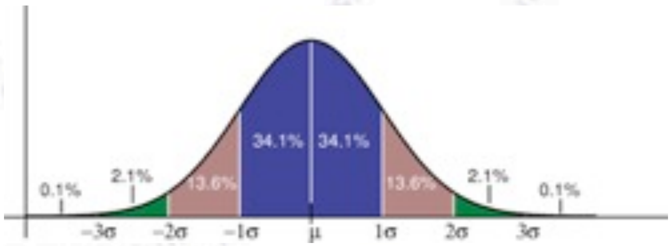
Applied Statistics

ChiSquare



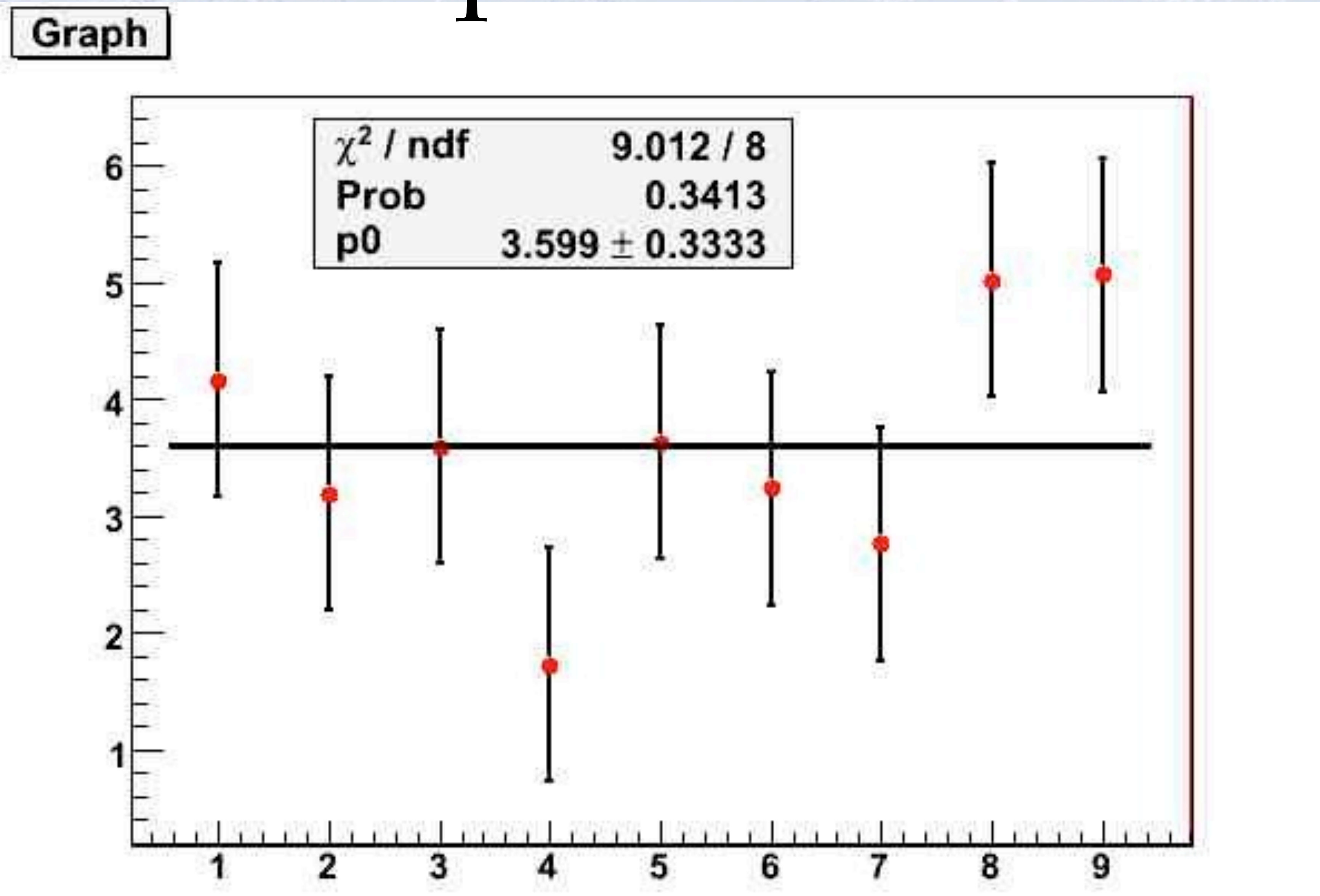



Troels C. Petersen (NBI)



"Statistics is merely a quantization of common sense"

ChiSquare method



ChiSquare method

Given data points (y_i as a function of x_i), how do you decide what fit/model best represents the data, and if this representation is good at all?

