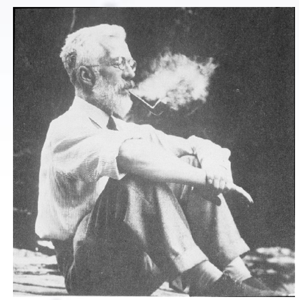
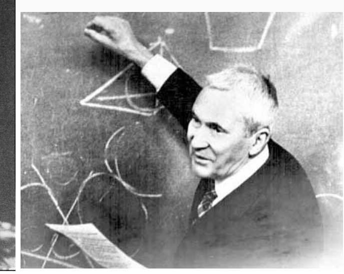
Applied Statistics

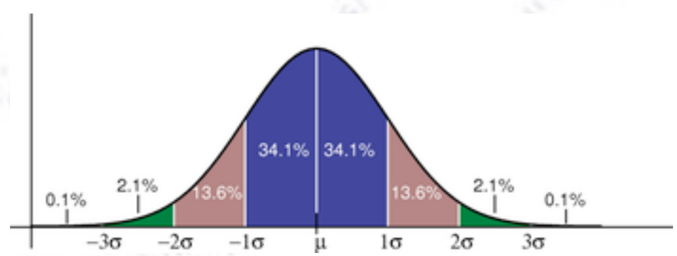
Central Limit Theorem





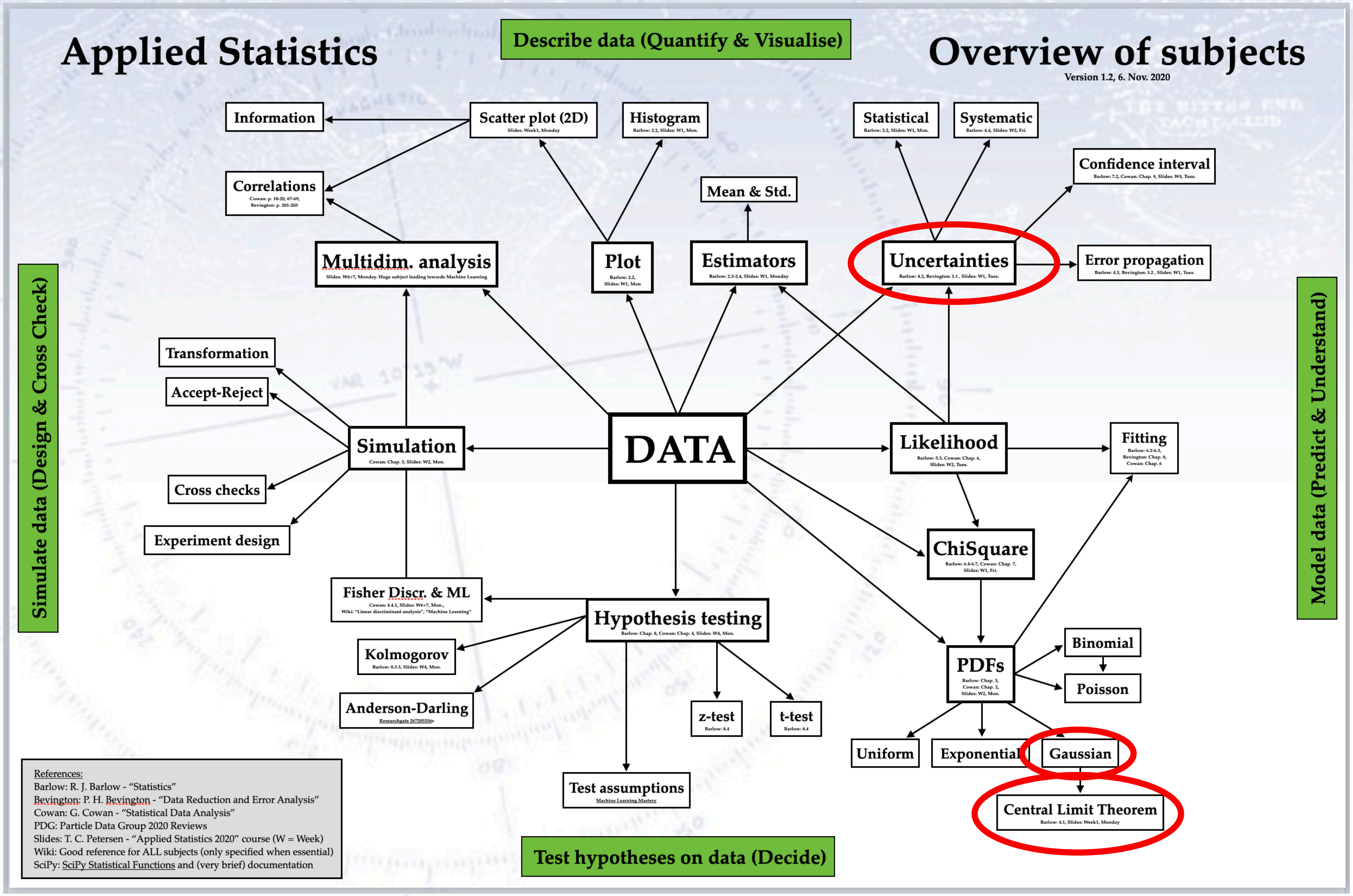



Troels C. Petersen (NBI)



