

Limited-Information Statistics When the Number of Variables is Large

Mark Reiser

School of Mathematical and Statistical Sciences

Arizona State University, USA *

Abstract

The Pearson and likelihood ratio statistics are commonly used to test goodness of fit for models applied to data from a multinomial distribution. When data are from a table formed by the cross-classification of a large number of variables, the common statistics may have low power and inaccurate Type I error level due to sparseness. Several statistics have been proposed that use components of the Pearson statistic obtained from marginal distributions. These statistics have mostly been applied to item response models or factor analysis of categorical variables and have very good performance for Type I error rate and power when the data table is formed from a moderate number of variables. However, there are limitations when the number of variables becomes larger than 20. This paper compares the performance of statistics based on marginal distributions as well as computational resources required when the number of variables is large. The comparison includes test statistics from Christoffersson (1975), Reiser (1996, 2008), Bartholomew and Leung (2002), Tollenaar and Mooijaart (2003), and Maydeu-Olivares and Joe (2005).

Key Words: multivariate discrete distribution, overlapping cells, orthogonal components, composite null hypothesis

1. Introduction

The goodness-of-fit test based on Pearson's chi-squared statistic is sometimes considered to be an omnibus test that gives little guidance to the source of poor fit when the null hypothesis is rejected. It has also been recognized that the omnibus test can often be outperformed by focused or directional tests of lower order.

The Pearson and likelihood ratio statistics are commonly used to test goodness of fit for models applied to data from a multinomial distribution. When data are from a table formed by the cross-classification of a large number of variables, the common statistics may have low power and inaccurate Type I error level due to sparseness. Several statistics have been proposed that use components of the Pearson statistic obtained from marginal distributions. These statistics have mostly been applied to item response models or factor analysis of categorical variables and have very good performance for Type I error rate and power when the data table is formed from a moderate number of variables. However, there are limitations when the number of variables becomes larger than 20. This paper compares the performance of statistics based on marginal distributions as well as computational resources required when the number of variables is large. The comparison includes test statistics from Christoffersson (1975), Reiser (1996, 2008), Bartholomew and Leung (2002), Tollenaar and Mooijaart (2003), and Maydeu-Olivares and Joe (2005).

2. Marginal Proportions

This section includes a presentation of transformations from joint proportions or frequencies to marginal proportions as a prelude to testing a model based on the fit to marginal

*Requests for reprints should be addressed to Mark Reiser, School of Mathematical and Statistical Sciences, Arizona State University, Box 871804, Tempe, Arizona, 85287-1804, USA (e-mail: mark.reiser@asu.edu).

frequencies.

2.1 First- and Second-Order Marginals

The relationship between joint proportions and first- and second-order marginals can be shown by using zeros and 1's to code the levels of dichotomous response variables. Then, a q -dimensional vector of zeros and 1's, sometimes called a response pattern, will indicate a specific cell from the contingency table formed by the cross-classification of q response variables. A T -dimensional set of response patterns can be generated by varying the levels of the q^{th} variable most rapidly, the $q^{th} - 1$ variable next, etc. Define \mathbf{V} as the T by q matrix with response patterns as rows.

For $q = 3$,

$$\mathbf{V} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} .$$

Let v_{is} represent element i of response pattern s , and let \mathbf{Y} be a vector of dichotomous variables. Also, define $\boldsymbol{\theta}$ as a parameter vector for a model of interest and $\pi_s(\boldsymbol{\theta})$ as the expected proportion for cell s as a function of the parameter vector $\boldsymbol{\theta}$. Then, under the model, the first-order marginal proportion for variable Y_i can be defined as

$$P_i(\boldsymbol{\theta}) = \text{Prob}(Y_i = 1 | \boldsymbol{\theta}) = \sum_s v_{is} \pi_s(\boldsymbol{\theta}),$$

and the true first-order marginal proportion is given by

$$P_i = \text{Prob}(Y_i = 1) = \sum_s v_{is} \pi_s .$$

The summation across the frequencies associated with the response patterns to obtain the marginal proportions represents a linear transformation of the frequencies in the multinomial vector $\boldsymbol{\pi}$ which can be implemented via multiplication by a certain matrix, denoted here generically by the symbol \mathbf{H} . The symbol $\mathbf{H}_{[t]}$ denotes the transformation matrix that would produce marginals of order t . The symbol $\mathbf{H}_{[t: u]}$, $t \leq u \leq q$, denotes the transformation matrix that would produce marginals from order t up to and including order u . Furthermore, $\mathbf{H}_{[t]} \equiv \mathbf{H}_{[t: t]}$, and $\mathbf{H} \equiv \mathbf{H}_{[t: u]}$. There will be occasions to delete certain rows from the matrix $\mathbf{H}_{[t: u]}$ due to collinearity, and the symbol $\mathbf{H}_{[t: u], -d}$ denotes the matrix $\mathbf{H}_{[t: u]}$ with d rows deleted.

Matrix $\mathbf{H}_{[1]}$ can be defined from matrix \mathbf{V} such that

$$\mathbf{H}_{[1]} = \mathbf{V}' .$$

Under the model, the second-order marginal proportion for variables Y_i and Y_j can be defined as

$$P_{ij}(\boldsymbol{\theta}) = \text{Prob}(Y_i = 1, Y_j = 1 | \boldsymbol{\theta}) = \sum_s v_{is} v_{js} \pi_s(\boldsymbol{\theta}),$$

and the true second-order marginal proportion is given by

$$P_{ij} = \text{Prob}(Y_i = 1, Y_j = 1) = \sum_s v_{is} v_{js} \pi_s .$$

For second-order marginals, where $j = 1, 2, \dots, k$; $i = j + 1, \dots, q$; $s = 1, 2, \dots, T$; and $\ell = i - j + \sum_{0 < r < j} (q - r)$, element ℓs of $\mathbf{H}_{[2]}$ is given by

