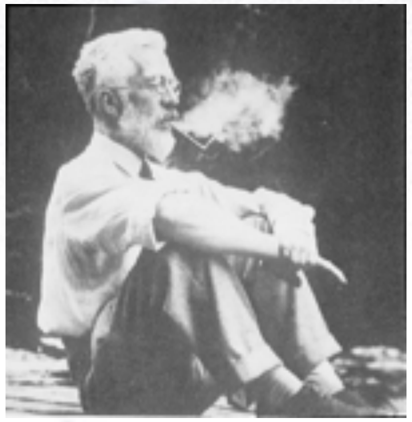
Applied Statistics

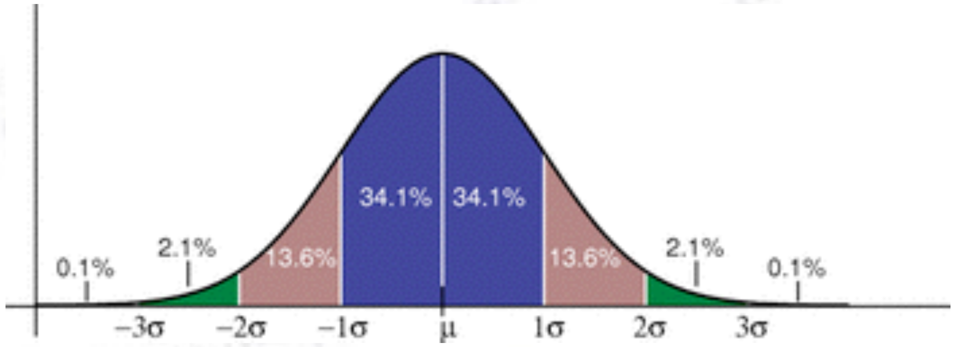
Principle of maximum likelihood








Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Likelihood function

"I shall stick to the principle of likelihood…"

[Plato, in Timaeus]

Likelihood function

Given a PDF f(x) with parameter(s) θ, what is the chance that with N observations, x_i falls in the intervals $[x_i, x_i + dx_i]$?

$$\mathcal{L}(\theta) = \prod_i f(x_i, \theta) dx_i$$

Likelihood function

Given a set of measurements **x**, and parameter(s) θ, the likelihood function is defined as:

$$\mathcal{L}(x_1, x_2, \ldots, x_N; \theta) = \prod_i p(x_i, \theta)$$

The **principle of maximum likelihood** for parameter estimation consist of maximising the likelihood of parameter(s) (here θ) given some data (here **x**).

The likelihood function plays a central role in statistics, as it can be shown to be:
- Consistent (converges to the right value).
- Asymptotically normal (converges with Gaussian errors).
- Efficient (reaches the Cramer-Rao lower bound for large N).

To some extend, this means that the likelihood function is "optimal", that is, if it can be applied in practice.

Likelihood vs. Chi-Square

For computational reasons, it is often much easier to minimise the logarithm of the likelihood function:

$$\left.\frac{\partial \ln \mathcal{L}}{\partial \theta}\right|_{\theta=\bar{\theta}} = 0$$

In problems with Gaussian errors, it turns out that the **likelihood function** boils down to the **Chi-Square** with a constant offset and a factor -2 in difference.

The likelihood function for fits comes in two versions:
- Binned likelihood (using Poisson) for histograms.
- Unbinned likelihood (using PDF) for single values.