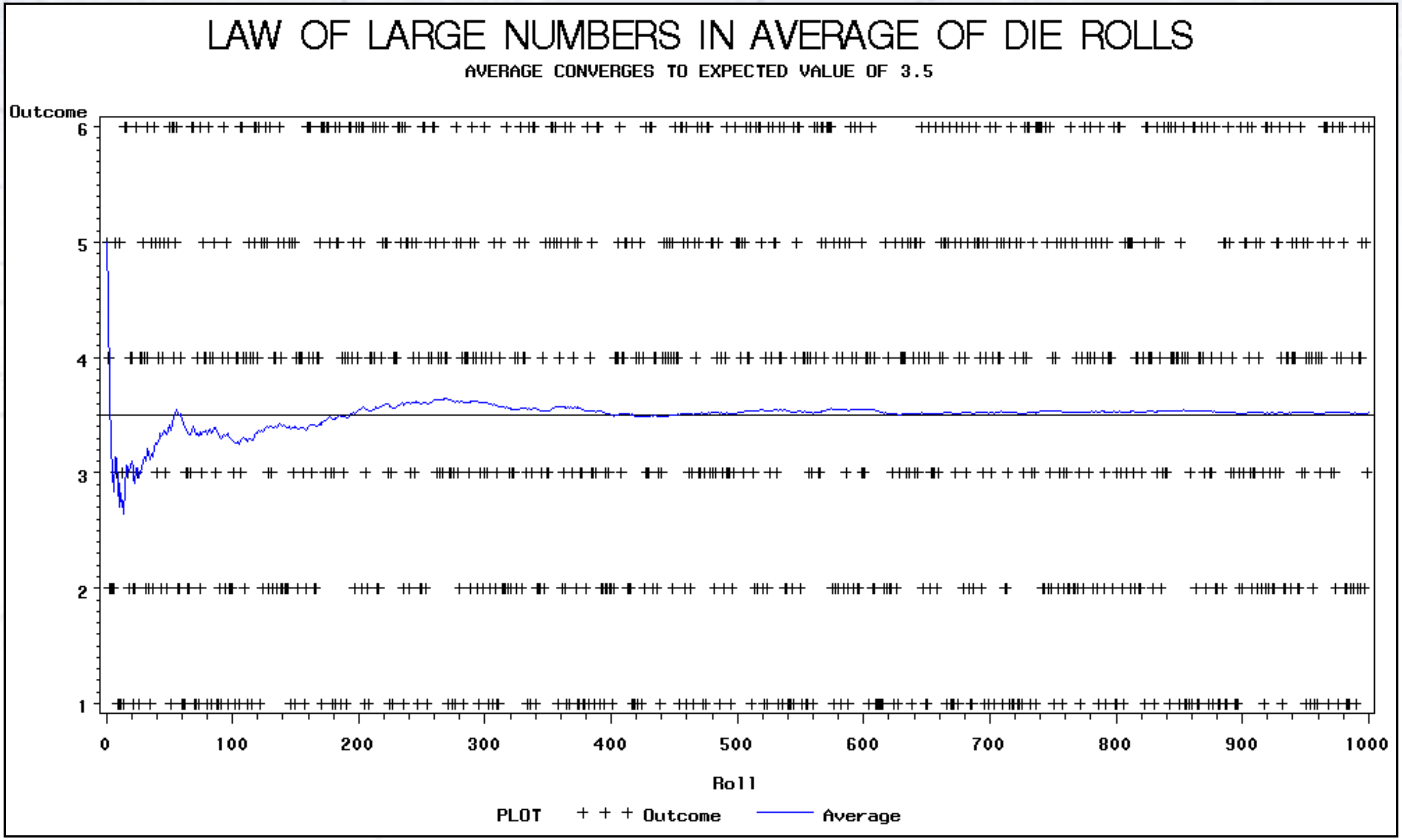
"Statistics is merely a quantisation of common sense"

Central Limit Theorem



Law of large numbers

When rolling a normal die and averaging the outcome, it is no surprise that this converges towards 3.5… with enough rolls, you can get as close as you want!



Adding random numbers

If each of you chose a random number from your own favorit distribution*, and we added all these numbers, repeating this many times…

What would you expect?

* OK - to be nice to me, you agree to have similar RMSs in these distributions!

Adding random numbers

If each of you chose a random number from your own favorit distribution*, and we added all these numbers, repeating this many times…

What would you expect?

Gaussian!!!

…by the central limit theorem!

* OK - to be nice to me, you agree to have similar RMSs in these distributions!

Adding random numbers

If each of you chose a random number from your own favorite distribution*, and we added all these numbers, repeating this many times…

Gaussian!!!

the central limit theorem!

<u>Central Limit Theorem:</u>

The sum of N *independent* continuous random variables x_i with means μ_i and variances σ_i^2 becomes a Gaussian random variable with mean $\mu = \Sigma_i \mu_i$ and variance $\sigma^2 = \Sigma_i \sigma_i^2$ in the limit that N approaches infinity.

