Power Calculations for Statistics Based on Orthogonal Components of Pearson's Chi-square

M.M.K Dassanayake[*] M. Reiser[†] J. Zhu

Abstract
The Pearson and likelihood ratio statistics are commonly used to test goodness of fit for models applied to count data from a multinomial distribution. When counts are from a table formed by the cross classification of a large number of variables, the traditional statistics may have lower power and inaccurate Type I error level due to sparseness. For a cross-classified table, Pearson's statistic can be decomposed into orthogonal components associated with the marginal distribution of observed variables, and an omnibus fit statistic defined on a sum of components for lower-order marginals has good performance for Type I error rate and statistical power, even when applied to a sparse table. In this study asymptotic power will be calculated for statistics based on orthogonal components and will be compared to results obtained by using Monte Carlo simulations. Power will be calculated for testing a confirmatory dichotomous variable factor analysis model and will be investigated for both individual components that can serve as lack-of-fit diagnostics and for omnibus statistics formed by summing orthogonal components.

Key Words: Item response model, Asymptotic power, Orthogonal components, Monte Carlo simulation

1. Introduction

Statistical modeling often involves finding a model that may have generated the data of interest and it is important to test the fit of the model because inferences drawn on poorly fitting models can be misleading. In multinomial models we often consider the null hypothesis $H_o\colon \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$, where $\boldsymbol{\pi}$ is a T-dimensional vector of multinomial probabilities, and $\boldsymbol{\pi}(\boldsymbol{\beta})$ is a vector of the multinomial probabilities as a function of parameters in the vector $\boldsymbol{\beta}$. When the model parameters $\boldsymbol{\beta}$ are unknown and estimated, the null hypothesis $H_o\colon \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$ is often tested with the Pearson-Fisher statistic:

$$\chi^2_{PF} = \sum_s z_s^2, \tag{1.1}$$

where

$$z_s = \sqrt{n}(\pi_s(\hat{\boldsymbol{\beta}}))^{-\frac{1}{2}}\left(\hat{\mathrm{p}}_s - \pi_s(\hat{\boldsymbol{\beta}})\right)$$

and where, $\hat{\mathrm{p}}_s$ is element s of $\hat{\mathbf{p}}$, vector of multinomial proportions, n is total sample size, $\hat{\boldsymbol{\beta}}$ parameter estimator vector, $\pi_s(\boldsymbol{\beta})$ is the expected proportion for cell s and $\pi_s(\hat{\boldsymbol{\beta}})$ is the estimated expected proportion for cell s.

[*]School of Mathematical and Statistical Sciences, Arizona State University, P.O. Box 871804, Tempe, AZ 85287, USA
[†]School of Mathematical and Statistical Sciences, Arizona State University, P.O. Box 871804, Tempe, AZ 85287, USA

The Pearson-Fisher statistic has an asymptotic chi-square distribution with T-g-1 degrees of freedom under the large sample theory conditions (Koehler and Larantz 1980), where T is the number of cells and g is the number of estimated model parameters. Thus, a usual assumption for the Chi-square approximation is that expected cell counts become large asymptotically. This assumption is not reasonable for analyzing a table where there are many cells with small counts and/or zeros. When the data are from a table formed by the cross classification of large number of variables, and thus many cells with small counts and/or zeros, Pearson's chi-square statistic test may have lower power and inaccurate Type I error due to sparseness. (Agresti and Yang 1987). Over the past years several statistics has been proposed to remedy this issue. Some of these statistics formed on lower-order marginals have been proven to overcome the deleterious effect of spareness. Tests based on these statistics also proven to have higher power under commonly encountered situations (Reiser 2008). The other issue related to Chi-square test statistic is that it gives little guidance about the source of poor fit when the null hypothesis is rejected. Dassanayake and Reiser (2015) conducted a simulation study using individual orthogonal components of Pearson's chi-square statistic. Results of this study suggests that the individual orthogonal components can be used to detect source of poor fit for models fit to binary cross-classified variables. This paper will be an extension of the Dassanayake and Reiser (2015) research. In this study, asymptotic power will be calculated for statistics based on orthogonal components and will be compared to results obtained by using Monte Carlo simulations. Power will be calculated for testing a confirmatory dichotomous variable factor analysis model and will be investigated for both individual components that can serve as lack-of-fit diagnostics and for omnibus statistics formed by summing orthogonal components.

2. Marginal Proportions

A traditional statistic such as Pearson's chi-square uses the joint frequencies to calculate goodness of fit for a model that has been fit to a cross-classified table. This section presents a transformation from joint proportions or frequencies to marginal proportions. Marginal proportions are used to develop test statistics presented in Section 3.2.