The answer: The ChiSquare…

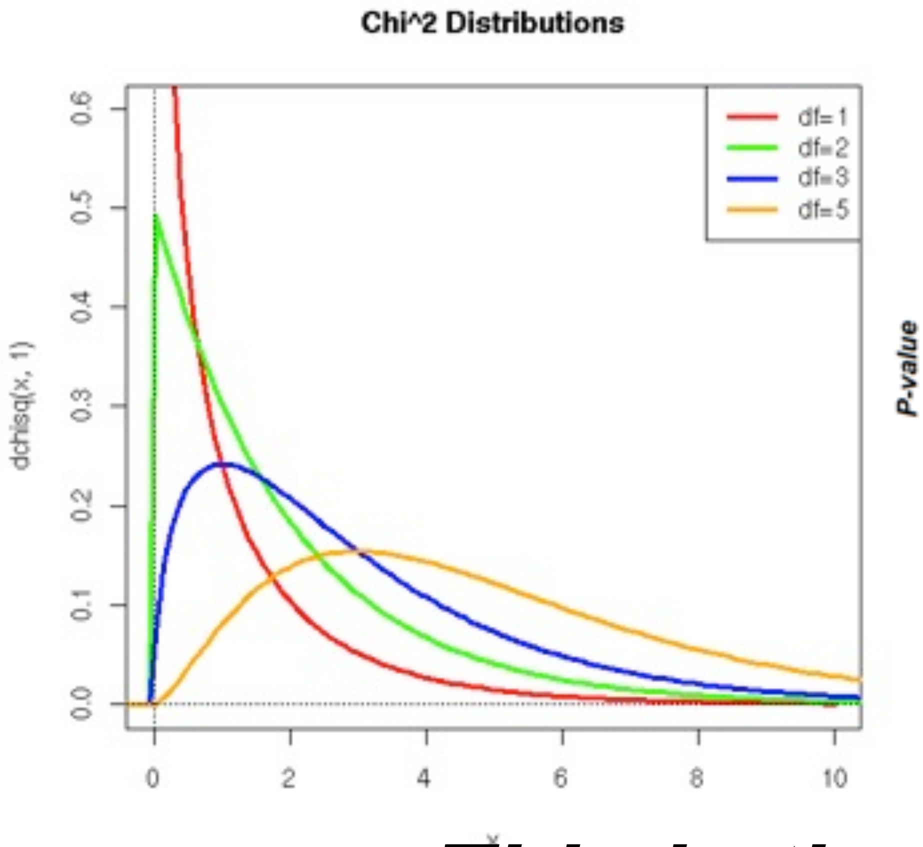
$$\chi^2(\bar{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i;\bar{\theta}))^2}{\sigma_i^2}$$