$$[\mathbf{H}_{[2]}]_{\ell s} = \begin{cases} 1 & \text{if } v_{is} = v_{js} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Alternatively, matrix $\mathbf{H}_{[2]}$ can be defined by forming Hadamard products among the columns of the matrix \mathbf{V} :

$$\mathbf{H}_{[2]} = \begin{pmatrix} (\mathbf{v}_1 \circ \mathbf{v}_2)' \\ (\mathbf{v}_1 \circ \mathbf{v}_3)' \\ \vdots \\ (\mathbf{v}_1 \circ \mathbf{v}_q)' \\ (\mathbf{v}_2 \circ \mathbf{v}_3)' \\ \vdots \\ (\mathbf{v}_2 \circ \mathbf{v}_q)' \\ \vdots \\ (\mathbf{v}_{q-1} \circ \mathbf{v}_q)' \end{pmatrix},$$

where \mathbf{v}_f represents column f of matrix \mathbf{V} , and $\mathbf{v}_f \circ \mathbf{v}_g$ represents the Hadamard product of columns f and g .

2.2 Higher-Order Marginals

The third-order marginal proportions for variables Y_i , Y_j , and Y_k can be obtained by employing the matrix $\mathbf{H}_{[3]}$, which can also be defined as Hadamard products among the columns of \mathbf{V} ,

$$\mathbf{H}_{[3]} = \begin{pmatrix} (\mathbf{v}_1 \circ \mathbf{v}_2 \circ \mathbf{v}_3)' \\ (\mathbf{v}_1 \circ \mathbf{v}_2 \circ \mathbf{v}_4)' \\ \vdots \\ (\mathbf{v}_1 \circ \mathbf{v}_2 \circ \mathbf{v}_q)' \\ (\mathbf{v}_2 \circ \mathbf{v}_3 \circ \mathbf{v}_4)' \\ \vdots \\ (\mathbf{v}_2 \circ \mathbf{v}_3 \circ \mathbf{v}_q)' \\ \vdots \\ (\mathbf{v}_{q-2} \circ \mathbf{v}_{q-1} \circ \mathbf{v}_q)' \end{pmatrix},$$

and then, for example,

$$\mathbf{H}_{[1:3]} = \begin{pmatrix} \mathbf{H}_{[1]} \\ \dots \\ \mathbf{H}_{[2]} \\ \dots \\ \mathbf{H}_{[3]} \end{pmatrix}.$$

For $q = 3$,

$$\mathbf{H}_{[1:3]} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ \dots & & & & & & & \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \dots & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} .$$

A general matrix $\mathbf{H}_{[t:u]}$ to obtain marginals of any order can be defined in a similar fashion by using Hadamard products among the columns of \mathbf{V} . $\mathbf{H}_{[1:q]}$ gives a mapping from joint proportions to the entire set of $(2^q - 1)$ marginal proportions:

$$\mathbf{P} = \mathbf{H}_{[1:q]} \boldsymbol{\pi} ,$$

where

$$\mathbf{P} = (P_1 \ P_2 \ P_3 \ \dots \ P_q \ P_{12} \ P_{13} \ \dots \ P_{q-1,q} \ P_{112} \ \dots \ P_{q-2,q-1,q} \ \dots \ P_{123\dots q})'$$

is the vector of marginal proportions

2.3 Residuals

Define the unstandardized residual $r_s = \hat{p}_s - \pi_s(\hat{\boldsymbol{\theta}})$, where

$\hat{p}_s = \frac{n_s}{n}$ is element s of $\hat{\mathbf{p}}$, the vector of multinomial proportions,

n_s = element s of \mathbf{n} , the vector of observed frequencies,

n = total sample size = $\sum_{s=1}^T n_s$,

$\hat{\boldsymbol{\theta}}$ = parameter estimator vector,

$\pi_s(\hat{\boldsymbol{\theta}})$ = estimated expected proportion for cell s ,

and denote the vector of unstandardized residuals as \mathbf{r} with element r_s .

A vector of simple residuals for marginals of any order may be defined such that

$$\mathbf{e} = \mathbf{H}(\hat{\mathbf{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) = \mathbf{H}\mathbf{r},$$

and a vector, $\boldsymbol{\xi}$, of differences between the marginals specified by the relevant model and the true population marginals may be defined for marginals of any order such that

$$\boldsymbol{\xi} = \mathbf{H}(\boldsymbol{\pi} - \boldsymbol{\pi}(\boldsymbol{\theta})).$$

3. Testing Fit on Marginal Distributions

3.1 Linear Combinations of Joint Frequencies

A traditional test of fit for of a multinomial model uses the null hypothesis $H_o: \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$, where $\boldsymbol{\pi}(\boldsymbol{\theta})$ is a vector of multinomial probabilities as a function of $\boldsymbol{\theta}$. Linear combinations of $\boldsymbol{\pi}$ may be tested under the null hypothesis $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$, or equivalently

$H_o: \boldsymbol{\xi}_{[t:u]} = \mathbf{0}$. \mathbf{H} may specify linear combinations that form marginal proportions as defined in the previous section. If \mathbf{H} is full rank, then $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ become equivalent to $H_o: \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$. If $\boldsymbol{\pi}$ has less than full rank, then $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ becomes specifies a test of components of the Pearson statistic, and the relationship between these null hypotheses works as follows: “Reject $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ ” is a sufficient but not necessary condition for “Reject $H_o: \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ ”; “Do not reject $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ ” is a necessary but not sufficient condition for “Do not reject $H_o: \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ ”. “Do not reject $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ ” is not a sufficient condition because it is possible that lack of fit may be manifest only in marginals of higher order than u , where u is defined as in $\mathbf{H}_{[t:u]}$. The null hypothesis $H_o: \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ can be partitioned into components corresponding to hypotheses $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$, with corresponding partition of degrees of freedom.