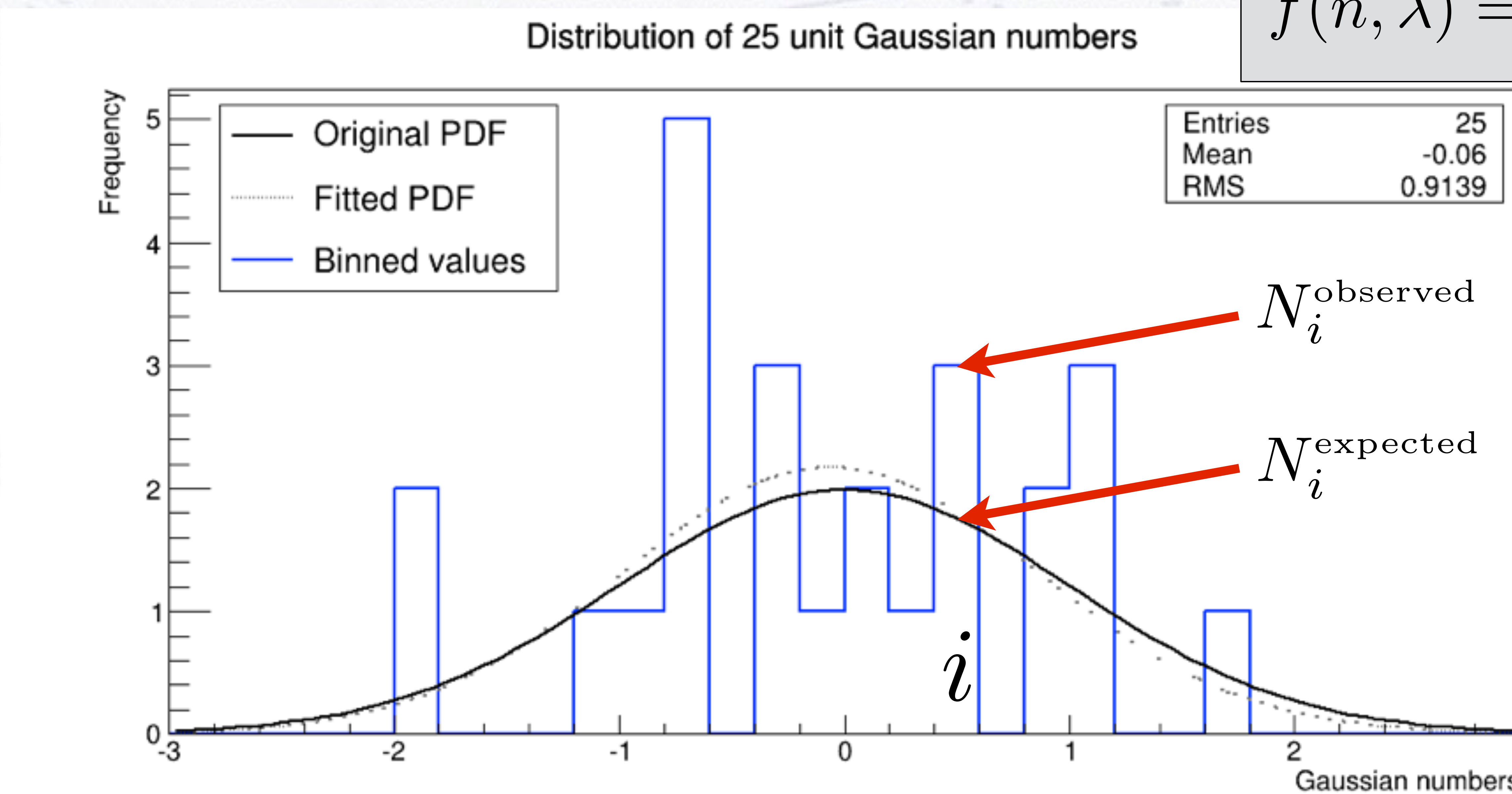
The “trouble” with the likelihood is, that it is unlike the Chi-Square, there is NO simple way to obtain a probability of obtaining certain likelihood value!

Binned Likelihood

The binned likelihood is a sum over bins in a histogram:

$$\mathcal{L}(\theta)_{\text{binned}} = \prod_{i}^{N_{\text{bins}}} \text{Poisson}(N_i^{\text{expected}}, N_i^{\text{observed}})$$