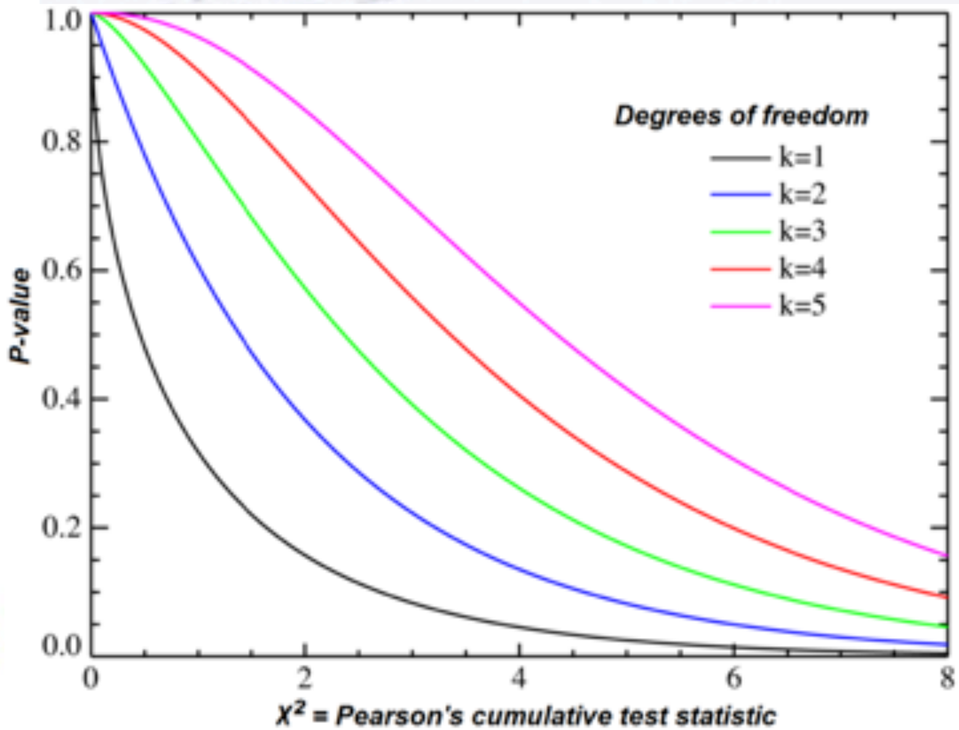
Take for every data point the difference between the value (y_i) and the fit (here λ) in terms of the uncertainty (σ_i) and square this. The goal is to ***MINIMIZE*** the sum of these terms!

From the value of the sum and the degrees of freedom, a fit quality can be calculated - ***This is the Chi2 test!***

ChiSquare test

For a given number of degrees of freedom, the PDF for expected Chi2 values is known! Thus one can compare the obtained value to this, and calculate the probability of observing something with the same Chi2 value or higher...





This is the Chi2 test!!!

Chi-square probability

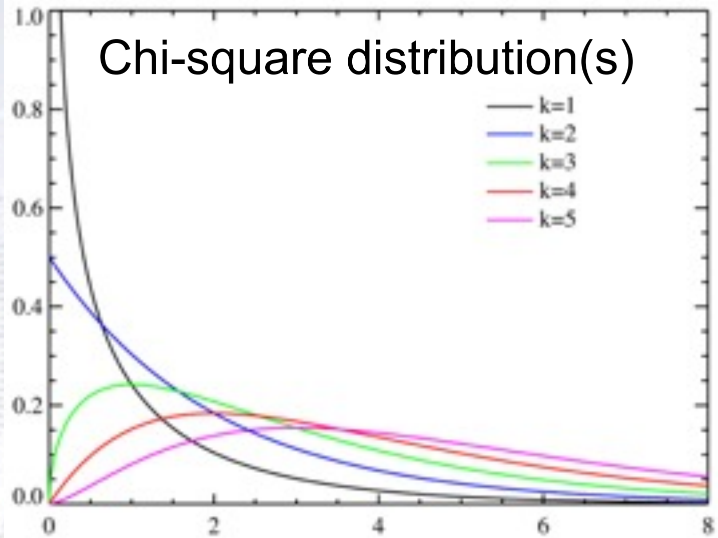
Given a **Chi-square value** and a **number of degrees of freedom** (Ndof), one can obtain a **"goodness-of-fit"**.

