Significance and the Likelihood Ratio

Roger Barlow

The Maximum Likelihood Ratio Test is often used to compare two models. If Model 1 is true under the null hypothesis, and Model 2 contains Model 1, i.e. it is Model 1 plus k extra parameters, then the maximum likelihood L_2 under model 2 will clearly be not less than that under model 1, L_1. The LR is twice the logarithm of the ratio.

$$LR = 2ln(L_2/L_1) = 2(lnL_2 - lnL_1)$$

However the improvement is unlikely to be large, and a large LR is an indication of an improvement ascribable to a real improvement in the modelling and can be used to reject the null hypothesis at some level of significance.

This approximately follows a χ^2 distribution with k degrees of freedom. [Cox and Hinkley, pp 313-314]. The approximation holds if the number of points is large, if the measurements are Gaussian, and if the model is linear. In a given problem these 3 assumptions will not be true, but may be approximately true. The validity of that approximation needs checking.

As an example we generate a uniform random variable over the range 0 to 1, and fit it with 4 different models. For each experiment we generates 10,000 values in 100 bins.

Model 1: A Uniform distribution. Only the normalisation is adjustable.

Model 2: A flat distribution plus a Gaussian: 4 free parameters altogether.

Model 3: A linear distribution with 2 free parameters.

Model 4: A cubic. This also has 4 free parameters.

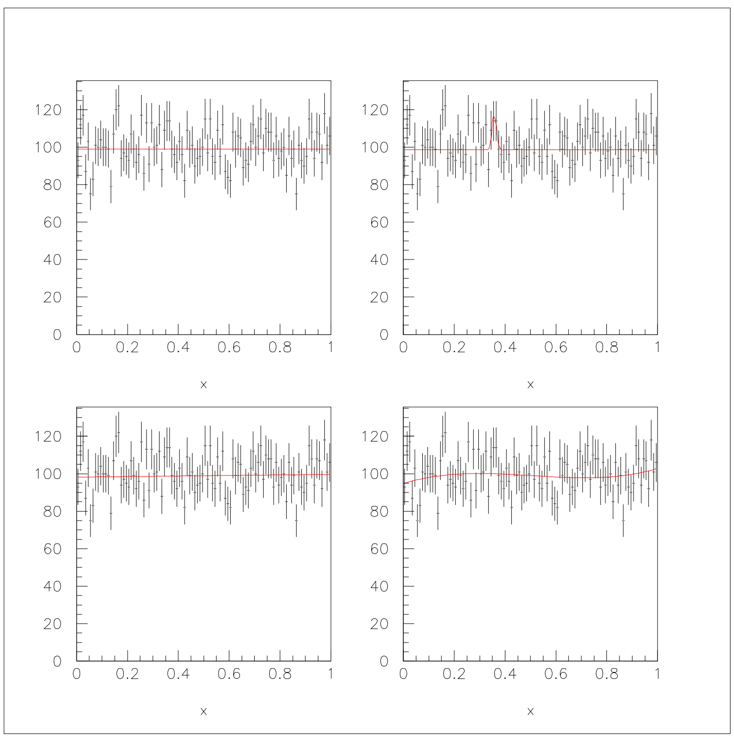

Figure 1: A typical experiment

Results of a typical 'experiment' are shown in Figure 1 with the fits of models 1 to 4.

For each model in each experiment we obtain a χ^2, and plot the probability distribution for the appropriate number of degrees of freedom (99, 96, 98 and 96 respectively). these are shown in Figure 2

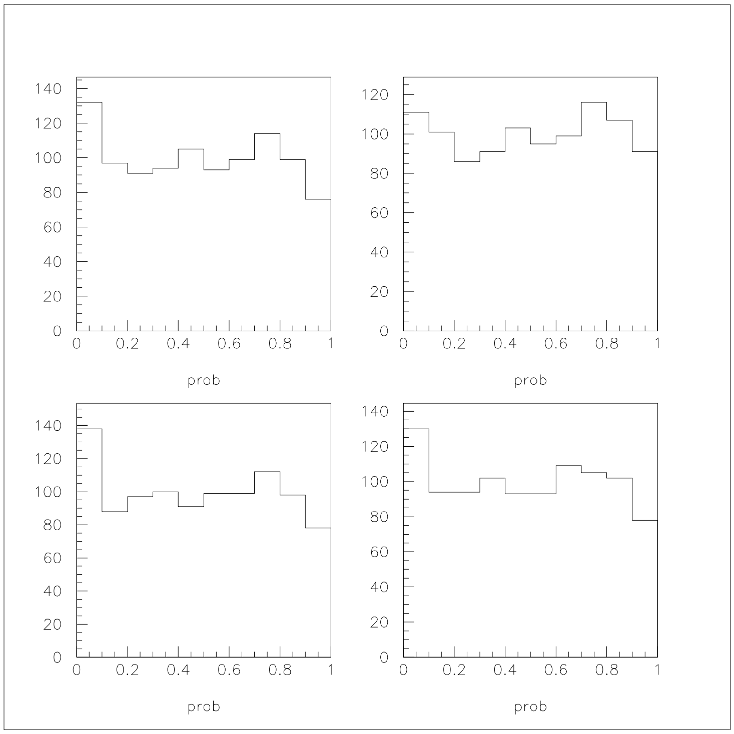

Figure 2: Probability distributions for χ^2 of models

Ideally these should be flat, as Model 1 is actually correct. While not perfect, this is actually pretty good. (If the number of values per bin is decreased then these are not flat. This is a check that the basic Gaussian errors are correct.)