2.1 First- and Second-Order Marginals

The relationship between joint proportions and marginals can be shown by using zeros and 1's to code the levels of dichotomous response random variables, $Y_i, i = 1, 2, \ldots, q$, where Y_i follow the Bernoulli distribution with parameter P_i . 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. For dichotomous response variables, a response pattern is a sequence of zeros and 1's with length q. The $T = 2^q$-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 $\boldsymbol{V}$ as the T by q matrix with response patterns as rows.

For instance when $q = 3$,

$$\boldsymbol{V} = \begin{pmatrix} 0\,0\,0 \\ 0\,0\,1 \\ 0\,1\,0 \\ 0\,1\,1 \\ 1\,0\,0 \\ 1\,0\,1 \\ 1\,1\,0 \\ 1\,1\,1 \end{pmatrix} .$$

Let v_{is} represent element i of response pattern s, $s = 1, 2, \ldots, T$. Then, under the model $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$, the first-order marginal proportion for variable Y_i can be defined as

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

and the true first-order marginal proportion is given by

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

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

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

where $j = 1, 2, \ldots, q-1$; $i = j+1, \ldots q$, and the true second-order marginal proportion is given by

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

2.2 Higher-Order Marginals

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 $\boldsymbol{V}$. 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, $\mathbf{H}_{[t]} \equiv \mathbf{H}_{[t:t]}$. $\mathbf{H}_{[1:q]}$ gives a mapping from joint proportions to the set of $(2^q - 1)$ marginal proportions:

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

where

$$\boldsymbol{P} = (P_1,\ P_2,\ P_3, \ldots P_q,\ P_{12},\ P_{13}, \ldots P_{q-1,q},\ P_{1,1,2} \ldots P_{q-2,q-1,q} \ldots P_{1,2,3\ldots q})'$$

is the vector of marginal proportions.

For example, when q=3,

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

2.3 Residuals

H matrix can also be used to create residuals for marginals. Define the unstandardized residual $\mathrm{r}_s = \hat{\mathrm{p}}_s - \pi_s(\hat{\boldsymbol{\beta}})$, and denote the vector of unstandardized residuals as **r** with element r_s . Then a vector of simple residuals for marginals of any order can be defined as

$$\boldsymbol{e} = \mathbf{H}(\hat{\mathrm{p}} - \boldsymbol{\pi}(\hat{\boldsymbol{\beta}})) = \mathbf{Hr}.$$

3. Testing Fit on Marginal Distributions

3.1 Linear Combinations of Joint Frequencies

A traditional composite null hypothesis for a test of fit on a multinomial model is H_o: $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$. Linear combinations of $\boldsymbol{\pi}$ may be tested under the null hypothesis H_o: $\mathbf{H}\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\pi}(\boldsymbol{\beta})$. $\mathbf{H}$ may specify linear combinations that form marginal proportions as defined in the previous section.

3.2 Test Statistic

The use of components of Pearson's chi-square statistic has a long history dating back at least to Lancaster (1969). The motivation for components has been the possibility that a directional test would have higher power for certain alternative hypotheses than the omnibus goodness-of-fit test (Rayner & Best, 1989). Reiser(1996, 2008) and Reiser and Lin (1999) proposed statistics that can be obtained from orthogonal components defined on marginal proportions. These statistics have higher power under some circumstances, and they usually perform well when applied to sparse frequency tables.

$\sqrt{n}\,\mathbf{r}$ has asymptotic covariance matrix $\boldsymbol{\Omega}_{\mathbf{r}}$, where

$$\boldsymbol{\Omega}_{\mathbf{r}} = (D(\boldsymbol{\pi}(\boldsymbol{\beta})) - \boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\boldsymbol{\beta})' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'),$$

and where

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

See Haberman (1973). Then consider the linear combination $\boldsymbol{e} = \mathbf{Hr}$. If $\mathbf{H}$ contains $2^q - g - 1$ linearly independent rows corresponding to marginals from order 1 to q, then define the statistic

$$X^2_{[1:q]} = n\mathbf{r}'\mathbf{H}'\boldsymbol{\Omega}_{\boldsymbol{e}}^{-1}\mathbf{Hr}.$$

Here the statistic is evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is now consistent and efficient for $\boldsymbol{\beta}$, such as the maximum likelihood estimator, and where $\boldsymbol{\Omega}_{\boldsymbol{e}} = \mathbf{H}\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{H}'$. With the added condition that the rows of $\mathbf{H}$ are linearly independent of the columns of $\mathbf{G}$, i.e., rank$(\mathbf{H}' \vdots \mathbf{G}) = T + g$, $X^2_{[1:q]}$ can be shown to be equivalent to X^2_{PF} due to the correspondence of the joint and marginal proportions. See also Reiser (2008). To obtain orthogonal components, define the upper triangular matrix $\boldsymbol{F}$ such that $\boldsymbol{F}'\boldsymbol{\Omega}_{\boldsymbol{e}}\boldsymbol{F} = \boldsymbol{I}$. $\boldsymbol{F} = (\boldsymbol{C}')^{-1}$, where $\boldsymbol{C}$ is the Cholesky factor of $\boldsymbol{\Omega}_{\boldsymbol{e}}$. Then writing $\boldsymbol{\Omega}_{\boldsymbol{e}}$ as $\boldsymbol{CC}'$,

