrt{n}$, а максимум, равный 0, расположен при $\mu = \bar{x}$.

$$\mathcal{L}(\mu) = -\frac{n}{2}(\mu - \bar{x})^2 \text{ при } n = 1.$$



Уровень $\mathcal{L} = -\frac{1}{2}$ характеризует доверительный интервал $(x - \sigma) \leq \mu \leq (x + \sigma)$, который можно записать в виде $\mu = x \pm \sigma$. Этому интервалу соответствует доверительная вероятность 68.3%. Аналогично, уровень $\mathcal{L} = -2$ соответствует 95.5% доверительному интервалу.

Достаточные статистики.

В примере с нормальным распределением мы видели, что $\widehat{\mu}$ и $\widehat{\sigma^2}$ являются функциями величин $\sum x_i$ и $\sum x_i^2$. Нельзя ли извлечь какую-либо полезную информацию из других функций от исходных величин, например, $\sum x_i^3$ или $\sum |x_i|$? С точки зрения параметров μ и σ — нет, если известны $\sum x_i$ и $\sum x_i^2$, то исходная выборка $\{x_i\}$ уже не нужна. Поэтому мы называем $\sum x_i$ и $\sum x_i^2$ *достаточными статистиками*, а любые оценки параметров, зависящие от $\sum x_i$ и $\sum x_i^2$, являются *достаточными оценками*.

Пусть имеется выборка $x_1, \ldots, x_n$ из генеральной совокупности с плотностью вероятности $f(x; \Theta)$. В качестве оценки параметра используется функция этих данных $\widehat{\Theta} = \widehat{\Theta}\ (\boldsymbol{x})$.

Пусть совместную плотность распределения выборки можно разложить на множители

$$\prod_{i=1}^{n} f(x_i; \Theta) = g(\boldsymbol{x}; \widehat{\Theta}) h(\widehat{\Theta}; \Theta),$$

где $g(\boldsymbol{x}; \widehat{\Theta})$ — функция только исходных данных $\boldsymbol{x}$ и не зависит от Θ. По существу, $g(\boldsymbol{x}; \widehat{\Theta})$ есть условное распределение $\boldsymbol{x}$ при заданном значении $\widehat{\Theta}(\boldsymbol{x})$. Раз сам параметр Θ уже не входит в это распределение, то функцию $g(\boldsymbol{x}; \widehat{\Theta})$ нельзя использовать для получения информации относительно Θ. Вся информация, касающаяся оценки Θ содержится в множителе $h(\widehat{\Theta}; \Theta)$, зависящего от $\boldsymbol{x}$ только посредством $\widehat{\Theta}(\boldsymbol{x})$.

Пример. $f(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$ — плотность распределения с одним параметром μ.

Плотность распределения для выборки

$$\prod_{i=1}^{n} f(x_i; \mu) = (2\pi)^{-n/2} e^{-\frac{\sum (x_i - \mu)^2}{2}} =$$

$$[(2\pi)^{-n/2} e^{-\sum x_i^2/2}] \cdot [e^{-\frac{(n\mu^2 - 2\mu \sum x_i)}{2}}].$$

Первый множитель не зависит от параметра μ, значит, *достаточной статистикой* для μ является $\sum x_i$.

$$\bar{x} = \frac{\sum x_i}{n}$$

представляет собой *достаточную оценку*.

Не в каждой задаче оценки параметров имеется достаточная статистика, но когда у параметра существует достаточная оценка, то метод максимального правдоподобия дает именно эту оценку.

Неравенство Крамера-Рао.

Пусть оценка параметра Θ есть $\hat{\Theta}(x)$. Какова ошибка оценки $\mathcal{E}[(\hat{\Theta}-\Theta)^2]$?

$$\mathcal{E}[(\hat{\Theta}-\Theta)^2] = \mathcal{E}[\left(\hat{\Theta}-\mathcal{E}(\hat{\Theta})\right)^2] + [\mathcal{E}(\hat{\Theta})-\Theta]^2.$$

Первое слагаемое здесь — дисперсия оценки, средний квадрат отклонения случайной величины $\hat{\Theta}$ от ее среднего значения. Второе — *смещение* оценки. Таким образом,

$$\mathcal{E}[(\hat{\Theta}-\Theta)^2] = \sigma^2_{\hat{\Theta}^2}(\Theta) + \beta^2(\Theta).$$