3.2 Test Statistics

When the model parameters $\boldsymbol{\theta}$ are unknown and estimated, the null hypothesis $H_o: \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ is often tested with the Pearson-Fisher statistic:

$$X_{PF}^2 = n \sum_s z_s^2,$$

where

$$z_s = (\pi_s(\hat{\boldsymbol{\theta}}))^{-\frac{1}{2}} (\hat{p}_s - \pi_s(\hat{\boldsymbol{\theta}})).$$

A test of $H_o: \mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\theta})$ may be considered to be a “limited-information” test since the full information in the joint distribution of \mathbf{Y} is not entirely employed. The original limited-information statistic with features of cell collapsing and cell focusing was given by Christofferson (1975). In our notation, this statistic would be written as

$$X_{Ch}^2 = \bar{\mathbf{r}}' \mathbf{H}'_{[1:2]} (D(\hat{\mathbf{p}}) - \hat{\mathbf{p}}\hat{\mathbf{p}}')^{-1} \mathbf{H}_{[1:2]} \bar{\mathbf{r}},$$

where $\bar{\mathbf{r}}$ is the residual calculated using a generalized least squares estimator of $\boldsymbol{\theta}$. X_{Ch}^2 has an asymptotic-square distribution with $2^q - g$ degrees of freedom, where g = number of model parameters to be estimated. The statistic could be generalized to include higher-order marginals, but even if marginals from first- to order q were included, this statistic would not be equivalent to the Pearson-Fisher statistic. Muthén (1978) developed a modified version of Christofferson’s statistic.

Reiser(1996) and Reiser and Lin (1999) proposed the following statistic. Let $\boldsymbol{\Sigma}_e$ represent the covariance matrix of the residuals, \mathbf{e} , where $\mathbf{e} = \mathbf{H}_{[t:u]} \mathbf{r}$. Using the matrix $\mathbf{H}_{[t:u]}$ as given above,

$$X_{[t:u]}^2 = \mathbf{e}' \hat{\boldsymbol{\Sigma}}_e^{-1} \mathbf{e}$$

where

$$\hat{\boldsymbol{\Sigma}}_e = n^{-1} \boldsymbol{\Omega}_e,$$

with $\boldsymbol{\Omega}_e$ evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\theta}}$, and where

$$\begin{aligned} \boldsymbol{\Omega}_e &= \mathbf{H}(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')\mathbf{H}', \\ D(\boldsymbol{\pi}) &= \text{diagonal matrix with } (s, s) \text{ element equal to } \pi_s(\boldsymbol{\theta}), \\ \mathbf{A} &= D(\boldsymbol{\pi})^{-1/2} \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\ \text{and } \mathbf{G} &= \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \end{aligned}$$

$\Omega_{\mathbf{r}} = (D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}')$ is the asymptotic covariance matrix of \mathbf{r} . Matrices are evaluated with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, which may be the maximum likelihood estimator.

The limiting distribution of $X_{[t:u]}^2$ as $n \rightarrow \infty$ can be shown to be the χ^2 -distribution because \mathbf{e} is a linear combination of the elements of \mathbf{r} , $n\hat{\boldsymbol{\Sigma}}\mathbf{e} \xrightarrow{P} \Omega_{\mathbf{e}}$, and $\mathbf{e} \xrightarrow{L} MVN(\boldsymbol{\xi}, \boldsymbol{\Sigma}_{\mathbf{e}})$. The regularity conditions for the asymptotic chi-square distribution are given by Birch (1964). $X_{[t:u]}^2$ can be seen as a special case of the score statistic given in Theorem 7.1.1 of Rayner and Best (1989).

The degrees of freedom are determined by the rank of $\Omega_{\mathbf{e}}$. In general $X_{[2]}^2$ will have degrees of freedom $\leq \min(2^g - g, 0.5q(q - 1))$, where g is the number of estimated parameters. Some model parameterizations may reduce the rank of $\Omega_{\mathbf{e}}$ and hence the degrees of freedom for $X_{[t:u]}^2$ in general. Any linear dependency among the rows of \mathbf{H} and the columns of \mathbf{G} will produce a marginal or sum of marginals that is perfectly fit when calculating $X_{[t:u]}^2$ and will lead to the loss of a degree of freedom. A second-order marginal that is perfectly fit under the model will reduce the degrees of freedom for $X_{[t:u]}^2$ by 1 if $t \leq 2 \leq u$. The statistic has been extended to ordinal response variables by Cagnone and Mignani (2007). Reiser (2008) defined orthogonal components of the Pearson statistic using this approach.

Bartholomew and Leung (2002) proposed the statistic Y based on only second order marginals:

$$Y = \mathbf{e}'_{[2]} \mathbf{D}_{[2]}^{-1} \mathbf{e}_{[2]},$$

where $\mathbf{e}_{[2]} = \mathbf{H}_{[2]}(\hat{\mathbf{p}} - \boldsymbol{\pi})$ and $\mathbf{D}_{[2]} = \text{diag}(\mathbf{H}_{[2]}\boldsymbol{\pi})(\mathbf{I} - \text{diag}(\mathbf{H}_{[2]}\boldsymbol{\pi}))$. Bartholomew and Leung gave a chi-square approximation for the distribution of

$$\frac{Y - a}{b}$$

on c degrees of freedom, where a, b and c are functions of the asymptotic moments of Y :

$$b = \frac{\mu_3(Y)}{4\mu_2(Y)}, \quad c = \frac{\mu_2(Y)}{2b^2}, \quad a = \mu_1(Y) - bc.$$

The statistic was presented in terms of known $\boldsymbol{\pi}$, but in an application, $\boldsymbol{\pi}$ is replaced by probabilities estimated from the model under consideration. In the original form, the statistic is simpler to calculate because it only requires estimates for $\boldsymbol{\pi}$. Cai, Maydeu-Olivares, Coffman and Thissen (2006) found that this statistic does not perform well with the degrees of freedom given by Bartholomew and Leung, and they proposed a modified version of the statistic, Y_2 , using both first- and second-order marginals, and revised degrees of freedom:

$$Y_2 = \mathbf{e}' \mathbf{D}_{[1:2]}^{-1} \mathbf{e},$$

where $\mathbf{e} = \mathbf{H}_{[1:2]}\mathbf{r}$ and $\mathbf{D}_{[1:2]} = \text{diag}(\mathbf{H}_{[1:2]}\boldsymbol{\pi}(\hat{\boldsymbol{\theta}}))(\mathbf{I} - \text{diag}(\mathbf{H}_{[1:2]}\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})))$, and $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. Since calculation of $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}$ is required for determination of the revised degrees of freedom, there is little computational advantage compared to $X_{[t:u]}^2$.