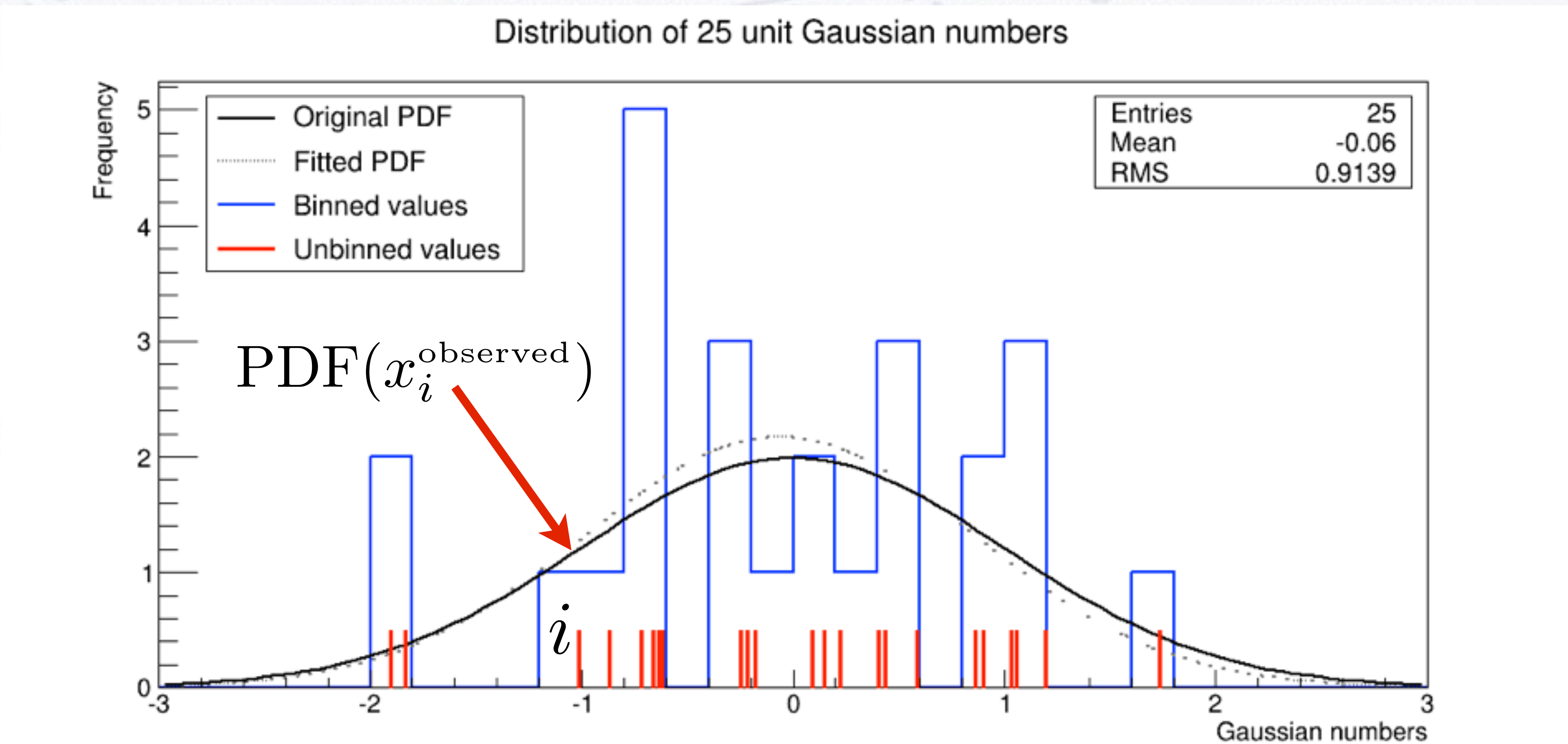
$$f(n, \lambda) = \frac{\lambda^n}{n!} e^{\lambda}$$



Unbinned Likelihood

The binned likelihood is a sum over single measurements:

$$\mathcal{L}(\theta)_{\text{unbinned}} = \prod_{i}^{N_{\text{meas.}}} \text{PDF}(x_i^{\text{observed}})$$



Notes on the likelihood

For a large sample, the maximum likelihood (ML) is indeed unbiased and has the minimum variance - that is hard to beat! Also, the binned LLH approaches the unbinned version. However...

For the ML, you have to know your PDF. This is also true for the Chi-Square, but unlike for the Chi-Square, you get no goodness-of-fit measure to check it!

Also, the small statistics, the ML is not necessarily unbiased, but still fares much better than the ChiSquare! But be careful with small statistics.
The way to avoid this problem is using simulation - more to follow.

The likelihood ratio test

Not unlike the Chi-Square, were one can compare χ^2 values, the likelihood between two competing hypothesis can be compared (SAME offset constant/factor!).

While their individual LLH values do not say much, their RATIO says everything!

As with the likelihood, one often takes the logarithm and multiplies by -2 to match the Chi-Square, thus the “test statistic” becomes:

$$\begin{aligned} D &= -2\ln\left(\frac{\text{likelihood for null model}}{\text{likelihood for alternative model}}\right) \\ &= -2\ln(\text{likelihood for null model}) + 2\ln(\text{likelihood for alternative model}) \end{aligned}$$

If the two hypothesis are simple (i.e. no free parameters) then the **Neyman-Pearson Lemma** states that this is the best possible test one can make.

If the alternative model is not simple, this difference approximately behaves like a Chi-Square distribution with $N_{dof} = N_{dof}(\text{alternative}) - N_{dof}(\text{null})$