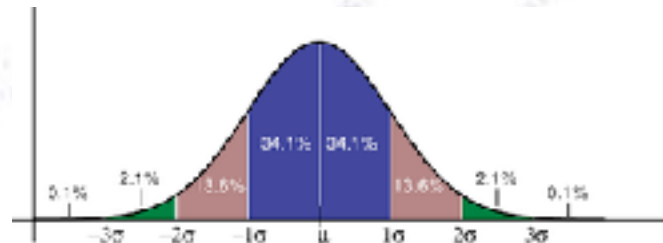


Applied Statistics

Notes on normalisation in fits



Troels C. Petersen (NBI)

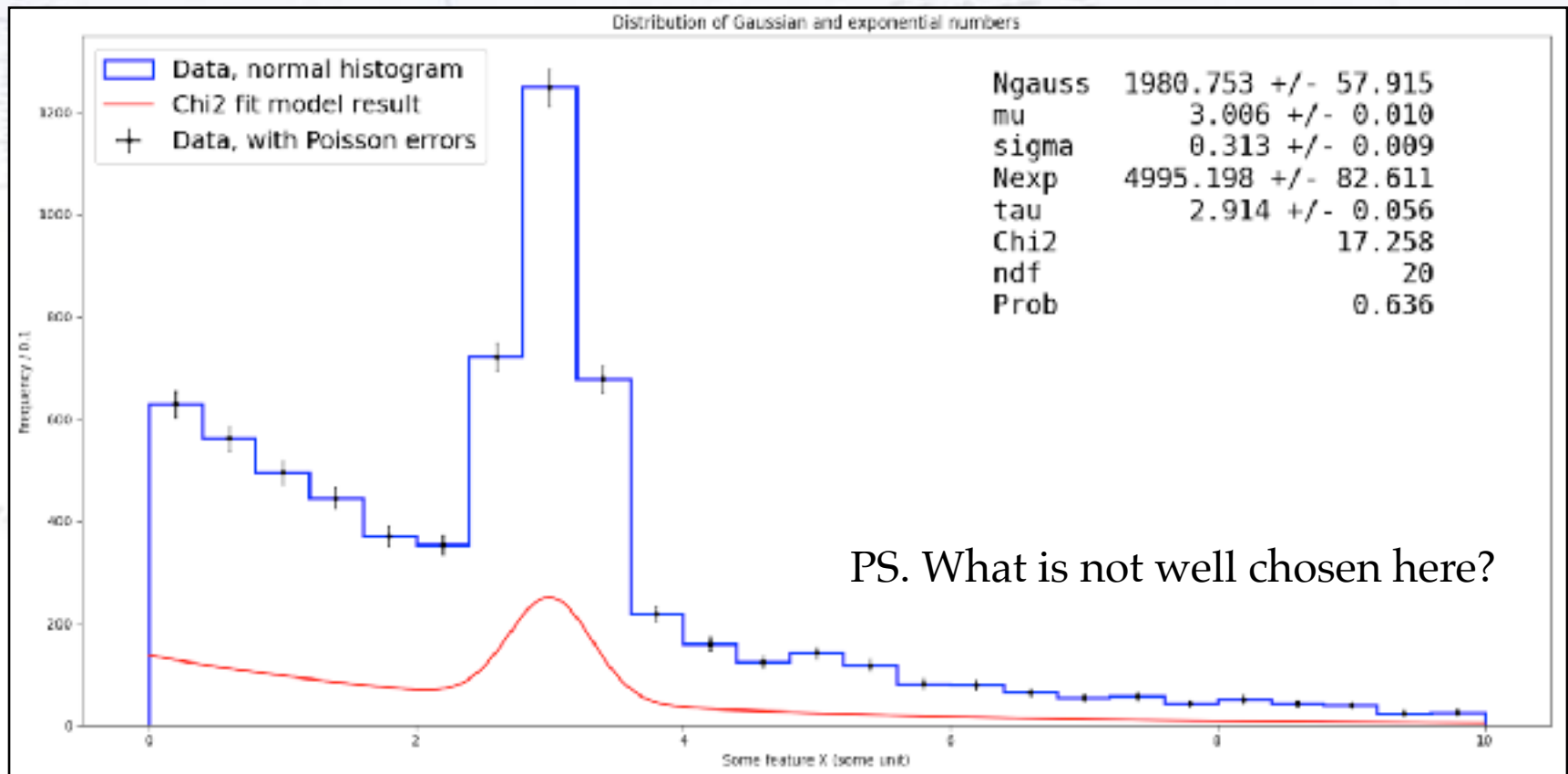


"Statistics is merely a quantisation of common sense"

Fit-histogram mismatches

You are fitting a function, and after the fit has converged, you (cleverly) decide that you want to draw the fitted function on top of the histogram with the data.