Предел точности, которая может быть достигнута при решении задачи об оценке параметров, устанавливается неравенством Крамера-Рао. Это неравенство не зависит от вида распределения случайных величин x в данном эксперименте. При объеме выборки n

$$\mathcal{E}[(\hat{\Theta}-\Theta)^2] \geq \frac{\gamma^2(\Theta)}{n}, \text{ где}$$

$$\gamma^2(\Theta) = \frac{[1+\frac{d\beta(\Theta)}{d\Theta}]^2}{\int[\frac{\partial ln f(x;\Theta)}{\partial\Theta}]^2 f(x;\Theta)dx}.$$

Оценку $\hat{\Theta}$ называют *эффективной*, если ее точность в действительности достигает верхней границы.

Теперь приведем свойства оценок, основанных на методе максимального правдоподобия:

- Если существует *достаточная* оценка, то МПП дает именно ее, тем самым исчерпывая всю имеющуюся информацию.

- Если существует *эффективная* оценка, то МПП дает именно ее, и более точной оценки найти нельзя.

- Оценка максимального правдоподобия имеет *асимптотически нормальное распределение* со средним значением Θ и дисперсией
$$\sigma^2_{\hat{\Theta}} = \frac{1}{-\mathcal{E}[\frac{\partial^2 \mathcal{L}}{\partial \Theta^2}]} \approx \frac{1}{-[\frac{\partial^2 \mathcal{L}}{\partial \Theta^2}]_{\Theta=\hat{\Theta}}},$$
где $\mathcal{L}$ — логарифмическая функция правдоподобия.

- Оценка максимального правдоподобия *инвариантна* относительно преобразований параметра. Если $g = g(\Theta)$, то $\widehat{g} = g(\widehat{\Theta})$. Например, если $s^2 = (1/n)\sum(x_i - \bar{x})^2$ — оценка максимального правдоподобия дисперсии σ^2, то оценкой максимального правдоподобия σ будет $\sqrt{s^2} = s$. Сравним с несмещенными оценками: несмещенной оценкой σ^2 является $(1/(n-1))\sum(x_i - \bar{x})^2$, а несмещенной оценкой σ служит $[((n-3)/2)!/\sqrt{2}((n-2)/2)!]\sqrt{\sum(x_i - \bar{x})^2}$.

- Главная трудность, возникающая при использованиии ММП в качестве критерия оценки, заключается в том, что при этом приходится предполагать известным вид распределения вероятности. Это допущение может наложить сильное ограничение. Для определенного класса задач критерий наименьших квадратов приводит к процедуре, которая дает точную оценку при любых распределениях случайной величины.

Вернемся к графическому методу получения доверительных интервалов по функции максимального правдоподобия. Ранее была показана процедура нахождения 68% и 95% доверительных интервалов для нормального распределения, где логарифмическая функция правдоподобия представляет собой параболу. Обобщим эту процедуру для любой непрерывной функции максимального правдоподобия $\mathcal{L}(\Theta)$ с единственным максимумом в области допустимых значений Θ.

- Добавим к функции $\mathcal{L}(\Theta)$ константу так, чтобы

$$\mathcal{L}_{max}(\Theta) = \mathcal{L}(\hat{\Theta}) = 0.$$

- На графике логарифмической функции правдоподобия $\mathcal{L}$ от Θ проведем прямую на уровне $\mathcal{L} = -1/2$, что даст 68% доверительный интервал, либо на уровне $\mathcal{L} = -2$, что даст 95% доверительный интервал.

