Statistical analysis of experimental data
Measurements and their uncertainties

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

Lecture 04

October 24, 2024

Measurements and their uncertainties

Probability distribution function (PDF)

also called "Probability density function" in some books

For given random variable X, probability distribution function, $f(x)$, describes the probability to obtain given numerical result x (in single experiment). For infinitesimal interval dx:

$$P(x < X < x + dx) = f(x)\, dx$$

For arbitrary interval $[x_1, x_2]$:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} dx\, f(x)$$

This can be considered an alternative definition of $f(x)$

Cumulative distribution function

We can also define cumulative distribution function $F(x)$, which is the probability that an experiment will result in a value not grater than x:

$$F(x) \;=\; P(X < x)$$

Probability distribution function can be then written as the derivative:

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

Probability that the X value observed is in the range from x_1 to x_2:

$$P(x_1 < X < x_2) \;=\; F(x_2) - F(x_1) \;=\; \int_{x_1}^{x_2} dx\, f(x)$$

Moment of order n (n^{th} moment) is defined as

$$\mu_n = \mathbb{E}(X^n) = \langle x^n \rangle = \int_{-\infty}^{+\infty} dx\, x^n f(x) \qquad = \sum_i x_i^n f(x_i)$$

Mean value is, by definition, the first ($n = 1$) moment of the probability distribution, $\mu \equiv \mu_1$.

Central moment of order n is defined as

$$m_n = \mathbb{E}((X-\mu)^n) = \langle (x-\mu)^n \rangle = \int_{-\infty}^{+\infty} dx\, (x-\mu)^n f(x)$$

$$= \sum_i (x_i - \mu)^n f(x_i)$$

By calculating moments of the probability distribution we can get information about its shape.

Moments of distribution functions

For the lowest order moments we have:

$$\mu_0 \equiv 1 \qquad m_0 \equiv 1 \qquad \text{normalization of } f(x)$$

$$\mu_1 \equiv \mu \qquad m_1 \equiv 0 \qquad \text{definition of mean value of } f(x)$$

The first moment which gives us information about the shape of $f(x)$ is **Variance**, which is the second central moment:

$$\sigma^2 = \mathbb{V}(X) = m_2 = \langle (x-\mu)^2 \rangle = \int_{-\infty}^{+\infty} dx \, (x-\mu)^2 \, f(x)$$

$$= \sum_i (x_i - \mu)^2 \, f(x_i)$$

The square root of the variance is referred to as the standard deviation σ. Describes the average difference between measurements x_i and the mean μ

Measurements and their uncertainties

Measurements

So far, we considered different properties of probability density functions and random variables assuming PDF parameters are known.

But this is not the case in most measurements.

We do experiments to extract parameters of the distribution!
Or even to reconstruct the shape of the PDF, if not predicted by theory.

We can constrain shape of the PDF by measuring its moments.

Mean μ and standard deviation σ are often of primary importance.
How do we reconstruct them? What is the precision of our estimate?

Measurements

So far, we considered different properties of probability density functions and random variables assuming PDF parameters are known.

But this is not the case in most measurements.

We do experiments to extract parameters of the distribution!
Or even to reconstruct the shape of the PDF, if not predicted by theory.

We can constrain shape of the PDF by measuring its moments.

Mean μ and standard deviation σ are often of primary importance.
How do we reconstruct them? What is the precision of our estimate?

Let us assume that random variable X is described by probability distribution $f(x)$.
We perform N experiments resulting in set of N measurements x_i, $i = 1 \ldots N$ (our sample).

Sample mean

is the arithmetic mean of the obtained numerical results:

$$\bar{x} = \frac{1}{N} \sum_i x_i$$

Expected value of the sample mean:

$$\mathbb{E}(\bar{x}) = \frac{1}{N} \mathbb{E}(x_1 + \ldots x_N) = \frac{1}{N} N \mu = \mu$$

$\Rightarrow$ The sample mean is an unbiased estimator of the true mean μ.

Sample mean

is the arithmetic mean of the obtained numerical results:

$$\bar{x} = \frac{1}{N} \sum_i x_i$$

Expected value of the sample mean:

$$\mathbb{E}(\bar{x}) = \frac{1}{N} \mathbb{E}(x_1 + \ldots x_N) = \frac{1}{N} N \mu = \mu$$

$\Rightarrow$ The sample mean is an unbiased estimator of the true mean μ.

Law of Large Numbers

"under suitable conditions"

The sample mean tends to the mean of the distribution

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} x_i = \mu$$

In the limit $N \to \infty$ sample mean is no longer a random variable...

Variance of sample mean

In the limit $N \to \infty$ we can obtain exact value of μ.

How precise is our estimate $\bar{x}$ of the true mean μ for finite N?

We can calculate variance of sample mean:

$$\begin{aligned}
\mathbb{V}(\bar{x}) &= \langle\, (\bar{x} - \mu)^2 \rangle \;=\; \langle\; (\overline{x - \mu})^{\,2}\; \rangle \\
&= \frac{1}{N^2} \left\langle \sum_i (x_i - \mu) \;\cdot\; \sum_j (x_j - \mu) \right\rangle \\
&= \frac{1}{N^2} \sum_i \langle (x_i - \mu)^2 \rangle \;+\; \frac{1}{N^2} \sum_{i,\, j \neq i} \langle (x_i - \mu)(x_j - \mu) \rangle \\
&= \frac{1}{N^2}\, N\, \sigma^2 \;+\; 0 \;=\; \frac{\sigma^2}{N}
\end{aligned}$$

Variance of sample mean

In the limit $N \to \infty$ we can obtain exact value of μ.

How precise is our estimate $\bar{x}$ of the true mean μ for finite N?

We can calculate variance of sample mean:

$$\begin{aligned}
\mathbb{V}(\bar{x}) &= \langle\, (\bar{x}-\mu)^2 \rangle \;=\; \langle\; (\overline{x-\mu})^{\,2}\; \rangle \\
&= \frac{1}{N^2} \left\langle \sum_i (x_i-\mu) \cdot \sum_j (x_j-\mu) \right\rangle \\
&= \frac{1}{N^2} \sum_i \langle (x_i-\mu)^2 \rangle \;+\; \frac{1}{N^2} \sum_{i,\, j\neq i} \langle (x_i-\mu)(x_j-\mu) \rangle \\
&= \frac{1}{N^2}\, N\, \sigma^2 \;+\; 0 \;=\; \frac{\sigma^2}{N}
\end{aligned}$$

Sample mean is a random variable with mean μ and standard deviation $\frac{\sigma}{\sqrt{N}}$

Variance estimate

Consider the value for the mean squared difference:

$$
\begin{aligned}
d^2 &\equiv \frac{1}{N}\sum_i (x_i - \bar{x})^2 = \frac{1}{N}\sum_i (x_i - \mu - (\bar{x} - \mu))^2 \\
&= \frac{1}{N}\sum_i \left[(x_i - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2\right] \\
&= \frac{1}{N}\sum_i (x_i - \mu)^2 - 2(\bar{x} - \mu)\frac{1}{N}\sum_i (x_i - \mu) + (\bar{x} - \mu)^2 \\
&= \frac{1}{N}\sum_i (x_i - \mu)^2 - (\bar{x} - \mu)^2
\end{aligned}
$$

Variance estimate

Consider the value for the mean squared difference:

$$
\begin{aligned}
d^2 &\equiv \frac{1}{N}\sum_i (x_i - \bar{x})^2 = \frac{1}{N}\sum_i (x_i - \mu - (\bar{x} - \mu))^2 \\
&= \frac{1}{N}\sum_i \left[(x_i - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2\right] \\
&= \frac{1}{N}\sum_i (x_i - \mu)^2 - 2(\bar{x} - \mu)\frac{1}{N}\sum_i (x_i - \mu) + (\bar{x} - \mu)^2 \\
&= \frac{1}{N}\sum_i (x_i - \mu)^2 - (\bar{x} - \mu)^2
\end{aligned}
$$

Expected value:

$$
\langle d^2 \rangle = \mathbb{V}(x_i) - \mathbb{V}(\bar{x}) = \sigma^2 - \frac{\sigma^2}{N} = \frac{N-1}{N}\,\sigma^2
$$

Sample mean variance

Sample variance for a sample of N measurements x_i is defined as

$$s^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

where Bessel's correction factor

$$\frac{\sigma^2}{\langle d^2 \rangle} = \frac{N}{N-1} > 1$$

is applied to obtain an unbiased estimator of the standard deviation:

$$\langle s^2 \rangle = \sigma^2$$

Sample mean variance

Sample variance for a sample of N measurements x_i is defined as

$$s^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

where Bessel's correction factor

$$\frac{\sigma^2}{\langle d^2 \rangle} = \frac{N}{N-1} > 1$$

is applied to obtain an unbiased estimator of the standard deviation:

$$\langle s^2 \rangle = \sigma^2$$

Bessel's correction is crucial for proper (unbiased) estimate of the sample mean variance:

$$\mathbb{V}(\bar{x}) = \frac{\sigma^2}{N} = \left\langle \frac{1}{N(N-1)} \sum_i (x_i - \bar{x})^2 \right\rangle$$

Variance estimate

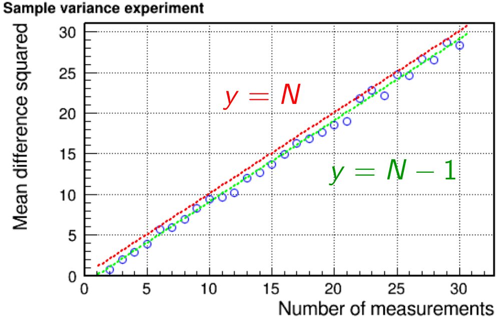
04_sample_variance.ipynb Open in Colab

Run multiple numerical experiments to calculate expected value for

$$y = \sum_i (x_i - \bar{x})^2 = N\, d^2$$

for x_i being gaussian random variables with $\mu = 0$ and $\sigma = 1$

100 experiments



Variance estimate

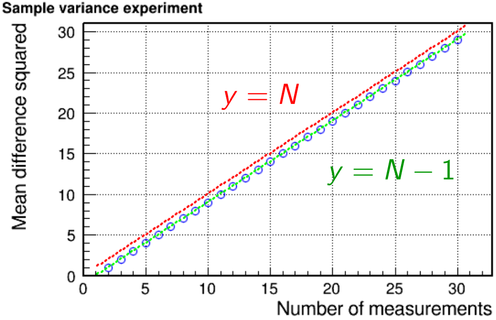
04_sample_variance.ipynb Open in Colab

Run multiple numerical experiments to calculate expected value for

$$y \;=\; \sum_i (x_i - \bar{x})^2 \;=\; N\, d^2$$

for x_i being gaussian random variables with $\mu = 0$ and $\sigma = 1$

100000 experiments



Expected value

$$\langle y \rangle \;=\; (N-1)\,\sigma^2$$

Measurements and their uncertainties

Probability distribution function $f(x)$ can be uniquely described by specifying all its moments. But this is an infinite set... We can also present full information about moments by introducing

Moment Generating Function

The moment generating function of a PDF of a random variable X can be written as

$$M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mu^n = \mu_0 + t\,\mu_1 + \frac{t^2}{2!}\mu_2 + \ldots$$

Probability distribution function $f(x)$ can be uniquely described by specifying all its moments. But this is an infinite set... We can also present full information about moments by introducing

