7.1 Goodness-of-fit

The question:

A goodness-of-fit statistic, for example χ^2, has a significance level p defined by

$$p = \int_C^\infty f(\chi^2)\, d\chi^2 \tag{1}$$

where C is the observed value of χ^2 and f is the appropriate distribution of χ^2. Show that p is uniformly distributed if C is indeed drawn from f.

We use the rule for changing variables in probability density distributions, essentially just conservation of probability:

$$\mathrm{prob}(p)\, dp = \mathrm{prob}(C)\, dC$$

and we know the distribution of C is f. Hence

$$\mathrm{prob}(p) = f(C)\frac{1}{dp/dC}$$

and by the definition of p

$$\begin{aligned} \frac{dp}{dC} &= \frac{d}{dC}\int_C^\infty f(\chi^2)\, d\chi^2 \\ &= f(C) \end{aligned}$$

by the Fundamental Theorem of Calculus. So

$$\mathrm{prob}(p) = \frac{f(C)}{f(C)} = 1$$

and p is uniformly distributed over its range from zero to 1.

Aa simple Monte Carlo example, we take a unit-area 1-sigma Gaussian as our goodness-of-fit statistic, and perform 10000 trials by calling C and integrating the Gaussian from C to ∞ each time. Binning the results in 100 bins from 0 to 1, we get the histogram shown, with its mean bin value of 100 as anticipated.

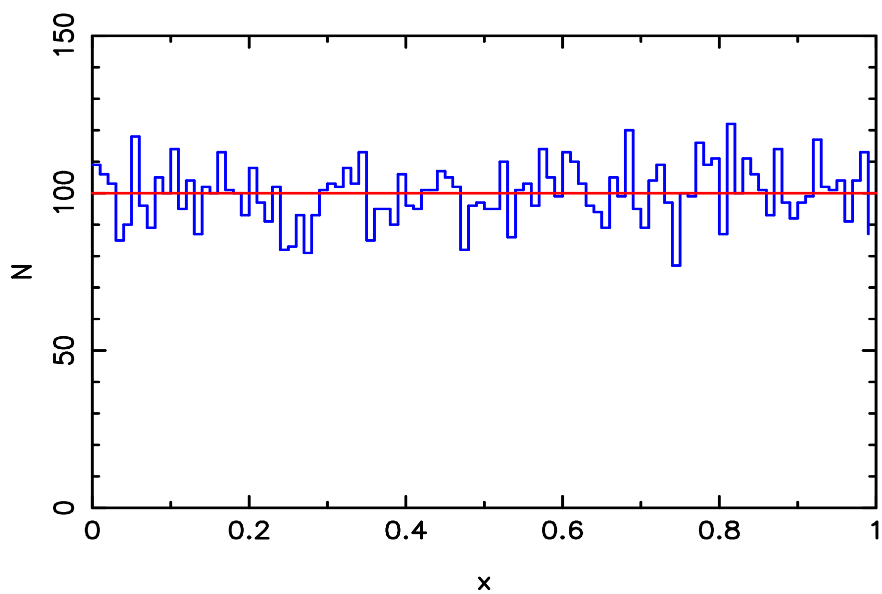

Figure 1: The binned results of 10000 integrations of the unit-area Gaussian, with the lower limit selected from such a Gaussian and the upper limit $+\infty$.