Central Limit Theorem

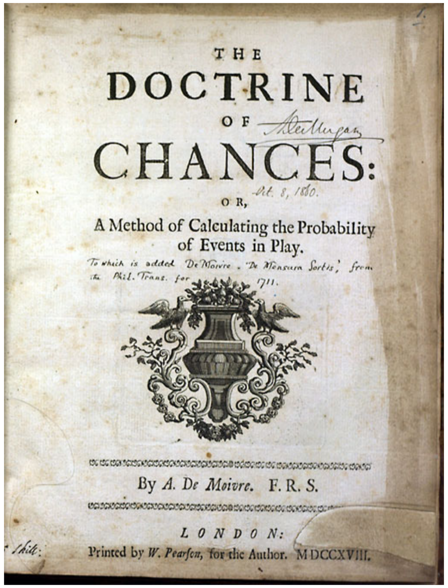
<u>Central Limit Theorem:</u>
The sum of N *independent* continuous random variables x_i with means μ_i and variances σ_i^2 becomes a Gaussian random variable with mean $\mu = \Sigma_i \mu_i$ and variance $\sigma^2 = \Sigma_i \sigma_i^2$ in the limit that N approaches infinity.

The Central Limit Theorem holds under fairly general conditions, which means that the Gaussian distribution takes a central role in statistics...

The Gaussian is "the unit" of distributions!

Since measurements are often affected by many small effects, uncertainties tend to be Gaussian (until otherwise proven!).

Statistical rules often require Gaussian uncertainties, and so **the central limit theorem is your new good friend..**



Central Limit Theorem

<u>Central Limit Theorem:</u>

The sum of N *independent* continuous random variables x_i with means μ_i and variances σ_i^2 becomes a Gaussian random variable with mean $\mu = \Sigma_i \mu_i$ and variance $\sigma^2 = \Sigma_i \sigma_i^2$ in the limit that N approaches infinity.

"The epistemological value of probability theory is based on the fact that chance phenomena, considered collectively and on a grand scale, create non-random regularity."

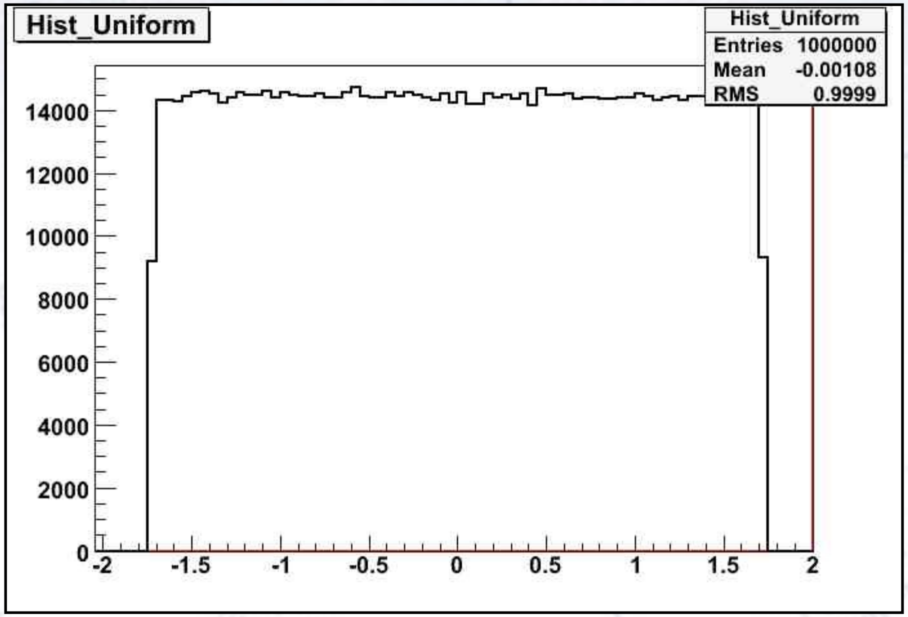
[Andrey Kolmogorov, Soviet mathematician, 1954]

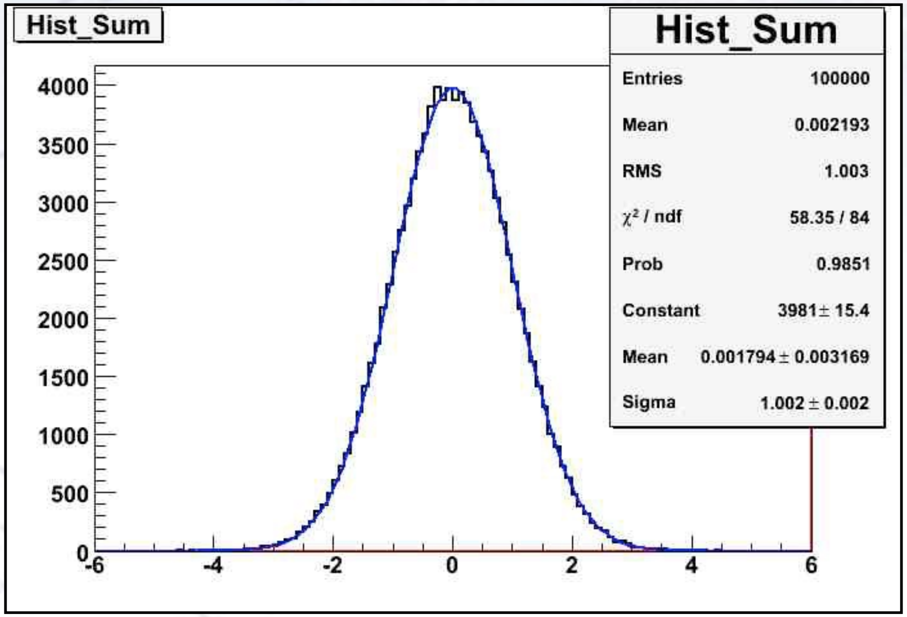
"Nowadays, the central limit theorem is considered to be the unofficial sovereign of probability theory."

