Reviewer's Report on the Paper "Non-parametric Goodness-of-Fit Tests for Discrete Null Distributions"

March 02, 2011

In this paper the authors present four non-parametric goodness-of-fit tests for discrete distributions: Kolmogorov-Smirnov test, and three Cramér-von Mises type tests.

Regarding the Kolmogorov-Smirnov test

Line 18 from bottom on page 2: the author said, " For large sample sizes, the null distribution can be approximated with the null distribution from the classical Kolmogorov-Smirnov test".

This statement is very ambiguous. What is the "null distribution from the classical Kolmogorov-Smirnov test"? Do the authors mean the corresponding test for continuous data? However, a discrete distribution function is near to a continuous function only if it has too many discontinuous points. So, if there are only a few discontinuous points for the null distribution function, the classical K-S test (for continuous data, as the reviewer understands) can not be applied, no matter how large the sample is. The authors should clarify this point.

Line 15 from bottom on page 2: When the sample size is small and an exact p-value is required, the authors' contribution is the implementation of the idea of Conover (1972) and the revision of the corresponding function *ks.test()*. The revised function *ks.test()* has obvious advantages over the existing version when the null distribution is discrete. However, it seems that the authors should mention that the number of discontinuous points should be small.

Regarding the three Cramér-von Mises Type tests

Asymptotically, the three tests have the general equivalent form of weighted sum of chi-squares:

$$Q = \sum_{i=1}^{p} \lambda_i \chi^2_{i,df=1},$$

and the p-value for a given value x of the test statistic is given by

$$\Pr\{Q \geq x\} = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin\theta(u,x)}{u\rho(u)}du$$

"for continuous function $\theta(\cdot, x)$ and $\rho(\cdot)$ depending on the weights λ_i". This result is due to Imhof (1961). The computation of the p-value is achieved numerically. This is fine in most situations. But when x is extremely large or the p-value is extremely small, the numerical process is instable, resulting weird results. To avoid the numerical instability, the authors' solution is to compute the upper bound of the p-value:

$$\Pr\left\{\chi^2_p \geq x/p\lambda_{\max}\right\} \equiv \Pr\left\{\lambda_{\max}\sum_{i=1}^{p}\chi^2_1 \geq x\right\} \geq \Pr\left\{\sum_{i=1}^{p}\lambda_i\chi^2_{i,df=1} \geq x\right\} \equiv \Pr\{Q \geq x\}.$$

According to the authors, this upper bound is a "conservative approximation".

The conservative approximation is useful when the p-value is extremely small, e.g. $< 10^{-5}$. However, for p-values in the range $(0.0001,\ 0.01)$, the authors should evaluate the performance of the conservative approximation by using the Monte-Carlo simulation. According to Figure 1 on page 3, it seems that the approximation may be poor in that range. In fact, in the computer age, one can easily use the Monte Carlo simulation to get a more accurate estimate for the p-value for a given value of x, (note the fact that all λ_i's are known!), unless the p-value is really tiny. The reviewer's experience is, when the p-value is larger than 10^{-4}, the Monte-Carlo simulation works very well.

In multiple testing, where a good portion of p-values are usually in the range $(0.00001, 0.001)$, it is very important to ensure the accuracy of the estimated p-values since it significantly impacts the set of null hypotheses to be rejected. The authors should consider how to improve the results for the p-values in that range. Perhaps, a more reasonable way to compute the p-value for a given value x of the test statistic is,

- to use the Imhof's method if the p-value is moderately large, (e.g., larger than 0.01);
- to apply the Monte-Carlo simulation if the p-value is in the range $(0.0001, 0.01)$; (This will give better results, unless the authors can give convincing evidence that this is not necessary.)
- to use the conservative approximation if the p-value is really tiny. (Monte Carlo simulation does not work and, the Imhof's method can not be applied due to numerical instability.)

Suggestions. The authors should revise the paper according to the above comments.