Then for each experiment we take the difference between the χ^2 from model 1 and from the more complicated models, and plot the probability distribution for this $\Delta\chi^2$ and the appropriate number of degrees if freedom (1 for the linear fit, 3 for the cubic and 3 for the Extra Gaussian). These are shown in Figure 3

Left to right, these are: the straight line, the cubic, and the Gaussian

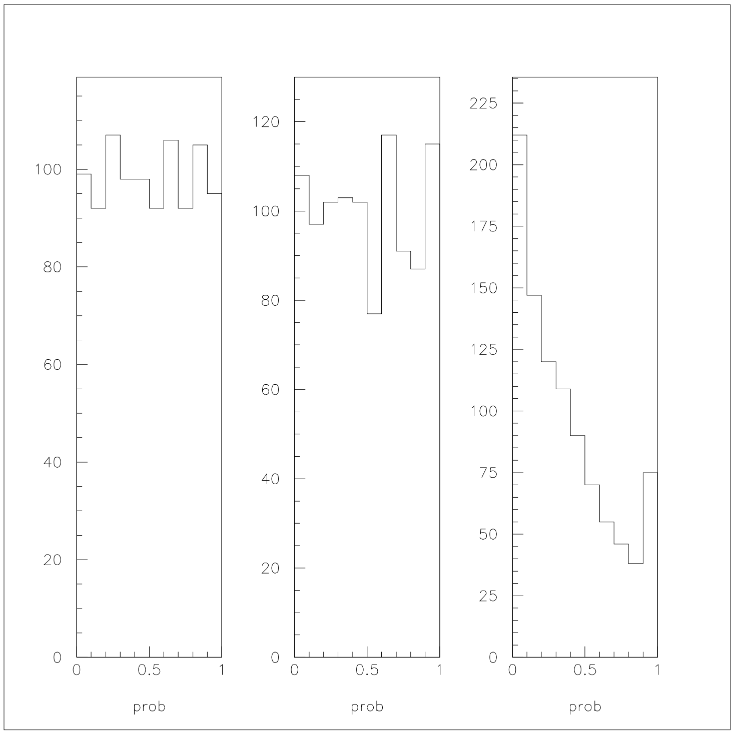

Figure 3: Probability distributions from $\Delta\chi^2$ for the linear function, the cubic, and the flat+gaussian.

The first two plots are acceptably flat, as expected. These are linear models, and the value of $n = 100$ is large, and the large numbers of events per bin means the Poisson variation is near enough Gaussian. The third plot is manifestly not flat. The 'Gaussian bump' model is *not* linear (in two of its 4 parameters) and this appears to matter. Even though the number of parameters is the same as the cubic, the Gaussian does better.

These low prbabilities correspond to large changes in χ^2 and thus might be claimed as evidence for the real need for the bump to improve the fit.

This is borne out by Figure 4, which shows the χ^2 of the Constant+Gaussian (horizontally) versus the χ^2 for the cubic (vertically). The scatter of points lie above the line of equality. In general the fit with the bump does better (i.e. has a smaller χ^2) than the cubic fit, even though the fits have the same number of degrees of freedom and even though the data are actually described by a flat distribution with no added featuers.

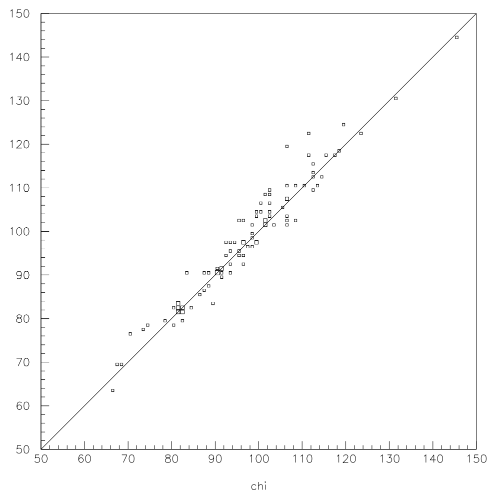

Figure 4: χ^2 for the Flat+Gaussian and the cubic

In conclusion: The use of the MLR test is restricted to Linear models, and does not work for nonlinear models, of which the addition of a bump of unspecified mass and width is one. Hence the $2ln(L_2/L_1)$ quantity, even if written as $\Delta\chi^2$, is not distributed according to a χ^2 distribution and cannot be used as a measure of significance.

Reference

D. R. Cox and D. V. Hinkley, 'Theoretical Statistics',Chapman and Hall (1974)