Tollenaar and Mooijaart (2003) proposed $X_{red}^2 = \mathbf{e}' \hat{\boldsymbol{\Gamma}}^{-1} \mathbf{e}$, where $\hat{\boldsymbol{\Gamma}} = D(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})'$. X_{red}^2 is a “reduced” version of $X_{[t:u]}^2$ in that the covariance matrix $\boldsymbol{\Gamma}$ does not include the term $\mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$. As noted by Tollenaar and Mooijaart, omitting this term may substantially reduce computations, depending on the model under investigation. Since X_{red}^2 and $X_{[t:u]}^2$ have different covariance matrices, the degrees of freedom are different. X_{red}^2 has an asymptotic-square distribution with $2^g - g$ degrees of freedom, where $g =$

number of model parameters to be estimated, so the chi-square approximation for the distribution of X_{red}^2 has the same degrees of freedom as the distribution of X_{Ch}^2 . X_{red}^2 can also be viewed as a similar form to Christofferson’s statistic, but with different estimators. X_{red}^2 uses the maximum likelihood estimator for $\boldsymbol{\theta}$ and $\boldsymbol{\pi}(\boldsymbol{\theta})$ instead of the GLS estimator, and $\widehat{\boldsymbol{\Gamma}}$ uses the maximum likelihood estimator $\widehat{\boldsymbol{\pi}}(\boldsymbol{\theta})$ instead of the observed proportions. Fitted versus observed proportions is a key difference. $\widehat{\boldsymbol{\Gamma}}$ is consistent for $\boldsymbol{\Sigma}$ as $n \rightarrow \infty$ even under the sparseness condition on the frequencies.

Maydeu-Olivares and Joe (2005) developed a statistic, M_r , that is closely related to $X_{[1:r]}^2$. The two statistics are not equivalent, however, due to a different covariance matrix in the quadratic form. For $\mathbf{e} = \mathbf{H}_{[1:r]}\mathbf{r}$,

$$M_r = \mathbf{e}'\widehat{\mathbf{C}}_r\mathbf{e}$$

where $\widehat{\mathbf{C}}_r = (\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}')^{-1} - (\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}')^{-1}\mathbf{H}\widehat{\mathbf{G}}(\widehat{\mathbf{G}}'\mathbf{H}'(\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}')^{-1}\mathbf{H}\widehat{\mathbf{G}})^{-1}\widehat{\mathbf{G}}'\mathbf{H}'(\mathbf{H}\widehat{\boldsymbol{\Gamma}}\mathbf{H}')^{-1}$, and where $\widehat{\boldsymbol{\Gamma}} = D(\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})) - \boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}})'$. \mathbf{H} is always equal to $\mathbf{H}_{[1:r]}$ when applied to the definition of M_r . $\widehat{\boldsymbol{\Sigma}}_e$ is a generalized inverse of $\widehat{\mathbf{C}}_r$. Because $\widehat{\mathbf{C}}_r$ appears in the quadratic form, the degrees of freedom for M_r do not match degrees of freedom for $X_{[1:r]}^2$, when $r < q$, and the two statistics are not equivalent under that condition. M_r has an asymptotic-square distribution with $2^q - g$ degrees of freedom, where $g =$ number of model parameters to be estimated, so the distribution of M_r has the same degrees of freedom as the distribution of X_{Ch}^2 and X_{red}^2 . $X_{[t:u]}^2$ is more general in two ways. First, the possibility that the statistic will not include some marginals of an entire lower order is allowed because power of a test may be reduced by including them. Second, Maydeu-Olivares and Joe adopt a condition whereby their statistic does not apply to certain circumstances in which $X_{[t:u]}^2$ would apply. This condition is $r \geq r_0$, where r_0 is the smallest integer r such that the model is (locally) identified from the joint moments up to order r . See Reiser (2008). The statistic has been extended to ordinal response variables by Maydeu-Olivares and Joe (2006).

4. Orthogonal Components

Consider the $T - g - 1$ by 2^q matrix $\mathbf{H}^* = \mathbf{F}'\mathbf{H}_{[1:q;-g]}$. \mathbf{H}^* has full row rank. \mathbf{F} is the upper triangular matrix such that $\mathbf{F}'\boldsymbol{\Omega}_e\mathbf{F} = \mathbf{I}$. $\mathbf{F} = (\mathbf{C}')^{-1}$, where \mathbf{C} is the Cholesky factor of $\boldsymbol{\Omega}_e$. Premultiplication by $(\mathbf{C}')^{-1}$ orthonormalizes the matrix $\mathbf{H}_{[1:q;-g]}$ in the matrix $D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$.

$$X_{PF}^2 = X_{[1:q;-g]}^2 = n^{-1}\mathbf{r}'(\widehat{\mathbf{H}}^*)'\widehat{\mathbf{H}}^*\mathbf{r}$$

where $\widehat{\mathbf{H}}^* = \mathbf{H}^*(\widehat{\boldsymbol{\theta}})$.

