Introduction to Monte Carlo methods

Marcin Chrząszcz
mchrzasz@cern.ch

University of Zurich[UZH]

Monte Carlo methods,
25 February, 2016

Literature

1. J. M. Hammersley, D. C. Hamdscomb, “Monte Carlo Methods”, London: Methuen & Co. Ltd., New York: J. Wiley & Sons Inc., 1964
2. I. M. Sobol, “The Monte Carlo Method”, Mir Publishers, Moscow, 1975.
3. M. H. Kalos, P. A. Whitlock, „Monte Carlo Methods”, J. Wiley & Sons Inc., New York, 1986
4. G. S. Fishman, „Monte Carlo: Concepts, Algorithms and Applications”, Springer, 1996.
5. R. Y. Rubinstein, D. P. Kroese, „Simulation and the Monte Carlo Method”, Second Edition, J. Wiley & Sons Inc., 2008.
6. R. Korn, E. Korn, G. Kroisandt, „Monte Carlo methods and models in finance and insurance”, CRC Press, Taylor & Francis Group, 2010.
7. S. Jadach, „Practical Guide to Monte Carlo”, arXiv:physics/9906056, http://cern.ch/jadach/MCguide/.

Course plan

⇛ On each lecture there will be set of problem (standard) to code and solve.

⇛ For each of them you will get points:

- 6 points if solved within 2 weeks.
- 5 points if solved within 3 weeks.
- 4 points if solved within 4 weeks.
- 3 points if solved within 5 weeks.
- etc.

⇛There will be points for additional problems.

⇛ The final mark:

$$\text{mark} = \frac{\sum \text{received points standard} + \sum \text{additional points}}{\sum \text{points standard}}$$

⇛ The final exam will be average of exam and class marks.

Mark	Range
6	$84 - 100$ %
5	$67 - 83$ %
4	$50 - 67$ %
3	$33 - 50$ %
2	$16 - 33$ %
1	$0 - 16$ %

Course plan

1. Definitions, history, applications.
2. Mathematical formulation.
3. Integration with MC methods, variance reduction.
4. Adaptive MC integration techniques.
5. Random number generators with $\mathcal{U}(0,1)$.
6. Generating random numbers for given PDF.
7. Application of MCMC in solving linear algebra systems, differential equations, integral equations, matrix algebra, function interpolation etc.
8. MC method in optimisation problems.
9. MC modelling.

Definitions

$\Rightarrow$ Basic definition:

Monte Carlo method is any technique that uses *random numbers* to solve a given mathematical problem.

$\rightarrowtail$ Random number: For the purpose of this course we need to assume that we know what it is, although the formal definition is highly non-trivial.

$\Rightarrow$ My favourite definition (Halton 1970): more complicated, but more accurate.

"Representing the solution of a problem as a parameter of a hypothetical population, and using a random sequence of numbers to construct a sample of the population, from which statistical estimates of the parameter can be obtained."

To put this definition in mathematical language:
Let F be a solution of a given mathematical problem. The estimate of the result $\hat{F}$:

$$\hat{F} = f(\{r_1, r_2, r_3, ..., r_n\}; ...),$$

where $\{r_1, r_2, r_3, ..., r_n\}$ are random numbers.

The problem we are solving doesn't need to be stochastic!

$\twoheadrightarrow$ One could wonder why are we trying to add all the stochastic properties to a deterministic problem. Those are the properties that allow to use all well known statistic theorems.