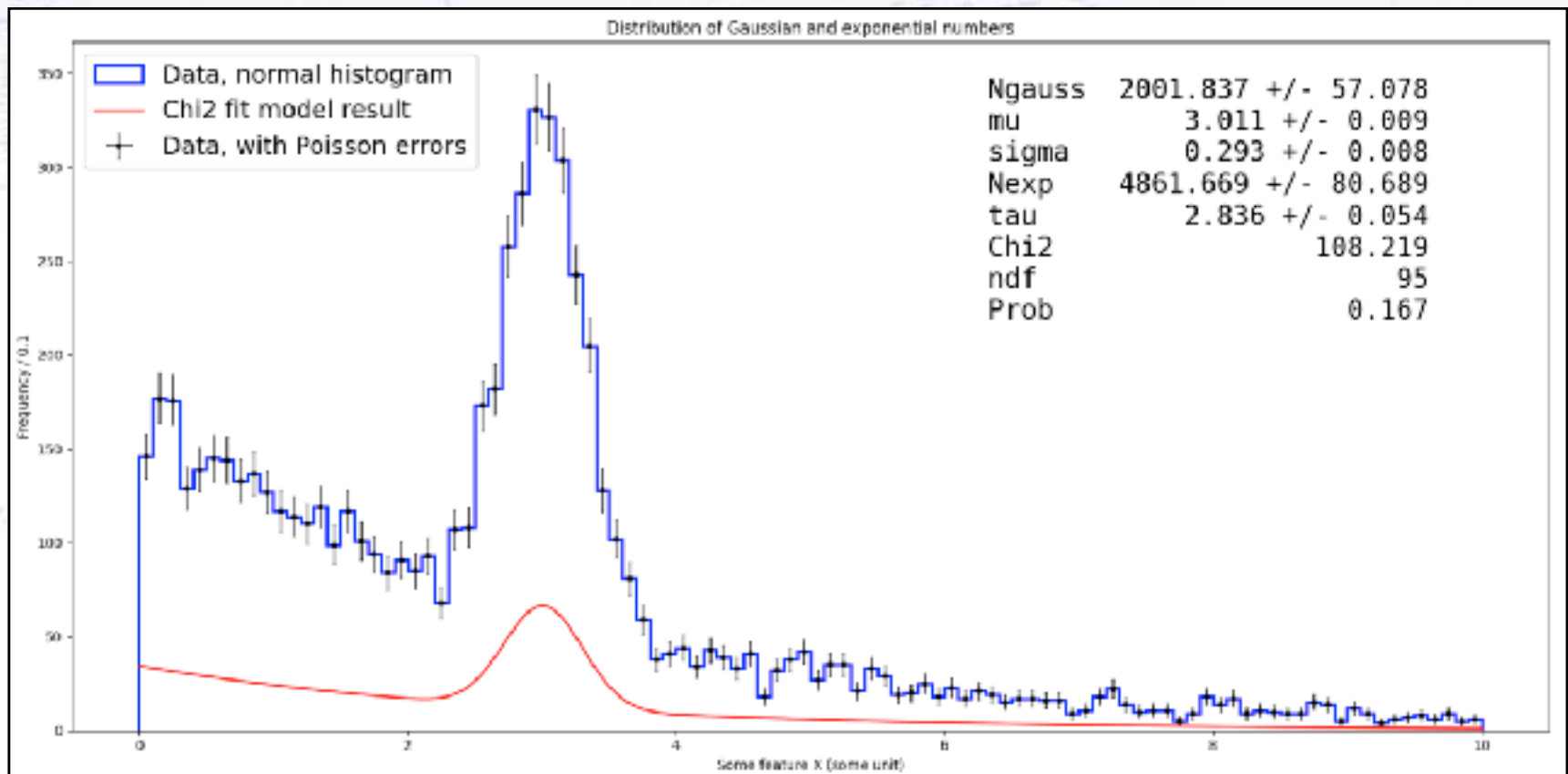
However, when you plot it, there is a clear mismatch!



Fit-histogram mismatches

You are fitting a function, and after the fit has converged, you (cleverly) decide that you want to draw the fitted function on top of the histogram with the data.

