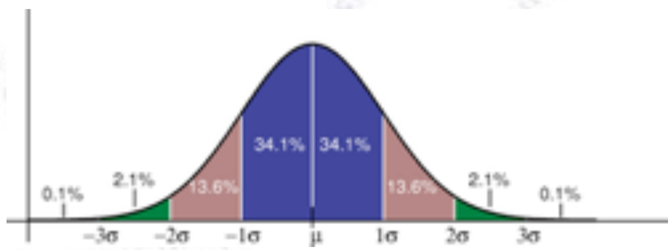
Applied Statistics

Advanced fitting






Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Defining the Chi-Square

<u>Problem Statement:</u> Given N data points (x,y), adjust the parameter(s) θ of a model, such that it fits data best.

The best way to do this, given uncertainties σ_i on y_i is by minimising:

$$\chi^2(\theta) = \sum_i^N \frac{(y_i - f(x_i, \theta))^2}{\sigma_i^2}$$

The power of this method is hard to overstate!

Not only does it provide a simple, elegant and unique way of fitting data, but more importantly it provides a **goodness-of-fit measure**.

This is the Chi-Square test!

Chi-Square probability interpretation

The Chi-Square probability can roughly be interpreted as follows:

- If χ^2 / Ndof ≃ 1 or more precisely if $0.01 < p(\chi^2, \text{Ndof}) < 0.99$, then all is good.
- If χ^2 / Ndof ≫ 1 or more precisely if $p(\chi^2, \text{Ndof}) < 0.01$, then your fit is bad, and your hypothesis is probably not correct.
- If χ^2 / Ndof ≪ 1 or more precisely if $0.99 < p(\chi^2, \text{Ndof})$, then your fit is TOO good and you probably overestimated the errors.

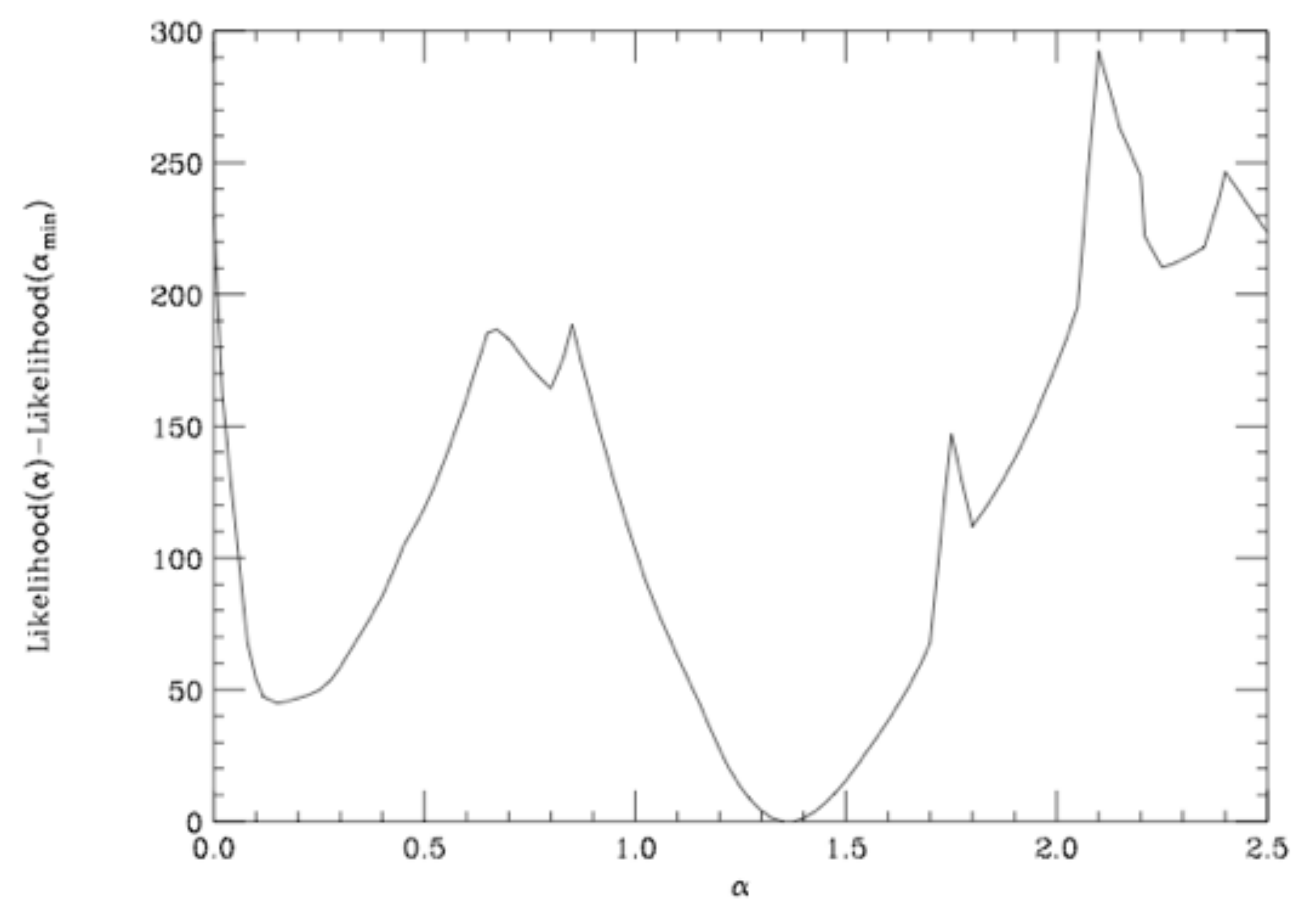
If the statistics behind the plot is VERY high (great than 10^6), then you might have a hard time finding a model, which truly describes all the features in the plot (as now tiny effects become visible), and one hardly ever gets a good Chi-Square probability.
However, in this case, one should not worry too much, unless very high precision is wanted.
Anyway, the Chi-Square still allows you to compare several models, and determine which one is the better.