$$\begin{aligned}
X^2_{PF} &= n\mathbf{r}'\mathbf{H}'(\hat{\boldsymbol{C}}')^{-1}\hat{\boldsymbol{C}}'(\hat{\boldsymbol{C}}\hat{\boldsymbol{C}}')^{-1}\hat{\boldsymbol{C}}(\hat{\boldsymbol{C}})^{-1}\mathbf{Hr} \\
&= n\mathbf{r}'\mathbf{H}'\widehat{\boldsymbol{F}}\widehat{\boldsymbol{F}}'\mathbf{Hr}
\end{aligned}$$

where $\widehat{\boldsymbol{F}}$ and $\hat{\boldsymbol{C}}$ are the matrices $\boldsymbol{F}$ and $\boldsymbol{C}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$.

Premultiplication by $(\boldsymbol{C}')^{-1}$ orthonormalizes the matrix $\mathbf{H}_{[1:q]}$ relative to the matrix $D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}' - \mathbf{G}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{G}'$. Let $\mathbf{H}^* = \boldsymbol{F}'\mathbf{H}_{[1:q]}$, then

$$X^2_{PF} = n\mathbf{r}'(\widehat{\mathbf{H}}^*)'\widehat{\mathbf{H}}^*\mathbf{r}$$

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

Define

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

Then

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

and the elements $\hat{\gamma}_j^2$ are orthogonal components of X^2_{PF}. Since $\widehat{\mathbf{H}}^*\mathbf{r}$ has asymptotic covariance matrix $\boldsymbol{F}'\boldsymbol{\Omega}_{\boldsymbol{e}}\boldsymbol{F} = \boldsymbol{I}_{T-g-1}$, the elements $\hat{\gamma}_j^2$ are asymptotically independent χ^2_1 random variables.

By summing these components one could obtain limited-information statistics. For example the statistc on first- and second-order marginals from Reiser (1996) is

$$X^2_{[1:2]} = \sum_{j=1}^{j=q(q+1)/2} \hat{\gamma}_j^2,$$

and the statistic on second-order marginals from Reiser and Lin (1999) is

$$X^2_{[2]} = \sum_{j=q+1}^{j=q(q+1)/2} \hat{\gamma}_j^2,$$

In general, using the matrix $\mathbf{H}_{[t:u]}$ as given above,

$$X^2_{[t:u]} = \sum_j \hat{\gamma}_j^2,$$

where the limits on the sum depend on t and u, the order of the selected marginals, and the statistic can also be expressed as

$$X^2_{[t:u]} = \boldsymbol{e}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1}\boldsymbol{e}$$

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = n^{-1}\boldsymbol{\Omega}_{\boldsymbol{e}}$, with $\boldsymbol{\Omega}_{\boldsymbol{e}}$ evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\beta}}$. However, depending on the fitted model, it may be difficult to calculate $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1}$ accurately due to collinearity. Direct calculation of components is considerably more stable.

Under the regularity conditions given by Birch (1964), the limiting distribution of $X^2_{[t:u]}$ as $n \to \infty$ can be shown to be the χ^2-distribution because $\boldsymbol{e}$ is a linear combination of the elements of $\mathbf{r}$, $n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} \xrightarrow{P} \boldsymbol{\Omega}_{\boldsymbol{e}}$, and $\boldsymbol{e} \xrightarrow{L} MVN(\boldsymbol{\xi}, \boldsymbol{\Sigma}_{\boldsymbol{e}})$. Another sparse asymptotic result from Simonoff (1986) is applicable here. Assuming $\hat{\boldsymbol{\beta}}$ is a consistent estimator, $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + O_p(n^{-\frac{1}{2}})$; if $\boldsymbol{\pi}(\boldsymbol{\beta})$ has bounded second partial derivatives with respect to $\boldsymbol{\beta}$, $sup_s \left|\pi_s(\hat{\boldsymbol{\beta}})/\pi_s - 1\right| = O_p(n^{-\frac{1}{2}})$. So, even under sparseness conditions, $\pi_s(\boldsymbol{\beta}) \xrightarrow{P} \pi_s$, $\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}) \xrightarrow{P} \boldsymbol{\pi}$, $n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{e}} \xrightarrow{P} \boldsymbol{\Omega}_{\boldsymbol{e}}$, and the asymptotic chi-square distribution for $X^2_{[t:u]}$ is valid. $X^2_{[t:u]}$ 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 for $X^2_{[t:u]}$ are known from theory and are determined by the rank of $\boldsymbol{\Omega}_{\boldsymbol{e}}$, which will be equal to the number of linearly independent rows in $\mathbf{H}$, assuming rank$(\mathbf{H}'\vdots\mathbf{G}) = m + g$ where m is the rank of $\mathbf{H}$, and assuming the model $\boldsymbol{\pi}(\boldsymbol{\beta})$ is identified. The statistic has been extended to ordinal response variables by Cagnone and Mignani (2007) and Reiser, Zhu, and Cagnone (2014).