History of MC methods

- G. Compte de Buffon (1777) - First documented usage of random numbers for integral computation (Buffon thrown niddle on the table with parrarel line; we will do a modern version of this exercise).
- Marquis de Laplace (1886) - Used the Buffon niddle to determine the value of π number.
- Lord Kelvin (1901) - Thanks to drawing randomly numbered cards he managed he managed to calculate some integrals in kinematic gas theorem.
- W. S. Gosse (better knows as Student) (1908) - Used similar way as Lord Kelvin to get random numbers to prove t-Student distribution.
- Enrico Fermi (1930) - First mechanical device (`FERMIAC`) for random number generations. Solved neutron transport equations in the nuclear plants.
- S. Ulam, R. Feynman, J. von Neumann et. al. - First massive usage of random numbers. Most applications were in Manhattan project to calculate neutron scattering and absorption.
  In Los Alamos the name Monte Carlo was created as kryptonim of this kind of calculations.

Applications of MC methods

$\Rightarrow$ Application of a MC method doesn't depend on the stochastic nature of the problem, but only on ability to represent a problem by a given hypothetical population so we can apply random numbers in that problem.

$\Rightarrow$ For example the calculations made in Los Alamos are so-called direct simulation. They really simulated neutron transportation in the material. The solutions are again to a non deterministic problem. Clearly 2 atomic bombs have equal energies...

$\Rightarrow$ The nature of the problem can be probabilistic, in which case we are performing so-called direct MC simulation. However when we are simulating not directly the problem but rather an abstract population we are using indirect MC method.

$\Rightarrow$ The application of a given method will depend only of mathematical structure of the problem.

Euler number determination

$\Rightarrow$ As mentioned before MC methods can be used to solve problems that **do not** have stochastic nature! All the integrals calculated in Los Alamos during the Manhattan project are nowadays solvable without any MC methods.
$\rightarrowtail$ Let's give a trivial example of solving a non stochastic problem: calculating Euler number e. We know that $e = 2.7182818....$ $\Rightarrow$ To calculate the $\hat{e}$ we will use the following algorithm:

- We generate a random number in range $(0, 1)$ (in stat. $\mathcal{U}(0, 1)$) until the number we generate is smaller then the previous one, aka we get the following sequence:

$$x_1 < x_2 < ... < x_{n-1} > x_n$$

- We store the number n. We repeat this experiment N times and calculate the arithmetic average of n. The obtained value is an statistical estimator of e:

$$\hat{e} = \frac{1}{N} \sum_{i=1}^{N} n_i \xrightarrow{N \to \infty} e.$$

$\Rightarrow$ Numerical example:

N	$\hat{e}$	$\hat{e} - e$
100	2.760000	0.041718
10000	2.725000	0.006718
1000000	2.718891	0.000609
100000000	2.718328	0.000046

Is this $\sim \sqrt{N}$?

$\Rightarrow$ (E1.1) Write this program (A1.1) and prove it mathematically.

Let's test the $\sqrt{N}$, Lecture1/Euler_number

$\Rightarrow$ In the last example we measured the Euler number using different number of pseudo-experiments.

$\rightarrowtail$ We compared the obtained value to the true and observed roughly a $\sqrt{N}$ dependence on the difference between the true value and the obtained one.

$\rightarrowtail$ Could we test this? YES! Lets put our experimentalist hat on!

$\rightarrowtail$ From the begging of studies they tooth us to get the error you need to repeat the measurements.

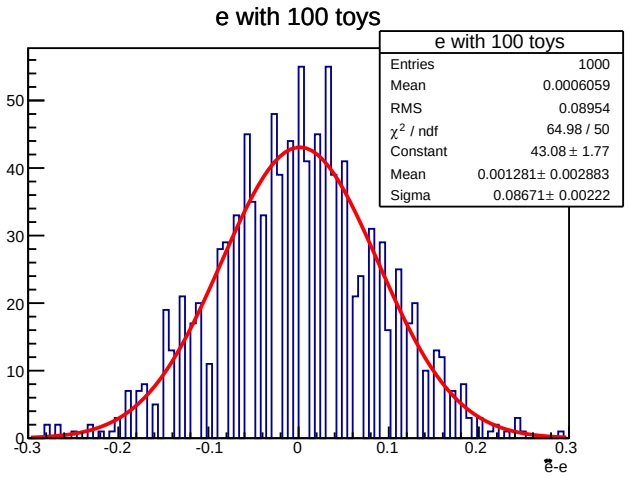
The algorithm:

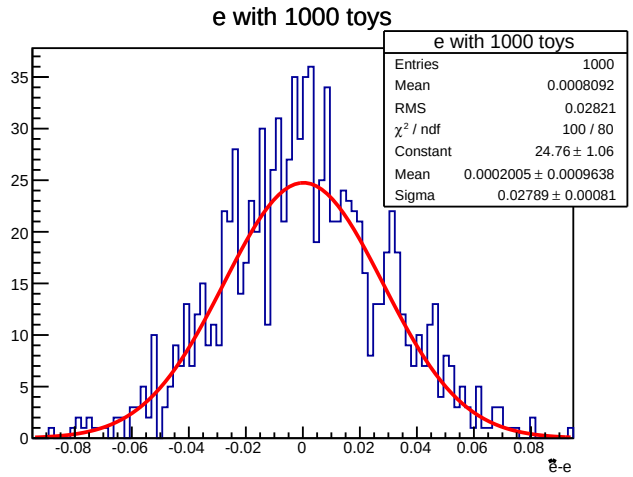
Previous time we measured Euler number using N events, where $N \in (100, 1000, 10000, 100000)$. Now lets repeat this measurement n_N times (of course each time we use new generated numbers). From the distribution of $\hat{e} - e$ we could say something about the uncertainty of our estimator for given N.

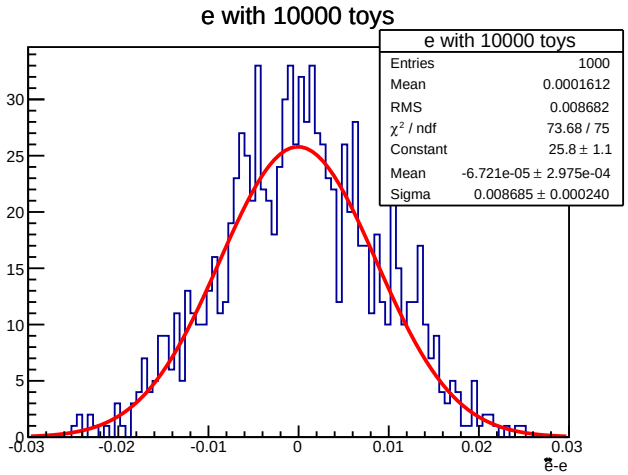
Let's test the $\sqrt{N}$, Lecture1/Euler_number

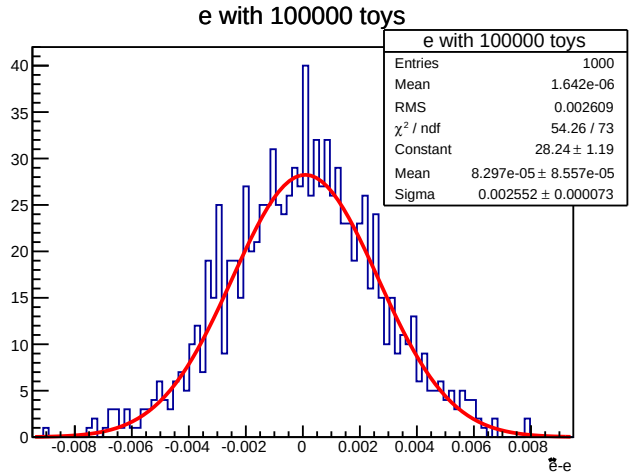
$\rightarrowtail$ Could we test this? YES! Lets put our experimentalist hat on!

$\rightarrowtail$ From the begging of studies they tooth us to get the error you need to repeat the measurements.





