Simulated Power of Discrete Goodness-of-Fit Tests for Likert Type Data

Steele, M. [1,2]**, C. Hurst** [3] **and J. Chaseling** [4]

[1] Faculty of Business, Technology & Sustainable Development, Bond University, Queensland
[2] Faculty of Health Science & Medicine, Bond University, Queensland
[3] Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland
[4] Griffith School of Environment, Griffith University, Queensland
Email: misteele@bond.edu.au

Keywords: *Goodness-of-fit, power, Likert, categorical*

EXTENDED ABSTRACT

Goodness-of-fit test statistics are widely used in surveys however little regard is given to the statistical power. This paper investigates the simulated power of a number of five categorical goodness-of-fit test statistics used on a 5-point Likert scale. The test statistics used in this power study are Pearson's Chi-Square, the Kolmogorov-Smirnov test statistic for discrete data, the Log-Likelihood Ratio, the Freeman-Tukey and the Power Divergence statistic with $\lambda=⅔$.

This paper aims to provide recommendations on which of these categorical goodness-of-fit test statistic is the most powerful overall and which is the most powerful for a uniform null distribution against alternative distributions with general shapes given in Figure 1.

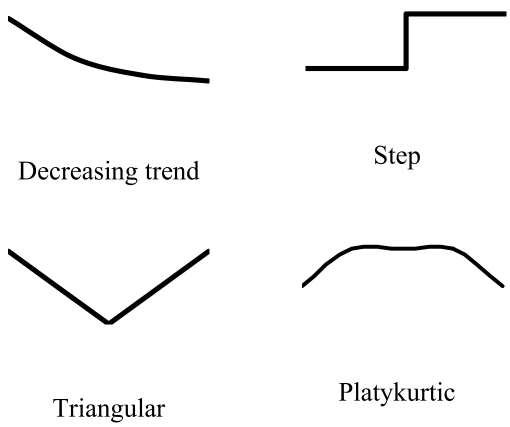

Figure 1. Type of alternative distributions used in the simulated power studies.

The results of the simulated power studies in this paper lead to the following conclusions for these goodness-of-fit test statistics when used on a 5-point Likert scale:

- for sample sizes less than six per cell under the uniform null distribution, the simulated powers of all four test statistics were very poor for all alternative distributions with the exception of the decreasing trend distribution
- the simulated power of the Freeman-Tukey test statistic is generally shown to be relatively less than the power of all the other investigated test statistics
- there is generally no improvement in the simulated power for the Power Divergence test statistic with $\lambda=⅔$ over either Pearson's Chi-Square or the Log-Likelihood Ratio test statistics.

1. INTRODUCTION

Goodness-of-fit test statistics for discrete data are used by many researchers from a very diverse range of disciplines however studies of their power are quite limited. The major contributors to the power of test statistics for discrete data have been made by Choulakian et al. (1994), Pettitt and Stephens (1977), From (1996) and Steele and Chaseling (2006). Although goodness-of-fit test statistics for discrete data are commonly used to analyse survey questions with data recorded on a Likert (1932) scale, published power studies for these situations are not available. The lack of published power studies and the general reliance of the Chi-Square test statistic (Pearson, 1900) in the research community means that in many situations a more powerful test statistic might be available.

This paper compares the simulated power of Pearson's Chi-Square test statistic (1), the Kolmogorov-Smirnov test statistic for discrete data (2) and the three Chi-Square type test statistics Log-Likelihood Ratio (3) (Wilks, 1935), Freeman-Tukey (4) (Freeman and Tukey, 1950) and the Power Divergence statistic with λ=⅔ (5) (Cressie and Read 1984). This particular Power Divergence statistic was identified by Cressie and Read (1984) as an alternative to Pearson's Chi-Square.

$$X^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i}, \tag{1}$$

$$KS = \max_{1 \le i \le k} \left| O_i - E_i \right|, \tag{2}$$

$$G = 2\sum_{i=1}^{k} O_i \ln\left(\frac{O_i}{E_i}\right), \tag{3}$$

$$FT = 4\sum_{i=1}^{k}\left(\sqrt{O_i} - \sqrt{E_i}\right)^2, \tag{4}$$

$$PD = \frac{2}{\lambda\left(1+\lambda\right)}\sum_{i=1}^{k} O_i\left[\left(\frac{O_i}{E_i}\right)^{\lambda} - 1\right], \tag{5}$$

where k is the number of cells, O_i is the observed frequency in cell i, E_i is the expected frequency in cell i.

For a number of sample sizes the simulated powers are calculated on a 5-point Likert scale (k=5) with a uniform null distribution against four alternative distributions given in Table 1.