[Henk Tijms, Dutch mathematician 2004]

Example of Central Limit Theorem

Take the sum of 100 uniform numbers!
Repeat 100000 times to see what distribution the sum has…



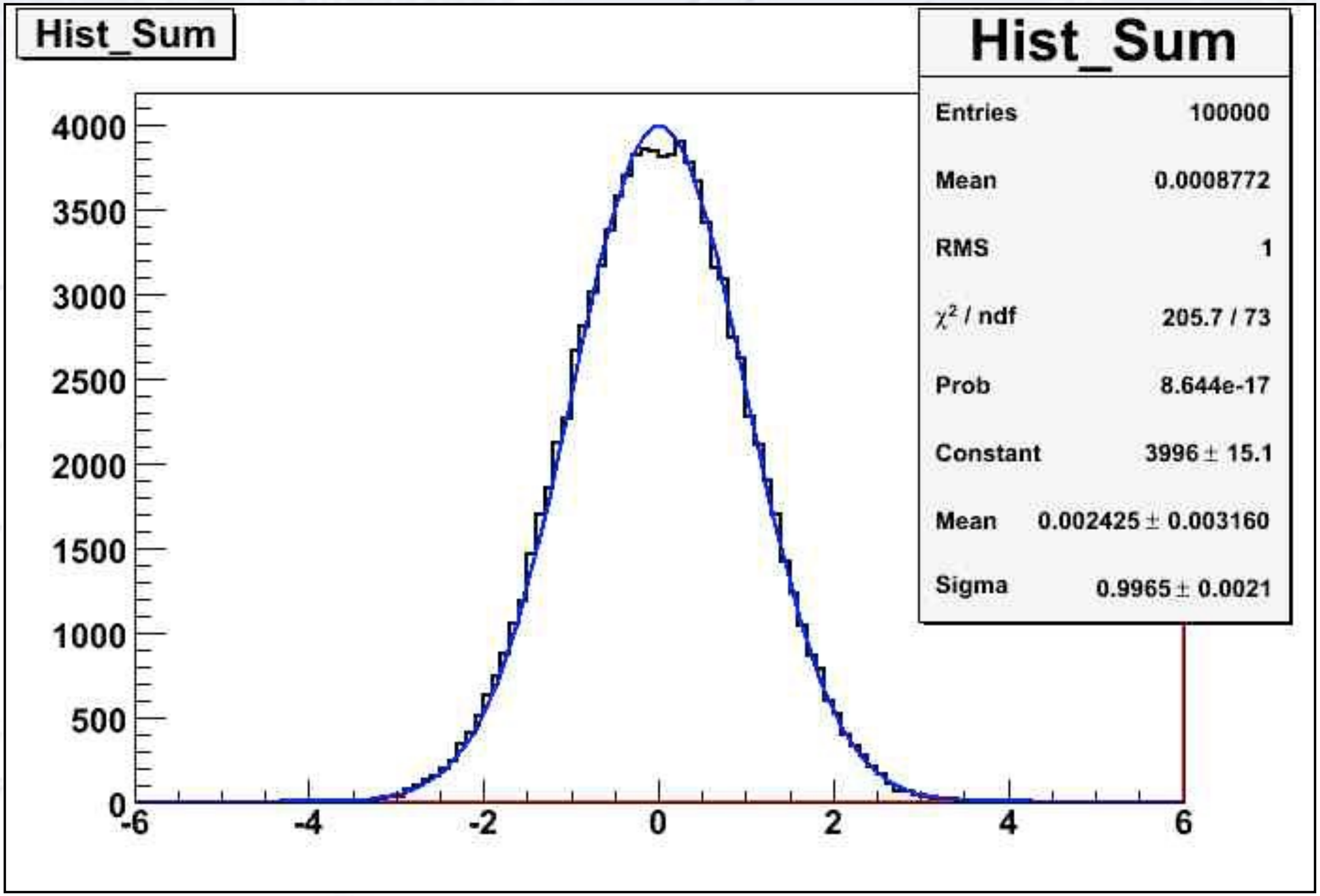


The result is a bell shaped curve, a so-called **normal** or **Gaussian** distribution.