Define

$$\hat{\boldsymbol{\gamma}} = n^{-\frac{1}{2}}\widehat{\mathbf{F}}'\mathbf{H}\mathbf{r} = n^{-\frac{1}{2}}\widehat{\mathbf{H}}^*\mathbf{r}$$

where $\widehat{\mathbf{F}}$ is the matrix \mathbf{F} evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. Then

$$X_{PF}^2 = \hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}} = \sum_{j=1}^{j=T-g-1} \hat{\gamma}_j^2$$

$\widehat{\mathbf{H}}^*\mathbf{r}$ has asymptotic covariance matrix $\mathbf{F}'\boldsymbol{\Omega}_e\mathbf{F} = \mathbf{I}_{T-g-1}$. The elements $\hat{\gamma}_j^2$ are asymptotically independent χ_1^2 random variables (Reiser, 2008).

The orthogonal components can be obtained as the sequential sum of squares from a weighted orthogonal regression as follows. Define

$$z_s = (\pi_s(\widehat{\boldsymbol{\theta}}))^{-\frac{1}{2}}(\hat{p}_s - \pi_s(\widehat{\boldsymbol{\theta}})).$$

The the regression of z_s on the columns of \mathbf{H} is

$$\mathbf{z} = \mathbf{H}\boldsymbol{\beta},$$

and

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}\widehat{\mathbf{W}}\mathbf{H}')^{-1}\mathbf{H}\widehat{\mathbf{W}}\mathbf{z}$$

where $\widehat{\mathbf{W}} = \widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}^{-\frac{1}{2}} = \widehat{\mathbf{D}}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}^{-\frac{1}{2}}$, and $\mathbf{D} = \text{diag}(\boldsymbol{\pi}(\boldsymbol{\theta}))$. $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = (\mathbf{I} - \boldsymbol{\pi}^{\frac{1}{2}}(\boldsymbol{\pi}^{\frac{1}{2}})' - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')$ is idempotent. There is no error term in the regression if \mathbf{H} contains columns representing marginals of all possible degree.

Let $\widehat{\mathbf{M}} = \widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}'$. Then

$$\hat{\boldsymbol{\beta}} = (\widehat{\mathbf{M}}'\widehat{\mathbf{M}})^{-1}\widehat{\mathbf{M}}'\mathbf{z}$$

$\hat{\boldsymbol{\gamma}}_j^2$, $j = 1, T - g - 1$ are the sequential SS from this regression. $\boldsymbol{\gamma} = \mathbf{C}'\boldsymbol{\beta}$ are the orthogonal coefficients. Components obtained as sequential sum of squares from the SWEEP operator (Goodnight, 1978; SAS PROC REG) are very accurate numerically when pivot values are checked for singularity. SAS PROC REG uses the tolerance value $1E^{-7}*\text{CSS}$, where CSS is the corrected sum of squares for the regressor, to check for singularities.

4.1 The $GF\text{fit}^{(ij)}$ Statistic

The matrices \mathbf{V} and \mathbf{H} defined earlier can be extended when the number of categories for observed variables is 3 or more. For example, 3 variables with 3 categories $\mathbf{H}_{[1]} = \mathbf{V}'$, where

$$\mathbf{V}_{27 \times 6} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

For higher-order marginal proportions, the columns of \mathbf{H} are again Hadamard products among the columns of \mathbf{V} . If $q = 3$ and $c = 3$, \mathbf{H} is an 18 by 27 matrix:

$$\mathbf{H}_{[2]} = \begin{pmatrix} (\mathbf{v}_1 \circ \mathbf{v}_3)' \\ (\mathbf{v}_1 \circ \mathbf{v}_4)' \\ \vdots \\ (\mathbf{v}_1 \circ \mathbf{v}_5)' \\ (\mathbf{v}_1 \circ \mathbf{v}_6)' \\ \vdots \\ (\mathbf{v}_3 \circ \mathbf{v}_5)' \\ \vdots \\ (\mathbf{v}_{i(c-1)} \circ \mathbf{v}_{j(c-1)})' \end{pmatrix}$$

Joreskog and Moustaki (2001) defined

$$GFfit^{(ij)} = n \Sigma_{ab} \frac{(\hat{p}_{ab}^{(ij)} - \hat{\pi}_{ab}^{(ij)})^2}{\hat{\pi}_{ab}^{(ij)}}$$

where $i = 1, \dots, q - 1$, $j = i + 1, \dots, q$; $a = 1, \dots, c$; $b = 1, \dots, c$. Define

$$\mathbf{H}_{[2]}^{(ij)} = \begin{pmatrix} \mathbf{h}'_{m+1} \\ \mathbf{h}'_{m+2} \\ \vdots \\ \mathbf{h}'_{m+(k-1)^2} \end{pmatrix}_{[2]}$$

where $m = (i - 1)(c - 1)^2 + (j - 2)(c - 1)^2$. Then $GFfit^{(ij)}$ is a special case of $X_{[t:u]}^2$ (Cagnone and Mignani, 2007):

$$GFfit^{(ij)} = \mathbf{e}' \widehat{\Sigma}_{\mathbf{e}}^{-1} \mathbf{e}$$

where $\widehat{\Sigma}_{\mathbf{e}} = n^{-1} \mathbf{\Omega}_{\mathbf{e}}$ with $\mathbf{\Omega}_{\mathbf{e}}$ evaluated at the MLE $\hat{\boldsymbol{\theta}}$. Now

$$\mathbf{\Omega}_{\mathbf{e}} = \mathbf{H}_{[2]}^{(ij)} (D(\boldsymbol{\pi}) - \boldsymbol{\pi} \boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1} \mathbf{G}') (\mathbf{H}_{[2]}^{(ij)})'$$