Разложим логарифмическую функцию максимального правдоподобия с одним параметром Θ в ряд Тейлора около значения максимально правдоподобной оценки $\hat{\Theta}$: $\mathcal{L}(\Theta) = \mathcal{L}(\hat{\Theta}) + \left[\frac{\partial \mathcal{L}}{\partial \Theta}\right]_{\Theta=\hat{\Theta}}(\Theta - \hat{\Theta}) + \frac{1}{2!}\left[\frac{\partial^2 \mathcal{L}}{\partial \Theta^2}\right]_{\Theta=\hat{\Theta}}(\Theta - \hat{\Theta})^2 + \ldots$

По определению $\hat{\Theta}$ мы знаем, что $\mathcal{L}(\hat{\Theta}) = \mathcal{L}_{max}$, а также, что $\left[\frac{\partial \mathcal{L}}{\partial \Theta}\right]_{\Theta=\hat{\Theta}} = 0$. Используя также

$$\sigma^2_{\hat{\Theta}} = \frac{1}{-\mathcal{E}[\frac{\partial^2 \mathcal{L}}{\partial \Theta^2}]} \approx \frac{1}{-[\frac{\partial^2 \mathcal{L}}{\partial \Theta^2}]_{\Theta=\hat{\Theta}}},$$

$$\text{получим } \mathcal{L} = \mathcal{L}_{max} - \frac{(\Theta - \hat{\Theta})^2}{2\sigma^2_{\hat{\Theta}}}.$$

Отсюда видно, что изменение параметра Θ на одно стандартное отклонение приводит к уменьшению логарифмической функции правдоподобия на $1/2$ относительно максимального значения.

Пример. Рассмотрим снова нормальное распределение с нулевым средним значением и неизвестной дисперсией $x \sim N(0; \sigma^2)$. Записать функцию максимального правдоподобия от параметра σ^2 через достаточную статистику $\omega = \sum x_i^2$ так, чтобы $\mathcal{L}_{max} = 0$.

$$L(\sigma^2|\mathrm{x}) = (\sigma^2)^{-n/2} e^{\sum x_i^2/2\sigma^2},$$

$$\mathcal{L}(\sigma^2|\omega) = -\frac{n}{2} ln\sigma^2 - \frac{\omega}{2\sigma^2} + \frac{n}{2} ln\frac{\omega e}{n}.$$

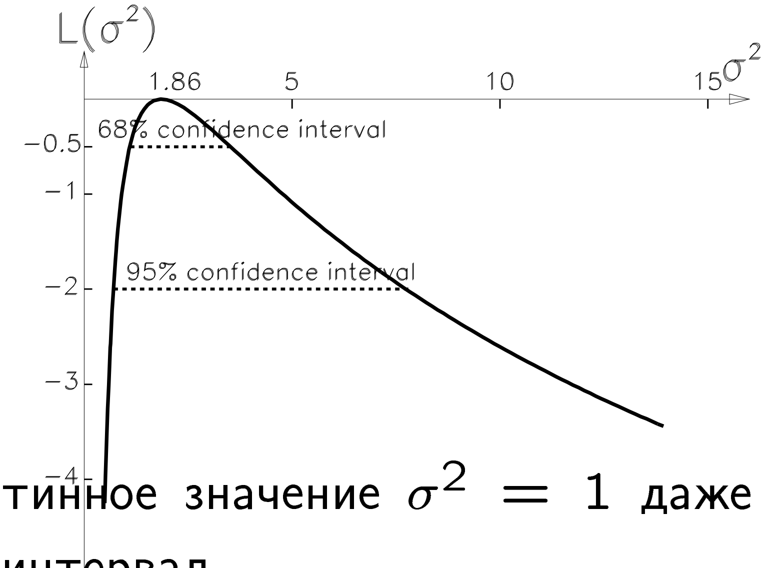
Случайная выборка малого объема. С помощью генератора случайных чисел произведена выборка объемом $n = 6$ из генеральной совокупности с распределением $N(0; 1)$. Получены следующие значения:

$$\mathbf{1.468,\ -2.097,\ 0.146,\ -0.295,\ -2.120,\ 0.017}$$

$\omega = \sum x_i^2 = 11.155;\ n = 6\ \Rightarrow$

$$\mathcal{L}(\sigma^2|\omega) = 4.860 - 3ln\sigma^2 - \frac{5.578}{\sigma^2}$$

Оценка максимального правдоподобия $\widehat{\sigma^2} = \frac{\omega}{n} =$ 1.86. Доверительные интервалы из графика функции