Moment Generating Function

The moment generating function of a PDF of a random variable X can be written as

$$M(t) \; = \; \sum_{n=0}^{\infty} \frac{t^n}{n!} \, \mu^n \; = \; \mu_0 + t\,\mu_1 + \frac{t^2}{2!}\,\mu_2 + \ldots$$

$\Rightarrow$ all moments can be now obtained as derivatives at $t = 0$:

$$\mu_k \; = \; \left. \frac{\partial^k M(t)}{\partial t^k} \right|_{t=0}$$

Probability distribution function $f(x)$ can be uniquely described by specifying all its moments. But this is an infinite set... We can also present full information about moments by introducing

Moment Generating Function

The moment generating function of a PDF of a random variable X can be written as

$$M(t) \;=\; \sum_{n=0}^{\infty} \frac{t^n}{n!}\, \mu^n \;=\; \mu_0 + t\,\mu_1 + \frac{t^2}{2!}\,\mu_2 + \ldots$$

$\Rightarrow$ all moments can be now obtained as derivatives at $t = 0$:

$$\mu_k \;=\; \left.\frac{\partial^k M(t)}{\partial t^k}\right|_{t=0}$$

By comparing with a Taylor series expansion of exponential function we realize that

$$M(t) \;=\; \int_{-\infty}^{+\infty} dx\; f(x)\, \left(1 + tx + \frac{(tx)^2}{2!} + \ldots\right) \;=\; \int_{-\infty}^{+\infty} dx\; f(x)\; e^{tx} \;=\; \mathbb{E}(e^{tX})$$

Moment Generating Function

If X and Y are two independent variables and $Z = X + Y$ then

$$M_z(t) = M_x(t) \cdot M_y(t)$$

where M_x, M_y and M_z are moment generating functions for X, Y and Z

Moment Generating Function

If X and Y are two independent variables and $Z = X + Y$ then

$$M_z(t) = M_x(t) \cdot M_y(t)$$

where M_x, M_y and M_z are moment generating functions for X, Y and Z

For the Gaussian distribution:

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

Moment Generating Function

If X and Y are two independent variables and $Z = X + Y$ then

$$M_z(t) \quad = \quad M_x(t) \cdot M_y(t)$$

where M_x, M_y and M_z are moment generating functions for X, Y and Z

For the Gaussian distribution:

$$\begin{aligned} M(t) &= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \\ \frac{\partial M(t)}{\partial t} &= (\mu + \sigma^2 t)\; M(t) & \rightarrow \mu \text{ for } t = 0 \\ \frac{\partial^2 M(t)}{\partial t^2} &= \left[\sigma^2 + (\mu + \sigma^2 t)^2\right]\; M(t) & \rightarrow \sigma^2 + \mu^2 \end{aligned}$$

Moment Generating Function

If X and Y are two independent variables and $Z = X + Y$ then

$$M_z(t) \quad = \quad M_x(t) \cdot M_y(t)$$

where M_x, M_y and M_z are moment generating functions for X, Y and Z

For the Gaussian distribution:

$$\begin{aligned} M(t) &= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \\ \frac{\partial M(t)}{\partial t} &= \left(\mu + \sigma^2 t\right) M(t) \\ \frac{\partial^2 M(t)}{\partial t^2} &= \left[\sigma^2 + (\mu + \sigma^2 t)^2\right] \; M(t) \end{aligned}$$

PDF described by generating function of this form is by definition gaussian

Central Limit Theorem

Consider the sum of N independent variables X_i:

$$Y = \sum_{i=1}^{N} X_i$$

For simplicity, let us assume mean values are zero, $\mathbb{E}(X_i) \equiv 0 \quad \Rightarrow \quad \mu_y = 0$

Central Limit Theorem

Consider the sum of N independent variables X_i:

$$Y = \sum_{i=1}^{N} X_i$$

For simplicity, let us assume mean values are zero, $\mathbb{E}(X_i) \equiv 0 \quad \Rightarrow \quad \mu_y = 0$

Variance of Y is given by

$$\sigma^2 = \mathbb{V}(Y) = \sum_{i=1}^{N} \sigma_i^2$$

and increases with increasing N...

Central Limit Theorem

Consider the sum of N independent variables X_i:

$$Y = \sum_{i=1}^{N} X_i$$

For simplicity, let us assume mean values are zero, $\mathbb{E}(X_i) \equiv 0 \Rightarrow \mu_y = 0$

Variance of Y is given by

$$\sigma^2 = \mathbb{V}(Y) = \sum_{i=1}^{N} \sigma_i^2$$

and increases with increasing N...

$\Rightarrow$ let us consider random variable scaled to unit variance $(\mu_z = 0,\ \sigma_z = 1)$

$$Z = \frac{1}{\sigma} Y = \sum_{i=1}^{N} \frac{X_i}{\sigma}$$

Distribution of Z should tend to a fixed distribution in $N \to \infty$ limit. What is it?

Central Limit Theorem

Moment generating function for Z:

$$M_Z(t) = \prod_{i=1}^{N} M_i\left(\frac{t}{\sigma}\right)$$

where moment generating function of variables X_i can be written as

$$M_i\left(\frac{t}{\sigma}\right) = 1 + \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{\mu_{3(i)}}{3!}\left(\frac{t}{\sigma}\right)^3 + \ldots$$

where the term linear in t vanishes because of $\mu_i \equiv 0$

Central Limit Theorem

Moment generating function for Z:

$$M_Z(t) = \prod_{i=1}^{N} M_i\left(\frac{t}{\sigma}\right)$$

where moment generating function of variables X_i can be written as

$$M_i\left(\frac{t}{\sigma}\right) = 1 + \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{\mu_{3(i)}}{3!}\left(\frac{t}{\sigma}\right)^3 + \ldots$$

where the term linear in t vanishes because of $\mu_i \equiv 0$

Logarithm of moment generating function for Z

$$\ln M_Z(t) = \sum_{i=1}^{N} \ln M_i\left(\frac{t}{\sigma}\right) = \sum_{i=1}^{N} \ln\left(1 + \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 + \ldots\right)$$

Central Limit Theorem

Shape of the PDF is defined by $M(t)$ shape for $t \approx 0$, we can expand $\ln(1+\varepsilon)$

$$\ln M_Z(t) = \sum_{i=1}^{N} \ln\left(1 + \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 + \ldots\right) = \sum_{i=1}^{N}\left(\frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 + \ldots\right)$$

ignoring terms $\left(\frac{t}{\sigma}\right)^3$ and higher: $$\approx \sum_{i=1}^{N} \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 = \frac{1}{2}\, t^2$$

Central Limit Theorem

Shape of the PDF is defined by $M(t)$ shape for $t \approx 0$, we can expand $\ln(1+\varepsilon)$

$$\ln M_Z(t) = \sum_{i=1}^{N} \ln\left(1 + \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 + \ldots\right) = \sum_{i=1}^{N}\left(\frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 + \ldots\right)$$

ignoring terms $\left(\frac{t}{\sigma}\right)^3$ and higher: $\approx \sum_{i=1}^{N} \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma}\right)^2 = \frac{1}{2}\,t^2$

We conclude that in the limit $N \to \infty$ (and $t \to 0$)

$$M_Z(t) = \exp\left(\frac{1}{2}\,t^2\right)$$

$\Rightarrow$ Z has gaussian distribution function with unit deviation ($\sigma = 1$)

Central Limit Theorem (2)

alternative proof

If all X_i have same variance σ_x^2 $\Rightarrow$ variance of Y: $\sigma^2 = \sigma_x^2 \cdot N$ $\Rightarrow$ $\sigma = \sigma_x \cdot \sqrt{N}$

$$M_i\left(\frac{t}{\sigma}\right) = 1 + \frac{\sigma_x^2}{2}\left(\frac{t}{\sigma_x\sqrt{N}}\right)^2 + \frac{\mu_{3(i)}}{3!}\left(\frac{t}{\sigma_x\sqrt{N}}\right)^3 + \ldots$$

$$= 1 + \frac{t^2}{2N} + \frac{\mu_{3(i)}}{3!\sigma_x^3} \cdot \frac{t^3}{N^{3/2}} + \ldots$$

Central Limit Theorem (2)

alternative proof

If all X_i have same variance σ_x^2 $\Rightarrow$ variance of Y: $\sigma^2 = \sigma_x^2 \cdot N$ $\Rightarrow$ $\sigma = \sigma_x \cdot \sqrt{N}$

$$M_i\left(\frac{t}{\sigma}\right) = 1 + \frac{\sigma_x^2}{2}\left(\frac{t}{\sigma_x\sqrt{N}}\right)^2 + \frac{\mu_{3(i)}}{3!}\left(\frac{t}{\sigma_x\sqrt{N}}\right)^3 + \ldots$$

$$= 1 + \frac{t^2}{2N} + \frac{\mu_{3(i)}}{3!\sigma_x^3} \cdot \frac{t^3}{N^{3/2}} + \ldots$$

Then the moment generating function for Z can be written as:

$$M_Z(t) = \prod_{i=1}^{N} M_i\left(\frac{t}{\sigma}\right) = \left(1 + \frac{t^2}{2N} + \ldots\right)^N$$

and in the limit $N \to \infty$ we get:

$$M_Z(t) = \exp\left(\frac{1}{2}\, t^2\right)$$

Example (1)

04_central_limit.ipynb Open in Colab

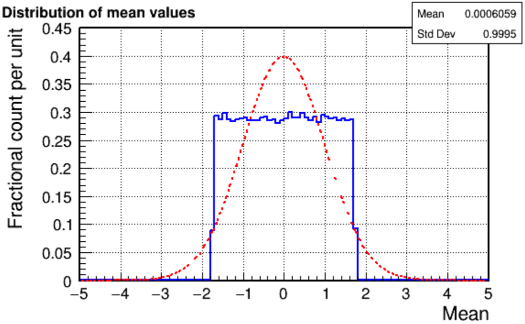
Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 1$



Example (1)

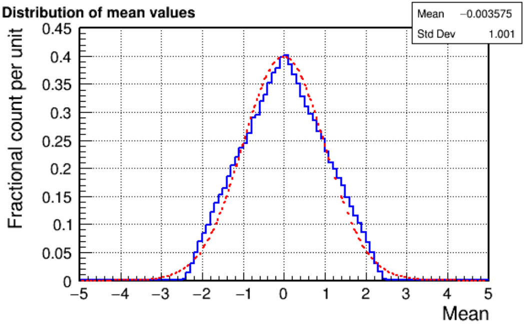
04_central_limit.ipynb Open in Colab

Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

N = 2



Example (1)

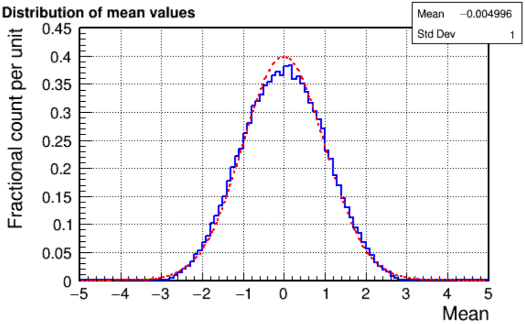
04_central_limit.ipynb   Open in Colab

Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

N = 3



Example (1)