3.3 Application to Factor Analysis

When categorical manifest variables are hypothesized to be associated with a continuous latent variable, the model is known as categorical variable factor analysis and sometimes as the item response theory model. In order to investigate the challenges of a large number of variables and intense computations, a comparison of the statistics reviewed in the previous section will be presented using this model with one factor.

According to the categorical factor 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:

$$P(Y_i = 1 \mid \boldsymbol{\beta}_i', X = x) = (1 + \exp(-\beta_{i0} - \beta_{i1}x))^{-1} \tag{3.1}$$

$$\begin{aligned} \text{where } Y_i & \text{ represents the response to item } i, \\ \beta_{i0} &= \text{intercept parameter for item } i \\ \beta_{i1} &= \text{slope parameter for item } i \\ \boldsymbol{\beta}_i' &= (\beta_{0i}, \beta_{1i}) \\ x &= \text{value taken on by latent random variable } X \end{aligned}$$

Since

$$P(Y_i = 0 \mid \boldsymbol{\beta}_i', X = x) = 1.0 - \pi(Y_i = 1 \mid \boldsymbol{\beta}_i', X = x),$$

it follows that

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

It is assumed that, *conditional* upon the latent variable, responses to the manifest variables are independent. Let $\boldsymbol{Y}$ represent a random vector of responses to the items, with element Y_i, and let $\mathbf{y}$ represent a realized value of $\boldsymbol{Y}$. Then

$$P(\boldsymbol{Y} = \mathbf{y} \mid \boldsymbol{\beta}, x) = \prod_{i=1}^{k} \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)^{y_i}[1 - \pi(Y_i = 1 \mid \boldsymbol{\beta}, x)]^{1-y_i} \tag{3.2}$$

$$\text{where } \boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{i1} \\ \beta_{02} & \beta_{12} \\ \beta_{03} & \beta_{13} \\ \vdots & \vdots \\ \beta_{0q} & \beta_{1q} \end{pmatrix}.$$

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:

$$\pi_s(\boldsymbol{\beta}) = \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}) = \int_{-\infty}^{\infty} \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}, x) f(x) dx \tag{3.3}$$

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

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

$$\pi(\boldsymbol{U} = \mathbf{n}) = n! \prod_{s=1}^{T} \frac{[\pi_s(\boldsymbol{\beta})]^{n_s}}{n_s!} \tag{3.4}$$

$$\begin{aligned} \text{where } \mathbf{n} =& \text{vector of observed frequencies} \\ n_s =& \text{element } s \text{ of } \mathbf{n} \\ n =& \text{total sample size} = \sum_{s=1}^{T} n_s \end{aligned}$$

4. Asymptotic power

In this section we will describe the theory behind the calculation of asymptotic power of individual orthogonal components of χ^2_{PF}.

Consider the situation of testing a hypothesis $H_o\colon \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta})$ against alternative $H_a\colon \boldsymbol{\pi} \neq \boldsymbol{\pi}(\boldsymbol{\beta})$ using Pearson-Fisher statistic. Suppose we have sequence of specific alternatives $\boldsymbol{\pi}_n$ satisfying $\sqrt{n}(\boldsymbol{\pi}_n - \boldsymbol{\pi}(\boldsymbol{\beta})) \to \boldsymbol{\delta}$ for some constant matrix $\boldsymbol{\delta}$. In this approach, the best fit of the model to the population gives $\boldsymbol{\pi}_s(\boldsymbol{\beta})$ as the probability for cell s, but the true probability differs from that value by $\boldsymbol{\delta}/\sqrt{n}$. Note the model lack of fit goes to zero at the rate $n^{\frac{1}{2}}$ as n approaches infinity. With this technique, Mitra (1958) shows that χ^2_{PF} has a limiting non-central chi-square distribution with non-centrality parameter λ, where

$$\lambda = \boldsymbol{\delta}' Diag[\boldsymbol{\pi}(\boldsymbol{\beta})]^{-1} \boldsymbol{\delta} \tag{4.1}$$

and $df = T - g - 1$, where $T = 2^q$. Under the condition $\mathbf{H} = \mathbf{H}_{[1:q],-g}$, and using a strategy similar to Reiser (2008), it can be shown that

$$\lambda = \boldsymbol{\delta}' \mathbf{H}' \boldsymbol{\Sigma}_e^{-1} \mathbf{H} \boldsymbol{\delta} \tag{4.2}$$