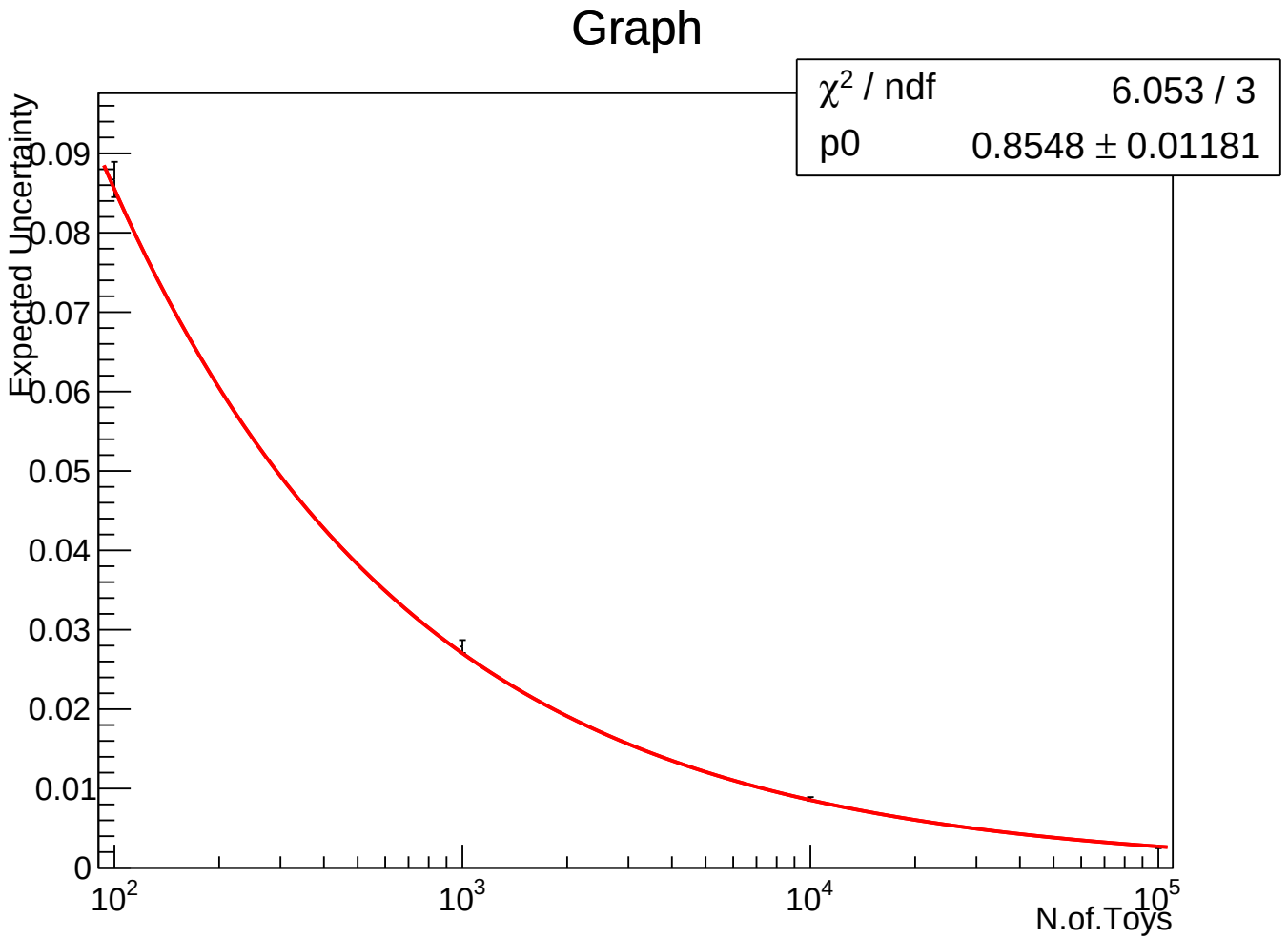




Let's test the $\sqrt{N}$, Lecture1/Euler_number

$\rightarrowtail$ Could we test this? YES! Lets put our experimentalist hat on!