$\mathbf{H}_{[2]}^{(ij)}$ is a partition of the general matrix $\mathbf{H}_{[1:q]}$. The extension to higher-order statistics is straightforward: Define

$$\mathbf{H}_{[3]}^{(ijk)} = \begin{pmatrix} \mathbf{h}'_{m+1} \\ \mathbf{h}'_{m+2} \\ \vdots \\ \mathbf{h}'_{m+(k-1)^3} \end{pmatrix}_{[3]}$$

where $m = (i - 1)(c - 1)^3 + (j - 2)(c - 1)^3 + (k - 3)(c - 1)^3$. Then

$$GFfit^{(ijk)} = \mathbf{e}' \widehat{\Sigma}_{\mathbf{e}}^{-1} \mathbf{e}$$

where $\widehat{\Sigma}_{\mathbf{e}} = n^{-1} \mathbf{\Omega}_{\mathbf{e}}$ with $\mathbf{\Omega}_{\mathbf{e}}$ evaluated at the MLE $\hat{\boldsymbol{\theta}}$. Now

$$\mathbf{\Omega}_{\mathbf{e}} = \mathbf{H}_{[3]}^{(ijk)} (D(\boldsymbol{\pi}) - \boldsymbol{\pi} \boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1} \mathbf{G}') (\mathbf{H}_{[3]}^{(ijk)})'$$

$\mathbf{H}_{[3]}^{(ijk)}$ is a partition of the general matrix $\mathbf{H}_{[1:q]}$. Now define an orthogonal components version of $GFfit$:

$$GFfit_{\perp}^{(ij)} = \sum_{\ell=m+1}^{\ell=m+(c-1)^2} \hat{\gamma}_{\ell}^2$$

where $m = q + (i - 1)(c - 1)^2 + (j - 2)(c - 1)^2$.

Then

$$X_{[2]}^2 = \sum_{i=1}^{i=q-1} \sum_{j=i+1}^{j=q} GFfit_{\perp}^{(ij)}$$

More general,

$$X_{PF}^2 = \sum_{\ell=1}^{\ell=q(c-1)} \hat{\gamma}_{\ell}^2 + \sum_{\ell=q(c-1)+1}^{\ell=\binom{q}{2}(c-1)^2} \hat{\gamma}_{\ell}^2 + \sum_{\ell=\binom{q}{2}(c-1)^2+1}^{\ell=\binom{q}{3}(c-1)^3} \hat{\gamma}_{\ell}^2 + \dots + \hat{\gamma}_{T-g-1}^2$$

Then

$$X_{PF}^2 = \sum_i GFfit_{\perp}^{(i)} + \sum_i \sum_j GFfit_{\perp}^{(ij)} + \sum_i \sum_j \sum_k GFfit_{\perp}^{(ijk)} + \dots + GFfit^{(1,2,\dots,q)}$$

because

$$X_{PF}^2 = \hat{\gamma}'\hat{\gamma} = \sum_{\ell=1}^{\ell=T-g-1} \hat{\gamma}_{\ell}^2$$

The extended $GFfit_{\perp}^{(ij)}$ are independent chi-squared statistics on $(c - 1)^2$ degrees of freedom due to the definition on orthogonal components. The original $GFfit^{(ij)}$ statistics are not necessarily independent and do not necessarily sum to $X_{[2]}^2$. Because the extended $GFfit_{\perp}^{(ij)}$ are defined on orthogonal components, they are order dependent.

Orthogonal components may be calculated using a Cholesky decomposition. Other methods for calculation are QR decomposition and sequential sum of squares from a weighted regression. The statistic $X_{[t:u]}^2$ may be calculated from a sum of orthogonal components.

It is also possible to form orthogonal components of Pearson's statistic from X_{red}^2 by the methods described above, but the components from X_{red}^2 are not distributed chi-square with 1 degree of freedom. X_{red}^2 yields $2^k - 1$ components to reproduce the Pearson-Fisher statistic, and the decomposition is more arithmetic than stochastic.

5. Comparison of Statistics When the Number of Variables is Large

5.1 Application to Factor Analysis

Factor analysis often involves a large number of variables. When the manifest variables are categorical, the model is known as categorical variable factor analysis and sometimes as the item response theory model. In an application of educational testing, the number of manifest variables could be 50 or more. Comparisons of the statistics reviewed in the previous section will be presented using this model with one factor. According to this model, the probability of the response to a manifest variable, sometimes also referred to as an item, can be given by a logistic item response function:

$$\pi(Y_i = 1 | \beta'_i, X = x) = (1 + \exp(-\beta_{i0} - \beta_{i1}x))^{-1}$$

where Y_i represents the response to item i ,

β_{i0} = intercept parameter for item i

β_{i1} = slope parameter for item i

$\beta'_i = (\beta_{0i}, \beta_{1i})$

x = value taken on by latent random variable X

Since

$$\pi(Y_i = 0 | \beta'_i, X = x) = 1.0 - \pi(Y_i = 1 | \beta'_i, X = x),$$

it follows that

$$\pi(Y_i = y_i | \beta'_i, x) = \pi(Y_i = 1 | \beta'_i, x)^{y_i} [1.0 - \pi(Y_i = 1 | \beta'_i, x)]^{1-y_i}$$

It is assumed that, *conditional* upon the latent variable, responses to the manifest variables are independent. Let \mathbf{Y} represent a random vector of responses to the items, with

element Y_i , and let \mathbf{y} represent a realized value of \mathbf{Y} . Then

$$\pi(\mathbf{Y} = \mathbf{y} \mid \boldsymbol{\beta}, x) = \prod_{i=1}^k \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)^{y_i} [1 - \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)]^{1-y_i}$$

$$\text{where } \boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{11} \\ \beta_{02} & \beta_{12} \\ \beta_{03} & \beta_{13} \\ \vdots & \vdots \\ \beta_{0k} & \beta_{1k} \end{pmatrix}.$$

Finally, the probability of response pattern s , say, is obtained by taking the expected value of the conditional probability over the distribution of X in the population, and is sometimes called the marginal probability:

$$\pi_s(\boldsymbol{\beta}) = \pi(\mathbf{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}) = \int_{-\infty}^{\infty} \pi(\mathbf{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}, x) f(x) dx$$

where $f(x)$ is the density function of X in the population of respondents.