Based on the right-hand side of this expression, it is possible to decompose the noncentrality parameter into orthogonal components associated with marginals. Consider the Cholesky decomposition in section 2.2 where $\boldsymbol{F}' \boldsymbol{\Omega_e} \boldsymbol{F} = \boldsymbol{I}$ and $\boldsymbol{F} = (\boldsymbol{C}')^{-1}$, where $\boldsymbol{C}$ is the Cholesky factor of $\boldsymbol{\Omega_e}$. Using the same decomposition, let

$$\boldsymbol{\zeta} = (\mathbf{F}') \mathbf{H} \boldsymbol{\delta} = \mathbf{H}^* \boldsymbol{\delta} \tag{4.3}$$

where $\mathbf{F}$and $\mathbf{H}^*$ are defined as in Section 2. Then $\lambda = \boldsymbol{\zeta}' \boldsymbol{\zeta}$, and orthogonal components are ζ_j^2 , where ζ_j is an element of $\boldsymbol{\zeta}$. These components can be used to calculate the power for tests based on marginals of differing order. For example, the non-centrality parameter for $\chi^2_{[1:2]}$ is given by

$$n \sum_{j=1}^{q(q+1)/2} \zeta_j^2 \tag{4.4}$$

As for our case we can calculate the power of each orthogonal component using non-central chi-square distribution and the non-centrality parameter for the j^{th} component is given by ζ_j^2.

4.1 Power calculation example

For the purposes of power calculations under fixed, finite n, cell proportions were generated from a known model, with two factors and q=8 manifest variables and fit with a one-factor model. Loadings for the first factor were (1,1,1,0,0,0,1,1) and the loadings of the second factor had two settings, (0.1, 0.1, 0.1, 1.2, 1.2, 1.2, 0.2, 0.2) which here after we will call *higher factor loading* and (0.1, 0.1, 0.1, 0.6, 0.6, 0.6, 0.2, 0.2) which here after we will call *lower factor loading*. Note, we have allocated higher weights for items 4,5 and 6. Thus we are expecting to see lack-of-fit on those places for the one factor model. The intercepts of the model were kept symmetrical at (-2.0, -1.5, -1, -0.5 , 0.5 ,1, 1.5, 2). In order to generate the proportions of this two factor model, we used Multivariate Gaussian quadrature. Note the equation 2.3 will now become,

$$\pi_s(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \int\int \pi(\boldsymbol{Y} = \mathbf{y}_s \mid \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, x_1, x_2) f(x_1, x_2) dx_1 dx_2 \tag{4.5}$$

and the integral must be evaluated by numerical quadrature. These proportions were then multiplied by a selected initial sample size n_0 to create the true cell frequencies. As the next step, the model of the null hypothesis was analyzed using resulting cell frequencies. Two parameter IRT model was built under the null hypothesis and expected proportions were calculated. Thereafter, the non-centrality parameter was calculated as described in the previous section. The non-centrality parameter for any other sample size, say simply n, can be approximated by using the expression $\lambda \approx \frac{n}{n_0}\lambda_0$. Some Monte Carlo simulations were conducted to cross-validate this approach to the calculation of power for fixed, finite n, and the results are given in the next section. Note the Cholesky factor method can be numerically unstable since it involves getting an inverse of a matrix. An alternative method would be to calculate orthogonal components using weighted regression. Calculating orthogonal components using weighted regression proven to be more stable in extreme situations compared to Cholesky factor method (Dassanayake and Reiser, 2015). The appropriate weight matrix, $\hat{\mathbf{W}}$, for the regression is given by,

$$\hat{\mathbf{W}} = (\mathbf{I} - \boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\boldsymbol{\beta})' - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}') \tag{4.6}$$

Once the regression model is fitted, the orthogonal components can be obtained using the sequential sums of squares.

4.2 Results

In the previous section we described the process of obtaining non-centrality parameters for orthogonal components. Once acquired, the non-centrality parameters are used to calculate the power of each orthogonal component using non-central chi-square distribution with one degree of freedom. Note the significance level was set to 0.05. Asymptotic power results for n=500 for lower and higher factor loadings is given in the Table 1. We compared these with empirical power results from Monte-Carlo simulations. Results were similar in most of the aspects, however the difference between asymptotic power and empirical power was not as small as we expected. Further investigations was carried out with small number of manifest variables to figure out the reasons behind these results. After considerable amount of simulations and tests we were able to figure out the rationale behind our results. When the intercepts move away from zero, some of the response pattern frequencies get small, when this happens the chi-square approximation is not as good as we expected. So the simulation results for empirical power are showing the effect of this *not so good* chi-square approximation. To remedy this issue and illustrate our theory we change all the intercepts to zero and re-ran the simulations. Results given in the the table 2.