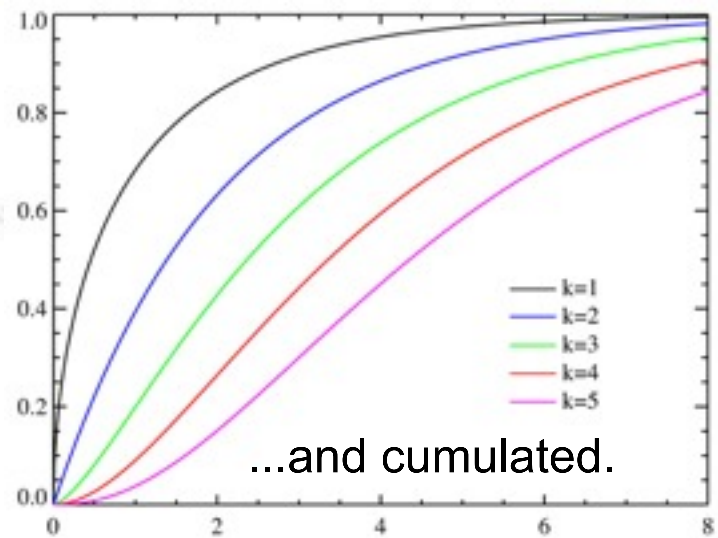
It is known, what Chi-square values to expect given the Ndof. One can therefore compare to this (Chi-square) distribution, and see...

what is the probability of getting this Chi-square value (or something worse!).

<u>Example:</u>

A fit gave the Chi-square 7.1 with 5 dof. The chance of getting this Chi-square or worse is... (reading the pink bottom curve (Ndof = k = 5) at 7.1)...





Chi-square probability

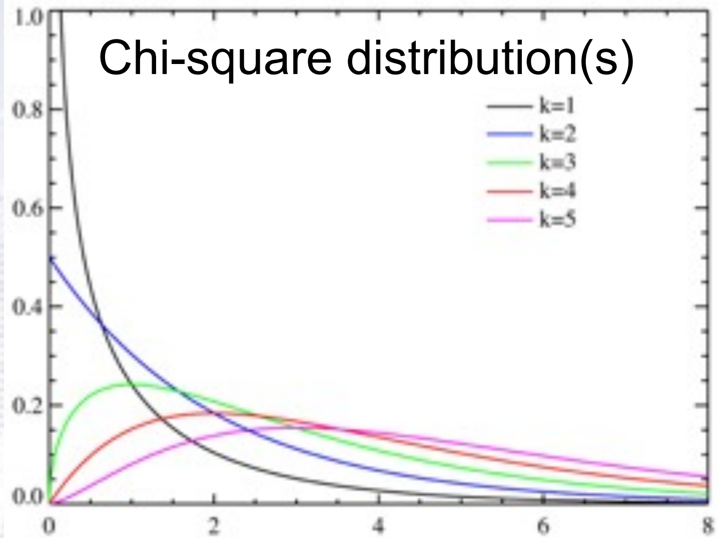
Given a **Chi-square value** and a **number of degrees of freedom** (Ndof), one can obtain a **"goodness-of-fit"**.

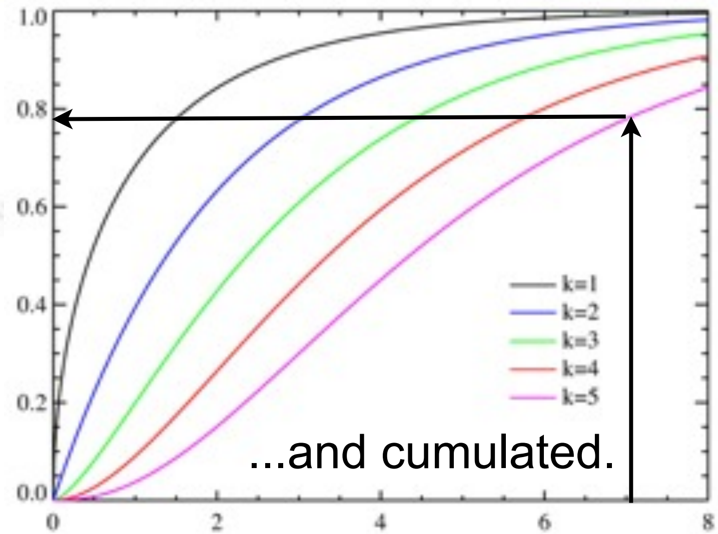
It is known, what Chi-square values to expect given the Ndof. One can therefore compare to this (Chi-square) distribution, and see...

what is the probability of getting this Chi-square value (or something worse!).