04_central_limit.ipynb Open in Colab

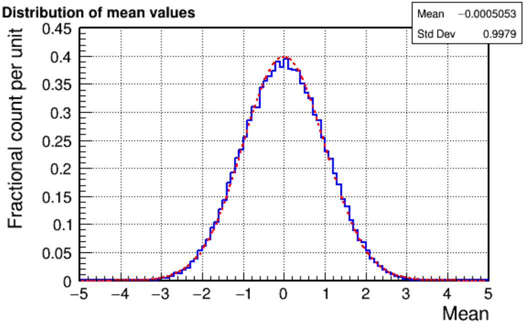
Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 5$



Example (1)

04_central_limit.ipynb Open in Colab

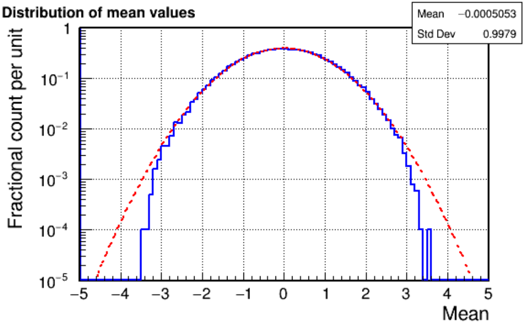
Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 5



Example (1)

04_central_limit.ipynb Open in Colab

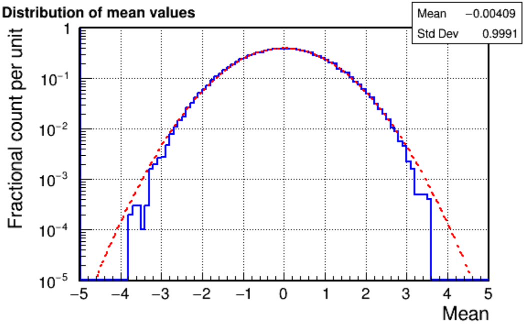
Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 7$



Example (1)

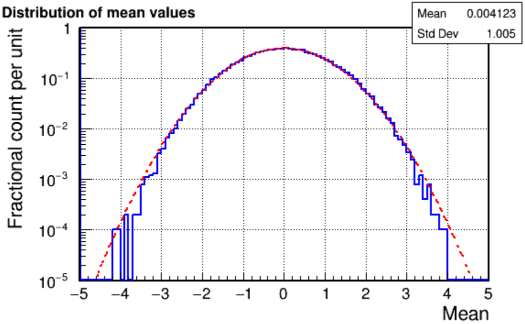
04_central_limit.ipynb   Open in Colab

Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:

compared with normal distribution   scaling to $\sigma = 1$ applied

$N = 10$



Example (1)

04_central_limit.ipynb Open in Colab

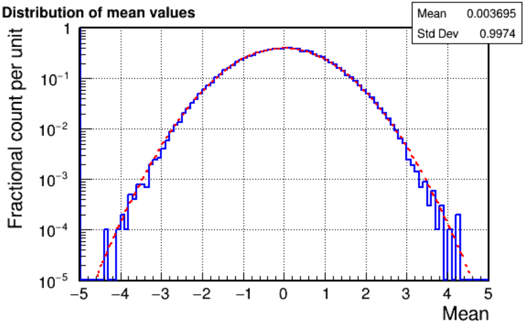
Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 20



Example (1)

04_central_limit.ipynb Open in Colab

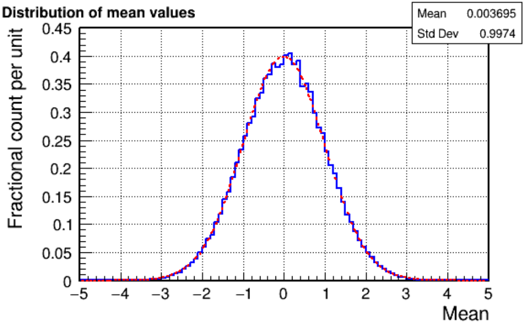
Simplest case: X_i are random numbers from uniform distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 20



Example (2)

04_central_limit2.ipynb Open in Colab

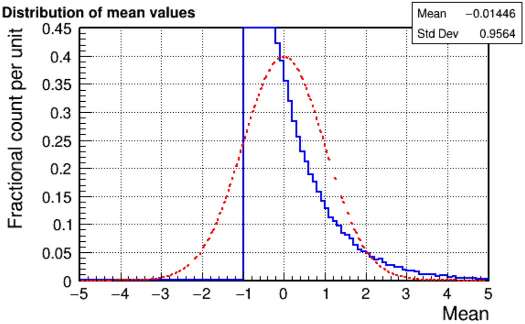
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 1



Example (2)

04_central_limit2.ipynb Open in Colab

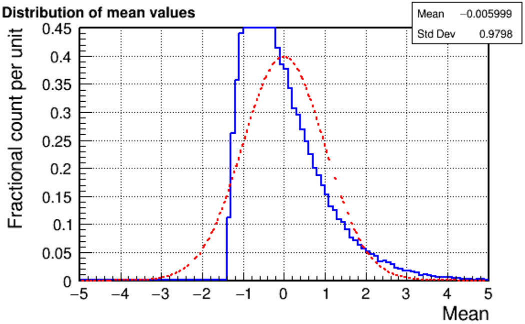
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 2$



Example (2)

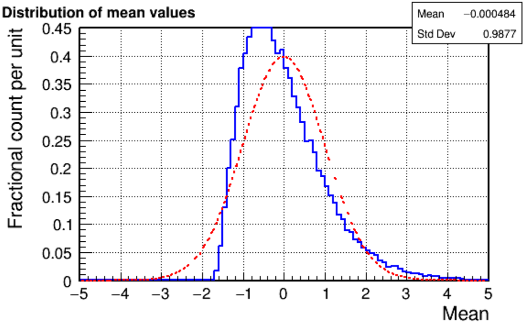
04_central_limit2.ipynb Open in Colab

More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

N = 3



Example (2)

04_central_limit2.ipynb Open in Colab

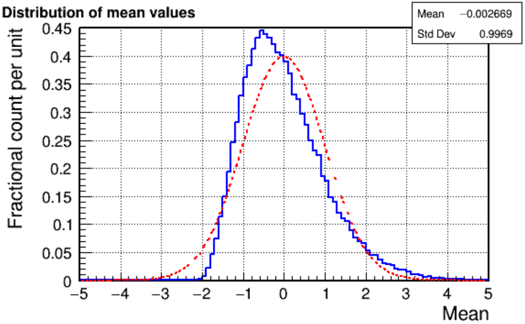
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 5



Example (2)

04_central_limit2.ipynb Open in Colab

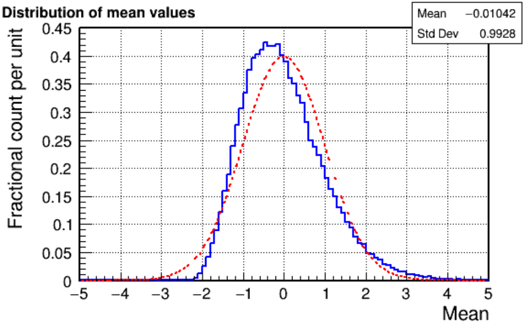
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 7$



Example (2)

04_central_limit2.ipynb Open in Colab

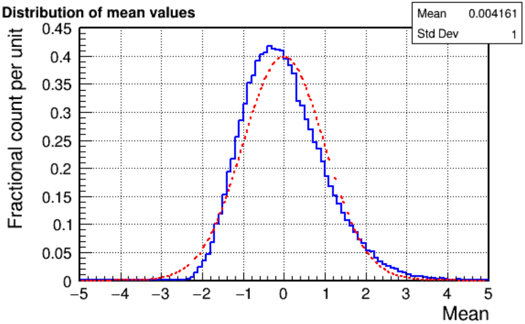
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 10



Example (2)

04_central_limit2.ipynb Open in Colab

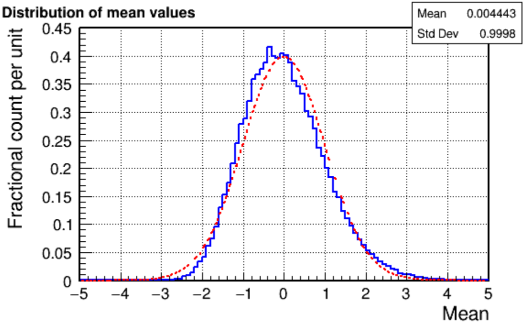
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 20$



Example (2)

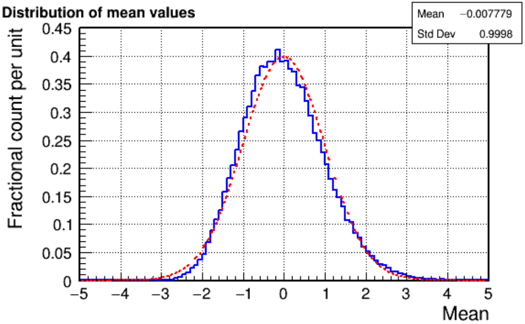
04_central_limit2.ipynb Open in Colab

More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 30$



Example (2)

04_central_limit2.ipynb Open in Colab

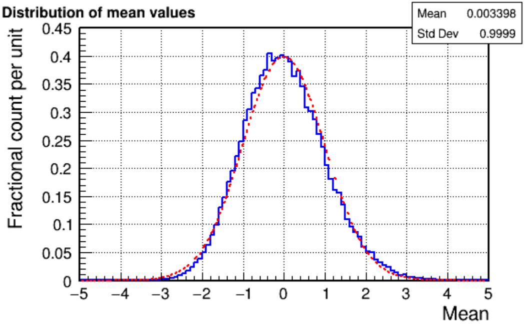
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 50$



Example (2)

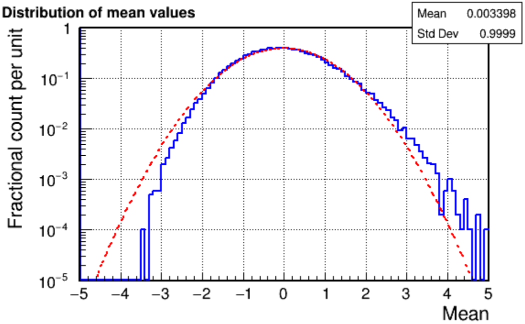
04_central_limit2.ipynb Open in Colab

More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

N = 50



Example (2)

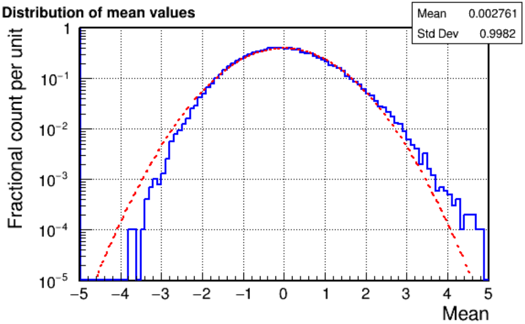
04_central_limit2.ipynb Open in Colab

More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 70$



Example (2)

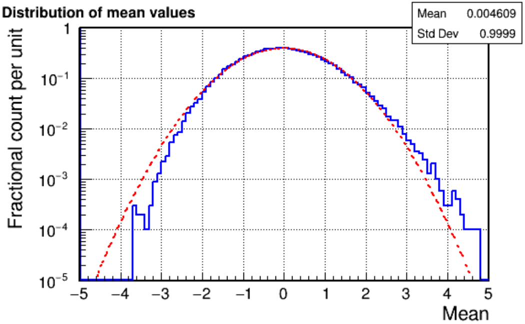
04_central_limit2.ipynb Open in Colab

More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 100$



Example (2)