Table 1: Asymptotic and simulated power comparison for model with symmetrical intercepts

	Lower factor loading (n=500)		Higher factor loading (n=500)	
Marginal	Simulated power	Asymptotic power	Simulated power	Asymptotic power
(1,2)	0.049147	0.05013	0.21988	0.06965
(1,3)	0.061184	0.05042	0.26406	0.09974
(1,4)	0.048144	0.05263	0.081325	0.11174
(1,5)	0.049147	0.05269	0.1004	0.12724
(1,6)	0.049147	0.05256	0.2239	0.14154
(1,7)	0.049147	0.05	0.074297	0.05001
(1,8)	0.053159	0.05001	0.10442	0.05007
(2,3)	0.049147	0.05137	0.31627	0.1766
(2,4)	0.055165	0.05504	0.12851	0.17265
(2,5)	0.060181	0.05513	0.12851	0.19965
(2,6)	0.054162	0.05486	0.31124	0.22633
(2,7)	0.035105	0.05	0.070281	0.05004
(2,8)	0.053159	0.05003	0.083333	0.05151
(3,4)	0.064193	0.06267	0.19478	0.36346
(3,5)	0.052156	0.06277	0.21486	0.42353
(3,6)	0.064193	0.06204	0.46787	0.48443
(3,7)	0.058175	0.05026	0.10643	0.05881
(3,8)	0.053159	0.05	0.077309	0.05055
(4,5)	0.32999	0.33781	0.66566	0.95282
(4,6)	0.2678	0.30097	0.68273	0.96247
(4,7)	0.053159	0.05002	0.037149	0.05106
(4,8)	0.05015	0.05005	0.052209	0.05006
(5,6)	0.29789	0.31903	0.84337	0.99147
(5,7)	0.047141	0.05	0.038153	0.05
(5,8)	0.043129	0.05006	0.080321	0.05051
(6,7)	0.059178	0.05	0.073293	0.05014
(6,8)	0.044132	0.05006	0.074297	0.05061
(7,8)	0.041123	0.05	0.059237	0.0503

Based on the results in the Table 2 it is clear that our theory was correct. Difference between asymptotic power and the empirical power is very small. We also did Monte-Carlo simulations to check Type I error of fitting a one factor model to the data with zero intercepts. If Type I error is not small then these power results do not have much meaning in terms of practical applications. Results (Table 3 and 4) shows that the Type I error rates are within 0.05 for all the simulations. Thus our power calculation indeed have meaningful results. As explained in previous sections each orthogonal component is distributed as chi-square distribution with one degree of freedom. To check this distributional assumption, chi-square Q-Q plots were built for the simulation values related to each component. None of the Q-Q plots showed deviations from the assumption. Some of theses results are shown in the Appendix.

Table 2: Asymptotic and simulated power comparison for zero intercept model

	Lower factor loading (n=500)		Higher factor loading (n=500)	
Marginal	Simulated power	Asymptotic power	Simulated power	Asymptotic power
(1,2)	0.052	0.05026	0.108	0.05694
(1,3)	0.066	0.05051	0.132	0.06275
(1,4)	0.056	0.05437	0.096	0.10705
(1,5)	0.068	0.05442	0.132	0.11064
(1,6)	0.057	0.05446	0.163	0.11458
(1,7)	0.034	0.05	0.049	0.05
(1,8)	0.04	0.05	0.059	0.05002
(2,3)	0.047	0.05131	0.126	0.07666
(2,4)	0.069	0.05717	0.151	0.14212
(2,5)	0.06	0.05725	0.154	0.14866
(2,6)	0.056	0.05734	0.215	0.15592
(2,7)	0.044	0.05	0.06	0.05
(2,8)	0.053	0.05	0.078	0.05
(3,4)	0.06	0.0643	0.23	0.2339
(3,5)	0.069	0.06451	0.254	0.24923
(3,6)	0.077	0.06473	0.352	0.26647
(3,7)	0.055	0.05	0.1	0.05
(3,8)	0.043	0.05	0.06	0.05002
(4,5)	0.347	0.37269	0.927	0.99343
(4,6)	0.381	0.37578	0.94	0.99556
(4,7)	0.053	0.05	0.043	0.05
(4,8)	0.043	0.05004	0.045	0.05007
(5,6)	0.372	0.3789	0.982	0.99713
(5,7)	0.045	0.05	0.035	0.05
(5,8)	0.053	0.05003	0.04	0.05001
(6,7)	0.06	0.05	0.053	0.05
(6,8)	0.05	0.05002	0.046	0.05001
(7,8)	0.056	0.05	0.062	0.05002

5. Conclusion

In this study, second order marginals related to orthogonal components were examined as lack-of-fit diagnostics. Simulations were based on the two parameter IRT model for a two latent variable model and were successful in indicating pair of variables for which the model does not fit well. When the sample size increases, ability to indicate pair of variables for which the model does not fit well increases significantly. The Asymptotic power results tally with empirical power results. But when the intercepts move away from zero, and the sample size is moderate, some of the response pattern frequencies may get small, and the chi-square approximation is not as good as we expected. Thus the simulated results can have deviations from asymptotic power.