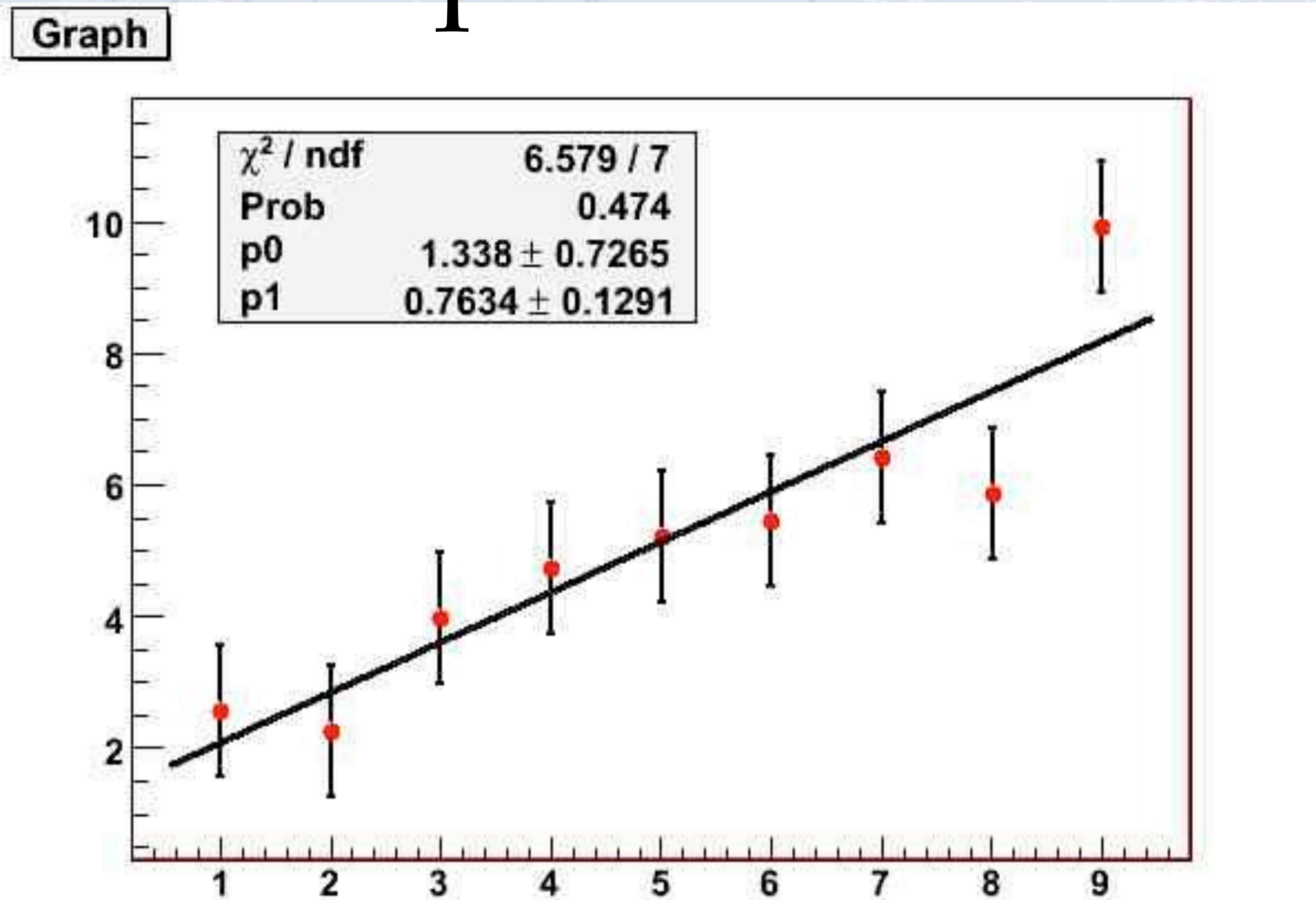
<u>Example:</u>

A fit gave the Chi-square 7.1 with 5 dof. The chance of getting this Chi-square or worse is... (reading the pink bottom curve (Ndof = k = 5) at 7.1)... 1 - 0.78 = **22%**