It turns out, that this is very general!!!

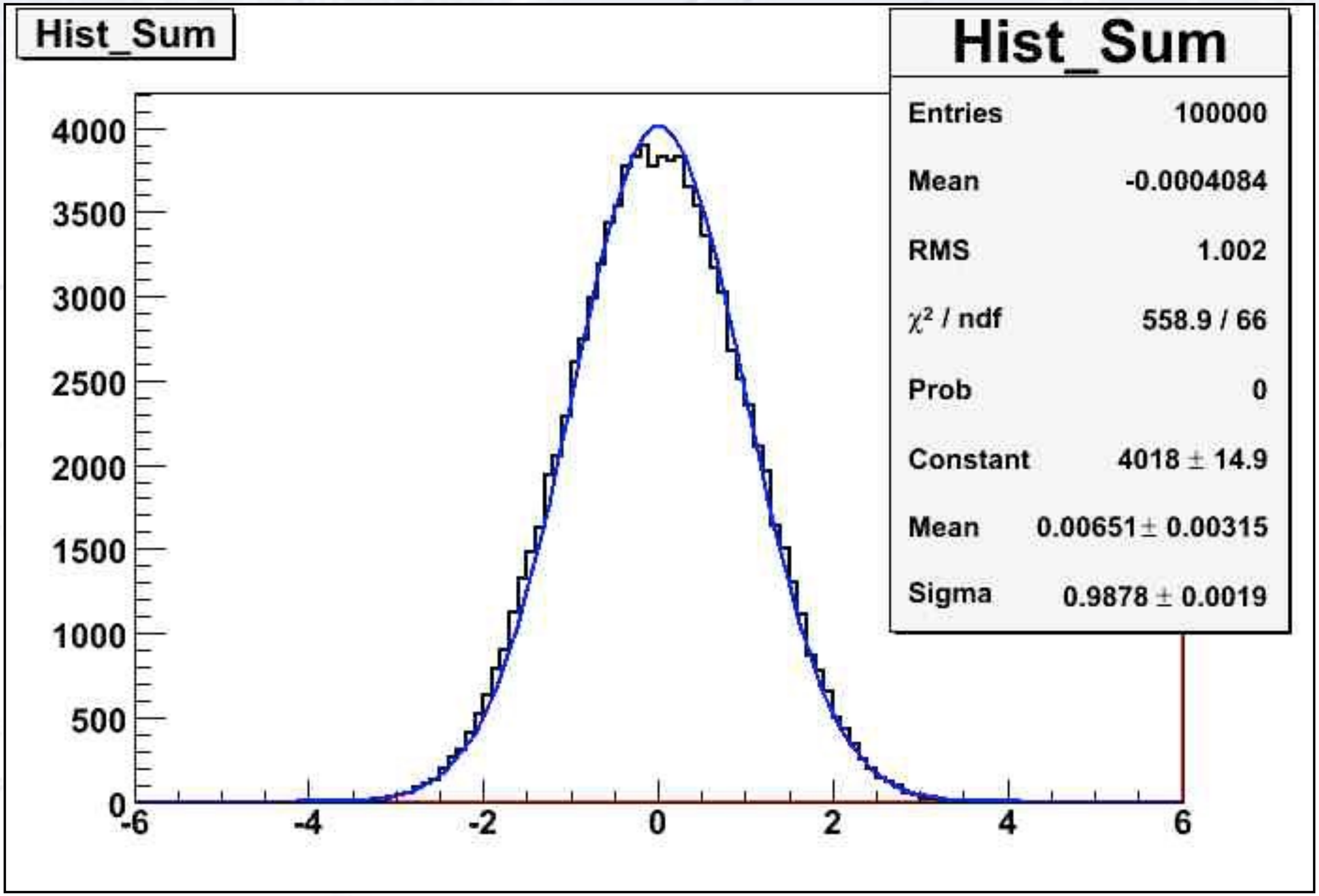
Example of Central Limit Theorem

Now take the sum of just **10** uniform numbers!