Table 1. Distributions used in the power study.

Distribution	Cell Probabilities				
	1	2	3	4	5
Uniform	0.20	0.20	0.20	0.20	0.20
Decreasing	0.45	0.19	0.14	0.12	0.10
Step	0.15	0.15	0.20	0.25	0.25
Triangular	0.27	0.18	0.10	0.18	0.27
Platykurtic	0.095	0.27	0.27	0.27	0.095

In Section 2 the calculation of the simulated power is discussed and the results of the power studies for each alternative distribution are given in Section 4. Concluding comments are made in Section 4 on which are the most powerful of these five test statistics for Likert type data.

2. CALCULATION OF THE SIMULATED POWER

For consistency this power study uses similar null and alternative distributions and sample size to those used by Steele and Chaseling (2006). The major difference being that the number of cells is five as is common in Likert type data. The sample sizes are 10, 20, 30, 50, 100 and 200, representing 2, 4, 6, 10, 20 and 40 observations per cell under the uniform null distribution.

The power of each test statistic is estimated from 10000 simulated random samples. In some cases the discrete nature of the data precludes a critical value at the nominated 5% level of significance. In such cases linear interpolation is used to enable meaningful power comparison.

3. RESULTS OF THE POWER STUDY

3.1. Decreasing alternative

Figure 2 confirms the common recommendations made in the power studies identified in Section 1 that Empirical Distribution Function (EDF) test statistics such as the discrete Kolmogorov-Smirnov are generally more powerful for these trend type alternative distributions. Although all five test statistics are shown to have very high simulated power for the larger sample sizes there are some other differences identified. It is clear for this alternative distribution that the Freeman-Tukey test statistic has relatively lower power than the other four test statistics and that the Log-Likelihood Ratio has slightly lower power than Pearson's Chi-Square and the Power Divergence test with λ=⅔. Based on these results for a 5-point Likert scale it appears that the discrete Kolmogorov-Smirnov is more powerful at identifying the trend alternative should one exist.

Should a Chi-Square type test statistic be desired, then either the Power Divergence with λ=⅔ or Pearson's Chi-Square should be used with higher power.

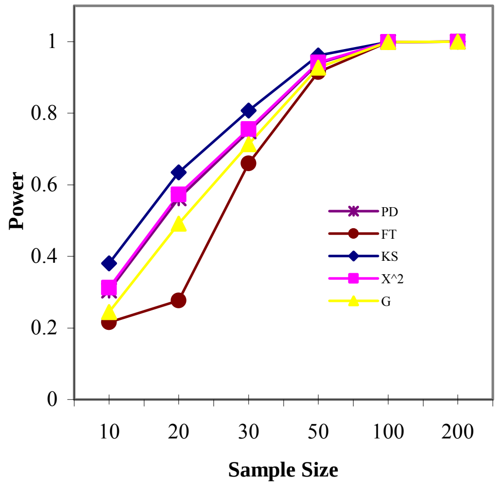

Figure 2. Simulated power for a uniform null against a decreasing alternative distribution.

3.2. Step alternative

As was shown to occur with the decreasing alterative in Section 3.1 the power of the Kolmogorov-Smirnov test statistic is shown in Figure 3 to be greater than the power of the other four test statistics for this step type increasing alternative distribution. Although the powers are calculated for a 5-point Likert scale it again agrees with the general comments about EDF test statistics when testing a uniform null against a trend type alternative distribution. It is interesting to note that all of the Chi-Square type test statistics produce approximately the same power. Such results were not observed in the more comprehensive power study of Chi-Square type test statistics with 10 cells by Steele and Chaseling (2007). It is important to note that even for the sample size of 50, that is 10 observations per cell under the uniform null distribution, the power of all five test statistics is below 0.35 which indicates that a very large sample size is required to correctly reject this uniform null in favour of this step type alternative when using a Likert scale.

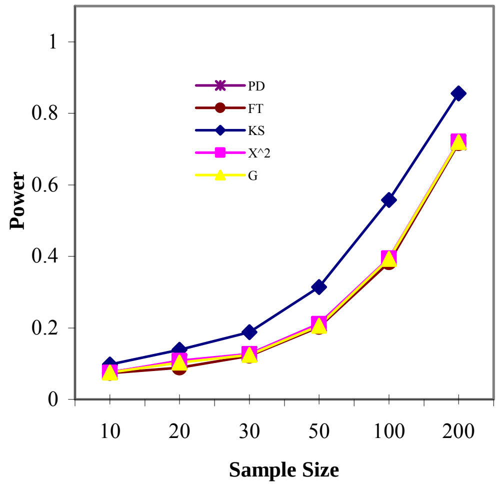

Figure 3. Simulated power for a uniform null against a step alternative distribution.