6. Appendix

Table 3: Type I error of the Orthogonal components for n=500 with lower factor loading

ortho(1,2)	**ortho(1,3)**	**ortho(1,4)**	**ortho(1,5)**	**ortho(1,6)**	**ortho(1,7)**	**ortho(1,8)**
0.043	0.042	0.03	0.045	0.041	0.038	0.042
ortho(2,3)	**ortho(2,4)**	**ortho(2,5)**	**ortho(2,6)**	**ortho(2,7)**	**ortho(2,8)**	**ortho(3,4)**
0.039	0.04	0.037	0.033	0.047	0.056	0.041
ortho(3,5)	**ortho(3,6)**	**ortho(3,7)**	**ortho(3,8)**	**ortho(4,5)**	**ortho(4,6)**	**ortho(4,7)**
0.033	0.034	0.047	0.061	0.069	0.075	0.038
ortho(4,8)	**ortho(5,6)**	**ortho(5,7)**	**ortho(5,8)**	**ortho(6,7)**	**ortho(6,8)**	**ortho(7,8)**
0.048	0.052	0.039	0.06	0.05	0.06	0.053

Table 4: Type I error of the Orthogonal components for n=500 with higher factor loading

ortho(1,2)	**ortho(1,3)**	**ortho(1,4)**	**ortho(1,5)**	**ortho(1,6)**	**ortho(1,7)**	**ortho(1,8)**
0.0501	0.059118	0.04509	0.06012	0.0501	0.046092	0.062124
ortho(2,3)	**ortho(2,4)**	**ortho(2,5)**	**ortho(2,6)**	**ortho(2,7)**	**ortho(2,8)**	**ortho(3,4)**
0.0501	0.053106	0.058116	0.041082	0.063126	0.053106	0.046092
ortho(3,5)	**ortho(3,6)**	**ortho(3,7)**	**ortho(3,8)**	**ortho(4,5)**	**ortho(4,6)**	**ortho(4,7)**
0.04008	0.044088	0.049098	0.048096	0.053106	0.051102	0.057114
ortho(4,8)	**ortho(5,6)**	**ortho(5,7)**	**ortho(5,8)**	**ortho(6,7)**	**ortho(6,8)**	**ortho(7,8)**
0.052104	0.0501	0.051102	0.053106	0.061122	0.058116	0.056112