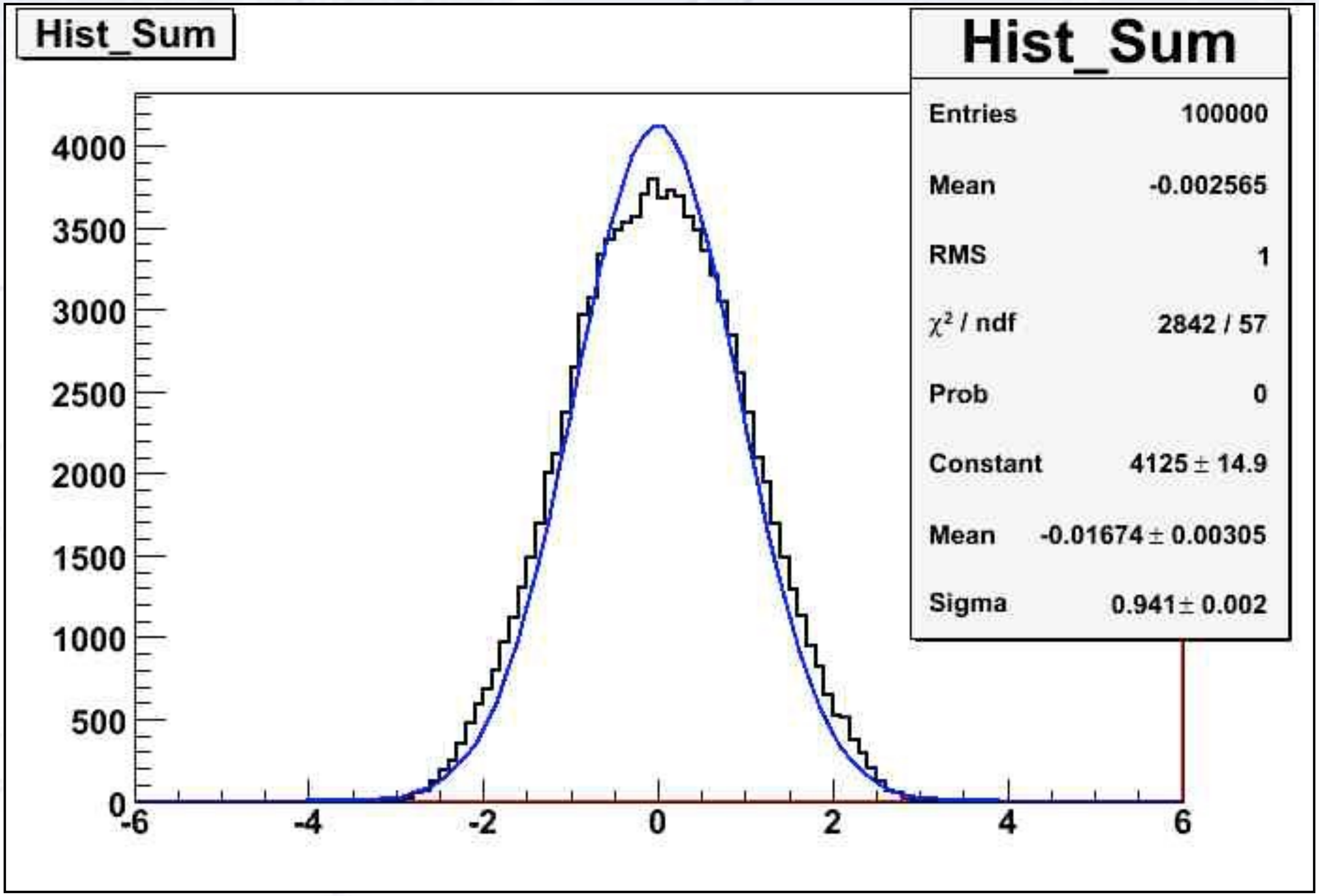
Example of Central Limit Theorem

Now take the sum of just **5** uniform numbers!