3.3. Triangular alternative

For this particular triangular alternative distribution Figure 4 shows that the powers of all the Chi-Square type test statistics are generally much higher than the discrete Kolmogorov-Smirnov test statistic. As was the case with the step type alternative in Section 3.2 the power of all the test statistics were very low even for sample sizes as high as 10 per cell under the uniform null distribution. Clearly a very large sample size is required to increase to power of these test statistics for a triangular alternative distribution when the data are collected on a 5-point Likert scale.

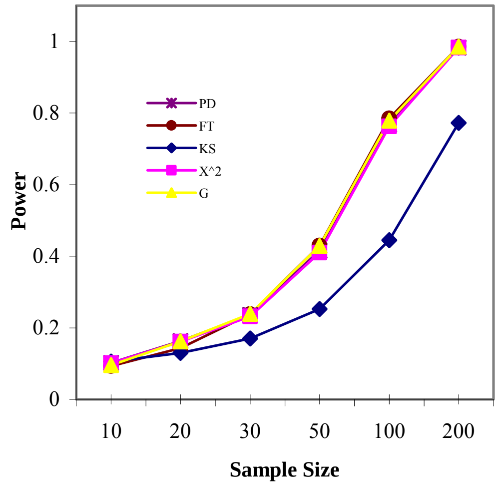

Figure 4. Simulated power for a uniform null against a triangular alternative distribution.

3.4. Platykurtic alternative

Figure 5 shows that the powers of all four Chi-Square type test statistics are relatively similar, and quite low for sample sizes up to and including 6 per cell under the uniform null distribution. The power of the discrete Kolmogorov-Smirnov only becomes competitive for the very large sample size of 40 per cell under the uniform null distribution.

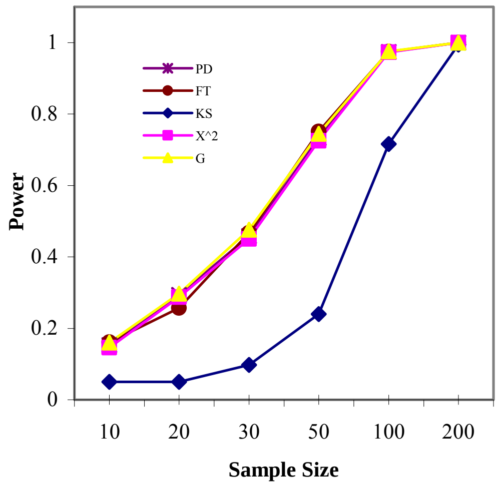

Figure 4. Simulated power for a uniform null against a triangular alternative distribution.

4. CONCLUSIONS

As shown by Steele and Chaseling (2006), albeit for a larger number of cells, it is difficult to make general recommendations as to the most powerful goodness-of-fit test statistic for the specific alternative distributions used in this study. Given the widespread use in the research community of Pearson's Chi-Square for Likert type data some comments relating to power are needed:

- for sample sizes less than six per cell under the uniform null distribution, the simulated powers of all four test statistics were very poor for all alternative distributions with the exception of the decreasing trend distribution
- the simulated power of the Freeman-Tukey test statistic is generally shown to be relatively less than the power of all the other investigated test statistics
- there is generally no improvement in the simulated power for the Power Divergence test statistic with $\lambda=2/3$ over either Pearson's Chi-Square or the Log-Likelihood Ratio test statistics.

5. REFERENCES

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

Cressie, N. and T.R.C. Read (1984), Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society Series B*, 46, 440-464.

Freeman, M.F. and J.W. Tukey (1950), Transformations related to the angular and the square root, *Annals of Mathematical Statistics*, 21, 607-611.

From, S.G. (1996), A new goodness of fit test for the equality of multinomial cell probabilities verses trend alternatives, *Communications in Statistics-Theory and Methods*, 25, 3167-3183.

Likert, R. (1932), A technique for the measurement of attitudes, *Archives of Psychology*, 140, 1-55.

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 Series 5*, 50, 157-175.

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

Steele, M. and J. Chaseling (2006), Powers of discrete goodness-of-fit test statistics for a uniform null against a selection of alternative distributions, *Communications in Statistics-Simulation and Computation*, 35, 1067-1075.

Steele, M., C. Hurst and J. Chaseling (2007), The power of Chi-Square type goodness-of-fit test statistics, *Communications in Statistics-Simulation and Computation*, (Under review).

Wilks, S.S (1935), The likelihood test of independence in contingency tables, *Annals of Mathematical Statistics*, 6, 190-196.