$\rightarrowtail$ From the begging of studies they tooth us to get the error you need to repeat the measurements.



Mathematical foundations of MC methods

$\Rightarrow$ All MC methods are mathematically equal to a integral estimator (if you don't remember this you will immediately fail the exam).
$\Rightarrow$ For simplicity lets assume that r_i are numbers from $\mathcal{U}(0,1)$.
$\Rightarrow$ The MC result:

$$F = F(r_1, r_2, ..., r_n)$$

is an unbias estimator of the integral:

$$I = \int_0^1 ... \int_0^1 F(x_1, ..., x_k) dx_1 ... dx_k$$

, or the expected value of F is I:

$$E(F) = I$$

$\Rightarrow$ This formal identity gives us the theoretical tool for application of the MC methods.

Mathematical foundations of MC methods

A random number is a number that can take more then one value (usually takes the values from continuum) and non of it's value can be predicted before hand.

$\Rightarrow$ Even thought we cannot predict the random number we can predict it's probability. $\Rightarrow$ For the continuous variables we define a *Probability Density Function* (PDF):

$$\rho(u)du = \mathcal{P}[u < u' < u + du],$$

where $\rho(u)$ is the PDF.
$\Rightarrow$ Cumulative Distribution Function (CDF):

$$R(u) = \int_{-\infty}^{u} \rho(x)dx, \qquad \rho(u) = \frac{dR(u)}{du}$$

$\Rightarrow$ The $R(u)$ is monotonically non decreasing function and takes the value in $[0, 1]$.

Mathematical foundations of MC methods

⇛ The expected value of a function $f(x)$ is defined as an average of the function

$$E(f) = \int f(u)dR(u) = \int f(u)\rho(u)du$$

⇛ If $x \in \mathcal{U}(a,b)$ then:

$$dR = \frac{du}{b-a} \qquad E(f) = \frac{1}{b-a}\int_a^b f(u)du$$

⇛ The variation is defined as:

$$V(f) = E([f - E(f)]^2) = \int [f - E(f)]^2 dR$$

⇛ The Standard deviation:

$$\sigma(f) = \sqrt{V(f)}$$

⇛ In practice we give the standard deviation not the varaition as it's the same dimension as measured quantities.

⇛ For 2 random variables we define:

$$E(cx+y) = cE(x) + E(b)$$
$$V(cx+y) = c^2V(x) + V(y) + 2cCov(x,y), \quad (1)$$

where $Cov(x,y) = E([x-E(x)][y-E(y)])$ is called the covariance.

$$cov(x,y) = \begin{cases} = 0 & x,y\ uncorrelated \\ > 0 & x,y\ correlated \\ < 0 & x,y\ anticorrelated \end{cases} \quad (2)$$

⇛ If x,y are independent then $cov(x,y) = 0$ and $V(x+y) = V(x) + V(y)$

⇛ If variables are independent then they are uncorrelated. If they are uncorrelated then can still be dependent†.

⇛ E(1.2) Give example for †. E(1.3) Prove Equations: 1.

Law of large numbers

The law of large numbers (LLN): let's take n numbers from $\mathcal{U}(a, b)$ and for each of them we calculate the $f(u_i)$. The LLN:

$$\frac{1}{n}\sum_{i=1}^{n} f(u_i) \xrightarrow{N\to\infty} \frac{1}{b-a}\int_a^b f(u)du \tag{3}$$

$\Rightarrow$ We say (in statistic terminology) that the left side is asymptotically equivalent to the value of the integration if $n \to \infty$.

$\Rightarrow$ Assumptions:

- f is integrative.
- Smooth in most of the points.
- Limited.

The LLN can be interpreted as the fact the MC estimator of the integration is approaching the true value if the n is increasing.

Convergence

$\Rightarrow$ There convergence problem is a bit different compared to our standard analysis.