04_central_limit2.ipynb Open in Colab

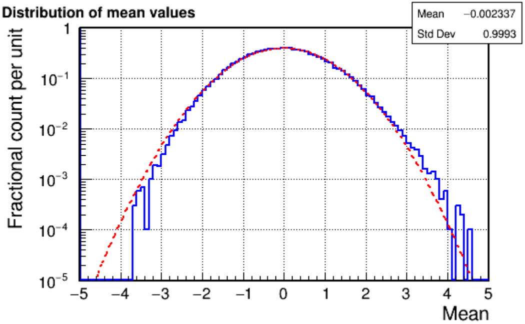
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

$N = 200$



Example (2)

04_central_limit2.ipynb Open in Colab

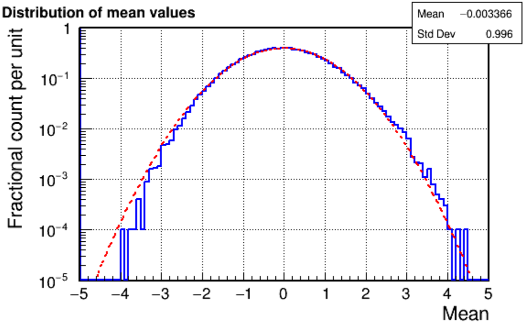
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 300



Example (2)

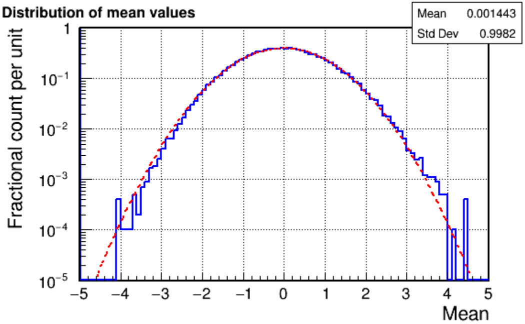
04_central_limit2.ipynb Open in Colab

More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:
compared with normal distribution

scaling to $\sigma = 1$ applied

N = 500



Example (2)

04_central_limit2.ipynb Open in Colab

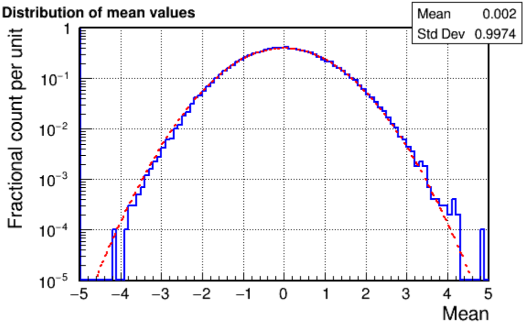
More problematic: X_i are random numbers from exponential distribution

Distribution of the mean of N values:

compared with normal distribution

scaling to $\sigma = 1$ applied

N = 1000



Central Limit Theorem as presented by M. Bonamente

The sum of a large number of independent random variables is approximately distributed as a Gaussian. The mean of the distribution is the sum of the means of the variables and the variance of the distribution is the sum of the variances of the variables. This result holds regardless of the distribution of each individual variable.

Central Limit Theorem

as presented by M. Bonamente

The sum of a large number of independent random variables is approximately distributed as a Gaussian. The mean of the distribution is the sum of the means of the variables and the variance of the distribution is the sum of the variances of the variables. This result holds regardless of the distribution of each individual variable.

This is a very strong statement!

It is true in most of the "physical" cases, but it is not true in general case...

In our proof we had to assume that $M(t)$ can be defined, that is all moments of the considered probability distribution can be calculated and are finite. This is not always the case...

Example (3)

04_central_limit3.ipynb Open in Colab

Consider random numbers x_i from Cauchy distribution

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$$

It is properly normalized, but already its mean is not well defined:

$$\mathbb{E}(X) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{2x\,dx}{1+x^2}$$

Example (3)

04_central_limit3.ipynb Open in Colab

Consider random numbers x_i from Cauchy distribution

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$$

It is properly normalized, but already its mean is not well defined:

$$\mathbb{E}(X) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{2x\,dx}{1+x^2} = \frac{1}{2\pi} \ln(1+x^2)\Big|_{-\infty}^{+\infty}$$

Higher moments diverge even faster...

Our basic assumption is not fulfilled $\Rightarrow$ theorem does not hold !

Example (3)

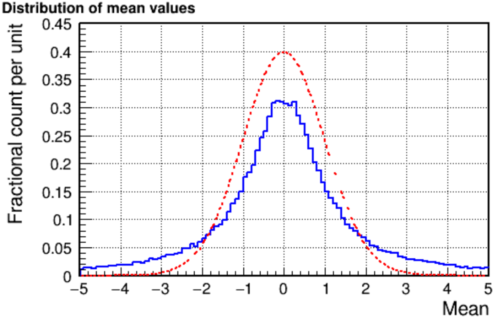
04_central_limit3.ipynb Open in Colab

Consider random numbers x_i from Cauchy distribution

Distribution of the mean of N values:

compared with normal distribution

$N = 1$



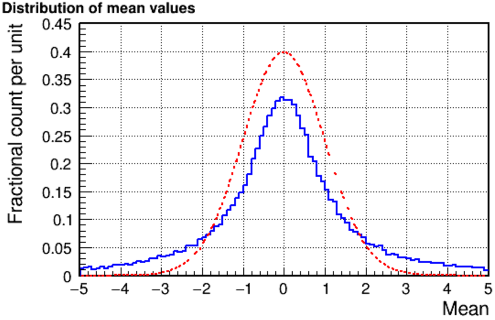
Example (3)

04_central_limit3.ipynb Open in Colab

Consider random numbers x_i from Cauchy distribution

Distribution of the mean of N values: compared with normal distribution

$N = 10$



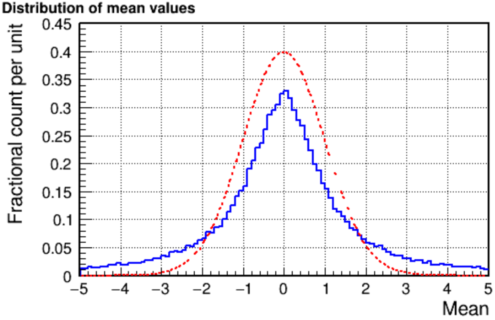
Example (3)

04_central_limit3.ipynb Open in Colab

Consider random numbers x_i from Cauchy distribution

Distribution of the mean of N values: compared with normal distribution

N = 100



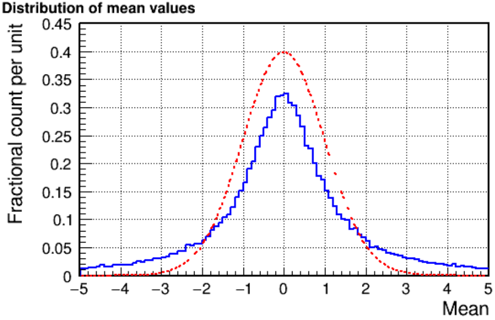
Example (3)

04_central_limit3.ipynb Open in Colab

Consider random numbers x_i from Cauchy distribution

Distribution of the mean of N values: compared with normal distribution

N = 1000



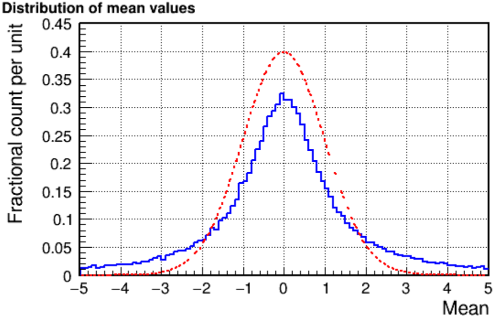
Example (3)

04_central_limit3.ipynb Open in Colab

Consider random numbers x_i from Cauchy distribution

Distribution of the mean of N values: compared with normal distribution

$N = 10000$



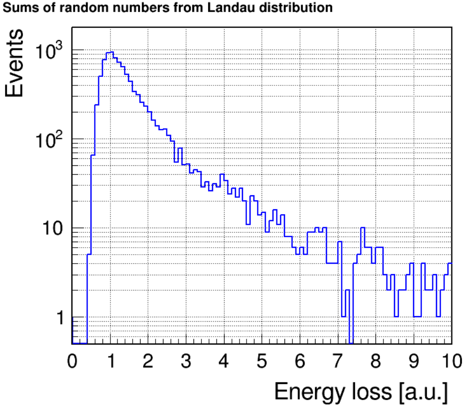
Mean of random numbers from Cauchy distribution has Cauchy distribution itself, even for $N \to \infty$

Example (4): Landau distribution

Describes the energy loss distribution for relativistic charged particles matter.

In the limit $\beta\gamma \to \infty$, mean and RMS of the energy loss distribution are not well defined!

Only the most probable value is...



single deposit $\Rightarrow \frac{\sigma}{Mean} \approx 0.76$

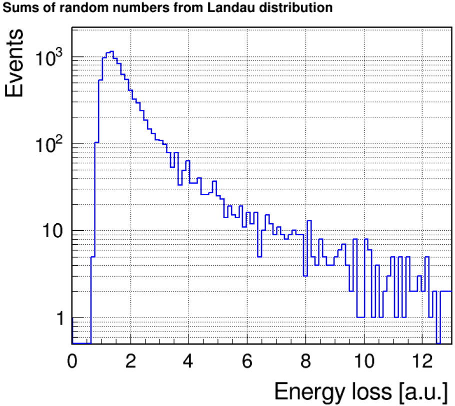
(from histogram range shown)

Example (4): Landau distribution

Describes the energy loss distribution for relativistic charged particles matter.

In the limit $\beta\gamma \to \infty$, mean and RMS of the energy loss distribution are not well defined!

Only the most probable value is...



Sum of Landau distributed deposits is also described by Landau distribution!

$N_{sum} = 4 \Rightarrow \frac{\sigma}{Mean} \approx 0.73$

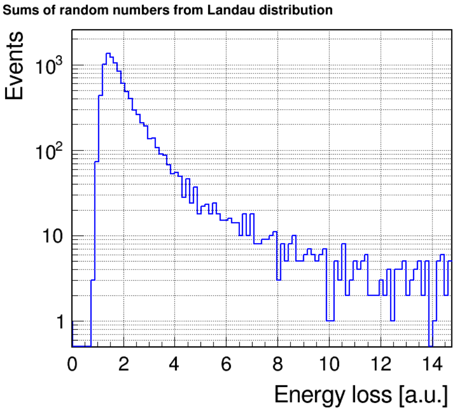
(from histogram range shown)

Example (4): Landau distribution

Describes the energy loss distribution for relativistic charged particles matter.

In the limit $\beta\gamma \to \infty$, mean and RMS of the energy loss distribution are not well defined!

Only the most probable value is...



Sum of Landau distributed deposits is also described by Landau distribution!

