A Power Study of Goodness-of-Fit Tests for Categorical Data

Michael Steele
School of Mathematical and Physical Sciences, James Cook University,
Mike.Steele@jcu.edu.au

Janet Chaseling
Australian School of Environmental Studies, Griffith University

Cameron Hurst
School of Information Technology and Mathematical Sciences, University of Ballarat.

1. Introduction

Goodness-of-fit (GOF) tests are used for the analysis of categorical data by applied researchers from many disciplines however studies of their relative powers are limited. Although the Chi-Square (χ^2) test is a popular choice for many researchers, power studies show that this may be at the expense of power in some instances. This paper compares the powers of two of the lesser known GOF test statistics based on the empirical distribution function with the χ^2 test to determine which is the more powerful for the investigated null and alternative distributions.

2. The test statistics used in the power study

The test statistics used are χ^2 (Pearson 1900), the discrete Kolmogorov-Smirnov *KS* (Pettitt and Stephens 1977) and the discrete Cramér-von Mises W^2 (Choulakian *et al*. 1994).

$$(1)\quad \chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$(2)\quad KS = \max_{1 \le i \le k} |Z_i|$$

$$(3)\quad W^2 = N^{-1} \sum_{i=1}^{k} Z_i^2 p_i$$

where k is the number of cells, N is the sample size, p_i is the probability for cell i, O_i and E_i and are the observed and expected frequencies for cell i, and Z_i is the cummulative sum of the differences between O_i and E_i up to and including cell i.

3. The power study

The power for each test statistic is approximated for a uniform null distribution over 10 cells against the increasing trend and triangular ∨ or 'bath-tub' type alternatives defined in Table 1. The total sample sizes range from 10 to 200 which represents expected frquencies under the uniform null distribution of 1 to 20 per cell. The power of each test statistic is estimated at the 5% significance level from 10000 simulated random samples. The simulated distributions of the test statistic are discrete. To overcome that there may not be a unique test statistic at the required significance level of 5%, linear interpolation of the powers about this level is used for consistency.

Table 1. Distributions used in the power study.

Description	Cell Probability (2 Decimal Places)									
	1	2	3	4	5	6	7	8	9	10
Uniform	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Increasing	0.03	0.06	0.07	0.09	0.10	0.11	0.12	0.13	0.14	0.15
Triangular ∨	0.17	0.13	0.10	0.07	0.03	0.03	0.07	0.10	0.13	0.17

4. Results from the power study

The powers for the increasing alternative distribution are given in Figure 1. The powers for the triangular ∨ or 'bath-tub' type alternative are given in Figure 2. A summary of which of the test statistics have the higher power for the two alternatives is given in Table 2.

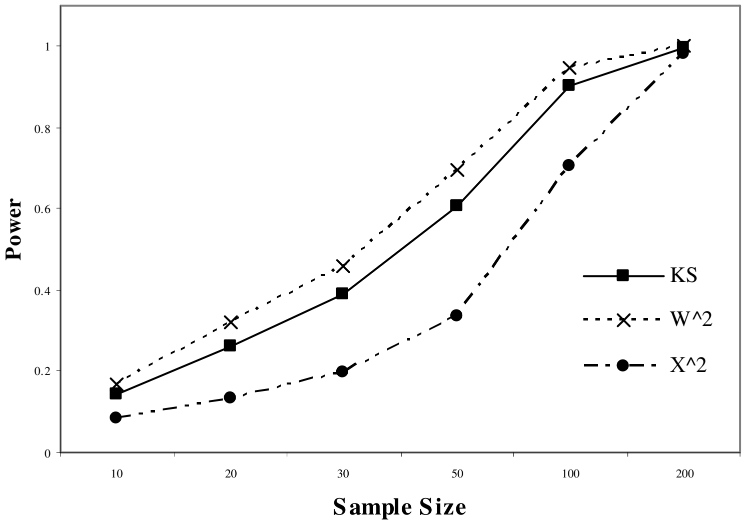

Figure 1. Powers for a uniform null and increasing alternative.

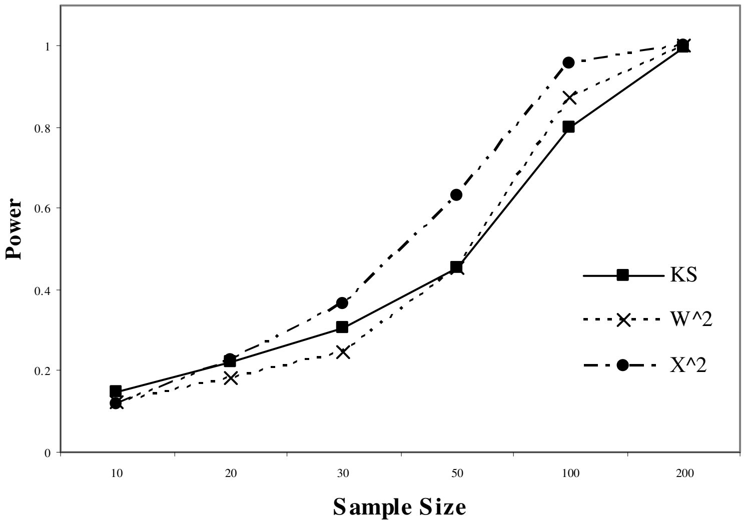

Figure 2. Powers for a uniform null and triangular ∨ or 'bath-tub' type alternative.

Table 2. Summary of the power of the three test statistics.

Alternative Distribution	General Ranking of Power from Highest to Lowest
Increasing	$W^2 > KS > \chi^2$
Triangular ∨	$\chi^2 > KS \approx W^2$

REFERENCES

Choulakian, V., Lockhart, R.A. and Stephens, M.A. (1994). Cramér-von Mises statistics for discrete distributions. The Canadian Journal of Statistics, 22 125-137.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, 5(50) 157-175.

Pettitt, A.N. and Stephens, M.A. (1977). The Kolmogorov-Smirnov goodness-of-fit statistic with discrete and grouped data. Technometrics, 19 205-210.

Steele, M.C. (2002). The power of categorical goodness-of-fit test statistics. Unpublished PhD Thesis, Griffith University.