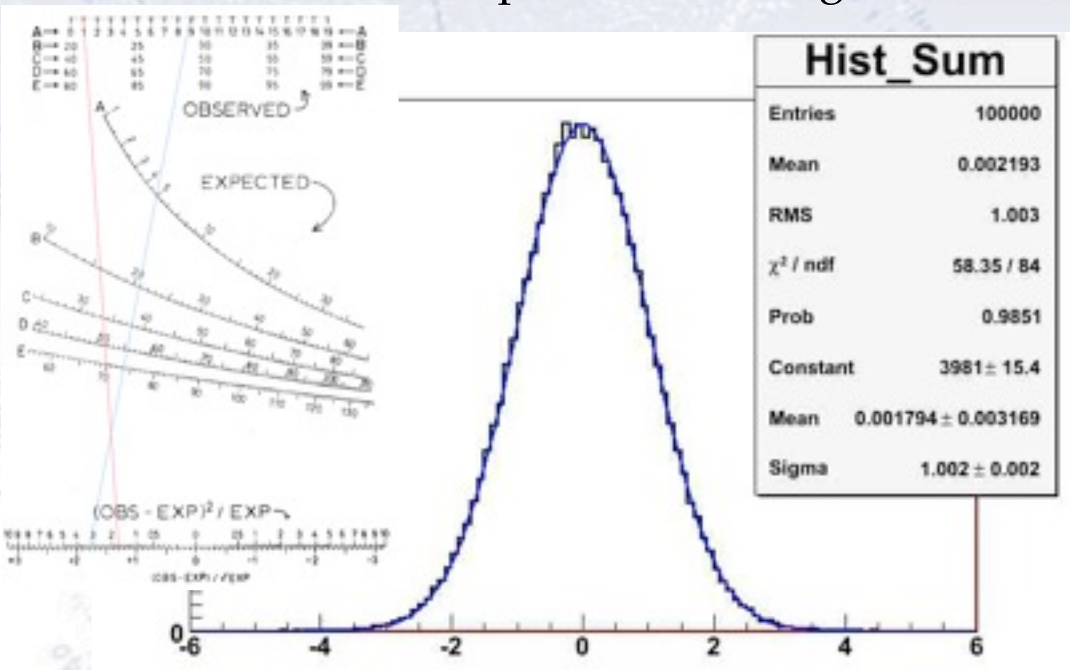
ChiSquare method



Note: The ChiSquare can also be used to determine size of uncertainty.