$N_{sum} = 9 \Rightarrow \frac{\sigma}{Mean} \approx 0.72$

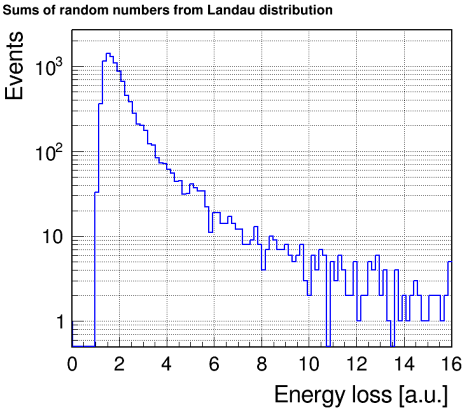
(from histogram range shown)

Example (4): Landau distribution

Describes the energy loss distribution for relativistic charged particles matter.

In the limit $\beta\gamma \rightarrow \infty$, mean and RMS of the energy loss distribution are not well defined!

Only the most probable value is...



Sum of Landau distributed deposits is also described by Landau distribution!

$N_{sum} = 16 \Rightarrow \frac{\sigma}{Mean} \approx 0.70$

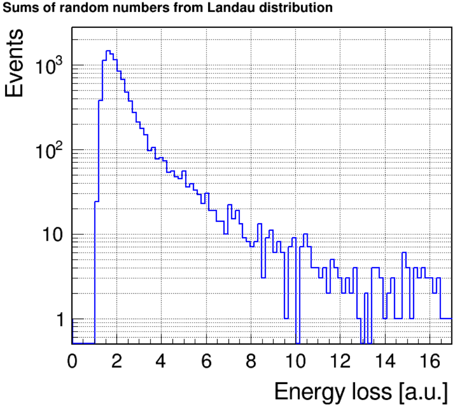
(from histogram range shown)

Example (4): Landau distribution

Describes the energy loss distribution for relativistic charged particles matter.

In the limit $\beta\gamma \to \infty$, mean and RMS of the energy loss distribution are not well defined!

Only the most probable value is...



Sum of Landau distributed deposits is also described by Landau distribution!

$N_{sum} = 25 \Rightarrow \frac{\sigma}{Mean} \approx 0.69$

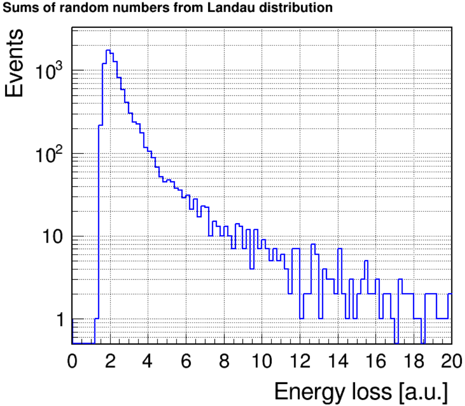
(from histogram range shown)

Example (4): Landau distribution

Describes the energy loss distribution for relativistic charged particles matter.

In the limit $\beta\gamma \rightarrow \infty$, mean and RMS of the energy loss distribution are not well defined!

Only the most probable value is...



Sum of Landau distributed deposits is also described by Landau distribution!

$$N_{sum} = 100 \Rightarrow \frac{\sigma}{Mean} \approx 0.67$$

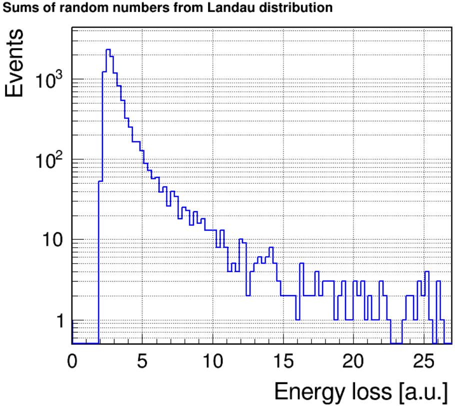
(from histogram range shown)

Example (4): Landau distribution

Describes the energy loss distribution for relativistic charged particles matter.

In the limit $\beta\gamma \rightarrow \infty$, mean and RMS of the energy loss distribution are not well defined!

Only the most probable value is...



Sum of Landau distributed deposits is also described by Landau distribution!

$N_{sum} = 2500 \Rightarrow \frac{\sigma}{Mean} \approx 0.62$

(from histogram range shown)

High energy tail significant even in large detector volumes!

Measurements and their uncertainties

Multiple measurements

It is often the case that we measure two or more variables at the same time and we are interested in joint probability of obtaining given result.

So far we assumed that variables were independent, eg. coming from different measurements or from repeated experiment...

Let us consider two random variables X and Y.
We can define joint probability function $h(x, y)$ by the relation:

$$P(x < X < x + dx \;\; \&\& \;\; y < Y < y + dy) \;\; = \;\; h(x, y)\, dx\, dy$$

which should be fulfilled for infinitesimal intervals dx and dy, or:

$$P(x_1 < X < x_2 \;\; \&\& \;\; y_1 < Y < y_2) \;\; = \;\; \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy\; h(x, y)$$

Joint probability function

Joint probability function gives full information on the measured set of variables. If we want to look only at one variable, we can consider marginal probability density functions defined as

$$f(x) = \int_{-\infty}^{+\infty} dy\ h(x,y)$$
$$g(y) = \int_{-\infty}^{+\infty} dx\ h(x,y)$$

Joint probability function

Joint probability function gives full information on the measured set of variables. If we want to look only at one variable, we can consider marginal probability density functions defined as

$$f(x) = \int_{-\infty}^{+\infty} dy \; h(x, y)$$

$$g(y) = \int_{-\infty}^{+\infty} dx \; h(x, y)$$

We can also directly extract parameters of X and Y distributions:

$$\mu_x = \mathbb{E}(X) = \iint dx \; dy \; x \; h(x, y)$$

$$\sigma_x^2 = \mathbb{E}((X - \mu_x)^2) = \iint dx \; dy \; (x - \mu_x)^2 \; h(x, y)$$

Joint probability function

Joint probability function gives full information on the measured set of variables. If we want to look only at one variable, we can consider marginal probability density functions defined as

$$
\begin{aligned}
f(x) &= \int_{-\infty}^{+\infty} dy \; h(x,y) \\
g(y) &= \int_{-\infty}^{+\infty} dx \; h(x,y)
\end{aligned}
$$

We can also directly extract parameters of X and Y distributions:

$$
\begin{aligned}
\mu_y = \mathbb{E}(Y) &= \iint dx \; dy \; y \; h(x,y) \\
\sigma_y^2 = \mathbb{E}((Y-\mu_y)^2) &= \iint dx \; dy \; (y-\mu_y)^2 \; h(x,y)
\end{aligned}
$$

Covariance

Covariance of two random variables X and Y is defined as:

$$\mathbb{C}(X,Y) \;=\; Cov(X,Y) \;=\; \mathbb{E}((X-\mu_x)(Y-\mu_y)) = \sigma_{xy}$$

Covariance

Covariance of two random variables X and Y is defined as:

$$\mathbb{C}(X, Y) = Cov(X, Y) = \mathbb{E}((X - \mu_x)(Y - \mu_y)) = \sigma_{xy}$$

We can also define the correlation coefficient:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_x \, \sigma_y} = \rho_{xy}$$

which is the number between -1 and $+1$.

Covariance

Covariance of two random variables X and Y is defined as:

$$\mathbb{C}(X,Y) \;=\; Cov(X,Y) \;=\; \mathbb{E}((X-\mu_x)(Y-\mu_y)) = \sigma_{xy}$$

We can also define the correlation coefficient:

$$\rho(X,Y) \;=\; \frac{Cov(X,Y)}{\sigma_x\,\sigma_y} = \rho_{xy}$$

which is the number between -1 and $+1$.

When two variables are independent, their correlation is zero:

$$h(x,y) \;=\; f(x)\cdot g(y) \qquad \Rightarrow \quad Cov(X,Y) \;=\; \rho(X,Y) \;=\; 0$$

But the opposite statement is not true!
Vanishing correlation does not guarantee that variables are independent!

Covariance

The role of covariance can be best shown on a simple example.

Let us consider random variable $Z = X + Y$. The expected value of Z:

$$\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y) = \mu_x + \mu_y$$

The variance of Z:

$$\begin{aligned}
\mathbb{V}(Z) &= \mathbb{E}((X + Y - \mu_x - \mu_y)^2) \\
&= \mathbb{E}((X - \mu_x)^2 + (Y - \mu_y)^2 + 2(X - \mu_x)(Y - \mu_y)) \\
&= \mathbb{V}(X) + \mathbb{V}(Y) + 2\,Cov(X, Y)
\end{aligned}$$

Covariance

The role of covariance can be best shown on a simple example.
Let us consider random variable $Z = X + Y$. The expected value of Z:

$$\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y) = \mu_x + \mu_y$$

The variance of Z:

$$\begin{aligned}
\mathbb{V}(Z) &= \mathbb{E}((X + Y - \mu_x - \mu_y)^2) \\
&= \mathbb{E}((X - \mu_x)^2 + (Y - \mu_y)^2 + 2(X - \mu_x)(Y - \mu_y)) \\
&= \mathbb{V}(X) + \mathbb{V}(Y) + 2\,Cov(X, Y) \\
\sigma_z^2 &= \sigma_x^2 + \sigma_y^2 + 2\rho_{xy}\sigma_x\sigma_y
\end{aligned}$$

Covariance

The role of covariance can be best shown on a simple example.
Let us consider random variable $Z = X + Y$. The expected value of Z:

$$\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y) = \mu_x + \mu_y$$

The variance of Z:

$$\begin{aligned}
\mathbb{V}(Z) &= \mathbb{E}((X + Y - \mu_x - \mu_y)^2) \\
&= \mathbb{E}((X - \mu_x)^2 + (Y - \mu_y)^2 + 2(X - \mu_x)(Y - \mu_y)) \\
&= \mathbb{V}(X) + \mathbb{V}(Y) + 2\,Cov(X, Y)
\end{aligned}$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\rho_{xy}\sigma_x\sigma_y$$

Correlation of input variables can have significant impact on result:

$$(\sigma_x - \sigma_y)^2 \leq \sigma_z^2 \leq (\sigma_x + \sigma_y)^2$$

Sample covariance

The true covariance/correlation between variables is usually unknown.

We want to estimate it from the set of N measurements (x_i, y_i).
For unbiased estimate we should calculate sample covariance:

$$s^2_{xy} \;=\; \frac{1}{N-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

The sample correlation coefficient is then defined as

$$r_{xy} \;=\; \frac{s^2_{xy}}{s_x \; s_y}$$

where s_x and s_y are sample variances (with Bessel's correction).