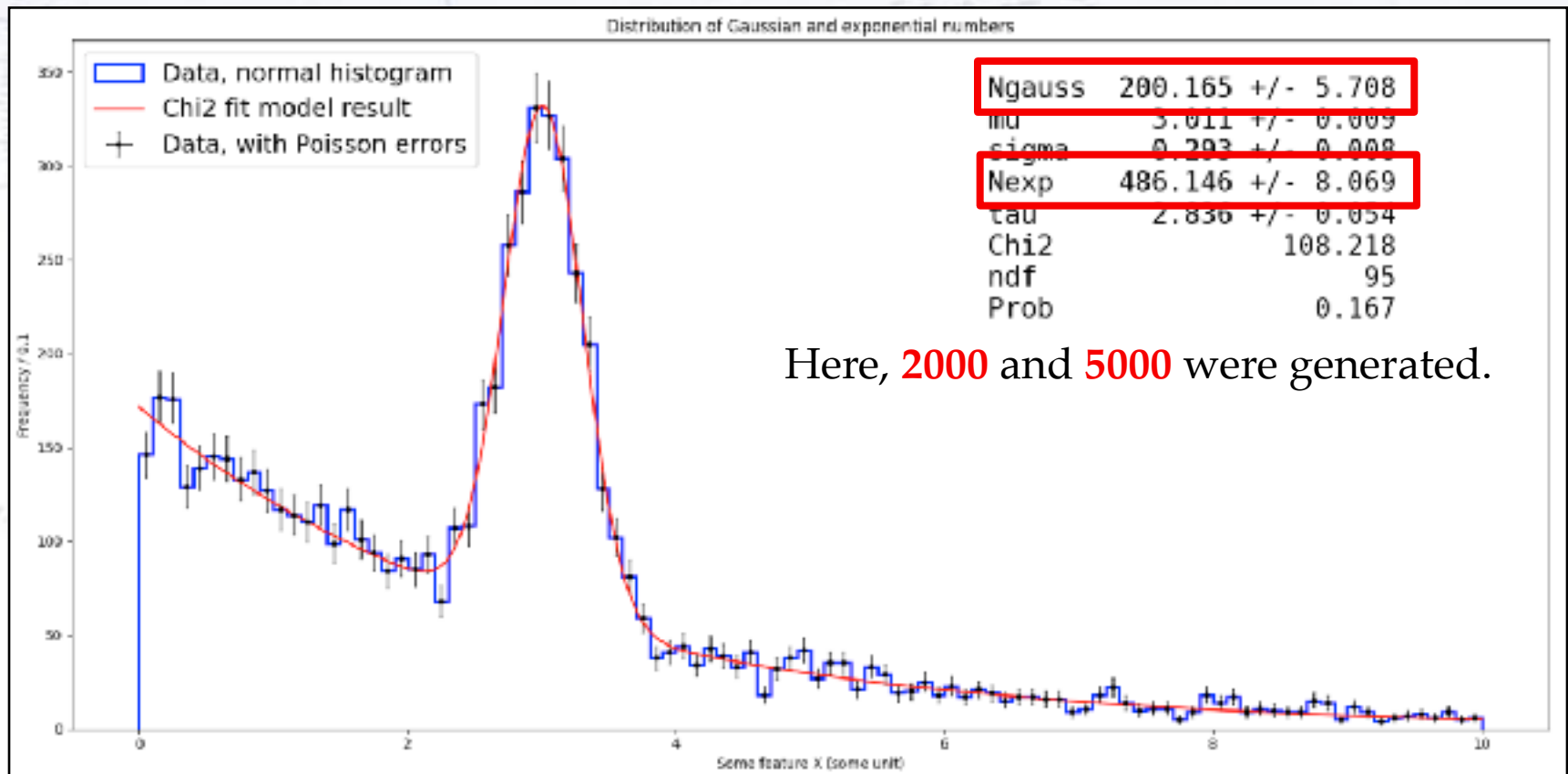
However, when you plot it, there is a clear mismatch!



Fit-histogram mismatches

You are fitting a function, and after the fit has converged, you (cleverly) decide that you want to draw the fitted function on top of the histogram with the data.

Or, the normalisations do not at all match what you would expect of the data!



Here, **2000** and **5000** were generated.

Fit-histogram mismatches

Three things are usually the source of this:

1. **You have not normalised your fitting function to be PDFs.**

This is not always a requirement (i.e. if you fit with a Pol3). But when writing papers (and master theses?), it should be so...

2. **You have not included the bin width.**

This part is essential, except for the unbinned likelihood fit case.

3. **When drawing an unbinned likelihood fit on top of a histogram.**

Here, you still need to include the bin width, as the histogram (not the fit) has this feature.

The bin width needs inclusion since the PDFs are normalised (?), but the histograms not. Therefore the usual “recipe” is a fitting function as:

$$f_{fit}(x, \theta) = N_{entries} \times \Delta(\text{bin width}) \times PDF(x)$$

Fit-histogram mismatches

Note how the uncertainties on the normalisations are approximately the square root of the count, however enlarged by overlaps between PDFs.