ChiSquare method for binned data

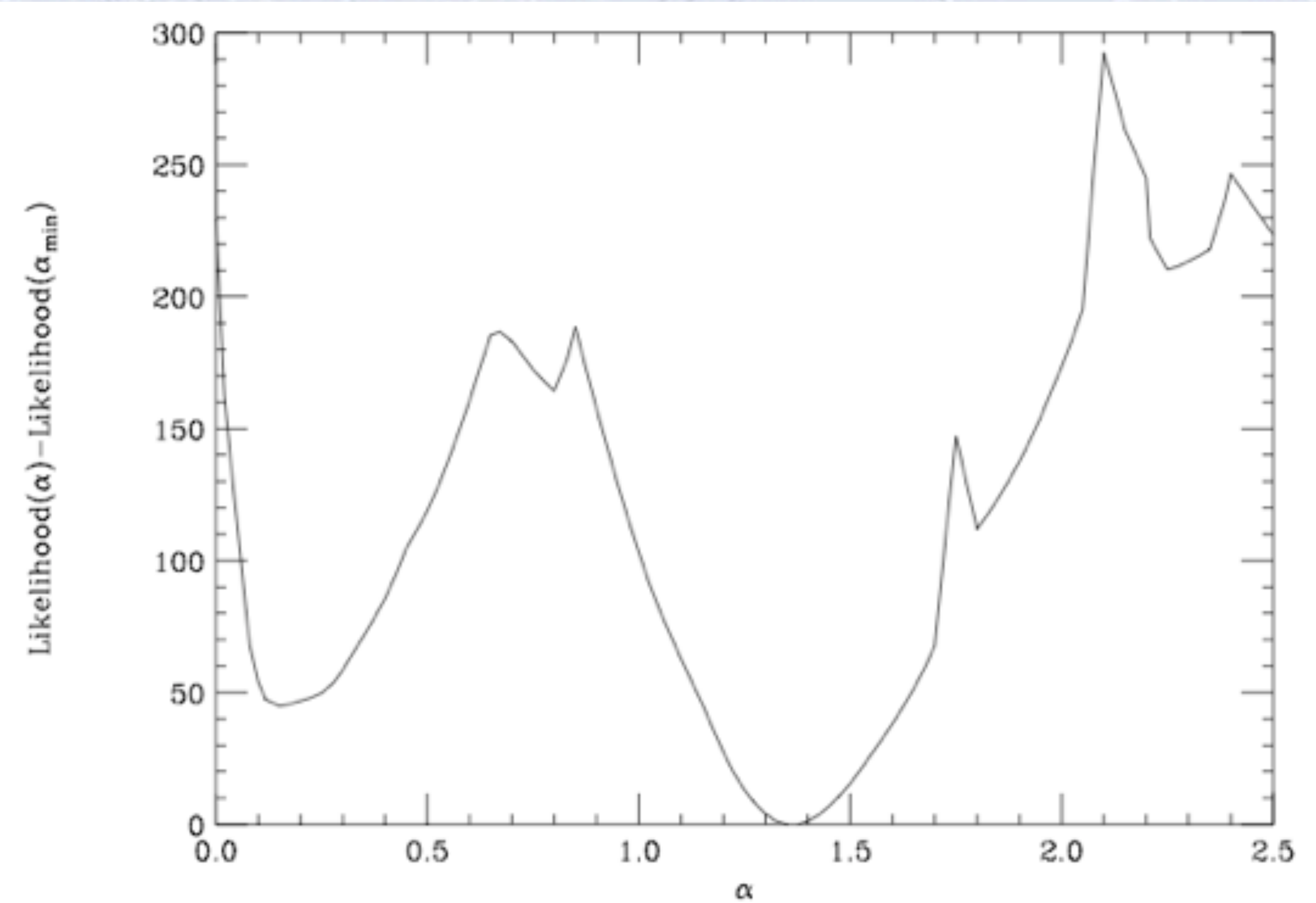
If the data is binned (i.e. put into a histogram), then Pearson's ChiSquare applies:



The formula (based on Poisson statistics) is:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Example of Chi2 fit