$\Rightarrow$*Math Analysis* For a series of a_n converges to a if for each δ we can find an element n that for each element $n+k$, $k>0$ $|a_{n+k}-a|<\delta$.

$\Rightarrow$*Statistics* We say that $a(n)$ is converging to a if for each probability $\mathcal{P}[0<\mathcal{P}<1]$ and each δ we can find $n+k$ that for each $k>0$ the provability of $a(n)$ to be within distance smaller then δ is grater then $\mathcal{P}$.

"For sure" $\longrightarrow$ "Probability"

$\Rightarrow$ Despite how big n is there is guaranty (we are not GM!) that $a(n)$ will be in a given interval around a.

$\Rightarrow$ The convergence is always with a given probability.

Central Limit theorem

A sum of large number random variables has always a Gaussian distributions, independent of what kind of PDF was used to generate the sample. The only requirement is that the PDFs have a finite expected value.

⇛ E(1.3) Using ROOT draw the Gauss distribution:

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

and calculate (for given μ and σ the:

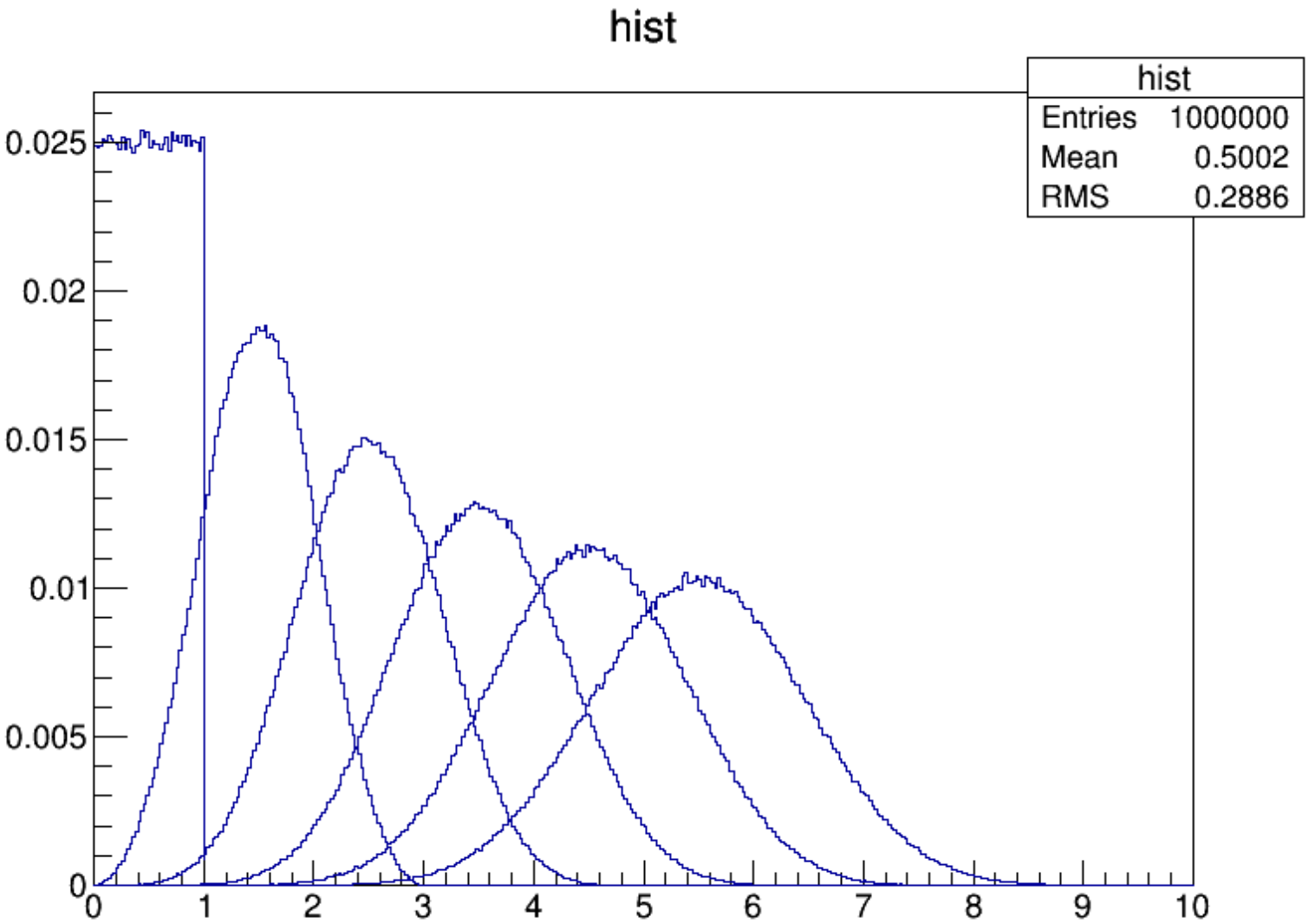
$$\mathcal{P}(-1.64 \leqslant \frac{x-\mu}{\sigma} \leqslant 1.64) = 0.9$$