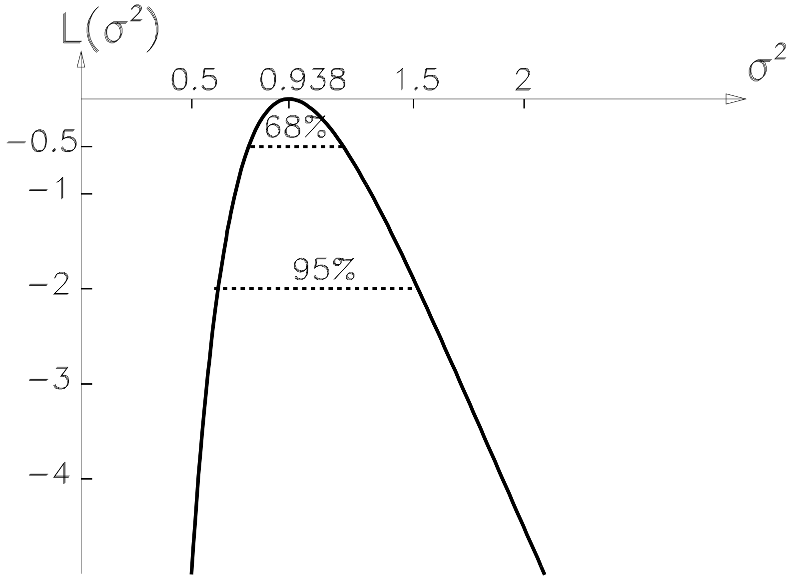
$68\% : 1.10 \leq \sigma^2 \leq 3$

$95\% : 0.71 \leq \sigma^2 \leq 7$

Видно, что истинное значение $\sigma^2 = 1$ даже не попадает в 68%-интервал.

Случайная выборка большого объема. С помощью генератора случайных чисел произведена выборка объемом $n = 40$ из генеральной совокупности с распределением $N(0;1)$. Пусть $\omega = \sum x_i^2 = 37.52$. Тогда $\widehat{\sigma^2} = 0.938$.

$$\mathcal{L}(\sigma^2|\omega) = 18.72 - 20ln\sigma^2 - \frac{18.76}{\sigma^2}$$



Эта кривая уже намного ближе к параболе, чем в случае $n = 6$.
$\sigma^2 = \widehat{\sigma^2}{}^{+0.24}_{-0.17}$

$$\mathcal{D}(\sigma^2) \approx \frac{1}{-\left[\frac{\partial^2\mathcal{L}}{(\partial\sigma^2})^2\right]_{\sigma^2=\widehat{\sigma^2}}} \approx \frac{2(\widehat{\sigma^2})^2}{n}$$

$$\sigma(\sigma^2) \approx \frac{\widehat{\sigma^2}}{\sqrt{20}} = 0.21$$

Метод наименьших квадратов. Связь с методом максимального правдоподобия.

Пусть имеется N независимых нормально распределенных случайных величин y_i, $i = 1, \ldots, N$, причем каждая связана с точно известной переменной x_i. Например, это может быть N измерений температуры $y_i = T(x_i)$ в различных точках с координатой x_i. Предположим, что каждая величина y_i имеет свое неизвестное среднее значение λ_i и известную, но различную для каждого i дисперсию, σ_i^2. Объединенная плотность вероятности для N измерений y_i представляет произведение нормальных распределений: $g(y_1, \ldots, y_N; \lambda_1, \ldots, \lambda_N, \sigma_1^2, \ldots, \sigma_N^2) =$

$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\Big(\frac{-(y_i - \lambda_i)^2}{2\sigma_i^2}\Big).$$

Предположим теперь, что истинное значение измеряемой величины является некоторой функцией x, $\lambda = \lambda(x;\, \boldsymbol{\Theta})$, зависящей от неизвестных параметров $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_m)$. Цель метода наименьших квадратов состоит в том, чтобы найти значения параметров $\boldsymbol{\Theta}$. Кроме того, метод позволяет найти критерий качества гипотезы о функции $\lambda = \lambda(x;\, \boldsymbol{\Theta})$.