2D Gaussian distribution

Normal distribution for two correlated variables X and Y:

$$h(x, y) = N \exp\left[\frac{-1}{2(1-\rho_{xy}^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right]$$

where:

$$N = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}$$

2D Gaussian distribution

Normal distribution for two correlated variables X and Y:

$$h(x, y) = N \exp\left[\frac{-1}{2(1-\rho_{xy}^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right]$$

where:

$$N = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}$$

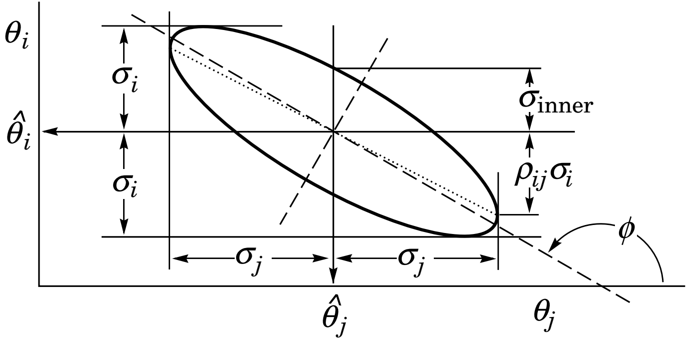
Due to correlation, distribution of X for fixed Y (and of Y for fixed X) becomes narrower than the marginal distribution !

$$\mathbb{V}(X|Y) = (1-\rho_{xy}^2)\cdot\mathbb{V}(X)$$

2D Gaussian distribution

Graphical presentation of variable correlation: "one sigma" contour

R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022)



"One sigma" contour: (x, y) values resulting in $h(x, y) = N \exp(-\frac{1}{2})$

Ellipse long axis direction:

$$\tan 2\phi = \frac{2\rho_{xy}\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$$

Example case of negative correlation

Multiple variable case

Definition of covariance can be generalized to the set of N variables X_i:

$$c_{ij} \; = \; Cov(X_i, X_j) \; = \; \mathbb{E}((X_i - \mu_i)(X_j - \mu_j))$$

Multiple variable case

Definition of covariance can be generalized to the set of N variables X_i:

$$c_{ij} \;=\; Cov(X_i, X_j) \;=\; \mathbb{E}((X_i - \mu_i)(X_j - \mu_j))$$

We can present it in a form of the covariance matrix:

$$\mathbb{C} \;=\; \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & c_{2N} \\ \vdots & & & \vdots \\ c_{N1} & c_{N2} & \cdots & c_{NN} \end{pmatrix}$$

where diagonal elements correspond to variances of the variables

$$c_{ii} \;=\; Cov(X_i, X_i) \;\equiv\; \mathbb{V}(X_i)$$

Multiple variable case

Let us now consider linear combination of N random variables

$$Y = \sum_{i=1}^{N} a_i \, X_i$$

Expected value is given by (from the linearity of the mean)

$$\mu_y = \sum a_i \, \mu_i$$

The variance can be calculated as:

$$\sigma_y^2 = \mathbb{V}\left(\sum a_i \, X_i\right) = \mathbb{E}\left[\left(\sum a_i \, X_i - \sum a_i \, \mu_i\right)^2\right]$$

Multiple variable case

Let us now consider linear combination of N random variables

$$Y = \sum_{i=1}^{N} a_i \, X_i$$

Expected value is given by (from the linearity of the mean)

$$\mu_y = \sum a_i \, \mu_i$$

The variance can be calculated as:

$$\sigma_y^2 = \mathbb{V}\left(\sum a_i \, X_i\right) = \mathbb{E}\left[\left(\sum a_i \, X_i - \sum a_i \, \mu_i\right)^2\right]$$

$$= \mathbb{E}\left[\left(\sum_i a_i \, (X_i - \mu_i)\right)\left(\sum_j a_j \, (X_j - \mu_j)\right)\right] = \sum_{i,j} a_i a_j \, c_{ij}$$

Multiple variable case

The variance of linear combination of variables

$$\sigma_y^2 = \sum_{i,j} a_i a_j \, c_{ij}$$

$$= \sum_i a_i^2 \sigma_i^2 + 2 \sum_{i,\, j>i} a_i a_j \, \sigma_i \sigma_j \, \rho_{ij}$$

factor 2 introduced due to $j > i$ condition

Multiple variable case

The variance of linear combination of variables

$$\sigma_y^2 = \sum_{i,j} a_i a_j \; c_{ij}$$

$$= \sum_i a_i^2 \sigma_i^2 \; + \; 2 \sum_{i,\, j>i} a_i a_j \; \sigma_i \sigma_j \; \rho_{ij}$$

factor 2 introduced due to $j > i$ condition

We can also write linear combination using vector notation:

$$y = \mathbf{a}^{\mathsf{T}} \cdot \mathbf{x}$$

where **a** and **x** are vectors $(a_1, \ldots, a_N)$ and $(x_1, \ldots, x_N)$, and then

$$\sigma_y^2 = \mathbf{a}^{\mathsf{T}} \; \mathbb{C} \; \mathbf{a}$$

Measurements and their uncertainties

Functions of Random Variables

So far, we have mainly discussed linear combinations of random variables...

Let us consider general case of functional dependence $Y = Y(X)$, where X is a random variable with distribution $f(x)$.

Probability density function for dependent variable Y is given by

$$g(y) \quad = \quad f(x) \cdot \left| \frac{dx}{dy} \right|_{y=f(x)}$$

assuming that y(x) is a single-valued function (one-to-one).

If the function considered is not single-valued (i.e. $y(x)$ can not be inverted), we need to divide X domain into subdomains, where this condition is met...

Functions of Random Variables

Change of Variables can be also considered in multi-dimensional case

$$\mathbf{y} \;=\; \mathbf{y}(\mathbf{x})$$

where components of vector **y** are given by functions $y_i(\mathbf{x})$, $i = 1, \ldots, N$

Functions of Random Variables

Change of Variables can be also considered in multi-dimensional case

$$\mathbf{y} = \mathbf{y}(\mathbf{x})$$

where components of vector **y** are given by functions $y_i(\mathbf{x})$, $i = 1, \ldots, N$

Probability density function for dependent variables **y** is given by

$$g(\mathbf{y}) = f(\mathbf{x}(\mathbf{y}))\,|J|$$

assuming that function $\mathbf{y}(\mathbf{x})$ is one-to-one and can be inverted, with

$$J = \left(\frac{\partial x_i}{\partial y_j}\right)$$

being the Jacobian of the variable transformation (square matrix)

Mean of Functions of Random Variables

It is important to notice that

$$\mu_y = \mathbb{E}(y(x)) \neq y(\mu_x)$$

this relation holds only for linear dependencies but not in general case

It is not sufficient to know μ_x (reconstructed value of $\bar{x}$)

$\Rightarrow$ we need to know the probability density function for the variable X or have access to individual measurements of X

Mean of Functions of Random Variables

It is important to notice that

$$\mu_y \; = \mathbb{E}(y(x)) \;\; \neq \;\; y(\mu_x)$$

this relation holds only for linear dependencies but not in general case

It is not sufficient to know μ_x (reconstructed value of $\bar{x}$)
⇒ we need to know the probability density function for the variable X
or have access to individual measurements of X

Variance of Functions of Random Variables

In general, one has to calculate it from first principles, integrating over independent variable x or summing over individual measurements.

However, if variance of X is small, approximate formula can be used...

Variance of Functions of Random Variables

We can use Taylor series expansion about the means of variables $\mathbf{x}$

$$y(\mathbf{x}) \approx \mu_y + \sum_i (x_i - \mu_i) \left.\frac{\partial y}{\partial x_i}\right|_{x_i=\mu_i} + \ldots$$

where we also approximate $y(\hat{\mu}_\mathbf{x}) \approx \mu_y$

Variance of Functions of Random Variables

We can use Taylor series expansion about the means of variables **x**

$$y(\mathbf{x}) \approx \mu_y + \sum_i (x_i - \mu_i) \left.\frac{\partial y}{\partial x_i}\right|_{x_i=\mu_i} + \ldots$$

and, assuming the higher order terms can be neglected, we get

$$\mathbb{E}[(y-\mu_y)^2] = \mathbb{E}\left[\left(\sum_i (x_i - \mu_i) \left.\frac{\partial y}{\partial x_i}\right|_{x_i=\mu_i}\right)\left(\sum_j (x_j - \mu_j) \left.\frac{\partial y}{\partial x_j}\right|_{x_j=\mu_j}\right)\right]$$

Variance of Functions of Random Variables

We can use Taylor series expansion about the means of variables **x**

$$y(\mathbf{x}) \approx \mu_y + \sum_i (x_i - \mu_i) \left.\frac{\partial y}{\partial x_i}\right|_{x_i=\mu_i} + \ldots$$

and, assuming the higher order terms can be neglected, we get

$$\begin{aligned} \mathbb{E}[(y-\mu_y)^2] &= \mathbb{E}\left[\left(\sum_i (x_i-\mu_i)\left.\frac{\partial y}{\partial x_i}\right|_{x_i=\mu_i}\right)\left(\sum_j (x_j-\mu_j)\left.\frac{\partial y}{\partial x_j}\right|_{x_j=\mu_j}\right)\right] \\ &= \sum_{i,j} \left.\frac{\partial y}{\partial x_i}\right|_{x_i=\mu_i} \left.\frac{\partial y}{\partial x_j}\right|_{x_j=\mu_j} \mathbb{E}\left[(x_i-\mu_i)(x_j-\mu_j)\right] \\ &= \sum_{i,j} \frac{\partial y}{\partial x_i}\,\frac{\partial y}{\partial x_j}\, c_{ij} \end{aligned}$$

General form

Covariance matrix for $\mathbf{y} = \mathbf{y}(\mathbf{x})$ can be approximate as:

$$Cov(Y_k, Y_l) \;=\; \sum_{i,j} \frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \, Cov(X_i, X_j)$$

General form

Covariance matrix for $\mathbf{y} = \mathbf{y}(\mathbf{x})$ can be approximate as:

$$Cov(Y_k, Y_l) \;=\; \sum_{i,j} \frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} Cov(X_i, X_j)$$

In matrix notation:

$$\mathbb{C}_{\mathbf{Y}} \;=\; A \, \mathbb{C}_{\mathbf{X}} \; A^{\mathsf{T}}$$

where A is a matrix of partial derivatives:

$$A_{i,j} \;=\; \left. \frac{\partial y_i}{\partial x_j} \right|_{\hat{\mu}_{\mathbf{x}}}$$

Example

For a product of independent random variable powers:

$$y = \prod_i x_i^{\alpha_i}$$

we have

$$\frac{\partial y}{\partial x_i} = \frac{\alpha_i}{x_i} \, y$$

Example

For a product of independent random variable powers:

$$y = \prod_i x_i^{\alpha_i}$$

we have

$$\frac{\partial y}{\partial x_i} = \frac{\alpha_i}{x_i}\, y$$

⇒ variance of Y can be calculated as

$$\sigma_y^2 = \sum_{i,j} \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j}\, Cov(X_i, X_j)$$

Example

For a product of independent random variable powers:

we have

$$y = \prod_i x_i^{\alpha_i}$$

$$\frac{\partial y}{\partial x_i} = \frac{\alpha_i}{x_i}\, y$$

$\Rightarrow$ variance of Y can be calculated as

$$\sigma_y^2 = \sum_{i,j} \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j}\, Cov(X_i, X_j) = \sum_i \left(\frac{\alpha_i}{x_i}\, y \right)^2 \sigma_i^2$$

as we assume variables are independent

Example

For a product of independent random variable powers:

$$y = \prod_i x_i^{\alpha_i}$$

we have

$$\frac{\partial y}{\partial x_i} = \frac{\alpha_i}{x_i}\, y$$

⇒ variance of Y can be calculated as

$$\sigma_y^2 = \sum_{i,j} \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j}\, Cov(X_i, X_j) = \sum_i \left(\frac{\alpha_i}{x_i}\, y \right)^2 \sigma_i^2$$

