Chi-squared two-samples test

The aim of this test is to compare two distributions, f and g, with the aim of deciding if they were extracted from the same population.
The test statistics χ^2 is the following:

$$\chi^2 = \sum_{i=1}^{n} \frac{(f(x_i) - g(x_i))^2}{\sigma^2(f(x_i)) + \sigma^2(g(x_i))}$$

where n is the number of bins and the denominator is the variance of the numerator. As in the case of distributions we have:

$$\sigma(f(x_i)) = \sqrt{f(x_i)}$$

and

$$\sigma(g(x_i)) = \sqrt{g(x_i)}$$

the test statistics becomes:

$$\chi^2 = \sum_{i=1}^{\nu} \frac{(f(x_i) - g(x_i))^2}{f(x_i) + g(x_i)}$$

If the two distributions f and g are independent, the number of degrees of freedom is $\nu = n$. If the two distributions were normalised $\nu = n - 1$.
The asymptotic distribution is:

$$P(\chi^2 > \chi^2_{obs}) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \int_{\chi^2_{obs}}^{\infty} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \quad \mathrm{dx}$$

This test can be applied in the case of binned data. In case of unbinned distributions, data can always be grouped into bins. It must be noticed that in this case the in grouping operation will be lost part of the information contained in the unbinned data. For this reason this test can be applied almost in any case, being only limited by on condition: countings in each bin must be at least five. When this is not the case data must be regrouped until the minimum frequency per bin is not reached. Otherwise Yates' correction can be applied[1] [2].

[1] F.Yates, *Contingency tables involving small numbers and the χ^2 test*, J. Roy. Statist. Soc. Suppl., 1934.

[2] F.Yates, *Tests of significance for 2×2 contingency tables (with discussion)*, J. Roy. Statist. Soc. Series A, 1984.

$$\chi^2 = \sum_{i=1}^{n} \frac{[|f(x_i) - g(x_i)| - 0.5]^2}{\sigma^2(f(x_i)) + \sigma^2(g(x_i))}$$

Yates' correction is not yet contained in the Goodness-of-Fit Toolkit.