If \mathbf{V} represents a T -dimensional multinomial random vector of frequencies associated with the response patterns, the distribution of \mathbf{V} is given by

$$\pi(\mathbf{V} = \mathbf{n}) = n! \prod_{s=1}^T \frac{[\pi_s(\boldsymbol{\beta})]^{n_s}}{n_s!}$$

where \mathbf{n} = vector of observed frequencies

n_s = element s of \mathbf{n}

$$n = \text{total sample size} = \sum_{s=1}^T n_s$$

The maximum likelihood solution for parameter estimation based on the marginal likelihood function, but with a probit function in place of the logit function, was first given by Bock and Lieberman (1970).

5.2 Calculations with 20 or More Variables

For testing the fit of a multinomial model, the Pearson and likelihood ratio statistics are common. When the number of manifest variables is as large as 20, the cross-classified table has 2^{20} , or 1,048,576 cells. If the relevant sample size for testing the fit of the model is on the order of a few thousand observations, then the data table will be sparse in the sense that many cells will have counts of zero or 1. It is well known that the asymptotic chi-square approximation for the distribution of the Pearson and likelihood ratio statistics may not be valid when the data table is sparse. Extensive simulations have shown that p -values obtained from the chi-square distribution for a test of the categorical factor analysis model on a sample of size 1000 start to become unreliable at about 6 to 8 manifest variables, depending on the skew of distribution of the frequencies.

Calculation of the Pearson statistic itself does not necessarily encounter memory limits because the contribution of each cell can be calculated individually and cumulated. Processing time is not a concern at 20 variables, but it becomes a concern at 25 variables. At 25

variables, calculation of the Pearson statistic requires 2^{25} expected values, each of which requires numerical evaluation of an integral by a method such as Gauss-Hermite quadrature. If the calculations are performed in RAM, processing time is several minutes, and if the calculations are performed in virtual memory, reading and writing to disk, processing time is several hours. When the number of manifest variables becomes larger, the fitted probabilities become very small, but a 64 bit machine can accurately store very small values. The primary hurdle for a large number of variables is obtaining a p-value by a method such as the parametric bootstrap. For 30 variables and 100 bootstrap samples, processing time would exceed 20 hours. An application to a test data set with $n=500$ and $k=20$ variables resulted in a Pearson chi-square value of 212969.73 on 1048535 degrees of freedom.

Given values for generalized least squares estimates of model parameters, X_{Ch}^2 is fairly straightforward since the covariance matrix can be calculated from the observed counts or proportions. However, use of observed proportions in the estimator of $\Sigma_{\hat{\tau}}$ also restricts the usefulness of the statistic. $\mathbf{H}'_{[1:2]}(D(\hat{\mathbf{p}}) - \hat{\mathbf{p}}\hat{\mathbf{p}}')\mathbf{H}_{[1:2]}$ is consistent for $\Sigma_{\hat{\tau}}$ under typical conditions, but it is not a consistent estimator under sparseness conditions. Simulations reported by Reiser and VandenBerg (1994) show that chi-square approximation for the distribution of X_{Ch}^2 is valid only up to 8 to 10 variables for typical sample sizes. An application to a test data set with $n=500$ and 20 variables resulted in a value of 441.396 on 170 degrees of freedom.

Computation of $X_{[t:u]}^2$ is more demanding and requires careful calculations for some models. $X_{[t:u]}^2$ requires calculation of $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ which requires $k2^{k+1}$ integrals to be evaluated by numerical quadrature for the factor analysis model. Using SAS PROC IML, these calculations can be accomplished entirely in random access memory for 20 variables, if 6 to 8 GB of RAM are available, in approximately 4 minutes of CPU time. If the calculations are done using virtual memory, reading and writing to disk, then processing time for 20 variables is on the order of 30 hours. The R system has a limit of 3.5 GB of RAM for 32 bit installation. Due to an increase in the size of matrices by an exponential factor, increasing the number of manifest variable to just 25, increases the amount of RAM required to avoid virtual memory to approximately 100 GB. Although it is possible to locate servers with 100GB of RAM, the limit for currently available desktop computers appears to be 20 manifest variables for calculations performed in SAS or R. Also, if the number of manifest variables is greater than 27, the SAS IML will be exceeded, and if the number of variables is greater than 30, R limits on array size may be exceeded, depending on the operating system.

Calculation of $X_{[t:u]}^2$ may be accomplished by directly inverting $\widehat{\Sigma}_{\mathbf{e}}$. However, because there is high collinearity among the columns of H, the covariance matrix $\widehat{\Sigma}_{\mathbf{e}}$ is highly ill-conditioned. Calculations of the inverse may be numerically unstable with even as few as five manifest variables. A better approach is to calculate the orthogonal components and then calculate $X_{[t:u]}^2$ as a sum of components. Calculation of components by finding the Cholesky factor of $\widehat{\Sigma}_{\mathbf{e}}$, is more stable, but may become inaccurate with fewer than 10 variables. A QR decomposition of $\widehat{\Sigma}_{\mathbf{e}}$ than the Cholesky factor, but also becomes inaccurate with more variables. A very accurate calculation of components can be accomplished by obtaining the sequential sum of squares from a weighted least squares approach. If SAS PROC REG is used for the weighted least squares, for example, the components in the form of sequential sum of squares can be obtained very quickly and accurately for 20 variables. An application to a test data set with $n=500$ and 20 variables using a direct generalized inverse of $\widehat{\Sigma}_{\mathbf{e}}$ resulted in a value for $X_{[1:2]}^2$ of 21035.86 on 210 degrees of freedom and a value for $X_{[2]}^2$ of 12267.347 on 210 degrees of freedom. Using the orthogonal components method resulted in a value for $X_{[1:2]}^2$ of 217.05 on 200 degrees of freedom. Using only

second-order marginals and the method of orthogonal components resulted in a value for $X_{[2]}^2$ of 209.056 on 190 degrees of freedom.