Relative variance of the product is a quadratic sum of relative variances:

$$\left(\frac{\sigma_y}{\mu_y} \right)^2 = \sum_i \left(\alpha_i \, \frac{\sigma_i}{\mu_{x_i}} \right)^2$$

Example (2)

Consider cuboid with dimensions a, b and c.
Assuming lengths of the edges are measured with precision σ, calculate the uncertainty on cuboid volume V and the total length L of the edges, and their correlation:

$$V = a \cdot b \cdot c \quad \text{and} \quad L = 4\,(a + b + c)$$

Example (2)

Consider cuboid with dimensions a, b and c.
Assuming lengths of the edges are measured with precision σ, calculate the uncertainty on cuboid volume V and the total length L of the edges, and their correlation:

$$V = a \cdot b \cdot c \quad \text{and} \quad L = 4\,(a + b + c)$$

Solution

Assume measurements of a, b and c (variables $\mathbf{X}$) are independent:

$$\mathbb{C}_{\mathbf{X}} = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$$

Example (2)

Consider cuboid with dimensions a, b and c.
Assuming lengths of the edges are measured with precision σ, calculate the uncertainty on cuboid volume V and the total length L of the edges, and their correlation:

$$V = a \cdot b \cdot c \quad \text{and} \quad L = 4\,(a + b + c)$$

Solution

Assume measurements of a, b and c (variables $\mathbf{X}$) are independent:

$$\mathbb{C}_{\mathbf{X}} = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$$

Matrix of partial derivatives for transformation to V and L (variables $\mathbf{Y}$)

$$A = \left(\frac{\partial y_i}{\partial x_j}\right) = \begin{pmatrix} bc & ac & ab \\ 4 & 4 & 4 \end{pmatrix}$$

Example (2)

Using the general formula:

$$\mathbb{C}_{\mathbf{Y}} \;=\; A\,\mathbb{C}_{\mathbf{X}}\;A^{\mathsf{T}} \;=\; \begin{pmatrix} b^2c^2 + a^2c^2 + a^2b^2 & 4(bc+ac+ab) \\ 4(bc+ac+ab) & 48 \end{pmatrix} \cdot \sigma^2$$

Example (2)

Using the general formula:

$$\mathbb{C}_{\mathbf{Y}} = A\,\mathbb{C}_{\mathbf{X}}\,A^{\mathsf{T}} = \begin{pmatrix} b^2c^2 + a^2c^2 + a^2b^2 & 4(bc+ac+ab) \\ 4(bc+ac+ab) & 48 \end{pmatrix} \cdot \sigma^2$$

resulting in:

$$\sigma_V = \sigma \cdot \sqrt{b^2c^2 + a^2c^2 + a^2b^2}$$

$$\sigma_L = 4\sqrt{3}\,\sigma$$

$$\rho(V, L) = \frac{bc + ac + ab}{\sqrt{3(b^2c^2 + a^2c^2 + a^2b^2)}} > \frac{1}{\sqrt{3}}$$

Correlation is 1 for the cube $(a = b = c)$!

Measurements and their uncertainties

Direct Method

Let us assume we have a large number of uniformly distributed random numbers corresponding to the uniform probability distribution:

$$u(r) = \begin{cases} 0 & \text{for } r < 0 \\ 1 & \text{for } 0 \leq r < 1 \\ 0 & \text{for } r \geq 1 \end{cases}$$

This type of uniform random number generators is commonly accessible in most systems and numerical libraries.

Can we use uniform random number generator to generate random numbers from arbitrary probability distribution $f(x)$?

Direct Method

Very simple procedure can be used.
Let us consider cumulative distribution function for X, $F(x)$:

$$P(X \leq x) = F(x)$$

Direct Method

Very simple procedure can be used.
Let us consider cumulative distribution function for X, $F(x)$:

$$P(X \leq x) = F(x)$$

By definition, probability distribution for $U = F(X)$ is uniform:

$$F(u) = P(U \leq F(x)) = F(x) = u$$

Direct Method

Very simple procedure can be used.
Let us consider cumulative distribution function for X, $F(x)$:

$$P(X \leq x) = F(x)$$

By definition, probability distribution for $U = F(X)$ is uniform:

$$F(u) = P(U \leq F(x)) = F(x) = u$$

$$f(u) = \frac{dF(u)}{du} = 1$$

Direct Method

Very simple procedure can be used.
Let us consider cumulative distribution function for X, $F(x)$:

$$P(X \leq x) \;=\; F(x)$$

By definition, probability distribution for $U = F(X)$ is uniform:

$$F(u) \;=\; P(U \leq F(x)) \;=\; F(x) \;=\; u$$

$$f(u) \;=\; \frac{dF(u)}{du} \;=\; 1$$

If cumulative distribution function $F(x)$ can be inverted, we can generate random numbers from $f(x)$ by using relation:

$$x \;=\; F^{-1}(r)$$

where r is uniformly distributed random number (from $u(r)$)

Direct Method example

04_generation.ipynb Open in Colab

Nice example is Cauchy distribution:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \Rightarrow \quad F(x) = \frac{1}{\pi} \cdot \arctan(x) + \frac{1}{2}$$

which can be easily inverted, resulting in:

$$x = \tan\left(\pi\left(r - \frac{1}{2}\right)\right)$$

where r is uniformly distributed random number

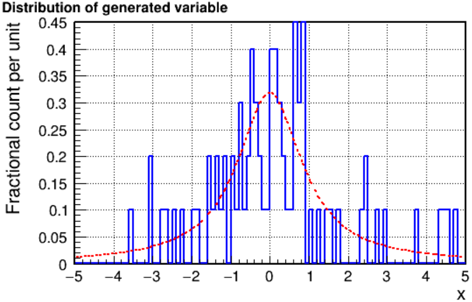
Direct Method example

04_generation.ipynb Open in Colab

Nice example is Cauchy distribution:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \Rightarrow \quad F(x) = \frac{1}{\pi} \cdot \arctan(x) + \frac{1}{2}$$

Example generation, $N = 100$



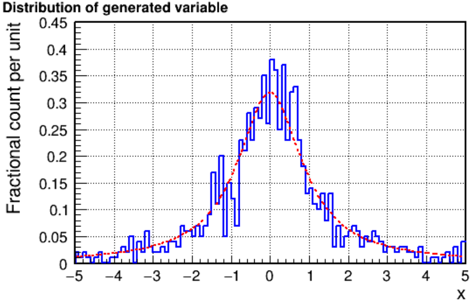
Direct Method example

04_generation.ipynb   Open in Colab

Nice example is Cauchy distribution:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \Rightarrow \quad F(x) = \frac{1}{\pi} \cdot \arctan(x) + \frac{1}{2}$$

Example generation, $N = 1000$



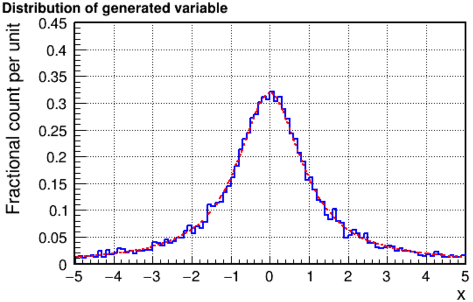
Direct Method example

04_generation.ipynb Open in Colab

Nice example is Cauchy distribution:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \Rightarrow \quad F(x) = \frac{1}{\pi} \cdot \arctan(x) + \frac{1}{2}$$

Example generation, $N = 10000$



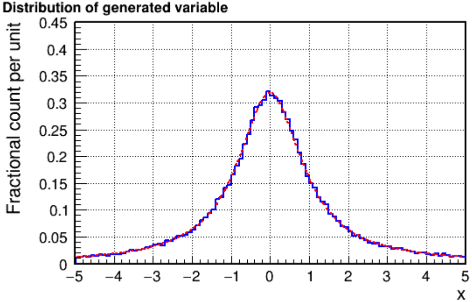
Direct Method example

04_generation.ipynb Open in Colab

Nice example is Cauchy distribution:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \Rightarrow \quad F(x) = \frac{1}{\pi} \cdot \arctan(x) + \frac{1}{2}$$

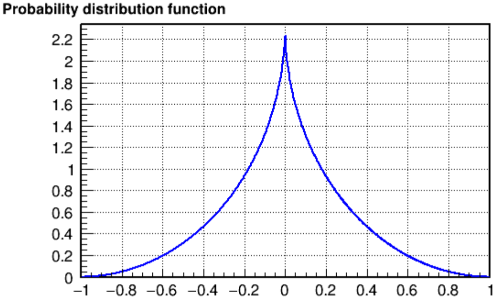
Example generation, $N = 100000$



von Neumann method

The direct method is very elegant and efficient, but requires that the invert of cumulative distribution is known. This is not the case in many problems - solution is not general.
We need an alternative...

Example problem:



von Neumann method

The direct method is very elegant and efficient, but requires that the invert of cumulative distribution is known. This is not the case in many problems - solution is not general. We need an alternative...

Example problem:

$$f(x) = \begin{cases} 0 & \text{for } x < -1 \\ N\left[1 - \sqrt{1-(1-|x|)^2}\right] & \text{for } -1 \leq x \leq 1 \\ 0 & \text{for } x > 1 \end{cases}$$

where normalization factor

$$N = \frac{2}{4-\pi}$$

von Neumann method

We can notice that

$$f_{\max} = \max_x f(x) = N$$

Assume $f(x)$ is non-zero only for $a \le x \le b$.

We can then apply the following procedure:

- generate value x uniformly distributed in $[a, b]$
- generate test variable r from uniform distribution $([0, 1[)$
- accept generated value of x, if $r \cdot f_{\max} < f(x)$
  otherwise repeat from the beginning

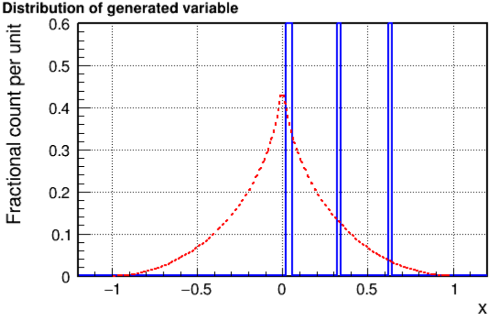
This procedure is called von Neumann Acceptance–Rejection Technique

04_generation2.ipynb Open in Colab

von Neumann method

Acceptance–Rejection Technique applied to the example problem:

Test generation, $N = 10$

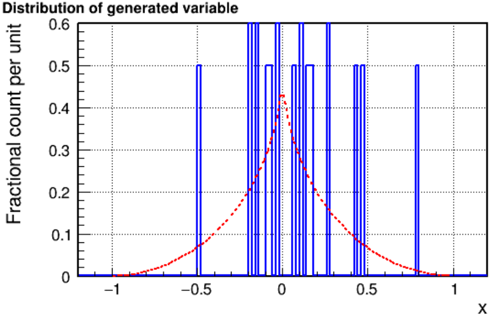


04_generation2.ipynb Open in Colab

von Neumann method

Acceptance–Rejection Technique applied to the example problem:

Test generation, $N = 100$

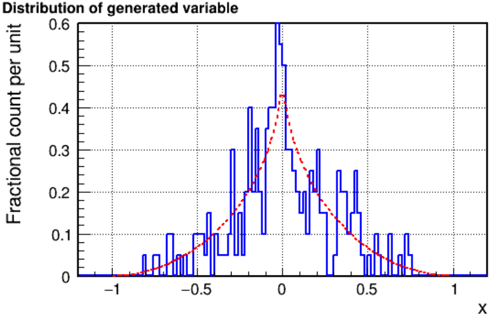


04_generation2.ipynb Open in Colab

von Neumann method

Acceptance–Rejection Technique applied to the example problem:

Test generation, $N = 1000$

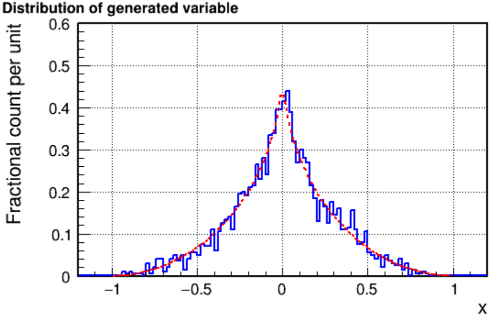


04_generation2.ipynb Open in Colab

von Neumann method

Acceptance–Rejection Technique applied to the example problem:

Test generation, N = 10000

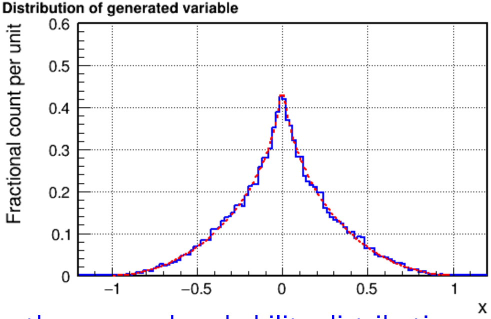


04_generation2.ipynb Open in Colab

von Neumann method

Acceptance–Rejection Technique applied to the example problem:

Test generation, $N = 100000$



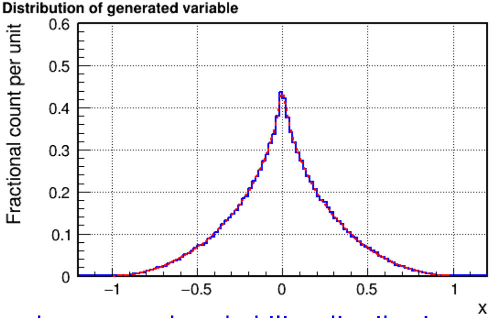
We can perfectly reproduce the assumed probability distribution

04_generation2.ipynb Open in Colab

von Neumann method

Acceptance–Rejection Technique applied to the example problem:

Test generation, N = 1000000



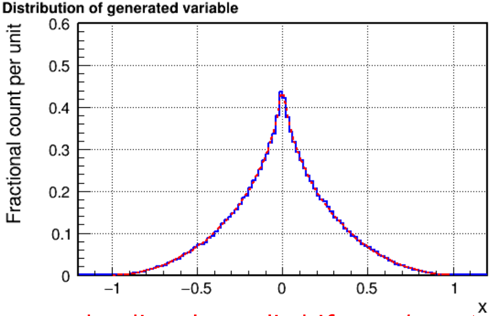
We can perfectly reproduce the assumed probability distribution

04_generation2.ipynb Open in Colab

von Neumann method

Acceptance–Rejection Technique applied to the example problem:

Test generation, N = 1000000



However, this procedure can not be directly applied if a or $b \to \pm\infty$

General method

For PDF distributions with infinite domain we can combine direct and von Neumann methods.

Example: generate photon scattering angles for diffractive scattering

$$f(x) \quad = \quad N \cdot \frac{\sin^2 x}{x^2}$$

General method

For PDF distributions with infinite domain we can combine direct and von Neumann methods.

Example: generate photon scattering angles for diffractive scattering

$$f(x) = N \cdot \frac{\sin^2 x}{x^2}$$

One can check that this function is always smaller than:

$$g(x) = \frac{3}{2} N \cdot \frac{1}{1+x^2}$$

and both functions have similar asymptotic behavior.

Also, cumulative distribution function for $g(x)$ can be inverted,
so we know how to generate random numbers with this distribution...

Variable simulation

General method

The following procedure can now be used:

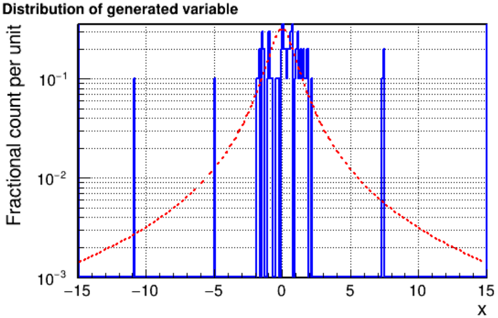
- generate value x distributed according to $g(x)$
- generate test variable r from uniform distribution $([0,1[)$
- accept generated value of x, if $r < f(x)/g(x)$
  otherwise repeat from the beginning

General method

04_generation3.ipynb Open in Colab

Described procedure applied to the diffractive problem:

Test generation, $N = 100$



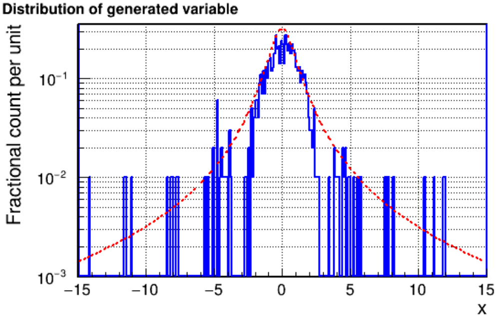
Variable simulation

General method

04_generation3.ipynb Open in Colab

Described procedure applied to the diffractive problem:

Test generation, $N = 1000$

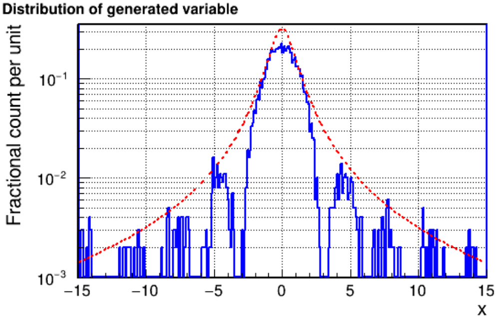


General method

04_generation3.ipynb Open in Colab

Described procedure applied to the diffractive problem:

Test generation, $N = 10000$

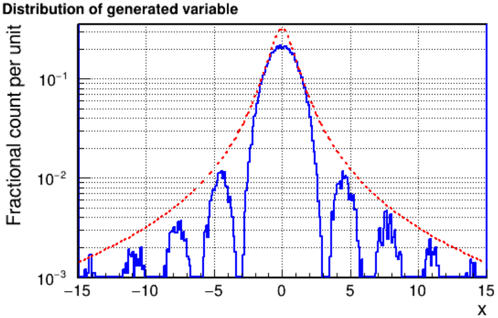


General method

04_generation3.ipynb Open in Colab

Described procedure applied to the diffractive problem:

Test generation, $N = 100000$

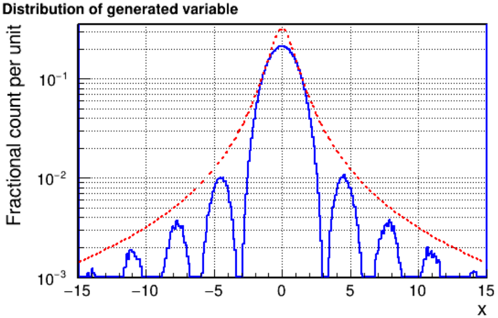


General method

04_generation3.ipynb Open in Colab

Described procedure applied to the diffractive problem:

Test generation, $N = 1000000$



We reproduce the expected probability distribution

Gaussian Variable

Cumulative distribution function can not be inverted analytically.
However, we can use a simple trick, consider 2-D distribution:

$$g(x,y) \;=\; \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2}(x^2+y^2)\right)$$

for simplicity we assume $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$
and change variables to polar coordinates:

$$g(R,\theta) \;=\; \frac{1}{2\pi}\, R\, \exp\left(-\frac{1}{2}R^2\right) \;=\; h(R)\cdot t(\theta)$$

Gaussian Variable

Cumulative distribution function can not be inverted analytically.
However, we can use a simple trick, consider 2-D distribution:

$$g(x, y) \;=\; \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

for simplicity we assume $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$

and change variables to polar coordinates:

$$g(R, \theta) \;=\; \frac{1}{2\pi}\, R\, \exp\left(-\frac{1}{2}R^2\right) \;=\; h(R) \cdot t(\theta)$$

Cumulative distribution functions can be now considered for R and θ:

$$H(R) = 1 - \exp\left(-\frac{1}{2}R^2\right) \qquad T(\theta) = \frac{1}{2\pi}\,\theta$$

where $R > 0$ and $0 \leq \theta \leq 2\pi$

Gaussian Variable

Cumulative distribution functions for r and θ can be inverted resulting in the following formula for generation of a pair of random variables:

$$R = \sqrt{-2\ln(1-r_1)}$$

$$\theta = 2\pi\, r_2$$

where r_1 and r_2 are two independent uniformly distributed random numbers

Gaussian Variable

Cumulative distribution functions for r and θ can be inverted resulting in the following formula for generation of a pair of random variables:

$$R = \sqrt{-2\ln(1-r_1)}$$

$$\theta = 2\pi\ r_2$$

where r_1 and r_2 are two independent uniformly distributed random numbers

We can now transform back to the (x, y) plane:

$$x = R\ \cos\theta = \sqrt{-2\ln(1-r_1)} \cdot \cos(2\pi\ r_2)$$

$$y = R\ \sin\theta = \sqrt{-2\ln(1-r_1)} \cdot \sin(2\pi\ r_2)$$

$\Rightarrow$ from pair (r_1, r_2) of uniform distributed numbers we get a pair (x, y) of independent (!) random numbers from normal distribution

Correlated 2D Gaussian

What, if we need to generate pair (x, y) of correlated random numbers?

For 2-D Gaussian distribution it can be done by smart "rotation" of variables. Rotation angle:

$$\phi = \frac{1}{2} \arcsin(\rho_{xy})$$

Correlated 2D Gaussian

What, if we need to generate pair (x, y) of correlated random numbers?

For 2-D Gaussian distribution it can be done by smart "rotation" of variables. Rotation angle:

$$\phi = \frac{1}{2} \arcsin(\rho_{xy})$$

Then we can generate correlated pair of variables using the formula:

$$x = g_1 \cos\phi + g_2 \sin\phi$$
$$y = g_1 \sin\phi + g_2 \cos\phi$$

where (g_1, g_2) is a pair of independent random numbers from Gaussian distribution (assuming $\sigma = 1$ in all cases)

Measurements and their uncertainties

Homework

Consider production of rectangular tables with length $\ell = 1.4\,\mathrm{m}$ and width $w = 1.0\,\mathrm{m}$.

Tolerance of production process (standard deviation) is $\sigma_\ell = 2\,\mathrm{mm}$ for the table length and $\sigma_w = 1\,\mathrm{mm}$ for the table width:

- what is the expected variation (standard deviation) of the surface of the table, assuming table length and width are independent variables?
- what is the expected variation of the total length of the table sides?
- what is the expected correlation between the surface and the total edge length?
- perform numerical experiment generating table lengths and widths from the Gaussian distribution to confirm the result.

Solutions should be uploaded until November 6.