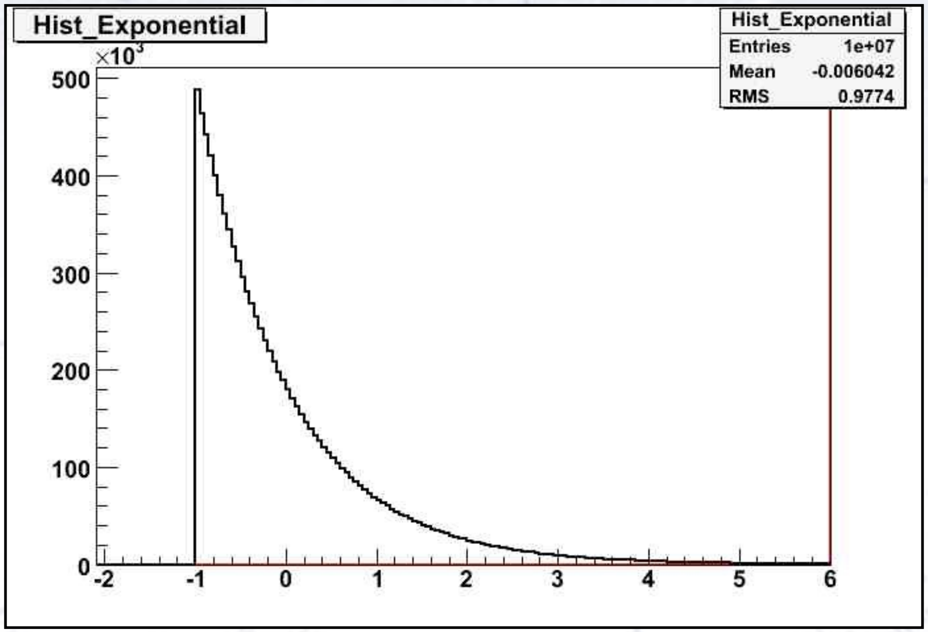
Example of Central Limit Theorem

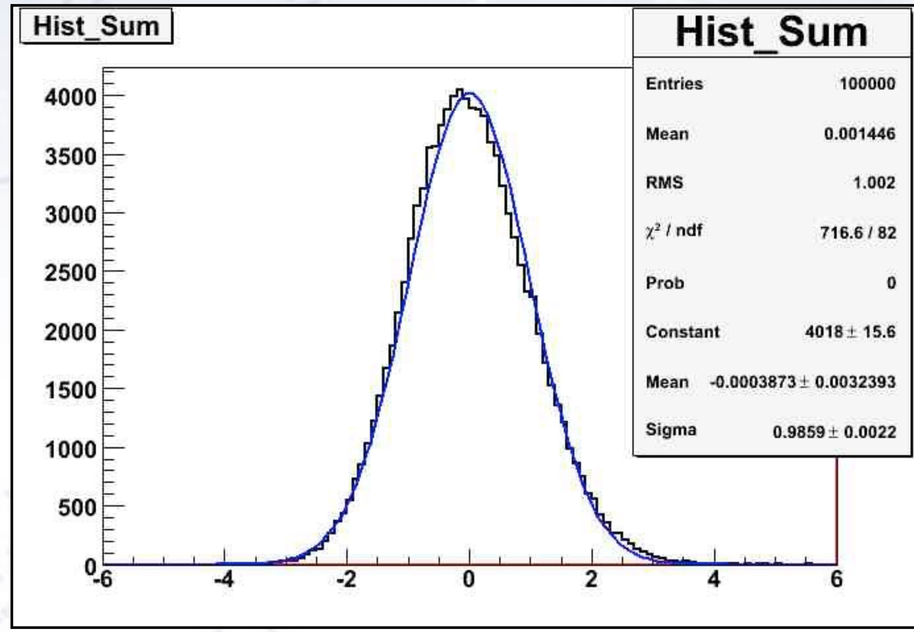
Now take the sum of just **3** uniform numbers!



Example of Central Limit Theorem

This time we will try with a much more "**nasty**" function. Take the sum of 100 *exponential* numbers! Repeat 100000 times to see the sum's distribution…





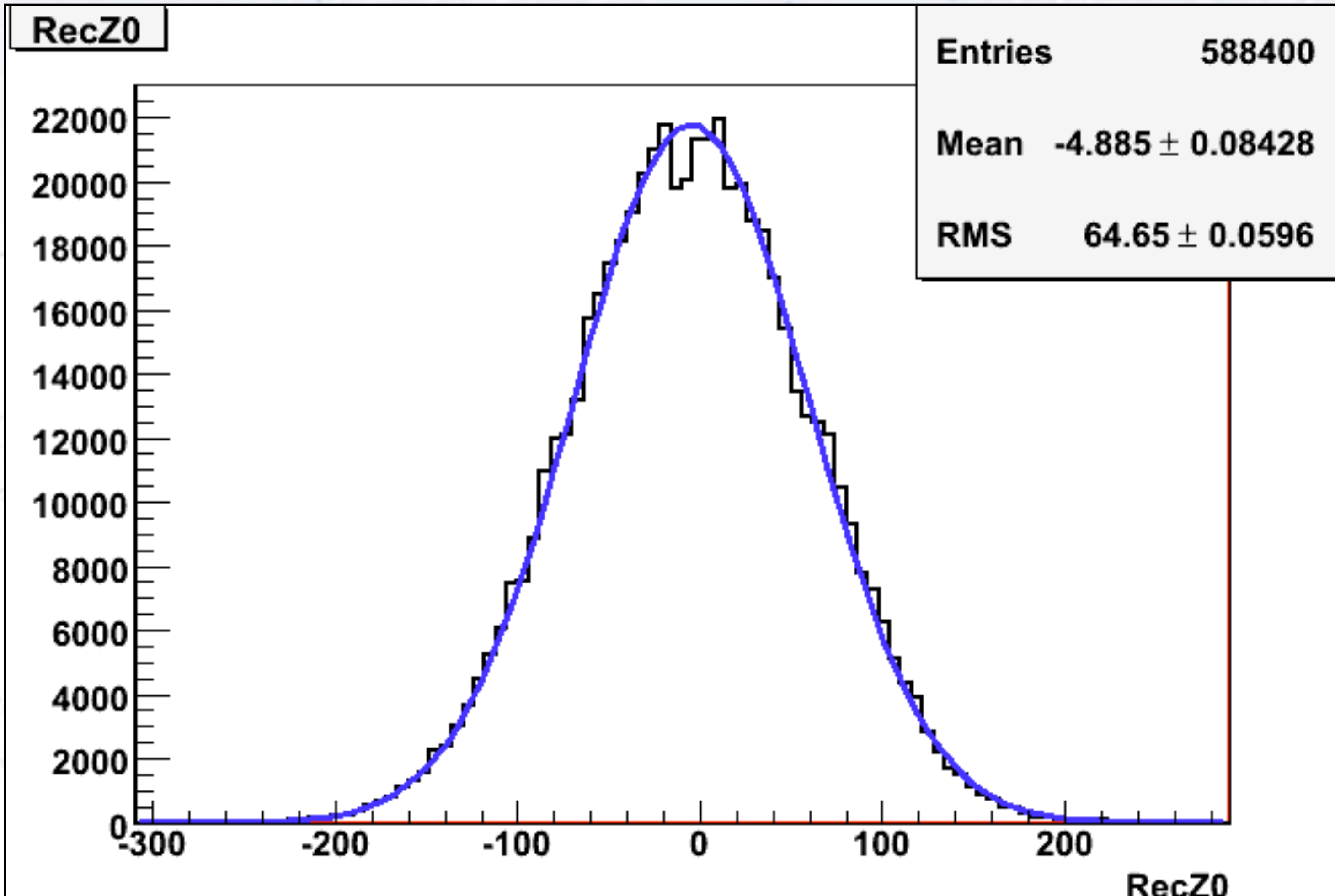
It doesn't matter what shape the input PDF has, as long as it has finite mean and width, which all numbers from the real world has! Sum quickly becomes:

Gaussian!!!