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 realised 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]

Example of Chi-Square

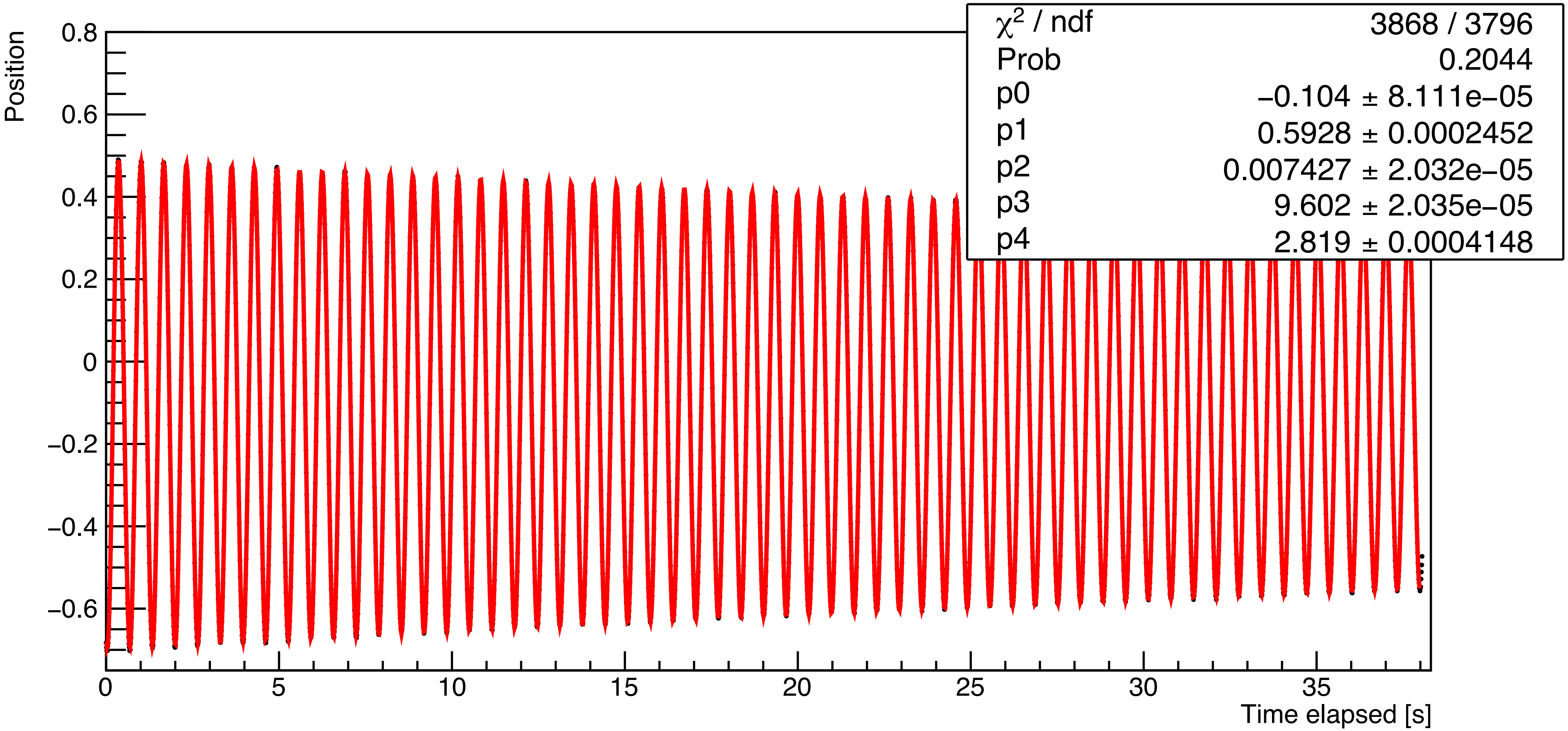


The fact that there are several minima makes fitting difficult/uncertain!

