Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

There is an ever present need in data analysis to estimate goodness of fit. That is, an experimentalist makes n observations

$$O_1, O_2, \ldots, O_n$$

and wishes to test how well a theory that predicts expected values

$$E_1, E_2, \ldots, E_n$$

fits the data. This leads naturally to the chi-square variable and chi-square tests.

1 Definitions

Given a normally distributed random variable x_i with mean μ and variance σ^2 it is possible to derive from it a standard normal variable z_i using

$$z_i = \frac{x_i - \mu}{\sigma}$$

which is normally distributed with mean 0 and variance 1. A sum of squares of n such independent variables defines a chi-square variable with n degrees of freedom. That is,

$$\chi^2 = z_1^2 + z_2^2 + \ldots + z_n^2$$

is chi-square distributed with n degrees of freedom, and has expectation n and variance $2n$. For $n = 1$ the density is infinite at $\chi^2 = 0$, for $n = 2$ it is that of the exponential distribution, while the distribution becomes asymptotically normal for large n.

In applications the actual distribution and its parameters are unknown and must be estimated, say from the sample. Tests based on chi-square usually require the estimation of $k \geq 0$ such parameters in order to asses the size of test statistics like C^2 defined by

$$C^2 = \frac{(O_1 - E_1)^2}{E_1^2} + \frac{(O_2 - E_2)^2}{E_2^2} + \cdots + \frac{(O_n - E_n)^2}{E_n^2}$$

which becomes asymptotically χ^2 distributed with $n - 1 - k$ degrees of freedom as $n \to \infty$. Instead of frequencies, the objective function from weighted nonlinear regression, namely

$$WSSQ = \sum_{i=1}^{n} \left\{ \frac{y_i - f(x_i, \hat{\theta})}{s_i} \right\}^2$$

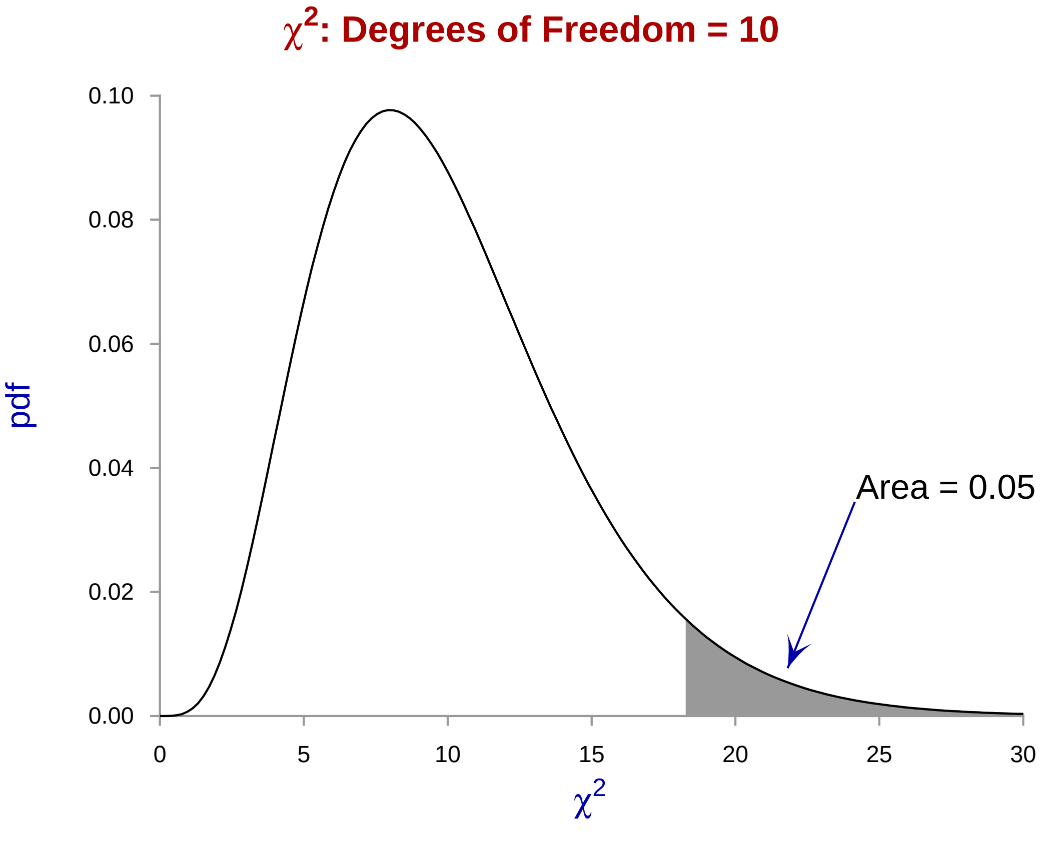
where parameters $\hat{\theta}$ have been estimated, converges to a χ^2 distribution as long as the model is correct and not over-determined, and the weights s_i are accurate.

2 Using the chi-square distribution

Choose [A/Z] from the main SIMFIT menu and open program **chisqd** when the following options will be available.

```
Input: number of degrees of freedom
Input: x-values then output pdf(x)
Input: x-values then output cdf(x)
Input: alpha then output x-critical
Input: sample then test for chi-square distribution
Input: O and E values for a chi-square test
Input: contingency table for chi-square test
Input: parameters for non-central chi-square distribution
```

After input of the number of degrees of freedom a graph like the following can be viewed.



The essence of chi-square testing is to see if test statistics such as C^2 or $WSSQ$ fall in the upper tail of the appropriate χ^2 distribution. For instance, in the above graph, the shaded region contains 5% of the probability, and a test statistic falling in this region would be considered as sufficiently extreme to support rejecting a null hypothesis, such as consistency of the data with the assumed model, at the 5% significance level. Of course it is always assumed that the sample size is sufficiently large to justify treating the test statistic as a χ^2 variable instead of an approximate χ^2 variable.