Direct calculation of Y_2 , X_{red}^2 , and M_r are usually numerically stable even for a large number of variables because the inverse of $\widehat{\Sigma}_e$ is not required. However, computational demands for the calculation of M_r and Y_2 are the same as for $X_{[t:u]}^2$ because M_r also requires the calculation of the derivatives $\mathbf{G} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and of degrees of freedom for Y_2 requires $\widehat{\Sigma}_e$. An application to a test data set with $n=500$ and 20 variables resulted in $Y_2 = 43.16$ on 43 degrees of freedom (rounding down from $c = 43.039$), $X_{red}^2 = 187.79$ on 170 degrees of freedom, and $M_r = 87.77$ on 170 degrees of freedom.

6. Simulation Study for 20 Variables

A Monte Carlo simulation study was performed under the following conditions: $q = 20$ manifest variables, unidimensional latent variable, $n = 300$ sample size, and $N = 1000$ pseudo samples. With 20 variables and sample size 300, the cross-classified table of the 20 variables is very sparse. The following parameter values were used for the simulation study: $\boldsymbol{\beta}'_0 = (0.40, 0.65, 0.60, 0.35, 0.25, 0.75, 0.50, 0.60, 0.40, 0.50, 0.50, 0.40, 0.60, 0.50, 0.75, 0.25, 0.35, 0.60, 0.65, 0.40)$; $\boldsymbol{\beta}'_1 = (-3.8, -3.4, -3.0, -2.6, -2.2, -1.8, -1.4, -1.0, -0.6, -0.2, 0.2, 0.6, 1.0, 1.4, 1.8, 2.2, 2.6, 3.0, 3.4, 3.8)$.

Simulation results for Type I error are shown in Table 1. Consistent with many previous simulation studies, a chi-square approximation for the Pearson and likelihood ratio statistics does not perform well under severe sparseness, although the Pearson statistic performs noticeably better than the likelihood ratio statistic. $X_{[1:2]}^2$ and $X_{[2]}^2$ do not perform well when a generalized inverse or Cholesky decomposition are used to calculate values. $X_{[1:2]}^2$ and $X_{[2]}^2$ perform well when values are obtained by using a sum of orthogonal components calculated from sequential sum of squares using the SWEEP operator with a check of pivot values for a singularity. Y_2 , X_{red}^2 , and M_r perform within an acceptable range. Some of the empirical Type I error rates are low, but the true nominal level is contained within a 95% confidence interval around the empirical value.

Simulation results for power are shown in Table 2. In the simulation, the model under H_0 is misspecified because it constrains all slopes equal. $X_{[2]}^2$, X_{red}^2 , and M_r show equivalent power levels. $X_{[1:2]}^2$ and Y_2 have somewhat lower power due to dilution of the test caused by inclusion of first-order marginals in the statistics.

7. Conclusion

X_{red}^2 requires substantially less computation than the other statistics considered and performs as well in terms of power and Type I error levels. When the number of cross-classified manifest variables is large, the difference in computation time is important. While $X_{[1:2]}^2$ and $X_{[2]}^2$ require more computation, they have the advantage that they can be decomposed into orthogonal components that have asymptotic chi-square distribution. These components can be used to investigate the source of lack of fit when the model is rejected by a more omnibus test statistic.

Table 1. Simulation Results Type 1 Error

Statistic	N	$\alpha = 0.05$	$\alpha = 0.01$
Pearson	1000	0.055	0.054
LR	1000	0	0
Christoff	1000	1.00	0.999
Y_2	1000	0.047	0.015
$X_{[1:2]}^2$ inv	953	0.953	0.939
$X_{[1:2]}^2$ SS	998	0.047	0.007
$X_{[2]}^2$ inv	998	0.212	0.143
$X_{[2]}^2$ chol	0	.	.
$X_{[2]}^2$ SS	998	0.047	0.007
X_{red}^2	1000	0.040	0.004
M_2	1000	0.038	0.004

Table 2. Simulation Results for Power

Statistic	N	$\alpha = 0.05$	$\alpha = 0.01$
Y_2	1000	0.190	0.061
$X_{[1:2]}^2$ SS	1000	0.196	0.067
$X_{[2]}^2$ SS	1000	0.217	0.065
X_{red}^2	1000	0.218	0.065
M_2	1000	0.218	0.065

REFERENCES

- Bartholomew, D. J., & Leung, S. O. (2002). A goodness of fit test for sparse 2^P contingency tables. *British Journal of Mathematical and Statistical Psychology*, 55, 1-15.
- Birch, M. W. (1964). A new proof of the Pearson-Fisher Theorem. *Annals of Mathematical Statistics*, 35, 818-824.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- Cagnone, S., & Mignani, S. (2007). Assessing the goodness of fit of a latent variable model for binary data. *Metron*, LXV, 337-361.
- Cai, L., Maydeu-Olivares, A., Coffman, D.L., & Thissen, D. (2006). Limited information goodness of fit testing of item response theory models for sparse 2^p tables. *British Journal of Mathematical and Statistical Psychology*, 59, 173-194.
- Goodnight, J. H. (1978). The sweep Operator: Its importance in Statistical Computing. SAS Technical Report R-106, SAS Institute, Cary, NC.
- Joreskog & Moustaki (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36, 347-387.
- Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and goodness-of-fit testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association*, 100, 1009-1020.
- Maydeu-Olivares, A. & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, 71, 713-732.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.
- Rayner, J. C. W., & Best, D. J. (1989). *Smooth Tests of Goodness of Fit*. Oxford: New York.
- Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika*, 61, 509-528.
- Reiser, M. (2008). Goodness-of-fit testing using components based on marginal frequencies of multinomial data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 331-360.
- Reiser, M., & Lin, G. (1999). A goodness-of-fit test for the latent class model when expected frequencies are small. In M. Sobel & M. Becker (Eds), *Sociological Methodology 1999*, 81-111. Boston: Blackwell.
- Tollenaar, N., & Mooijaart, A. (2003). Type I errors and power of the parametric bootstrap goodness-of-fit test: Full and limited information. *British Journal of Mathematical and Statistical Psychology*, 56, 271-288.