Always give good starting values!!!

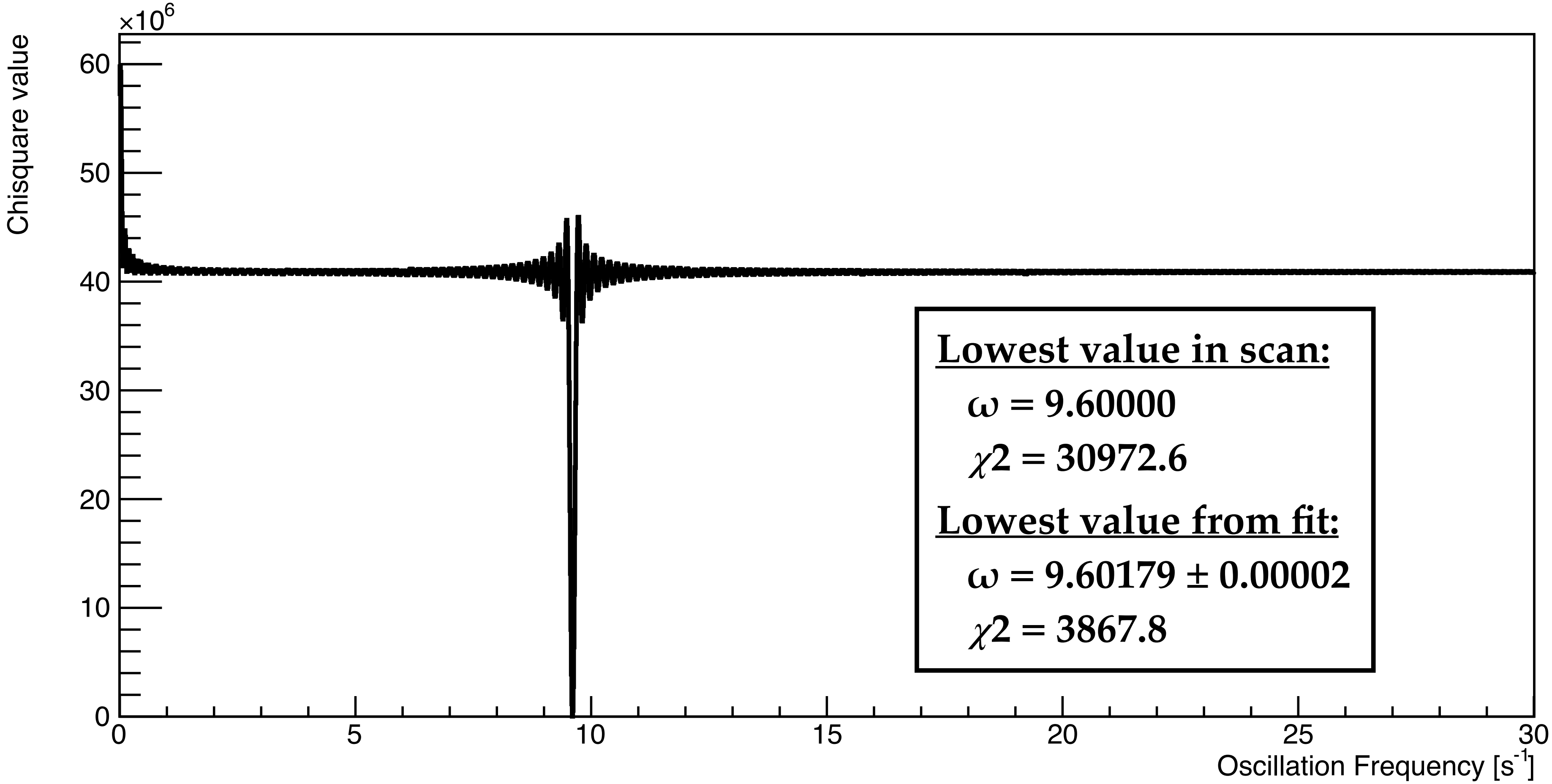
Example of Chi-Square

Especially fitting oscillatory data requires a good starting value for omega.
Even a small offset may result in a Chi2, which is not sensitive to omega!



Example of Chi-Square

Especially fitting oscillatory data requires a good starting value for omega.
Even a small offset may result in a Chi2, which is not sensitive to omega!



Fitting tips & tricks

There are a few tips & tricks that will make your (fitting) life a bit easier:

- Always give good initial values!!!
- Never start with an advanced fit - make a simple one work and expand!
- Try to make your parameters as little correlated as possible.
- Let the parameters represent the quantities of interest.
- Always start with a ChiSquare fit, as these usually has better convergence.

When a fit refuses to work, try the following:

- Draw function on top of data to check formula and quality of initial values.
- Check the correlations between the parameters.
- Try to fix one or more parameter to a value you find reasonable.

Even with all of this advice, there is no guaranty that your fit will work.
It is after all a bit of an art….