It turns out, that this fact saves us from much trouble: Makes statistics "easy"!

Example of Central Limit Theorem

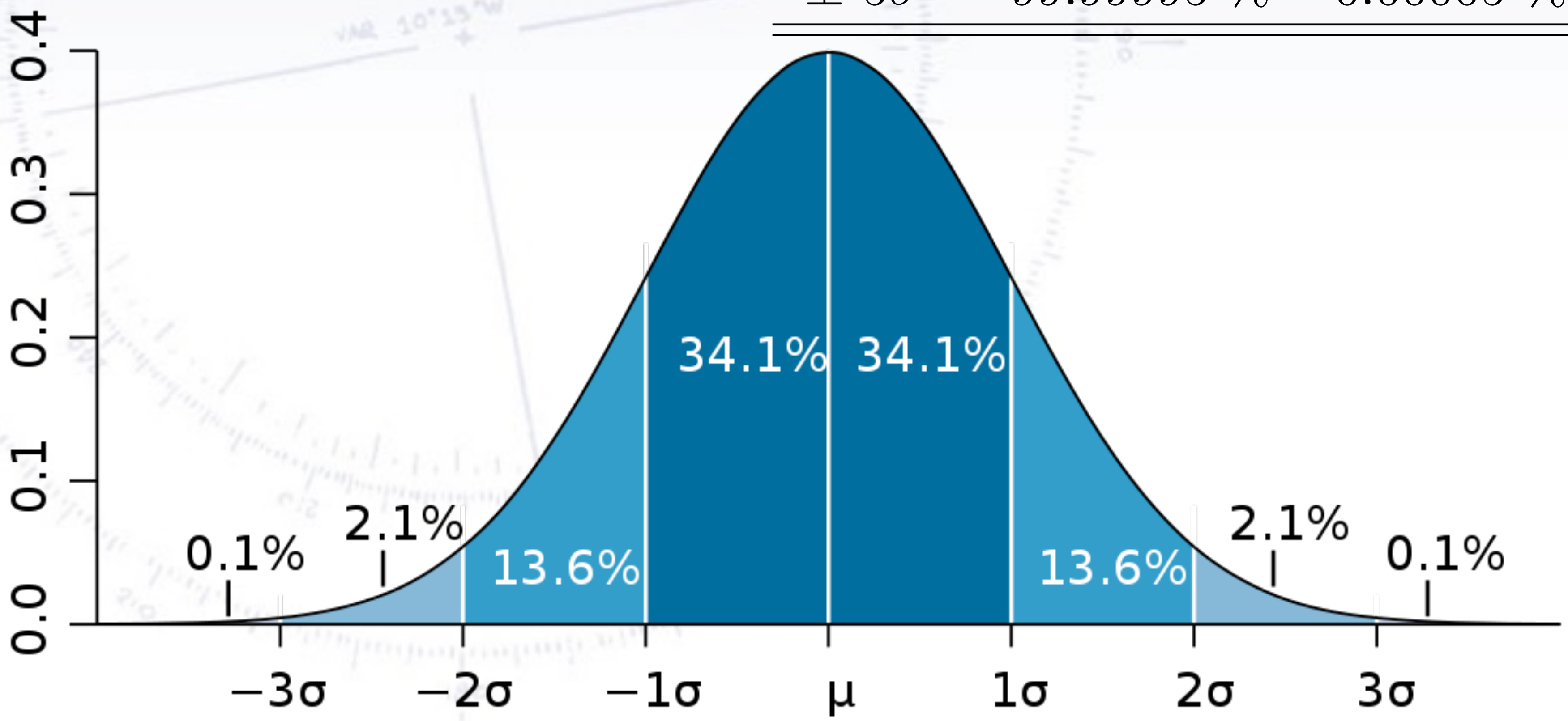
Looking at z-coordinate of tracks at vertex from proton collisions in CERNs LHC accelerator by the ATLAS detector, this is what you get:



The Gaussian distribution

It is useful to know just a few of the most common Gaussian integrals:

Range	Inside	Outside
$\pm\,1\sigma$	**68** %	32 %
$\pm\,2\sigma$	**95** %	5 %
$\pm\,3\sigma$	**99.7** %	0.3 %
$\pm\,5\sigma$	99.99995 %	0.00005 %



Summary

The Central Limit Theorem

...is your good friend because it…

ensures that uncertainties tend to be Gaussian

…which are the easiest to work with!