Опуская слагаемые, независящие от параметров, логарифм объединенной плотности вероятности даст:

$$\mathcal{L}(\boldsymbol{\Theta}) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - \lambda(x_i; \boldsymbol{\Theta}))^2}{\sigma_i^2}.$$

Нахождение параметров $\boldsymbol{\Theta}$ как максимум функции правдоподобия эквивалентно нахождению минимума величины

$$\chi^2(\boldsymbol{\Theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i; \boldsymbol{\Theta}))^2}{\sigma_i^2}.$$

Такая сумма квадратов отклонений измеренного значения от истинного с весами, равными обратной величине дисперсии, является базисом метода наименьших квадратов. Она используется и в случаях, когда распределение величин y_i не является нормальным, требуется лишь их независимость. В более общем случае при известной ковариационной матрице V и неизвестных средних значениях

$$\chi^2(\mathbf{\Theta}) = \sum_{i,j=1}^{N} (y_i - \lambda(x_i; \mathbf{\Theta}))^2 (V^{-1})_{ij} (y_j - \lambda(x_j; \mathbf{\Theta}))^2.$$

Метод наименьших квадратов для линейной функции.

Хотя метод может применяться для любого вида функции $\lambda(x; \mathbf{\Theta})$, наилучшим образом его свойства проявляются в случае, когда $\lambda(x; \mathbf{\Theta})$ представляет собой линейную функцию параметров $\mathbf{\Theta}$,

$$\lambda(x; \mathbf{\Theta}) = \sum_{j=1}^{m} a_j(x) \Theta_j,$$

где $a_j(x)$ — линейно независимые функции от x.

Для этого случая оценки параметров и их дисперсии можно вычислить аналитически. Кроме того, оценки получаются несмещенные и с минимально возможной дисперсией. Значение функции $\lambda(x; \boldsymbol{\Theta})$ в точке x_i можно записать

$$\lambda(x_i; \boldsymbol{\Theta}) = \sum_{j=1}^{m} a_j(x_i)\Theta_j = \sum_{j=1}^{m} A_{ij}\Theta_j,$$

где $A_{ij} = a_j(x_i)$. Матричный вид записи выражения для χ^2

$$\chi^2 = (\boldsymbol{y} - \boldsymbol{\lambda})^T V^{-1}(\boldsymbol{y} - \boldsymbol{\lambda}) = (\boldsymbol{y} - A\boldsymbol{\Theta})^T V^{-1}(\boldsymbol{y} - A\boldsymbol{\Theta}).$$

Чтобы найти минимум χ^2, требуем равенство нулю производных по параметрам Θ_i,

$$\nabla\chi^2 = -2(A^T V^{-1}\boldsymbol{y} - A^T V^{-1} A\boldsymbol{\Theta}) = 0.$$

Решение этого уравнения

$$\widehat{\boldsymbol{\Theta}} = (A^T V^{-1} A)^{-1} A^T V^{-1}\boldsymbol{y} \equiv B\boldsymbol{y}.$$

Таким образом, оценки параметров $\widehat{\boldsymbol{\Theta}}$ представляют собой линейную функцию исходных измерений $\boldsymbol{y}$.

Используя формулы переноса ошибок, можно найти ковариационную матрицу оценок параметров $U_{ij} = cov[\widehat{\Theta}_i, \widehat{\Theta}_j]$:

$$U = BVB^T = (A^T V^{-1} A)^{-1}.$$

Это эквивалентно $(U^{-1})_{ij} = \frac{1}{2}\Big[\frac{\partial^2\chi^2}{\partial\Theta_i\partial\Theta_j}\Big]_{\mathbf{\Theta}=\hat{\mathbf{\Theta}}}$.

В случае линейной функции $\lambda(x; \mathbf{\Theta})$ χ^2 является квадратичной функцией $\mathbf{\Theta}$:

$$\chi^2 = \chi^2(\hat{\mathbf{\Theta}}) + \frac{1}{2}\sum_{i,j=1}^{m}\Big[\frac{\partial^2\chi^2}{\partial\Theta_i\partial\Theta_j}\Big]_{\mathbf{\Theta}=\hat{\mathbf{\Theta}}}(\Theta_i - \hat{\Theta}_i)(\Theta_j - \hat{\Theta}_j)$$

Аппроксимация полиномом.

Частным случаем линейной функции, когда коэффициенты $a_j(x)$ равны соответствующим степеням x, является аппроксимация функции $\lambda(x; \mathbf{\Theta})$ полиномом порядка m ($m + 1$ параметр):

$$\lambda(x; \Theta_0, \ldots, \Theta_m) = \sum_{j=0}^{m} x^j \Theta_j.$$