Figure 1: QQ plots for the simulation n=300

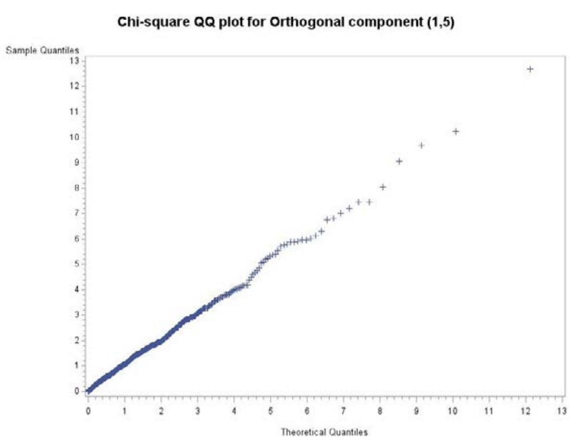


Figure 2: QQ plots for the simulation n=300

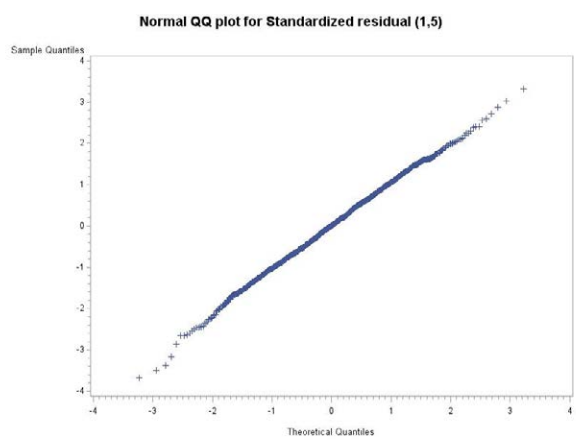


Figure 3: QQ plots for the simulation n=500

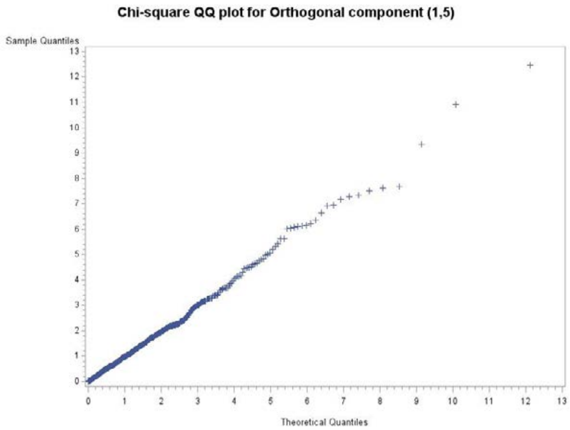


Figure 4: QQ plots for the simulation n=500

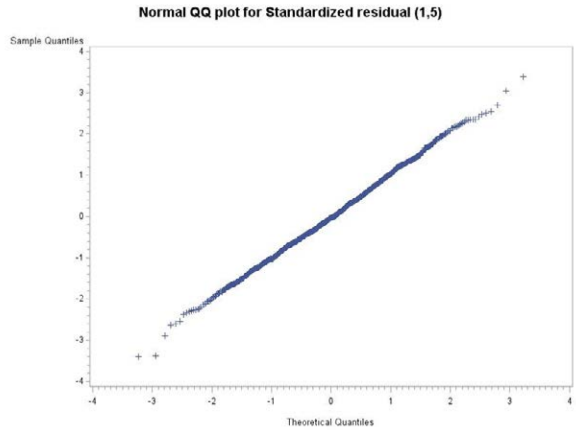


Figure 5: QQ plots for the simulation n=1000

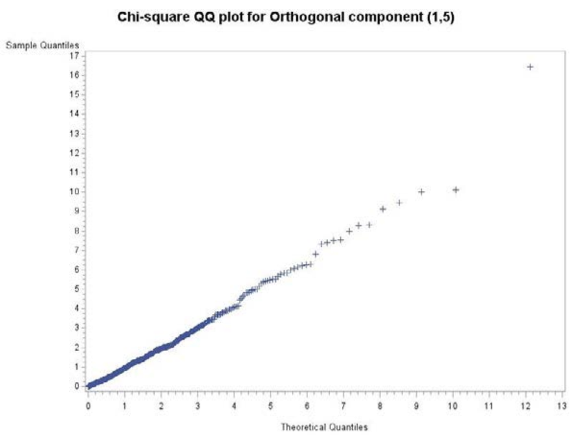
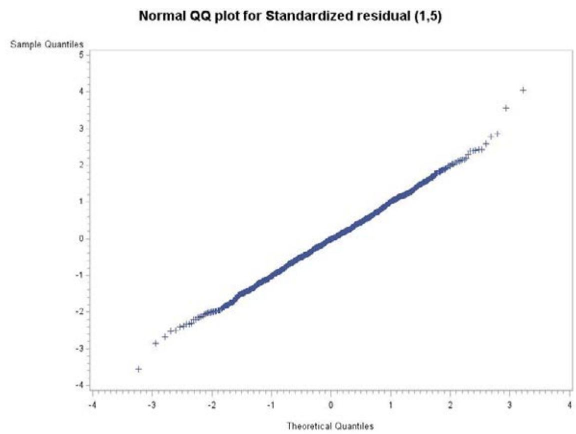


Figure 6: QQ plots for the simulation n=1000



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.

Bartholomew, D.J., Knott, M. & Moustaki (2011). *Latent variable models and factor analysis: A unified approach, 3rd Edition.* New York: Wiley.

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 2p tables. *British Journal of Mathematical and Statistical Psychology*, 59, 173-194.

Dassanayake, M.K., and M. Reiser. 2015. Lack-of-Fit Diagnostics Based on Standardized Residuals and Orthogonal Components of Pearson's Chi-Square. In JSM Proceedings, Social Statistics Section. Alexandria, VA: American Statistical Association. 614-628.

Goodnight, J. H. (1978). The sweep Operator: Its importance in Statistical Computing. SAS Technical Report R-106, SAS Institute, Cary, NC.

Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics, 29*, 205-220.

Joreskog & Moustaki (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36, 347-387.

Maydeu-Olivares, A. (2001). Limited information estimation and testing of thurstonian models for paired comparison data under multiple judgement sampling. *Psychometrika, 66*, 209-228.

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.

Maydue-Olivares, A., Garcia-forero, c., Gallardo-Pujol, D., & Renom, J. (2009). Testing categorized bivariate normality with two-stage polychoric correlation estimates. *Methodology: European Journal of Research Methods for the behavioral and social sciences,* 5, 131-136.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.

Muthén, L.K & Muthén, B.O. (1998-2010). Mplus User's Guide. Sixth edition. Los Angeles, CA.

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.

Takane, Y., and de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 292-408.

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.