$$\mathcal{P}(-2.58 \leqslant \frac{x-\mu}{\sigma} \leqslant 2.58) = 0.99$$

$$\mathcal{P}(-3.29 \leqslant \frac{x-\mu}{\sigma} \leqslant 3.29) = 0.999$$

Central Limit Theorem

Large independent random numbers assembly has always Gaussian distribution no matter from what distribution they were generated from as far as they have finite variances and expected values and the assembly is sufficiently large.



Gauss Generator

⇛ Let's take $x_i \in (0,1)$, $i = 1, ..., n$. The $R_n = \sum_{i=1}^{n} x_i$. Then:

$$E(x_i) = 1/2 \Rightarrow E(R_n) = n/2$$
$$V(x_i) = 1/12 \Rightarrow V(R_n) = n/12$$

⇛ E(1.4) Calculate the above.
⇛ From above we get:

$$\frac{R_n - n/12}{\sqrt{n/12}} \xrightarrow{N\to\infty} N(0,1). \tag{4}$$

aka we get a Gaussian distribution. ⇛ For $n = 12 \Rightarrow \ (R_{12} - 6)$ "practical" Gauss generator.
⇛ Attention: This kind of generators are not good in reproducing the tails of the Gauss!
⇛ E(1.5) Construct the Gauss generator based on the math above in ROOT.

Wrap up

⇛ From LLN:

$$\frac{1}{n}\sum_{i=1}^{n} f(u_i) \xrightarrow{N\to\infty} \frac{1}{b-a}\int_a^b f(u)du$$

⇛ Mathematical properties of MC estimator:

- If $V(f) < \infty$ then estimator is a good estimator, aka it converges to the true value.
- The estimator is an unbias estimator, aka the expected value is the true value.
- The estimator has a normal distribution for large n (CLT).
- The standard deviation of the estimator:

$$\sigma = \frac{1}{\sqrt{n}}\sqrt{V(f)}$$

In practice

$$I \equiv \frac{1}{b-a}\int_a^b f(x)dx = E(f).$$

⇛ The estimator:

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_i),\ x_i \in \mathcal{U}(a,b)$$

⇛ The varaince:

$$V(\hat{I}) = \frac{1}{n}V(f) = \frac{1}{n}[E(f^2) - E^2(f)] = \frac{1}{n}[\frac{1}{b-a}\int_a^b f^2 2(x)dx - I^2].$$

⇛To calculate the varaince you need to know the integral I!

⇛ in practice we use the following:

$$\hat{V}(f) = \frac{1}{n-1}\sum_{i=1}^{n}[f(x_i) - \frac{1}{n}\sum_{i=1}^{n} f(x_i)]^2$$

⇛ And the uncertainty estimator: $\sigma = \sqrt{\hat{V}}$