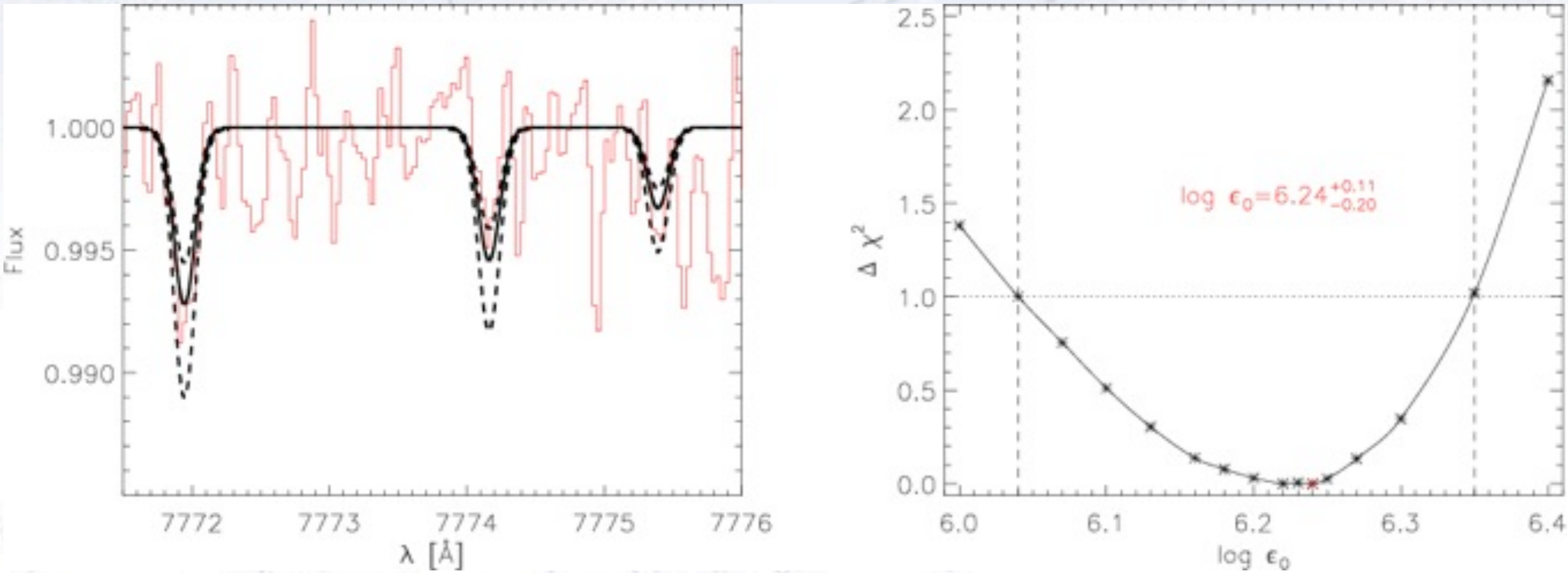


The fact that there are several minima makes fitting difficult/uncertain!

Always give good starting values!!!

Example of Chi2 fit

Uncertainties need not always be symmetric (though that is usually better!)



The uncertainty on a parameter is found where the Chi2 has increased by 1 from the minimum.

Notes on the ChiSquare method

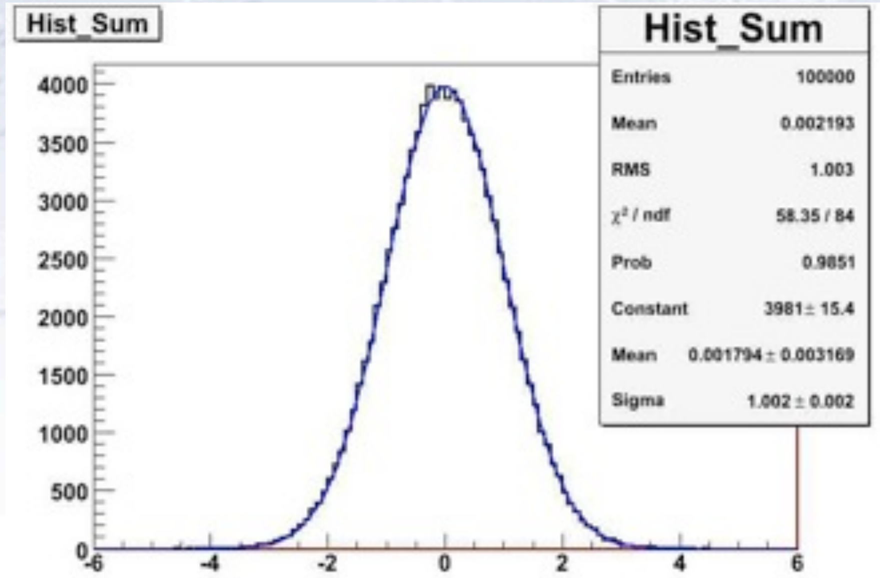
"It was formerly the custom, and is still so in works on the theory of observations, to derive the method of least squares from certain theoretical considerations, the assumed normality of the errors of the observations being one such.
It is however, more than doubtful whether the conditions for the theoretical validity of the method are realized in statistical practice, and the student would do well to regard the method as recommended chiefly by its comparative simplicity and by the fact that it has ***stood the test of experience****"*.

[G.U. Yule and M.G. Kendall 1958]

ChiSquare method for binned data

Alternatively, one can use the G-test:

$$G = 2\sum_{ij} O_{ij} \cdot \ln(O_{ij}/E_{ij}),$$



The advantage is that this responds better, when the number of events is low.

However, it is much less known, and should perhaps only be used as a way of testing what the systematic uncertainty is between methods.

(I have not personally seen it used anywhere!)

ChiSquare method